# Essays on Consumption and Labour Supply over the Life-cycle

*Peter Levell*

A dissertation submitted in partial fulfillment

of the requirements for the degree of

**Doctor of Philosophy**

of

**University College London**.

Department of Economics

University College London

May 3, 2019

I, Peter Levell, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

No part of this thesis has been presented before to any university or college for submission as part of a higher degree. Chapters 3 and 5 are sole-authored research papers of my own. Chapter 2 was undertaken as joint work with Kevin Milligan and Garry Barrett. Chapter 4 was undertaken as joint work with James Banks, Richard Blundell and James Smith. Chapter 6 was undertaken as joint work with Thomas Crossley and Hamish Low. Chapter 7 was undertaken as joint work with Orazio Attanasio, Hamish Low and Virginia Sanchez-Marcos.

Chapter 2 is published in Christopher Carroll, Thomas F. Crossley and John Sabelhaus, Eds. (2015), “Improving the Measurement of Consumer Expenditures”, Studies in Income and Wealth, Volume 74. Chicago: University of Chicago Press. Chapter 3 is published in the Journal of the Royal Statistical Society: Series A (Vol. 178, Part 2, pp. 1-34). Chapter 4 is forthcoming in the American Economic Journal: Economic Policy. Chapter 7 is published in Econometrica (Vol. 86, pp. 2049-2082)

# Abstract

This thesis contains six self-contained chapters on the microeconomics of household consumption and labour supply behaviour, an introduction and a conclusion.

Chapter 1 provides an introduction. Chapters 2 and 3 address issues of measurement. The first of these considers the quality of household budget surveys relative to the national accounts. The second considers how we measure the inflation (with a focus on how price changes should be calculated across goods for which there is no corresponding spending data).

The following two chapters discuss consumption patterns at older ages. Chapter 4 discusses spending declines in two countries - the US and the UK - and the role of medical expenses in accounting for these differences. Chapter 5 attempts to shed light on long-standing puzzles surrounding consumption around retirement using non-parametric, 'revealed preference' tests of different models of consumption behaviour over the life-cycle.

The Chapters 6 and 7 examine how consumers' spending and labour supply choices are affected by changes in their economic environment. Chapter 6 looks how households responses to house price changes are affected by their initial leverage. Chapter 7 looks at how women's labour supply responds to changes in wages along both intensive and extensive margins.

Chapter 8 concludes.

# Acknowledgements

The completion of this PhD would not have been possible without a great deal of support and encouragement. I would like to thank my supervisors Professor Richard Blundell and Professor Ian Preston for their guidance and advice. I would also like to extend particular thanks to those who co-authored some of the chapters in this thesis: Orazio Attanasio, James Banks, Richard Blundell, Garry Barret, Thomas Crossley, Hamish Low, Kevin Milligan, and Virginia Sanchez-Marcos. It has been a privilege to work with you all.

For the duration of my PhD, I have been employed as an economist at the Institute for Fiscal Studies. I am indebted to the IFS for providing me with the funding and time to continue my studies, as well as for providing a stimulating, supportive and collegial research environment.

A number of other friends and colleagues have also at various times gone above and beyond the call of duty to inspire and challenge me, patiently explain concepts and methods to me, and most importantly, provide me with support and encouragement at the times I most needed it. This category includes far too many individuals to list individually, but you know who you are. I will be forever grateful to every one of you.

Finally, I would like to thank my parents, Bruce and Sally, and my brothers, Daniel and Jack, for many years of forbearance and for being the best family one could wish for.

# Impact statement

There are a number of ways that the contents of this thesis either has, or could be, put to beneficial use.

By documenting changes in the coverage of household surveys relative to the national accounts, and exploring possibly causes, Chapter 2 should be of interest to those designing household consumption surveys and to the users of those surveys. This includes national statistical agencies such as the Office for National Statistics (ONS) and the US Bureau of Labor Statistics.

Chapter 3 has direct relevance for current debates about the use of the Carli index in the UK Retail Prices Index. A draft version of this chapter fed into the ONS consultation on the use of the Carli index.

Chapter 4 documents reasons for differences in the decline in spending at older ages in the US and the UK, while Chapter 5 conducts revealed preference tests on consumption behaviour at retirement. Both have relevance for debates around the adequacy of households' retirement savings.

Chapter 6 discusses how leverage affects households' spending decisions. The contents of this chapter could usefully contribute to debates around the use of macro-prudential policies. These policies aim to promote macroeconomic stability by limiting the growth of leverage among households during credit booms.

Chapter 7 estimates labour supply elasticities and questions the idea that aggregate labour supply responses can be summarised by a single parameter. Labour supply elasticties should of interest to those interested in understanding employment and hours changes over the business-cycle or labour supply changes to changes in tax rates.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introductory Material

How do consumers make allocate their labour supply and consumption expenditures as they age? How do these aspects of consumer choice change in response to shock to their economic environment, such as wages or house prices?

This thesis examines questions such as these in six self-contained chapters looking at household spending and labour supply decisions over the life-cycle.

### 1.0.1 Measurement of prices and expenditures

Before moving on to discuss empirical questions on consumer behaviour, we begin by discussing two preliminary issues of measurement. Chapter 2 looks at household budget surveys and their quality as assessed using comparisons with spending totals from the national accounts. Chapter 3 looks at how we measure changes in the overall price level (and in particular how we should aggregate price movements for goods for which we have no corresponding expenditure data). The subsequent chapters in this thesis, and a great deal of the research literature in this field more generally, rely heavily on both of these data sources. These two chapters discuss important potential limitations of the measures of prices and expenditure that we use, which will be important to bear in mind in any empirical analysis of consumer behaviour.

#### 1.0.1.1 A comparison of micro and macro expenditure measures across countries using differing survey methods

Chapter 2 presents a comparative assessment of the performance of the household expenditure survey programs in Australia, Canada, the UK and US. It uses cross-country and time series variation in survey methodology to assess the role of factors influencing the performance of household surveys.

The chapter starts by examining the coverage of aggregate expenditure as measured in each of the surveys relative to national account data. The chapter describes how coverage rates are highest in Canada and the UK. It also shows that, over the past three decades, coverage remained fairly stable in Canada and Australia, while in the UK and US coverage rates declined sharply. It then goes on to consider survey response rates and top income shares in tandem with coverage rates. Falls in response rates are found to be predictive of changes in coverage rates. Furthermore, the change in coverage rates over time coincided with the growing concentration of income, suggesting that growing inequality may have contributed to declining coverage rates.

Finally the chapter moves on to examine the coverage of specific expenditure components. There are no clear differences in levels or changes in coverage by collection method. It is worth however noting the high and stable coverage of regularly purchased items (e.g. food), along with the more volatile coverage of irregular and larger expenditure items (e.g. vehicles, furniture and household equipment). The chapter concludes that aggregate patterns in coverage cannot be attributed to specific expenditure components or collection methods.

#### 1.0.1.2 Is the Carli Index flawed?: assessing the case for the new retail price index RPIJ

Chapter 3 discusses the relative merits of different formulae for use at the elementary aggregate level of price indices (i.e the level at which expenditure weights are not observed). It does so in the context of the decision by the UK's Office for National Statistics to replace the controversial Carli index with the Jevons index in a new version of the Retail Prices Index - the RPIJ.

The literature on price indices has identified three ways to select index numbers. In effect, we can ask:

1. Does the index respond appropriately when prices change in different situations or does it give answers we might consider perverse?

2. Is the index a good *statistical* estimator of the general price change as distinct from relative price changes across goods according to some measure (as we will discuss below, exactly which measure is a matter of some debate)?

3. Does the index provide a good measure of how the *cost of living* is changing for consumers (i.e the costs of obtaining a given level of welfare) ?

The first of these is called the *test* approach. This is because we determine which index has the best properties by setting out a list of criteria ('tests') and then asking which indices satisfy them. The second is called the *statistical* or stochastic approach. The third is the interpretation of price indices that is used by most economists and as such is referred to as the *economic* approach.

In this chapter, I explain what each of these approaches are and use them to make my own assessment on the suitability of the Carli index for use at the elementary aggregate level of a price index. In doing so I make a number of contributions not only to the current debate on the new RPIJ index but also to the way that elementary indices should be selected more generally. A primary concern of the ONS was the Carli's sensitivity to so-called price-bouncing which could lead to an upward bias. I formalise these concerns in a new price-bouncing test for the test approach. For the statistical approach, I present some evidence on the relative performance of the Carli and Jevons. I find no clear evidence for the superiority of one index over the other, and that the relative performances of the Carli and Jevons are not invariant to factors such as the month the index is calculated; the sample size and the choice of base month against which price changes can be compared; and the type of goods included in the elementary aggregate. I also argue that the economic approach *can* be applied to the elementary level, and moreover that it favours the Jevons index, by appealing to something analogous to the principle of insufficient reason from information theory. Overall, I conclude that there is indeed a case for replacing the Carli index with the Jevons.

### 1.0.2 Consumption spending at retirement

The next two papers discuss consumption patterns at older ages. Chapter 4 discusses spending declines in two countries - the US and the UK - and the role of medical expenses in accounting for these differences. Chapter 5 attempts to shed light on long-standing puzzles surrounding consumption around retirement using non-parametric, ‘revealed preference’ tests of different models of consumption behaviour over the life-cycle.

#### 1.0.2.1 Life-cycle consumption patterns at older ages in the US and the UK: can medical expenditures explain the difference?

Chapter 4 documents how nondurable expenditures decline significantly more at older ages in the UK compared to the US, in spite the fact that income paths are similar. It then explores several possible causes, including: differential cohort effects in the two countries that may distort average life-cycle age profiles, differences in timing of retirement in the presence of separabilities with employment, differential paths of housing expenditures possibly driven by institutional differences in housing markets between countries, level and path differences in health status and mortality, and finally the levels, prices and volatility of medical spending.

The chapter shows that, among all the potential explanations considered, those relating to healthcare  differences in levels and age paths in medical expenses and medical expenditure risk  can fully account for the steeper declines in nondurable consumption in the UK compared to the US. This is because deteriorating health with age in the US leads to higher spending there while this is not true in the UK where healthcare is provided free by the National Health Service.

We show this in two ways. Firstly, we quantify differences in the paths of different variables for different cohorts as they age in the two countries. While there are some differences in the way health, housing tenure, employment and mortality evolve in the two countries, these differences do not seem large enough to account for the cross-country difference in spending patterns. Secondly, we examine the role of different variables in explaining consumption changes in a regression context. We find that controlling for health, housing, mortality and employment only marginally reduces the cross-country difference in the decline in nondurable consumption spending with age when medical expenditure is included. We then turn to model non-medical consumption conditional on health status and real medical expenditures. This approach allows preferences for non-medical consumption to change in a non-separable way with health and the consumption of medical goods. It also captures any substitution effects driven by the change in the relative price of medical consumption. We also consider the role medical expense uncertainty may play in explaining consumption profiles in the US, partly by exploiting differences in the institutional environments in the two countries. We find suggestive evidence that precautionary savings against medical expense risk play an important role in US consumption decisions. Controlling for both medical uncertainty and relative prices fully explains the cross-country difference in spending declines.

Our regression estimates imply that medical uncertainty increases consumption growth at older ages in the US by around 0.90 percentage points per year on average for the ages we consider. Precautionary motives against medical expense risk in the UK are, by contrast, negligible.

#### 1.0.2.2 Revealed preference and consumption behaviour at retirement

Simple versions of the life-cycle model predict that households' consumption should not respond to anticipated changes in economic circumstances. However, a number of studies have documented falls in consumption as workers retire. Since retirement should be largely foreseeable for most workers, and a failure of consumers to smooth their consumption around this event violates a central prediction of the standard life-cycle model, this tendency is referred to in the literature as the "retirement consumption puzzle".

In Chapter 5 I probe the performance of the life-cycle model around the retirement period by applying non-parametric, "revealed preference" tests to households in a Spanish consumption panel. These tests allow us to examine whether the behaviour of retirees can be rationalised by specific variants of the life-cycle model. The tests themselves are non-parametric and avoid making specific assumptions on the form of the utility function (beyond it satisfying standard properties such as concavity, continuity, transitivity and so on).

I test whether household's behaviour can be rationalised under the standard life-cycle model, and the extent to which this affected when we tighten assumptions on consumer's foresight, allow for non-separabilities in preferences over consumption and labour force participation, or allow for revisions in the marginal utility of wealth at the point of retirement. Tests of the life-cycle model without perfect foresight are equivalent to tests of the Generlised Axiom of Revealed Preference (GARP) and have little probative force. The perfect foresight life-cycle model performs poorly in the sense that it is largely rejected in our data. Results for this model are not substantially improved when one allows for a revision in expectations at the point of retirement. I also find that minimising deviations from the perfect-foresight life-cycle model suggests rising marginal utility of wealth over time. Thus the smoothest possible paths of marginal utility that rationalise the data, are associated with wealth decreasing faster than we would otherwise expect. I show that this particular result is unlikely to be due to changing family composition over the retirement period, aggregate shocks over the period we consider or credit constraints.

This final conclusion cannot be interpreted as strong evidence of consumer myopia however. The path of marginal utility could be more stable over time if one allowed larger deviations from the perfect foresight model (due to either uncertainty or perhaps measurement error in the prices and interest rates used). Perhaps one key lesson that we can draw from the results of these exercises is the limits on what we can learn given only data on prices and quantities. The same data can often be rationalised by increasing or decreasing marginal utilities of wealth, depending on exactly what one assumes about the nature of the utility function. It is therefore not easy to conclude whether consumption patterns over retirement are consistent with smoothed marginal utility or not on the basis of expenditure data alone.

### 1.0.3 Household responses to economic shocks

In a final section, we examine how consumers' spending and labour supply choices are affected by changes in their economic environment. Chapter 6 looks at how households responses to house price changes are affected by their initial leverage. Chapter 7 considers how women's labour supply responds to changes in wages along both intensive and extensive margins.

#### 1.0.3.1 Consumption spending, housing investments and the role of leverage

Do house price booms induce households to borrow and re-leverage their balance sheets? And can leverage make households more sensitive to future house price shocks (whether good or bad)? It is now widely believed that increases in debt and leverage that accompanied the international house price boom prior to the 2008 financial crisis both deepened and prolonged the length of the slump in consumption spending that followed. As a result, recent years have seen policy-makers showing increasing interest in macro-prudential measures that limit leverage growth among households during boom periods, and relax them when economic conditions weaken.

In Chapter 6, we use two-sample IV methods to combine panel data on household wealth and leverage with detailed household spending survey data to examine the borrowing, spending and investment decisions of existing homeowners in response to house price increases. In particular, we consider whether these responses differ according to households' initial leverage (where leverage is defined as the ratio of households' mortgage debts to their net housing wealth). We use variation in credit conditions over time to isolate exogenous differences in households' leverage positions. We find that households who were initially more leveraged are more likely to both purchase other residential properties and invest in their own homes in response to local house price increases than other households. However they do not disproportionately increase their consumption spending as house prices rise, as would be expected if household spending were driven by traditional housing wealth effects, or if changes in house values helped to relax consumers' credit constraints.

We show using a simple model how this behaviour can be rationalised in a framework where households treat leverage as a portfolio choice, choosing leverage to optimise the risk and return on their assets. Households respond to house price increases by borrowing and investing in housing (including their own homes) in order to maintain their desired loan-to-value ratios. More leveraged households experience a larger reduction in the portfolio share of housing for a given price increase, and so we would expect that their residential investment spending should respond more strongly than other households'.

#### 1.0.3.2 Aggregating elasticities: Intensive and extensive margins of women's labour supply

For a long time economists have disagreed on how much hours of work and labour market participation rates respond to changes in wages. Labour economists have tended to estimate small labour supply elasticities from individual level data, while macroeconomists, who use business cycle fluctuations of wages and hours, have tended to find that that labour supply elasticities are considerably larger.

Chapter 7 aims to show that to understand labour supply behaviour and to calculate aggregate labour supply elasticities, it is crucial to both account for heterogeneity across individuals and to quantify responses along both intensive and extensive margins of labour supply. To make these points precisely, we estimate a life-cycle model of intratemporal and intertemporal choices over consumption, saving and work and characterise the response of women's labour supply to different types of wage changes. We consider both intertemporal and intratemporal choices, and identify intensive and extensive responses using data from the US Consumer Expenditure Survey. In estimating such a model, we use a flexible specification of preferences that allows us to test some of the assumptions commonly used both in the macro and labour literature on labour supply. We show that there is substantial heterogeneity in women's labour supply elasticities at the micro level and highlight the implications for aggregate behaviour.

We find substantial heterogeneity in labour supply responses, and that this heterogeneity is prevalent at both the intensive and extensive margins. The median static Marshallian elasticity is 0.18, but has a 90-10 range of -0.14 to 0.79. The corresponding Hicksian elasticity is 0.54, with 90-10 range of 0.38 to 1.16; and the corresponding Frisch wage elasticity is 0.87, with 90-10 range of 0.8 to 1.92. Responses at the extensive margin explain about 54% of the total labour supply response for women under 30, although this declines with age. Aggregate elasticities are higher in recessions, and increase with the length of the recession.

The heterogeneity that we find at the level of individual households implies that the aggregate labour supply elasticity is not a *structural* parameter: any aggregate elasticity will depend on the demographic structure of the economy as well as on factors such as the distribution of wealth and the particular point in the business cycle.

# Chapter 2

# A Comparison of Micro and Macro Expenditure Measures Across Countries Using Differing Survey Methods

Household-level consumption lies at the centre of research into many important economic questions. The measurement of microeconomic phenomena such as household poverty requires the observation of consumption choices made by households to provide useful information on economic hardship. At the macroeconomic level, the understanding of responses to booms and busts is enhanced by observing household consumption responses. Reliable consumption data are necessary to engage in meaningful empirical research in these areas.

However, there are ongoing concerns about the reliability of expenditure surveys in many countries. These concerns have led to efforts to renew expenditure survey methodology. In the United States, this activity centres on the 'Gemini Project' of the Bureau of Labor Statistics, tasked with improving the Consumer Expenditure Survey.[1] In Canada, the Survey of Household Spending was revised in 2010 with similar goals in mind.[2]

In this paper, we aim to contribute to these discussions by providing an international comparison of the performance of household expenditure survey data

---

[1]Edgar and Safir (2011) provide an overview of the Gemini Project.

[2]Tremblay et al. (2010) report results from a pilot project from 2007 evaluating several changes. Many of these changes have been implemented for the 2010 Survey of Household Spending.

across four 'Anglosphere' countries: Australia, Canada, the United Kingdom, and the United States. Our international comparison is a useful way to gather some evidence on the potential sources of problems with expenditure surveys, as differences in experience and methodology provide sources of variation that may give insights into the importance of factors influencing the performance of expenditure surveys.

Our strategy is to compare household expenditure survey data to expenditure measured in the national accounts of each country. While this 'coverage' approach is frequently adopted in country-specific studies of expenditure behaviour, the novelty of our contribution is to produce comparable results across four countries.[3] Attanasio et al. (2006), in assessing the expenditure behaviour of poor households in the US and UK, provide an assessment of micro survey evidence benchmarked against the national income. In comparison to their paper, we provide more recent years of data, two more countries with differing methodology, and a more detailed accounting of the survey differences. Deaton (2005) provides a comparison of a vast array of countries, with analysis of the same kind of 'survey vs. national accounts' comparison we perform here.

The paper proceeds first by reviewing the survey methods employed in the four target countries. We then discuss in more detail the construction and interpretation of household survey vs. national account comparisons, and examine the trends in aggregate ratios of survey to national account data across countries. Next, we consider how survey response rates have varied across countries and relate them to our aggregate coverage measures. We then compare the coverage measures to high income concentration through time and across our countries. Finally, we look at selected subcategories of expenditure to observe how trends vary across countries.

## 2.1 Expenditure survey methodology

In this section, we provide some background on the methodology employed for the household expenditure surveys in each of the four countries in our focus. We

[3]Some well-known examples of this measurement approach are Slesnick (1992), Garner et al. (2006), Garner et al. (2009) for the United States. Adler and Wolfson (1988) perform a similar exercise for Canada. Passero et al. (2015) also provide an updated approach to comparing the CE survey with PCE.

describe the target population, survey design, and other special features for each country. We begin with Australia, and proceed through Canada, the United Kingdom and the United States. At the end of these descriptions, we provide a summary table of the key elements of the survey methodologies.

## 2.2 Australia: Household Expenditure Survey

Our analysis draws on seven waves of the Australian Household Expenditure Survey (HES): 1975-76, 1984, 1988-89 and 1993-94, 1998-99, 2003-2004 and 2009-10. The HES is conducted over a 12 month period, typically coinciding with the financial (July-June) rather than calendar year, with households enumerated evenly over the survey period. The primary purpose of the HES is to collect comprehensive information on household expenditures, along with household income and, since 2003/ 04, wealth. The original objective of the HES program was to provide information for the construction of commodity weights in the consumer price index (CPI) - for more details on the HES background and methodology see Australian Bureau of Statistics (2011).

Expenditures are recorded in HES on an acquisition basis, with details on most regular expenditures collected using diary methods. Regular expenditure items for each household member aged 15 years or older are recorded in a personal diary covering a two week reporting period.[4] The fineness of the expenditure categories used in the survey has increased over time, with 660 items separately recorded in the 2003-04 and 2009-10 surveys.

Expenditures on infrequent, irregular or expensive items are recorded by personal interview with each household member aged 15 years or older. The recall period for irregular purchases varies, ranging from three months for major household furniture and appliances, 12 months for motor vehicle registrations, and three years for house purchases. Items such as insurance, rent payments and utility bills are recorded in the interview with respondents asked the value of the last payment and the period of time that payment covered. Given the recall periods for items

[4]The HES records regular expenditures using one-week diaries for two consecutive weeks. In the 1975-76 and 1984 HES, the reporting period for rural respondents was four weeks.

recorded in the household interview questionnaire, some of these expenditures will refer to time periods prior to the reference year. The public release HES reports average weekly expenditures for all items, with expenditures on some items converted to average weekly amounts. Additional information on household demographics and income are also collected during the household interview on a recall basis.

The scope of the HES includes "usual residents of private dwellings in urban and rural areas of Australia." Excluded from the survey are residents of non-private dwellings such as hotels, boarding schools, boarding houses and institutions. Further exclusions are residents of very remote districts (or Indigenous Communities).[5] In addition, "non-usual" residents of a private dwelling (e.g. visitors) are not included in the survey. Approximately 97-98 percent of the Australian population are within the scope of HES.

Sampling is based on a stratified multistage cluster design. The strata are based on census collector districts. Individual household records are weighted according to the probably of initial selection into the survey adjusted according to population benchmarks based on the demographic characteristics of household size and age composition, geographic location and labour force status.[6] The sample size of the individual HES collections is typically 7,000 households, though has ranged from 4,492 in 1984 to 9,774 in 2009-10. For the most recent survey we use, the response rate in the HES was 71.9 percent.

## 2.3 Canada: The Survey of Household Spending

The Survey of Household Spending (SHS) has been the primary household expenditure survey in Canada since it replaced the Family Expenditure Survey (FAMEX) starting in 1997.[7] The methodology is described in detail in Statistics Canada (2001). When relevant, we also referred to the methodological description in the

[5]Non-Australian defence forces stationed in Australia and the diplomatic personnel of overseas governments located in Australia are also excluded.

[6]The two initial HES did not use population benchmarking.

[7]The differences between the SHS and FAMEX are outlined in Statistics Canada (2000). The sample size increased, the survey became annual, population coverage broadened, and some minor changes to survey content were implemented. We include some data from the FAMEX in our work here, but primarily focus on the SHS.

User Guide from 2009 Statistics Canada (2011). A detailed comparison of the SHS with the American Consumer Expenditure Survey is provided in Brzozowski and Crossley (2011). These sources provide the foundation for the description of the SHS below. We also use the FAMEX surveys for some of our analysis, but the primary focus is on the more recent SHS.[8]

The SHS targets individuals living in Canadian private households, as well as residents of Indian reserves and Crown Lands. This definition excludes those who are official representatives of foreign countries living in Canada, as well as those who are representing Canada abroad. It also leaves out residents of institutions, hotels and rooming houses, religious orders and members of the Canadian Forces living in camps. For the lower 10 provinces, the coverage is around 98 per cent of the population. For the sparsely populated northern territories, coverage is over 80 per cent.

Sampling is based on the Labour Force Survey sample design, which uses stratified clusters. The strata are based on geography within each province. Special strata of households in areas with geographical concentrations of high and low income residents are also used. Clusters are chosen, and then a sample of households is chosen from each cluster. Extensive follow-up is engaged for households who refuse to comply, including further phone calls, visits, and letters. Sample size started at 18,031 in 1997. From then until 2007, the number of observations slid down to 13, 939. For 2008 and 2009, budget cutbacks meant a jump down to samples of 9,787 and 10,811.

The SHS attempts to gather information on the 12 month period from January 1st to December 31st. The information is gathered via a face-to-face recall survey of one household member in the January, February, or March following the end of the target calendar year. The survey respondents are encouraged to gather source documents such as credit card statements, mortgage statements, and their income tax records. The average survey takes one hour and forty minutes to complete.

[8]FAMEX surveys were conducted in 1969, 1974, 1978, 1982, 1984, 1986, 1990, 1992, and 1996. The 1984 and 1990 surveys are less comparable because in those years only residents of certain large cities were surveyed.

A 'balance edit' is applied when the difference between expenditure, income, and savings, exceeded a 20 percent tolerance level.[9] Item non-response is countered by imputing data based on 'nearest neighbor' imputations.

For 2009, weights are provided to account for non-response according to cells defined by province, age, household size, and family income as measured by administrative tax data. This weighting strategy has changed several times. Importantly, starting in 1999 tax-filing data from the Canada Revenue Agency were used to match on wage and salary income.[10] This is helpful if there is a concern that lower response rates are particular to certain parts of the income distribution, as the weights can account for such systematic patterns.[11]

## 2.4 United Kingdom: Living Costs and Food Survey

The information in this section is drawn from Office for National Statistics (2010). The Office for National Statistics (ONS) in the UK has carried out some form of annual survey of household expenditures since 1957. From 2008 this survey has been known as the Living Costs and Food Survey (LCFS). Prior to this it was known as the Expenditure and Food Survey, which brought together what were two separate surveys for food and expenditure  the Family Expenditure Survey and the National Food Survey  in 2001. The survey is conducted continuously throughout the year.

Participation in the survey is voluntary. In 2009, the survey selected over 12,000 addresses, but only 5,825 of these were included in the survey. The remaining addresses were either ineligible to be included (because, for instance, the addresses were for businesses), refused to participate or were not possible to contact. Households in Northern Ireland are sampled separately and oversampled relative to the rest of the UK in order to achieve the sample size required for separate analyses. The response rate among eligible addresses was 56% in Northern Ireland and 50.4% in the rest of the UK.

---

[9] Brzozowski and Crossley (2011) look into the impact of this balance edit in detail, by examining the data from 2006 when no balance edit was imposed.

[10] The income weights account for incomes in the following percentile ranges: 0-25th percentile, 25th-50th, 50th-65th, 65th-75th, 75th-95th,95th-100th).

[11] Sabelhaus et al. (2015) show that response rates in the U.S. Consumer Expenditure Survey are in fact much lower at the top of the income distribution.

Households who are surveyed are first asked a series of questions on income, demographic characteristics, large purchases over the last year or so (on white goods, vehicles, holidays etc.) and committed expenditures such as magazine subscription costs. Each household member over 16 is then given a spending diary in which they record all purchases made over the next two weeks. Simplified diaries for children aged 7-15 have also been included since 1998. At the end of the two weeks, each adult who kept a diary is paid £10 ($16) for completing the survey (children who kept a diary are paid £5 ($8)). Spending is grossed-up using weights from the most recent population Census (which have in the past been carried out once every 10 years - although 2011 may be the last). Weights are chosen so that the sample distribution matches the population in terms of region, age group and sex.

Data collected in the LCFS are used for a number of official purposes. As well as being used for the construction of the National Accounts, the LCFS is used to calculate expenditure weights for headline inflation measures.

## 2.5 United States: Consumer Expenditure Survey

The Consumer Expenditure Survey (CE) has been collecting information about American expenditure patterns on an ongoing basis since 1980.[12] The Bureau of Labor Statistics publishes the Handbook of Methods, of which Chapter 16 is devoted to the Consumer Expenditure Survey (Bureau of Labor Statistics (2011a)). A short summary is also provided in Bureau of Labor Statistics (2011b). A review of changes to methodology through time is provided by Goldenberg and Ryan (2009). We draw on these sources in forming our description of the CE survey in this section.

The CE survey combines two one-week diaries of around 7,000 households with a series of five quarterly recall surveys of another 7,000 households. The target is the total US civilian non-institutional population, which excludes military personnel living on base, nursing home residents, and people in prisons. Sampling

[12]There were antecedents to the ‘modern’ CE in 1960-61 and 1972-73, as well as earlier years.

takes place by choosing households from a list within each of 91 clusters. The list of addresses comes from the most recent Census file, augmented by new construction permits. For the 2010 survey, the response rate was 73.4 percent.

The diary component starts with an interview for demographic information on the first day. The diary of expenditure is to be completed every day during the week. The diary is collected at the end of the first week and a second diary is delivered. When the second diary is picked up, more questions are used to collect information on work and income from the previous year. The data are put through edits and adjustments when being processed. Some imputations are performed as well.

The quarterly recall survey component aims to gather information on less-frequently purchased items, with a three-month recall window. The raw data from the surveys is put through various checks, with imputed values being imposed for missing data. With the switch from pencil and paper to Computer Assisted Personal Interview in 2003, the time to complete the interview survey fell from about 90 minutes to around 65 minutes. In what follows we report results for the coverage of the interview survey only.

The survey is available annually from 1980 to 2010. For several quarters in the early 1980s, rural households were not surveyed. In our analysis below, we retain these years but they do stand out on several of the graphs for this reason.

Weights in the CE survey are calibrated to 24 population counts, including age, race, household tenure, region, and rural/urban. The target population counts are updated quarterly, and the demographics of the sample are assigned weights to match the population on these 24 factors. Of note, there is no adjustment for income.

**Table 2.1:** Features of the data sets

| | Australia | Canada (SHS) | United Kingdom | United States |
|---|---|---|---|---|
| Recall vs. Diary | Diary (regular) | Recall (irregular or large items) | Recall Diary (regular) Recall (irregular or large items) | Recall/ (and diary) |
| Interview recall period | Varies; up to 3 years | Annual | About a year | Five quarterly surveys |
| Balance Edit? | No | Yes | No | No |
| Weighting benchmarks | Age, household size, state, labour force status, income source | Age, province, earnings, household size | Age, region and sex | Age, race, region, urban/rural status |
| Typical Sample Size | 7,000 | 10,000 to 18,000 | 6,000 | 7,000/(7,000) |

Notes: Source is the documentation for the surveys in each of the four countries, as referenced in the text.

## 2.6 Comparison

In Table 2.1, we summarise the main features of the survey data in each country. The data from Canada are different in a number of ways, including the annual focus, having no diary, weighting based on administrative income data, and featuring a balance edit. In Australia, there is some weighting by income but just the source of income is used. The recall window for the surveys varies across countries. In Australia, it goes back up to three years for some items. In the UK, one interview goes back for a period of a year. For the United States, the survey is a sequence of five quarterly-focused questionnaires.

## 2.7 Aggregate coverage rates

The first step in our assessment of the performance of the household expenditure surveys is the examination of coverage rates of aggregate expenditure for each of the four countries. That is, we take the ratio of expenditures observed in the household survey, grossed up to the aggregate level, to the total expenditures taken from the national accounts. We compare this ratio across time and across countries.

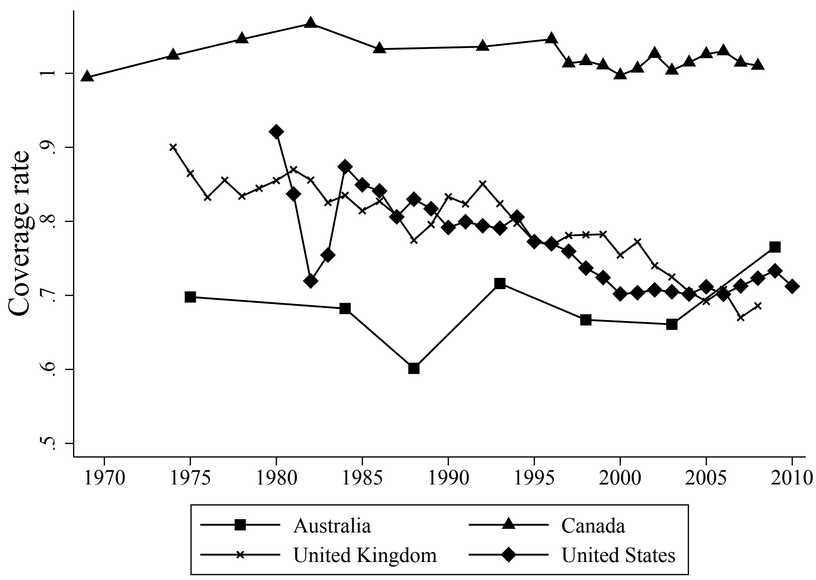
There are several well-known reasons to expect this ratio not to be 100 percent. (See, for example, Meyer and Sullivan (2009)) The population covered by each source may differ. For example, foreigners living in the host country and nationals living out of the country receive different attention in the national accounts and the expenditure surveys, as do military personnel and native peoples. In addition, the categories of expenditure available in the national accounts may not match those available in the expenditure surveys. For example, imputed housing rent is included in the national accounts, but not in the expenditure surveys. Finally, expenditure in the household sector of the national accounts includes spending by non-profit institutions serving households (such as charities), which does not appear in the expenditure surveys.

To make the best possible comparison, we adjust both the national accounts data and the expenditure survey data to remove items where they do not overlap.[13]

[13] For a detailed description of the methodology used for our UK sample, please see Crossley and O'Dea (2010).

For example, non-cash items such as imputed rent and food grown and consumed at home are taken out of the national accounts measure of household expenditure. Similarly, some items from the expenditure surveys, such as insurance purchases, are removed. With these adjustments made, we calculate the ratio of the grossed up expenditures from the household expenditure survey to the aggregate from the national accounts. This ratio is referred to as the 'coverage rate.' This coverage calculation is performed for expenditures in aggregate (as we do here in this section) as well as category by category comparisons (some of which are presented in a later section).

**Figure 2.1:** Coverage rates




Notes: Coverage rate is the proportion of consumer expenditure in the national accounts that is accounted for in the household surveys. Calculations by authors.

The coverage rates are graphed in Figure 2.1 for each of the four countries. In order to emphasise the nature of the decline, we have adjusted the y-axis to start at 0.5. Both the levels and trends differ sharply across countries. The Australian coverage rate stays in the 60 to 75 percent range, with no discernible trend. For Canada, the coverage rate is close to 1.0 for both the FAMEX (1969-1996) and the SHS (1997-2009) periods. There is no sign of an aggregate drop in coverage.

The coverage rate for the UK drops steadily over the years, from 90 to 67 percent. Finally, the United States shows coverage rates lower than Canada, but follows a very similar trend to the United Kingdom.[14]

In the two following two sections of the chapter, we investigate two aspects of this decline in our four countries. First, we look at the impact of declining response rates and increasing income inequality for the expenditure surveys on coverage. Following that, we compare different categories of expenditures, looking across diary and survey categories, as well as frequently and less frequently purchased items.

## 2.8 Candidate explanations for declining coverage

Many possible explanations for declining coverage rates have been offered. Here, we use our four countries to explore two possibilities. First, we look at declining survey participation rates.

A decline in survey participation rates has been observed around the developed world, a trend that began accelerating around 1990.[15] This trend coincides with the decline in the coverage rate in the CE survey in the United States, making non-response a candidate explanation for the decline in coverage. Response rates are relevant for the representativeness of samples, and reliability of micro-level evidence on expenditures. In particular, if the decline in response rates is not random across the population (and cannot be corrected adequately by sampling weights) then the results of the survey will no longer be representative of the population. For example, if high expenditure households have become increasingly less likely to respond, and if weighting did not account for this change, then coverage rates would be expected to decline.[16]

[14]The extra dip down in 1982-1983 is likely related to the discontinuation of rural data collection from the 3rd quarter of 1981 to the first quarter of 1985. We have checked our calculations against those in Meyer and Sullivan (2009) and found our coverage rates to be very similar.

[15]See de Leeuw and de Heer (2002) for international evidence. Tourangeau (2004) provides a discussion of the trends.

[16]Tourangeau (2004) reviews the evidence on the causes of declining survey participation, but does not discuss how non-participation is correlated with characteristics such as income. D'Alessio and Faiella (2002) find that non-response in the Bank of Italys Survey of Household Income and Wealth is more frequent among wealthier households. Finally, Sabelhaus et al. (2015) show that

The second possibility we examine is the impact of income inequality on survey accuracy.[17] The large trends in the concentration of income are documented across countries in Atkinson et al. (2011). This concentration has been especially acute in the 'Anglosphere' countries on which we focus. None of the four countries we study oversamples high income households for the expenditure surveys.[18] If increasing concentration of income is leading to an increasing concentration of expenditures, an increasing share of expenditure may be missed if the upper tail of expenditure is not adequately included in the survey sample. We also investigate this possibility. In addition, it is possible that the income inequality effect interacts with survey non-response. If the change in non-response is occurring more at the top of the income distribution, then the two effects (declining response rates and increasing income distribution) would reinforce each other.

With either survey response rates or income inequality, we will be comparing time series trends that happen to coincide with the change in coverage rates. It is prudent to be cautious in the interpretation of these results as causal. That said, we do get some mileage out of our cross-country comparison by including in our regression specification common time trends, allowing us to exploit the cross-country variation in the coverage, response rate, and inequality trends.

### 2.8.1 Response rates

Figure 2.2 shows the basic response rates for the expenditure surveys for the four countries, with the y-axis starting at 0.5. Each country exhibits declining responses rates, with the steepest occurring in the United Kingdom - where the drop is from 72 percent to 53 percent. The decline begins in the early 1990s in Australia, the United Kingdom, and the United States, but is not observable in Canada until the 2000s. While Canada was later starting downward, the decline exceeded 10 percentage points over the last decade.

To compare coverage and response rates, we graph the data from Figure 2.1

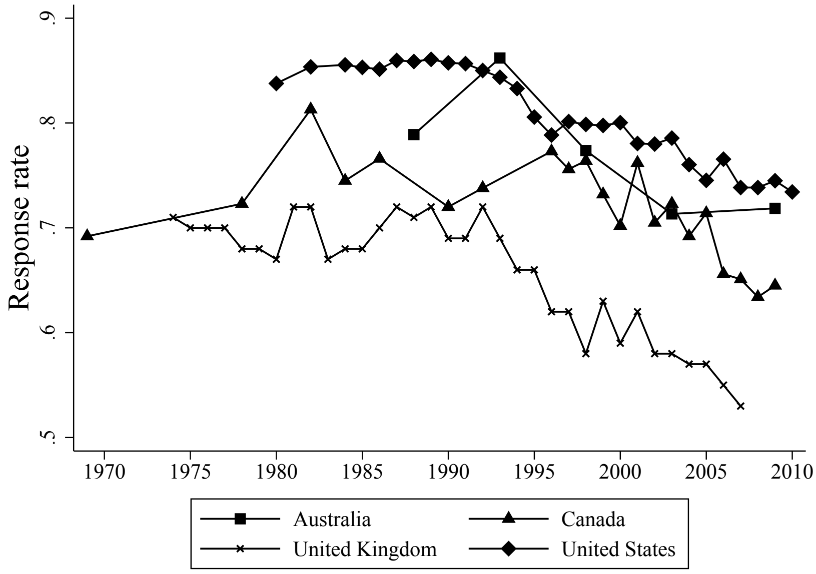
---

response rates in the CE are much lower at the top of the income distribution.

[17] We thank Angus Deaton for suggesting this possibility to us.

[18] Canada uses the Labour Force Survey sampling frame, which does target certain high income areas when choosing strata from which to sample. However, there is not explicit oversampling of high income households within the survey.

**Figure 2.2:** Response rates

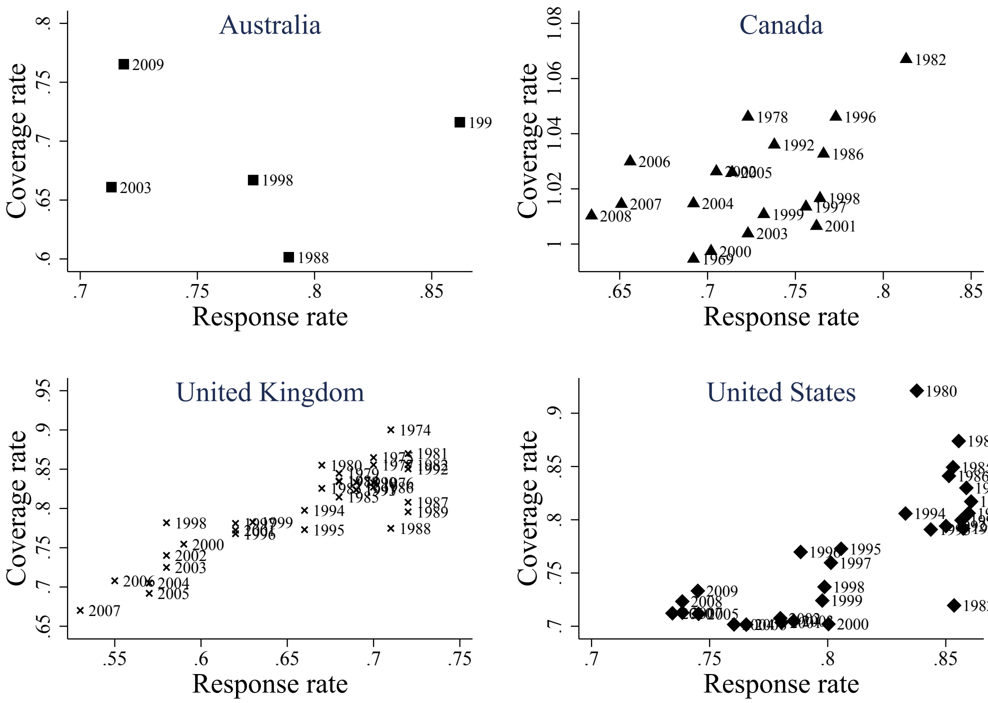



Notes: Response rate is the proportion of contacted households with completed surveys. Source is the documentation for the surveys in each of the four countries, as referenced in the text.

and Figure 2.2 together for each country as a scatter plot in Figure 2.3. The axes are different for each country in order to highlight the nature of the relationship in each country. For Canada, the United Kingdom, and the United States, there does appear to be a positive relationship between the response rate and the coverage rate. For Canada, the positive relationship in the figure is perhaps deceptive - the variation in the coverage rate is quite small - it ranges only from just under 1.0 to just under 1.07. The UK shows a fairly tight positive relationship across the 35 years available. In the US, the data are clustered in two groups that together suggest a similar positive relationship between coverage rates and response rates. For Australia, in contrast, there is no apparent relationship between response rates and coverage rates, although the limited number of surveys makes any conclusion difficult.

Figure 2.4 stacks together the data for all four countries in one plot with common axes. Looking across countries, the data display little clear relationship. However, within-country the United States and United Kingdom reveal positive relation-

**Figure 2.3:** Response rates vs. coverage rates




Notes: Coverage rate is the proportion of consumer expenditure in the national accounts that is accounted for in the household surveys. Response rate is the proportion of contacted households with completed surveys. Calculations by authors.

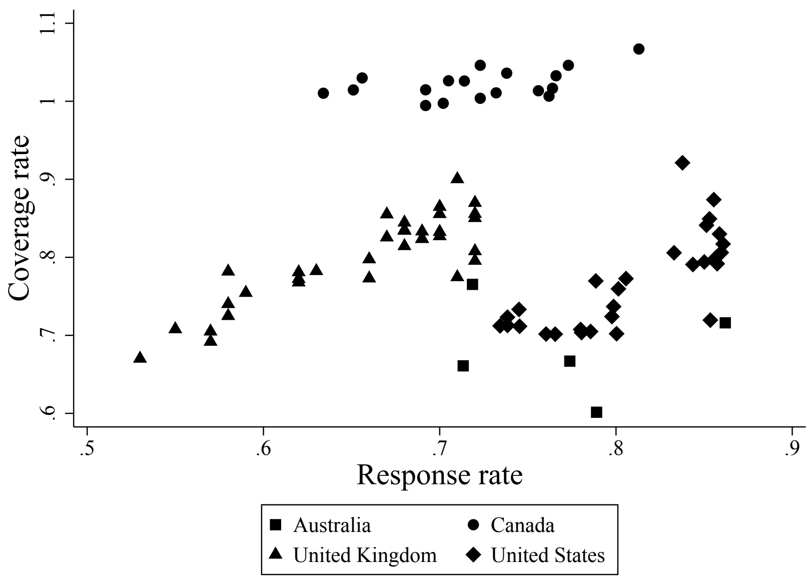
ships. Later in the paper, we can confirm these relationships in regressions.

### 2.8.2 Trends in high income concentration

The other trend we compare to declining coverage rates is the increase in income inequality. We draw on data from the high incomes database maintained at the Paris School of Economics (Alvaredo et al. (2012)). We use the proportion of income earned by those in the top one percent of the income distribution for our analysis here, although other high income measures showed similar results.

Figure 2.5 shows how the top one percent income shares have evolved in our four countries. Through the mid-1980s, there is little to be seen - although the top income share does start to rise in the UK following 1980. From around the beginning of the 1990s, there is a strong upward trend in each of the countries. The weakest of these upward trends is in Australia and the strongest is in the US. This

**Figure 2.4:** Response rates vs. coverage rates, all countries




Notes: Coverage rate is the proportion of consumer expenditure in the national accounts that is accounted for in the household surveys. Response rate is the proportion of contacted households with completed surveys. Calculations by authors.

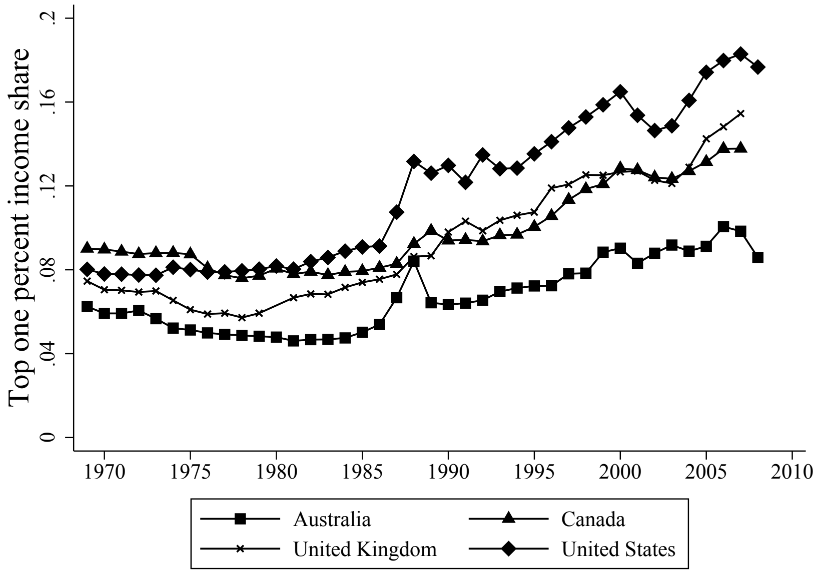
timing does correspond to the decline in coverage rates which accelerated in the 1990s.

We compare the trends in top income shares to the trends in coverage rates across all four countries in Figure 2.6, with separate scales for each country's axes. All four countries show signs of a negative relationship. Canada, again, has little variation in the coverage rate across years, so looks a bit different from the others. In the United States and the United Kingdom, there is a clear negative relationship between income inequality and the coverage rate.

Some parallels may be drawn here between our findings and those of Deaton (2005). In that paper, he finds that the coverage rate across countries is declining in the log of GDP, so higher income countries are experiencing worse coverage.[19] One of Deaton's explanations is that higher income countries tend to have higher income concentration, which may be captured less well in surveys. This is consistent with

[19] When comparing to our results, though, it must be remembered that much of the impact Deaton finds is concentrated among those countries with very low incomes.

**Figure 2.5:** Top 1 percent income shares, all countries




Notes: Top one percent income share is the share of total income received by those in the top one percent. Source is Paris School of Economics World Top Incomes Database, Alvaredo et al. (2012).

our findings here.

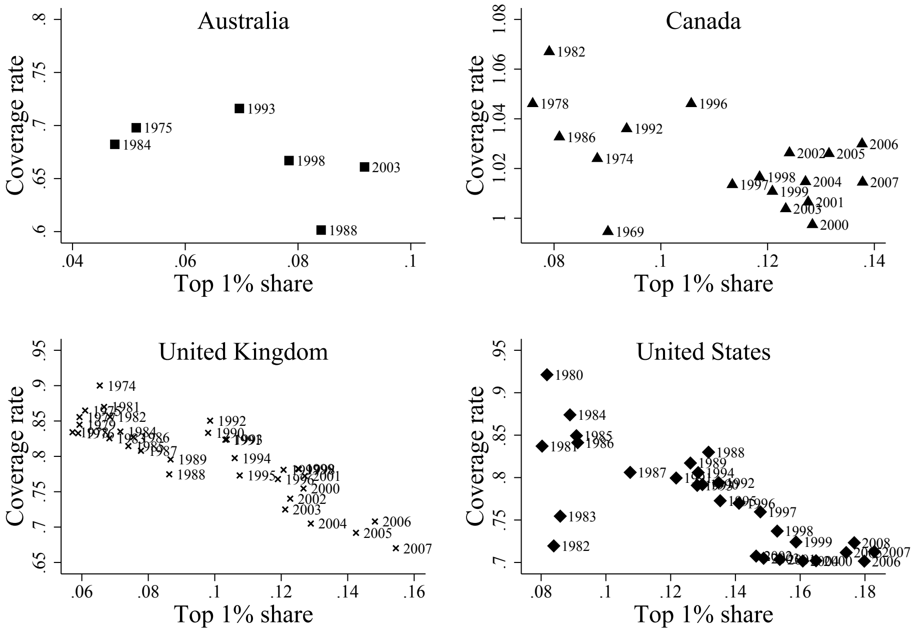
### 2.8.3 Regression analysis

The relationships from these figures can be summarised with some basic regressions. The coverage rate is regressed on the response rate, with country and time controls using Ordinary Least Squares. The equation takes the following form.

$$Coverage_{it} = \beta_0 + ResponseRate_{it}\beta_1 + Top1\%share\beta_2 + Country_i\beta_3 + Year_t\beta_4 + e_{it} \tag{2.1}$$

We report these results in Table 2.2. The dependent variable in all cases is the country-year coverage rate, and each column reports the results from a different specification. We report the regression coefficient, with the standard error beneath in parentheses. In column (1), we include no controls other than the constant term and the response rate variable. This effectively estimates a best-fit line through the data points as seen in Figure 2.4. The small and insignificant estimated coefficient

**Figure 2.6:** Coverage rates vs. top income shares




Notes: Top one percent income share is the share of total income received by those in the top one percent. Source is Paris School of Economics World Top Incomes Database, Alvaredo et al. (2012). Coverage rate is the proportion of consumer expenditure in the national accounts that is accounted for in the household surveys. Source is calculations by the authors.

confirms the lack of relationship across countries. The second column of the table includes country fixed effects. Here, the within-country relationships are used in the estimation, essentially taking an average of the relationships shown in the country-specific scatter plots in Figure 2.3. The coefficient swings strongly positive, at 0.779. This suggests that for every percentage point increase in the response rate, there is a 0.779 percentage point increase in the coverage rate. Taking the US as an example, the response rate dropped by 11.86 points from 1990 to 2008, so this coefficient explains a (0.779*11.86) 9.24 percentage point drop in coverage, which is larger than the 6.85 percent drop that occurred. On this basis, we interpret this coefficient as large.

**Table 2.2:** Coverage and response rates

Dependent Variable: country-year coverage rate

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | no controls | add country fixed effects | add top 1% | add interaction | with linear trend | add year fixed effects |
| Response Rate | 0.084 | 0.779 | 0.342 | 0.407 | 0.345 | 0.337 |
| | (0.154) | (0.078) | (0.111) | (0.255) | (0.112) | (0.207) |
| Top 1% income share | | | -1.006 | -0.642 | -1.232 | -1.026 |
| | | | (0.203) | (1.307) | (0.362) | (0.688) |
| Response X top share | | | -0.487 | | | |
| interaction | | | (1.731) | | | |
| Canada | | 0.406 | 0.413 | 0.414 | 0.420 | 0.416 |
| | | (0.020) | (0.017) | (0.018) | (0.020) | (0.028) |
| United Kingdom | | 0.237 | 0.197 | 0.199 | 0.204 | 0.191 |
| | | (0.021) | (0.020) | (0.021) | (0.022) | (0.036) |
| United States | | 0.086 | 0.153 | 0.154 | 0.166 | 0.147 |
| | | (0.019) | (0.021) | (0.022) | (0.027) | (0.047) |
| Linear trend | | | | | 0.0007 | |
| | | | | | (0.0010) | |
| Year fixed effects | no | no | no | no | no | yes |
| Adjusted R-Squared | -0.009 | 0.912 | 0.933 | 0.932 | 0.933 | 0.928 |
| Number of Observations | 81 | 81 | 81 | 81 | 81 | 81 |

Notes: Unit of observation is a country-year. Excluded country dummy is Australia.

In column (3) of Table 2.2 we include the top one percent income share variable. The coefficient on the response rate drops, but remains statistically significant and positive. The coefficient on the top income share is -1.006, which suggests that a one percentage point increase in the top income share is associated with a 1.006 percentage point decrease in the coverage rate, all else equal. To interpret these magnitudes differently, consider that the top one percent share in the US increased by 4.69 percentage points from 1990 to 2008. Over that same period, coverage dropped by 6.85 percentage points. The -1.006 coefficient means that the decline in top income share predicts a (1.006*4.69) 4.72 point drop in coverage, which is 68.9 percent of the 6.85 point drop.

Column (4) includes an interaction of the response rate and the top income share. This change leads to negative (but insignificant) coefficients on the top one percent share term and on the interaction term. The large standard errors on both estimated effects indicate that the interaction term is not well identified from the linear effect of the top one percent share on coverage rates. Indeed, a joint test for significance of these two variables shows they are highly jointly significant.[20] Further, the magnitude of the partial effect of an increase in the top one percent share evaluated at the mean US survey response rate of 0.81 based on the estimates in column (4) is numerically very similar to the linear partial effect of -1.006 in column (3). Together, the insignificance of the interaction terms and the comparable estimated partial of income inequality on coverage rates with the two specifications indicate no evidence that the effects seen in column (3) were driven by an interaction of the two factors.

In the last two columns of Table 2.2 we try alternative controls for time. Column (5) has a linear time trend. This time trend accounts for any global trend that is common to the four countries in our study. The coefficient on the response rate changes slightly to 0.345, while for the top income share the coefficient jumps up to a larger (in absolute value) magnitude. Finally, the last column includes dummies

[20]The calculated F-statistic for the interaction of the top one percent variable and the response rate and the one percent variable itself is 12.19. For both main effects and the interaction, the calculated F-statistic is 50.83. For the response rate and the interaction the joint test yields an F-statistic of 4.69. All of these are highly significant.

for each year of the sample, which controls flexibly for any common calendar time effects across countries. This is a fairly demanding specification given the number of observations we have and the nature of the variation we are using. Since there are only four observations per year, it may be difficult to detect any effect in this specification. The resulting coefficients remain fairly stable—but both lose statistical significance in this final specification.

This graphical and regression analysis shows that the trend downward in response rates is common across all four countries, and that the decline in expenditure coverage in surveys compared to national accounts has a positive relationship with changes in survey response rates. Top Income shares are also shown to be negatively related to coverage rates. Taken together, our results suggest that declining survey response rates and increasing income inequality may prove to be important determinants of the decline in expenditure coverage rates.

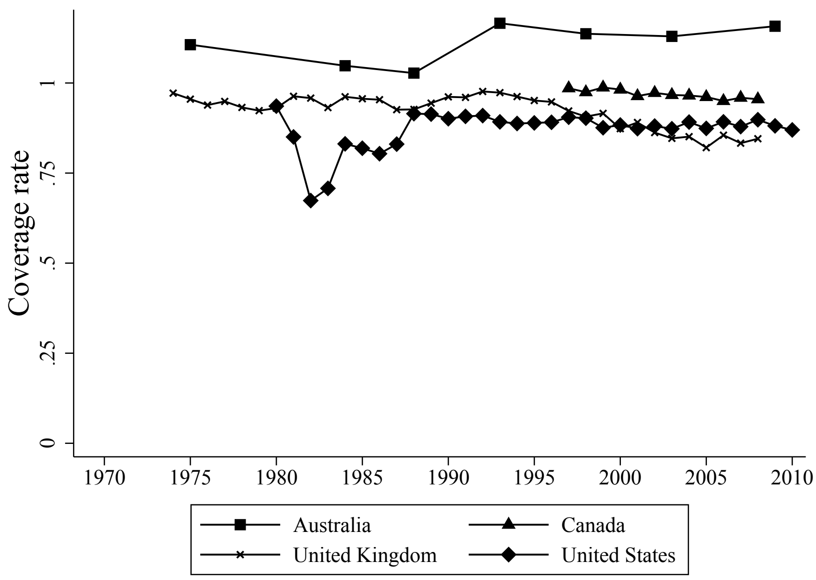
## 2.9 Coverage rates within expenditure categories

The next step in our analysis is to compare different categories of expenditure across countries, looking for evidence that aligns with differences in survey methodology. Canada here is the most noticeable outlier in survey methodology, as the SHS uses an annual recall survey for both frequently purchased and infrequently purchased items - with no diary component. There is also a balance edit, and substantial income weighting. The four categories we consider are food at home, alcohol purchased in stores, new and used motor vehicles, and furniture appliance and household equipment.

The first category we examine is food consumed in the home. These data are collected through a diary in Australia, the UK, and the US, but with recall in Canada. Food for consumption at home is a basic non-durable commodity that has been used as a summary measure of household welfare, and has been the focus of many studies testing predictions of consumption smoothing at the household and aggregate level. We graph the coverage rates in Figure 2.7.[21] The UK shows

[21] For Canada, we now show only the SHS results, as the category-by-category analysis tends to exhibit seams between the FAMEX and SHS survey years.

**Figure 2.7:** Coverage rates, food in the home




Notes: Coverage rate is the proportion of consumer expenditure in the national accounts that is accounted for in the household surveys in this category. Source is calculations by the authors.

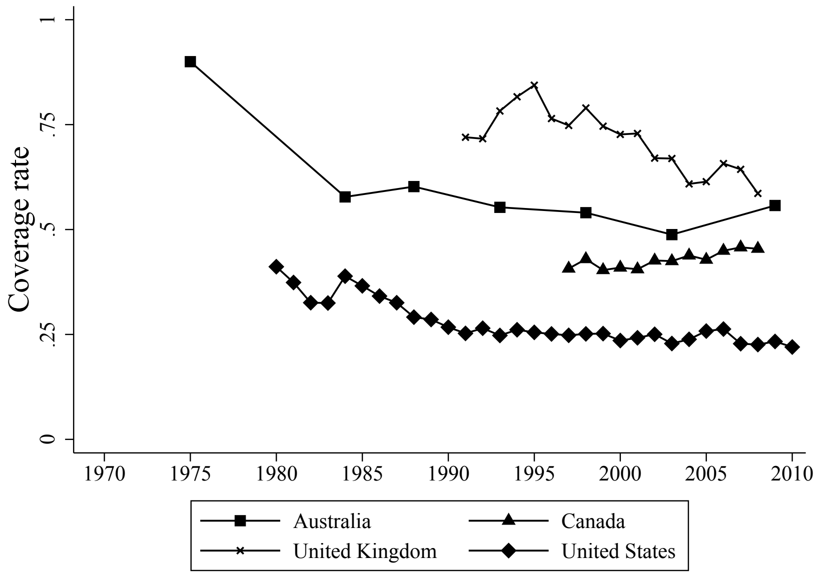
a decline of 10 percentage points since the early 1990s. However, there is little evidence of a similar trend in the other three countries.

The second expenditure category considered is alcohol purchased in stores. This category is collected using the same methods as for food consumed at home. This category is of interest because alcohol consumption is generally viewed as socially undesirable which may lead individuals to underreport these expenditures in household survey. As Figure 2.8 shows, it is the case that survey coverage of this item is very low - being around 50% for Australia and Canada, and substantially less for the US.[22] However, conditional on the lower level of coverage, the coverage rates for this item are remarkably stable in each of these three countries. For the UK, the coverage rate is higher and has declined through time.

The final two graphs show more infrequently purchased items. In all countries, these data are collected with recall surveys. Figure 2.9 shows new and used vehicles, while Figure 2.10 has household equipment, furniture and appliances. There are

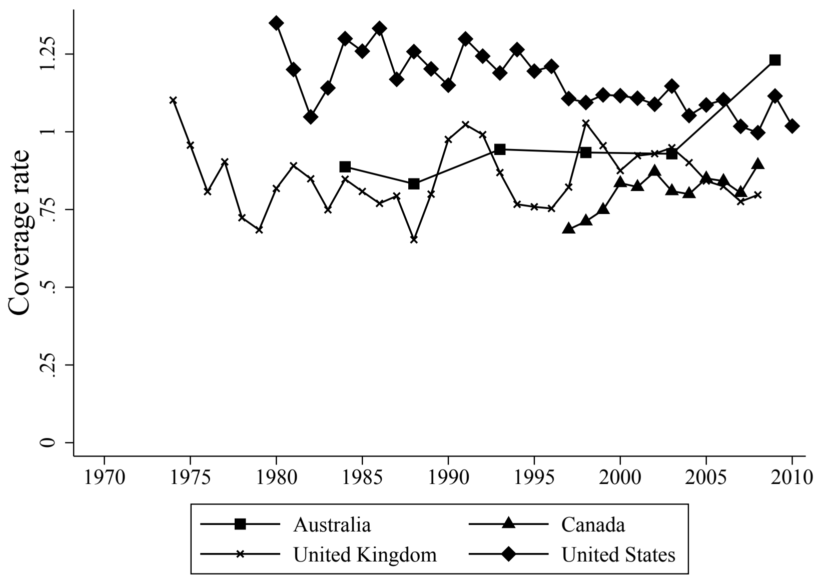
[22] Apart from atypically high coverage in the Australian HES in 1975-76.

**Figure 2.8:** Coverage rates, alcohol purchased in stores



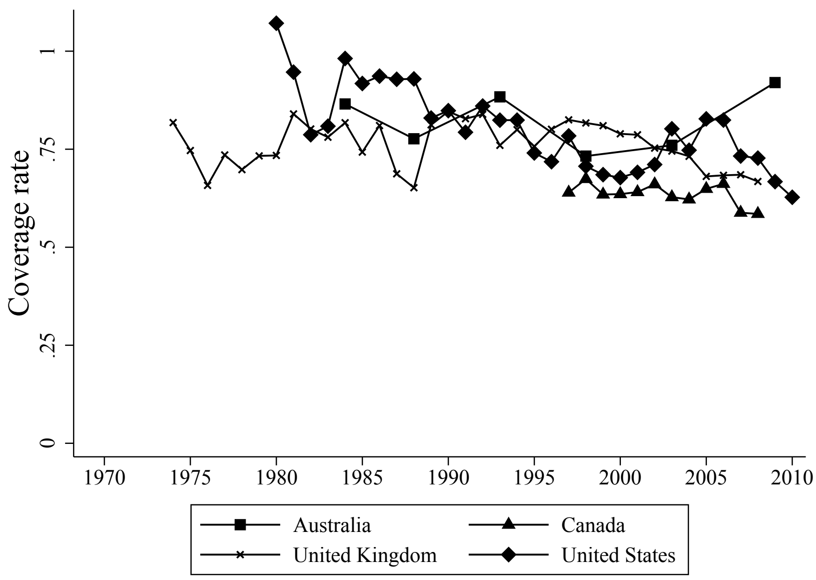

Notes: Coverage rate is the proportion of consumer expenditure in the national accounts that is accounted for in the household surveys in this category. Source is calculations by the authors.

**Figure 2.9:** Coverage rates, new and used vehicles




Notes: Coverage rate is the proportion of consumer expenditure in the national accounts that is accounted for in the household surveys in this category. Source is calculations by the authors.

**Figure 2.10:** Coverage rates, furniture, household equipment, and appliances




Notes: Coverage rate is the proportion of consumer expenditure in the national accounts that is accounted for in the household surveys in this category. Source is calculations by the authors.

no easily discernible patterns for vehicles. For Australia, coverage rates for these categories are neither consistently rising nor falling. In Canada, there is an upward trend for new and used vehicles, and perhaps small downward movement for the other two. For the UK, coverage of vehicles appears quite cyclical, but do not show a long term decline. The series for furniture, household equipment, and appliances shows a fairly slow and steady decline, although Australia does rebound at the end.

This examination of category-by-category patterns has revealed little clear evidence about differences across countries. In all countries, the frequently consumed product (food) seems to have high coverage. In contrast, the less frequent bigger purchases appear to be much more volatile year to year, and have a more pronounced downward trend on average. This is consistent with the evidence shown previously in Meyer and Sullivan (2009) and elsewhere for the US. The income elasticity of demand for the goods likely plays a role as well. As income concentration increases, coverage rates for goods consumed more by higher income households may decline.

## 2.10 Conclusions

In this paper we provide a comparative assessment of the performance of the household expenditure survey programs in Australia, Canada, the UK and US. The survey methodologies employed in each country share a number of common features while containing distinct elements. There are also differences in survey response rates and income concentration across the countries. We use this variation across countries to assess the implications for the performance of the household surveys.

After first outlining the key features of the household expenditure surveys for each country, we assess the coverage of aggregate expenditure relative to the national account benchmark. Both the survey expenditure aggregate and the national accounts data are adjusted to ensure the expenditure concepts are comparable. Coverage rates are highest in Canada and the UK; for Canada, and Australia coverage remained fairly stable over the past three decades. In contrast, in the UK and the US coverage rates have sharply declined over the past three decades.

Next, survey response rates and top income shares were considered in tandem with coverage rates. This analysis is motivated by the widely observed decline in participation rates for household surveys over time, and the strong concentration of income that has occurred in many countries. From a series of graphical comparisons and regression models it is found that the fall in response rates over time are predictive of changes in coverage rates within countries. Further, the pattern of changes in coverage rates over time within our sample of Anglosphere countries coincided with the timing of the growing concentration of income. The prima facie evidence is that the growing concentration of income has been associated with an increasing concentration of expenditures which has not been captured well by the micro surveys, hence contributing to declining coverage.

The last component of the analysis examined coverage rates for specific components of expenditure. Individual expenditure items considered were food at home, alcohol purchased in stores, new and used motor vehicles, and furniture appliance and household equipment. This list included categories which were collected using divergent methodologies (e.g. food by diary in Australia, UK and US; by annual recall in Canada) and by comparable methods (e.g. motor vehicles, furniture and recreational equipment collected by recall in interviews in all countries). From this, there was no clear pattern across countries by collection method. Rather, most evident is the high and stable coverage of regularly purchased items (such as food), along with the more volatile coverage of irregular and larger expenditure items (such as vehicles, furniture and hold equipment). Therefore the aggregate patterns in coverage cannot be readily attributed to specific expenditure components or collection methods.

Overall, our comparative assessment of the household expenditure surveys across the four Anglosphere countries studied has shown the sharpest differences between Canada and Australia vs. the United Kingdom and the United States. However, the many unique aspects of the Canadian survey methodology make it difficult to identify specific features of the methodology that are pivotal to its performance. Given the Canadian methodology changes that were put in place for 2010, further information may now be available on the reasons for the relative success of the Canadian data.

# Chapter 3

## Is the Carli Index flawed?: assessing the case for the new retail price index RPIJ

In March 2013, the UK's Office for National Statistics (ONS) started publishing a new inflation index - the RPIJ. This index is identical to the long-standing Retail Prices Index (RPI), except that it uses a geometric mean of price relatives (known in inflation circles as the *Jevons* index) rather than an arithmetic mean (the *Carli* index) to calculate price changes of goods at the so-called 'elementary' level - where expenditures on individual goods are not observed and so only price survey data are used. The Jevons index has long been used in the UK's other measure of consumer price inflation, the Consumer Prices Index (CPI). Around the same time as the RPIJ was introduced, the United Kingdom Statistics Authority, UKSA, decided that concerns the ONS had raised about a potential for upward bias in the Carli index meant that the old RPI would no longer be recognised as a National Statistic.[1] The old RPI is still published but it is now clearly marked with a health warning in the ONS's inflation publications.

These decisions have been quite controversial. Over the period 1998-2013, the RPI gave an average inflation rate of 2.9% compared to 2.5% that would have been given by the RPIJ. The use of the Jevons in the CPI, and the fact that this tended to mean it gave a lower measure of inflation than the RPI, has already generated some amount of mistrust in official numbers - particularly as in the last few years the

[1] To be classified as a 'National Statistic' published numbers must meet certain standards set out on the Code of Practice for Official Statistics.

government has replaced the RPI with the CPI for the indexation of state benefits, government pensions, and tax thresholds (all measures which are predicted to save the government money). This chapter will look at the original reasons for replacing the Carli index. It seems clear that this index has fallen out of favour with national statisticians in the UK, but what are the concerns with this particular index and how should we select index numbers at the elementary level more generally? Did the ONS and UKSA make the right decision? Answering these questions is important not just to decide whether a new RPIJ should have been created in the first place, but also for the more current debate about the whether the existing RPI should be considered "deficient" and demoted in the way it has been. This will require us to delve into the theoretical and practical reasons for preferring one index over another.

The literature on price indices has identified three ways to select index numbers. In effect, we can ask:

1. Does the index respond appropriately when prices change in different situations or does it give answers we might consider perverse?

2. Is the index a good *statistical* estimator of the general price change as distinct from relative price changes across goods according to some measure (as we will discuss below, exactly which measure is a matter of some debate)?

3. Does the index provide a good measure of how the *cost of living* is changing for consumers (i.e the costs of obtaining a given level of welfare) ?

The first of these is called the *test* approach. This is because we determine which index has the best properties by setting out a list of criteria ('tests') and then asking which indices satisfy them. The second is called the *statistical* or stochastic approach. The third is the interpretation of price indices that is used by most economists and as such is referred to as the *economic* approach. The three approaches are essentially separate and can on occasion come to conflicting conclusions. As part of a consultation on the future of the RPI which led to the creation of the new RPIJ index, Diewert (2012b) pointed out that the Carli index failed some important axioms of the test approach, that the statistical approach favoured the

Jevons, and that the economic approach was inapplicable at the elementary level (when quantity weights were unobserved). These conclusions were essentially endorsed by the ONS and underlay the decision to replace the Carli index in the new RPIJ.

In this chapter, I explain what each of these approaches are and use them to make my own assessment on the suitability of the Carli index for use at the elementary aggregate level of a price index. In doing so I make a number of contributions not only to the current debate on the new RPIJ index but also to the way that elementary indices should be selected more generally. A primary concern of the ONS was the Carli's sensitivity to so-called price-bouncing which could lead to an upward bias. I formalise these concerns in a new price-bouncing test for the test approach. For the statistical approach, I present some evidence on the relative performance of the Carli and Jevons. I find no clear evidence for the superiority of one index over the other, and that the relative performances of the Carli and Jevons are not invariant to factors such as the month the index is calculated; the sample size and the choice of base month against which price changes can be compared; and the type of goods included in the elementary aggregate. I also argue that the economic approach *can* be applied to the elementary level, and moreover that it favours the Jevons index, by appealing to something analogous to the principle of insufficient reason from information theory. Overall, I conclude that there is indeed a case for replacing the Carli index with the Jevons.

The remainder of the chapter is structured as follows. In Section 3.1, I discuss the historical and technical background to the 'formula effect' difference between the RPI and CPI in the UK. In Sections 3.2, 3.3 and 3.4 I discuss the test, statistical and economic approaches. Section 3.5 concludes.

## 3.1 The RPI and the CPI

The UK has for a long time been blessed with two headline measures of consumer price inflation - the RPI and the CPI. The RPI is the older of the two, dating back to an 'Interim index' that was introduced in June 1947 based on an expenditure survey carried out in 1937/38, and for most of its history was the UK's principal measure of consumer prices. The CPI is the UK's version of a Harmonised Index of Consumer Prices (HICP), which was developed by the European Union to ensure that member states published comparable measures of inflation.

These two indices differ in a number of ways which mean they can give quite different measures of price changes from year to year. For instance, in 2011 CPI growth averaged 4.5% compared to 5.2% for RPI growth. The differences are a result of the data they draw from, the coverage of the indices, and the methods used to calculate average price changes at the so-called elementary level. In recent years, the CPI replaced the RPI for a number of policy purposes, including the uprating of state benefits and pensions and the indexation of tax thresholds. Consequently, the large gap between the two measures, and particularly the factors that mean that the CPI tends to give a lower measure of inflation than the RPI, came increasingly under scrutiny.

Since 2010, the largest factor contributing to the gap between the RPI and CPI has been the differences in the way price changes are calculated at the lowest level of aggregation in the two indices (which became known as the 'formula effect'). In October 2012, questions about the formula effect culminated in the opening of a consultation on changes to the methods used in the RPI. In January 2013, the consultation concluded that the Carli index "did not meet international standards" (Office for National Statistics (2012b)) and that consequently a new RPIJ index would be published in which the Carli was replaced with the Jevons.

### 3.1.1 The formula effect

The formula effect is the difference between the RPI and CPI that results from the different indices they use at the first stage of aggregation. Aggregation refers to the process by which an overall index such as the RPI or CPI is calculated in successive stages. The calculation of both the RPI and CPI starts with (essentially the same) sample of prices collected across the country in each month. [2] This

[2] There are a few differences. The CPI uses a different approach to gathering car prices to the RPI.

**Table 3.1:** Examples of goods at different levels of aggregation in the UK

| Level | Price |
|---|---|
| Elementary aggregate | 800g white unsliced bread sold in the south east of England |
| Item | 800g white unsliced bread |
| Section/class | Bread |
| Group | Food |

sample is then used to produce weighted averages of price changes relative to a base month (in the UK, January, for details, see section 2 of Office for National Statistics (2012a)). In the very first stage, where the ONS does not have expenditure information, an unweighted average of price changes for particular products is taken within different 'strata', defined by either region, type of shop (independent or chain retailer), or both. These give what are known as *elementary aggregate* indices. An expenditure-weighted average of these elementary aggregates is then taken to give an overall national average price index for an 'item'. These different item indices are then aggregated further through expenditure-weighted averages into 'sections' or 'classes', which are in turn aggregated into 'groups'. Finally, an overall price index is calculated from the different group indices. Some examples of the 'goods' at each stage of aggregation are given in table 3.1.

There are various indices which can be used to calculate elementary aggregate price changes. These include

1. The Carli index (Carli, 1764): $P_C(\mathbf{p}_0, \mathbf{p}_1) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{p_1^i}{p_0^i}\right)$
2. The Dutot index (Dutot,1738): $P_D(\mathbf{p}_0, \mathbf{p}_1) = \frac{\frac{1}{N}\sum p_1^i}{\frac{1}{N}\sum p_0^i}$
3. The Jevons index (Jevons, 1865): $P_J(\mathbf{p}_0, \mathbf{p}_1) = \prod_{i=1}^{N}\left(\frac{p_1^i}{p_0^i}\right)^{\frac{1}{N}} = \frac{\prod_{i=1}^{N}(p_1^i)^{\frac{1}{N}}}{\prod_{i=1}^{N}(p_0^i)^{\frac{1}{N}}}$
4. The Harmonic index (Coggesshall, 1887):[3] $P_H(\mathbf{p}_0, \mathbf{p}_1) = \left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{p_1^i}{p_0^i}\right)^{-1}\right]^{-1}$
5. The Carruthers-Sellwood-Ward-Dalèn index (CSWD, proposed as an elementary index by Carruthers et al. (1980), and also Dalén (1992)): $P_{CSWD}(\mathbf{p}_0, \mathbf{p}_1) = \sqrt{P_C(\mathbf{p}_0, \mathbf{p}_1) \times P_H(\mathbf{p}_0, \mathbf{p}_1)}$

[3]Diewert (2012a) points out that it was also mentioned earlier in passing in Jevons (1865).

The Carli is an arithmetic mean of price changes (or price relatives), while the Jevons is a geometric mean. The Dutot is the ratio of average prices in the base year and the current year. The Harmonic index is simply the harmonic mean of price relatives. The CSWD index is a geometric mean of the Carli (arithmetic) and the Harmonic indices.

The RPI uses the two arithmetic averages: the Carli and the Dutot indices. The CPI by contrast makes use of the Dutot and the Jevons. This puts the CPI closer into line with international practice. Indeed, the RPI's use of the Carli is quite unusual. None of the other 27 European countries that reported a Harmonised Index of Consumer Prices (HICP) to Eurostat surveyed in Evans (2012) made use of the Carli index in their national price indices. Indeed there seems to be have been a general move away from the Carli index. Evans (2012) lists some countries that have abandoned the Carli index in favour of either the Jevons or the Dutot over the last few decades including Canada (in 1978), Luxembourg (in 1996), Australia (in 1998), Italy (in 1999) and Switzerland (in 2000). In 1996, the Boskin Commission in the USA recommended that a Carli-like index used in the US CPI should be replaced with a the Jevons (Boskin, 1996) - a change that was put into effect in 1999. Eurostat regulations also do not allow the use of the Carli in the construction of members' HICP indices except in "exceptional cases" (see Section 3, pg 180 of Eurostat, 2001).

The proportions of elementary aggregates that use of each of these formulae in the RPI and CPI are shown in Table 3.2. For the remaining 'other' goods in the table, no elementary aggregates are calculated and weights are used at every stage in the calculation of prices. The new RPIJ index uses the Jevons in place of the Carli but continues to use the Dutot for the same goods as the old RPI. The reason for the even split between the Carli and Dutot in the old RPI is that both indices can be distorted in particular situations. The Carli can be too sensitive to situations where individual goods see large price changes (such as when a sale for some items ends). The Dutot on the other hand can be dominated by the price movements of a single good, if that good is much more expensive than others included in the calculation

**Table 3.2:** Importance of different formulae used in the RPI and CPI

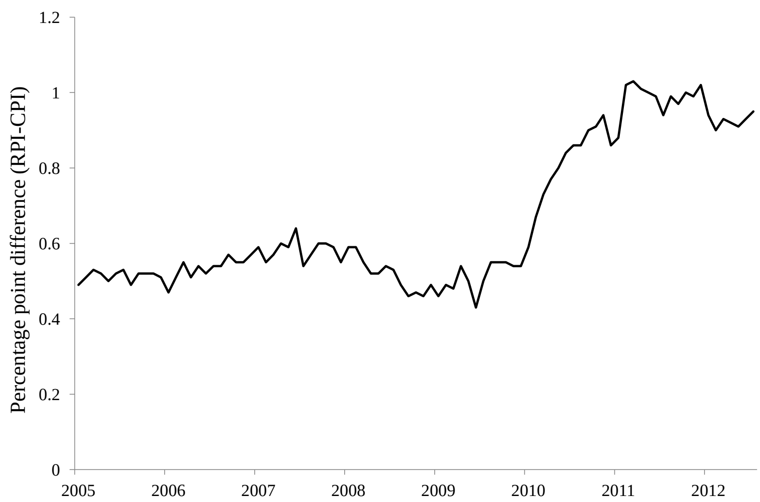
| Index | RPI | CPI |
|---|---|---|
| Carli | 27% | 0% |
| Dutot | 29% | 5% |
| Jevons | 0% | 63% |
| Other (weighted) formula | 43% | 33% |

Note: Source is ONS (2012c)

(see Section 9.3 of Office for National Statistics (2012c)).

What impact do these differences have in practice? Figure 3.1 shows the formula effect over time from 2005-2012. It shows that the formula effect consistently works to substantially reduce the growth of the CPI relative to the RPI. The effect averaged 0.5 percentage points over the years 2005 -2009, which increased to 0.9 since 2010 (for comparison, the average annual increase in the RPI over the same period was 3.4%). The sudden increase in the formula effect can be almost entirely attributed to a change in the sampling of clothing prices that came into effect in that year (Morgan and Gooding (2010)).

**Figure 3.1:** The size of the formula effect, 2005-2012




Note: Data from Office for National Statistics.

This systematic difference is primarily driven by the fact that the Carli and the

Jevons - being the geometric mean and the arithmetic mean of the price relatives - satisfy the classic inequality:

$$P_J(\mathbf{p}_0, \mathbf{p}_1) \leq P_C(\mathbf{p}_0, \mathbf{p}_1) \tag{3.1}$$

with equality if and only if $\frac{p_1^i}{p_0^i} = \pi$ for all $i$ (Hardy et al. (1934), p.26): that is the Jevons will *always* give either the same or a lower price increase than the Carli. It is not possible to establish a similar general result for the relationship between the Dutot and the other indices (and thus the formula effect needn't necessarily always be positive). Depending on the circumstances the Dutot could be greater or less than the Carli and greater or less than the Jevons. To understand more precisely what drives the formula effect we will need to delve a littler deeper into the mathematical relationships between the different indices. I do this by presenting the following three facts which will be useful when we turn to evaluating the different indices.[4] Proofs can be found in the supplied references and are also provided in the Appendix to this chapter.

**Fact 1.** *The difference between the Carli index and the Jevons index is bounded from below by the variance of the price-relatives (proof given in Hardy et al. (1934)):*

$$P_C(\mathbf{p}_0, \mathbf{p}_1) - P_J(\mathbf{p}_0, \mathbf{p}_1) \geq Var\left(\frac{p_1^i}{p_0^i}\right)$$

This fact helps to explain the growth in the size of the formula effect in Figure 3.1. In 2010, a change to the methods used to sample clothing prices led to an increase in the variance of price relatives, which is what led to an increase in the difference between the Carli used for clothing prices in the RPI and the Jevons used in the CPI.

**Fact 2.** *The difference between the Dutot and the Carli equals the covariance of base period prices and price relatives divided by the mean base period price (proof given in Carruthers et al. (1980)):*

[4]Some useful results on the relationships between these and the other indices discussed above using a theorem from Ladislaus von Bortkiewicz can be found in von der Lippe (2012).

$$P_D(\mathbf{p}_0, \mathbf{p}_1) - P_C(\mathbf{p}_0, \mathbf{p}_1) = \frac{Cov\left(p_0^i, \left(\frac{p_1^i}{p_0^i}\right)\right)}{E[p_0^i]} \tag{3.2}$$

Finally, by writing the price of the $i$th good in the $t$th period as a multiplicative deviation from its expected value $e_t^i$ where $E\left(e_t^i\right) = 0$

$$p_t^i = E\left(p_t^i\right)\left(1 + e_t^i\right) \tag{3.3}$$

we obtain the following useful approximation between the Jevons and Dutot

**Fact 3.** *The difference between the Jevons and the Dutot depends on the change in the variance of the prices (a result first obtained by Carruthers et al. (1980), proof also given in Diewert (2012b)):*

$$P_J(\mathbf{p}_0, \mathbf{p}_1) \approx P_D(\mathbf{p}_0, \mathbf{p}_1)\left(1 + \frac{1}{2}\left[Var(e_0^i) - Var(e_1^i)\right]\right)$$

The preceding discussion indicates that the choice of elementary index clearly matters a great deal. This makes it all the more important that in any inflation measure the most appropriate indices are chosen - the question to which we now turn. Since there is no definitive set of criteria that we can use to pick one index over another, we try to form judgements applying each of the three approaches used to select elementary indices in the literature (test, statistical and economic). We start with the test approach.

## 3.2 The test approach

This approach posits a number of desirable properties for index numbers. These form tests (or 'axioms') against which alternative index number formulae can be ranked - with index number formulae which satisfy the most, or the most important, axioms being ranked highest. This approach does not consider any behavioural interdependence between the price and quantity data unlike the economic approach which I discuss below. The test approach has its roots in the mathematical literature on functional equations, the general problem being that of determining an unknown functional form (i.e. what is the functional form for the price index?) given a set of requirements on the function. The properties are selected to be reasonable given the context.

When we have data on prices and quantities from two periods $t$ in $\{0,1\}$ the problem is to determine the forms of the price index linking the two period $P(\mathbf{p}_0,\mathbf{p}_1,\mathbf{q}_0,\mathbf{q}_1)$ and the corresponding quantity index $Q(\mathbf{p}_0,\mathbf{p}_1,\mathbf{q}_0,\mathbf{q}_1)$ such that nominal growth rate is (multiplicatively) decomposable in that part reflecting price changes and that part reflecting real changes:

$$P(\mathbf{p}_0,\mathbf{p}_1,\mathbf{q}_0,\mathbf{q}_1)\,Q(\mathbf{p}_0,\mathbf{p}_1,\mathbf{q}_0,\mathbf{q}_1) = \frac{x_1}{x_0}.$$

This decomposition property is sometimes called *weak factor reversal* and often isn't counted as a 'test' but as a defining property of bilateral index numbers. If this holds and neither are zero then once we have chosen one index number, the other is chosen implicitly. For example, given a price index we can recover the quantity index implicitly:

$$Q(\mathbf{p}_0,\mathbf{p}_1,\mathbf{q}_0,\mathbf{q}_1) = \frac{x_1}{x_0}\frac{1}{P(\mathbf{p}_0,\mathbf{p}_1,\mathbf{q}_0,\mathbf{q}_1)}$$

In the case of elementary price aggregates, quantity weights are not observed. Thus the bilateral index number problem is restated slightly as that of finding a price index $P(\mathbf{p}_0,\mathbf{p}_1)$ (and implicitly a quantity index $Q(\mathbf{q}_0,\mathbf{q}_1) = \frac{x_1/x_0}{P(\mathbf{p}_0,\mathbf{p}_1)}$), which satisfies certain tests, such that

$$P(\mathbf{p}_0,\mathbf{p}_1)\,Q(\mathbf{q}_0,\mathbf{q}_1) = \frac{x_1}{x_0}$$

The tests themselves have been developed over the course of well over a century mainly for the case in which prices*and* quantities are observed (for an authoritative discussion of these, see Section 2 of Diewert (1992b)). In most cases the tests relating to the price index do not depend on the quantity vectors and so will have obvious analogs for the elementary aggregates case where quantities are not known.[5] I divide the tests into 4 groups: those that simply establish basic properties for an index, those that consider the effects of scalar transformations to prices, a test that bounds the price index, and some tests that establish invariance properties. In the rest of this section, I will discuss each of these groups in turn. Throughout I assume that $\mathbf{p}_t \in \mathbb{R}^K_{++}$. If I want to set $\mathbf{p}_1 = \mathbf{p}_0$ I call the common vector $\mathbf{p}$.

### 3.2.1 Basic properties

The first set of tests establish some basic properties which we would expect any price index to have. The first of these is the *positivity test*: we want our price index to be positive for any set of prices, $P(\mathbf{p}_0, \mathbf{p}_1) > 0$ (Diewert (2012c) attributes this to Eichhorn and Voeller (1976)). A non-positive price index would cause all sorts of problems (for instance with chaining). A second basic property is that if no prices change between two periods, we would expect our price index to simply equal one, i.e $P(\mathbf{p}, \mathbf{p}) = 1$. This is the *identity test* which is sometimes called the constant prices test (Diewert (2012c), points out that this test was suggested by Laspeyres (1981), Walsh (1901), and Eichhorn and Voeller (1976)). Finally there are two further tests which establish that certain changes to the prices we feed into the index should always result in a greater or smaller index. The first of these is the *monotonicity in current prices test* which states that if we increase one of our current prices, the index as a whole should be greater than it was. That is $P(\mathbf{p}_0, \mathbf{p}_1) < P(\mathbf{p}_0, \mathbf{p})$ *if* $\mathbf{p}_1 < \mathbf{p}$. Similarly, if we increased any base period price, then the index as a whole should decrease $P(\mathbf{p}_0, \mathbf{p}_1) > P(\mathbf{p}, \mathbf{p}_1)$ if $\mathbf{p}_0 < \mathbf{p}$ which gives us the *monotonicity in base prices test*.[6] These last two tests are due to Eichhorn and Voeller (1976) (who actually include these two properties in a single 'monotonicity' axiom). Fortunately, all the elementary indices mentioned in Section 3.1 satisfy all four of these tests.

[5] This is not always true. There is, for example, no obvious parallel to the *Tabular Standard/Basket/Constant Quantities Test* $P(\mathbf{p}_0, \mathbf{p}_1, \mathbf{q}, \mathbf{q}) = \mathbf{p}_1'\mathbf{q}/\mathbf{p}_0'\mathbf{q}$ or the *Invariance to Proportional Changes in Current Quantities Test* $P(\mathbf{p}_0, \mathbf{p}_1, \mathbf{q}_0, \lambda\mathbf{q}_1) = P(\mathbf{p}_0, \mathbf{p}_1, \mathbf{q}_0, \mathbf{q}_1)$ for $\lambda > 0$ in the context of elementary aggregates. The approach of adapting those tests which are independent of quantities to use for elementary indices was used by Diewert (1995a), who drew on the work of Dalén (1992) and Eichhorn (1978).

[6] Here I adopt the notation of Eichhorn and Voeller (1976) for vector inequalities. If $x = (x_1, x_2 \ldots x_n)$ and $y = (y_1, y_2 \ldots y_n)$, then $x \geq y$ if $x_1 \geq y_1, \ldots x_n \geq y_n$ but $x \neq y$. In other words, all elements of $x$ are equal or greater than those in $y$, with at least one strictly greater.

### 3.2.2 Scalar transformations

This next few tests consider the effects of scalar transformation of the price data. The first two of these demand that the price index should be homogenous of degree 1 with respect to current prices and homogenous of degree -1 with respect to base prices, i.e $P(\mathbf{p}_0, \lambda\mathbf{p}_1) = \lambda P(\mathbf{p}_0, \mathbf{p}_1)$ for $\lambda > 0$ and $P(\lambda\mathbf{p}_0, \mathbf{p}_1) = \lambda^{-1}P(\mathbf{p}_0, \mathbf{p}_1)$ for $\lambda > 0$. These two requirements are called the *linear homogeneity* and *homogeneity of degree minus one* tests (Eichhorn and Voeller (1976)). These two tests further imply the *dimensionality* test (Eichhorn and Voeller (1976)) which states that the index should not be affected by common changes in scaling to all prices, i.e $P(\lambda\mathbf{p}_0, \lambda\mathbf{p}_1) = P(\mathbf{p}_0, \mathbf{p}_1)$. This means among other things that calculating price changes using pence rather than pounds should make no difference to our indices. The identity and linear homogeneity tests together imply the *proportionality* test (Eichhorn and Voeller (1976)) which states that increasing all prices by a common positive scalar $\lambda$ should give a price index equal to $\lambda$: or $P(\mathbf{p}_0, \lambda\mathbf{p}_0) = \lambda$ for $\lambda > 0$. All our indices satisfy these requirements.

### 3.2.3 Bound on indices

The next test states that the index itself should fall somewhere in between the price relative of the good with the smallest price increase and the price relative of the good with the largest price increase, or

$$\min_k \left\{ \frac{p_1^1}{p_0^1}, \ldots, \frac{p_1^K}{p_0^K} \right\} \leq P(\mathbf{p}_0, \mathbf{p}_1) \leq \max_k \left\{ \frac{p_1^1}{p_0^1}, \ldots, \frac{p_1^K}{p_0^K} \right\}$$

This is the *mean value test* (Eichhorn and Voeller (1976)). Whilst this last test is a fairly intuitive requirement it can be shown that it is implied by monotonicity in current and base prices, linear homogeneity, and identity tests (Eichhorn and Voeller (1976), pg 10). Since all our indices satisfy these weaker axioms, they all satisfy this test.

### 3.2.4 Invariance properties

The final group of tests are concerned with invariance properties of various kinds. This group of tests will help us discriminate between our three elementary indices and so I will discuss them in a bit more detail.

The first states that the index should be invariant to the ordering of goods. This implies for instance that if we took price quotes from the same outlets in a different order (but still keeping the order the same in base and current periods), then this would have no effect on the index.

*Commodity Reversal Test/Symmetric treatment of outlets:* rearranging the order of the components of both current and base period price vectors in the same way should have no effect on the index. That is,

$$P\left(\mathbf{A}\mathbf{p}_0, \mathbf{A}\mathbf{p}_1\right) = P\left(\mathbf{p}_0, \mathbf{p}_1\right)$$

where $\mathbf{A}$ denotes some permutation matrix which we use to reorder our price vector. Diewert (1992a) attributes this test to Fisher (1922). It can easily be shown that all of our indices pass this test. The next test concerns invariance to the units in which goods are defined.

*Commensurability Test*: Multiplying prices in both periods by a vector $\lambda$ should not affect the index.

$$P\left(\lambda^1 p_0^1, \ldots \lambda^M p_0^M; \lambda^1 p_1^1, \ldots \lambda^M p_1^M\right) = P\left(p_0^1, \ldots p_0^M; p_1^1, \ldots p_1^M\right) = P\left(\mathbf{p}_0, \mathbf{p}_1\right)$$

For all $\lambda^1 > 0 \ldots \lambda^M > 0$.

Diewert (1992a) attributes this test to Fisher (1911). This test implies that, ignoring quantity discounts and the like, a change in the units defining an individual item (such as switching from a single item of fruit to a bunch of fruit) should not affect the index. The Dutot index fails this test as it is not in general invariant to changes in the units in which individual goods are sold. If we were to double the base and current period price of one particular item (by for instance, measuring the price of a pair of gloves rather than a single glove), then the Dutot index would change, while our other indices would be unaffected. This comes about because the level of the Dutot index depends on the value of base period prices relative to their mean.[7] As Diewert (2012b) points out, this means that the Dutot will not be appropriate for elementary aggregates where there is a great deal of heterogeneity and items are measured in different units, as in these situations "the price statistician can change the index simply by changing the units of measurement for some of the items." However, in most cases goods at the elementary level are typically fairly homogeneous, and so a sleight of hand involving an arbitrary change in the units in which one brand is defined while leaving the definition of other nearly identical products the same may be more easily noticed. If we trust that this is the case, we may be satisfied that the index satisfies the weaker dimensionality test referred to above (which the Dutot passes).

The next test states that if prices go up one period and return to their previous level the next, a chained index should record no price increase.

*Time Reversal Test:* if the data for the base and current periods are interchanged, then the resulting index is the reciprocal of the original

$$P(\mathbf{p}_0, \mathbf{p}_1) = \frac{1}{P(\mathbf{p}_1, \mathbf{p}_0)}$$

Diewert (1992a) attributes this to Fisher (1922). A sufficient (but not necessary) condition for an index to satisfy this property is that it can be expressed in a form $f(\mathbf{p}_t)/f(\mathbf{p}_0)$ as it is of course always true that

$$\frac{f(x)}{f(y)} = \frac{1}{f(y)/f(x)}$$

The Jevons and the Dutot can be written in this form and so pass the test. The CSWD cannot but still satisfies time reversal. The Carli and Harmonic indices on other the hand do not. In fact, if prices go up and then return to their former level the Carli will record an increase in prices (unless all prices increase in the same proportion), since it can be shown that

$$P_C(\mathbf{p}_0, \mathbf{p}_1) P_C(\mathbf{p}_1, \mathbf{p}_0) \geq 1$$

[7] The Dutot can be rewritten as $E\left[\frac{p_0^i}{E[p_0^i]}\left(\frac{p_1^i}{p_0^i}\right)\right]$ and so can be thought of as an index where price relatives are weighted by base prices.

Similarly it can be shown that

$$P_H(\mathbf{p}_0,\mathbf{p}_1)P_H(\mathbf{p}_1,\mathbf{p}_0) \leq 1$$

Thus, these indices both fail the time reversal test in a *biased* way. The use of the term bias here is due to Fisher (1922) who uses it to refer to a "foreseeable tendency [of an index] to err in a particular direction" (pg. 86). In this case it refers to the difference between the value of $P(\mathbf{p}_0,\mathbf{p}_1)P(\mathbf{p}_1,\mathbf{p}_0)$ and the 'correct' value of unity. It is important to note that it does not refer to bias in the statistical sense of the difference between the expected value of a statistic and its population value. Two indices may both be 'unbiased' yet differ considerably in terms of the price changes they record.

If we then consider price changes over more than two periods, we get the following test.

*Circularity Test:* The product of a chain of indices over successive periods should equal the total price change over the whole period.

$$P(\mathbf{p}_0,\mathbf{p}_1)P(\mathbf{p}_1,\mathbf{p}_2) = P(\mathbf{p}_0,\mathbf{p}_2)$$

This is a transitivity test.[8] If this test were not satisfied, then different inflation rates over a given period could be obtained by chaining the index over different subperiods. One consequence of this is that an index could go up or down even if prices had not changed. For instance, consider a case where prices increased from $\mathbf{p}_0$ to $\mathbf{p}_1$ between periods 0 and period 1, but in period 2 returned to $\mathbf{p}_0$. In this case, a chained index that didn't satisfy circularity could potentially record inflation over the three periods when there had in fact been none.

Eichhorn and Voeller (1976) prove that, unlike in the case of the time reversal test, it is both necessary and sufficient that the index can be written in the form $f(\mathbf{p}_t)/f(\mathbf{p}_0)$ to pass circularity. The circularity test therefore implies the time re-

[8] A related test that is sometimes included is the *multiperiod identity test* that requires that the index satisfies $P(\mathbf{p}_0,\mathbf{p}_1)P(\mathbf{p}_1,\mathbf{p}_2)P(\mathbf{p}_2,\mathbf{p}_0) = 1$ (Diewert (2012c), attributes this to Walsh (1901)). The author is grateful to an anonymous referee for demonstrating that this is in fact just an implication of the circularity test.

versal test. The Carli does not satisfy the circularity test since it is $E\left[p_1^i/p_0^i\right]$ and in general

$$E\left[\frac{x}{y}\right] \neq E\left[\frac{x}{z}\right] \times E\left[\frac{z}{y}\right]$$

The Harmonic and CSWD indices also fail this test.

A final test we could add to this list concerns so-called price 'bounces'. This is concerned with how an index would change if different outlets merely exchanged prices from one period to the next. One test of this property, attributed to Dalén (1992) is as follows

$$P\left(\mathbf{A}\mathbf{p}_0, \mathbf{B}\mathbf{p}_1\right) = P\left(\mathbf{p}_0, \mathbf{p}_1\right)$$

where **A** and **B** are *different* permutation matrices. The Jevons and Dutot pass this test, but the Carli, Harmonic and CSWD indices fail it. This test has been criticised (Diewert (2012b), for instance calls it "suspect") on the grounds that prices should be matched to outlets in a one to one manner across periods (that is, that $\mathbf{p}_0$ and $\mathbf{p}_1$ should not be permuted in different ways), for the simple reason that outlets vary by quality, and so even when the same good is bought in different places it ought to be considered a different product. This is true, but a sensitivity to price bouncing may still be a problem, as it is possible for an index to register a price increase if outlets exchanged prices with one another but then swapped *back* - a property which is somewhat harder to justify. Indeed, the problem of the Carli's sensitivity to price bouncing was highlighted by the ONS in its consultation (Office for National Statistics (2012c)). This concern suggests the following revised price bouncing test:

*Price bouncing test:* The price index should not change over three periods if prices are just rearranged from the first to the second period and then returned to their original order in the third

$$P\left(\mathbf{p}_0, \mathbf{A}\mathbf{p_0}\right) P\left(\mathbf{A}\mathbf{p_0}, \mathbf{p}_0\right) = 1$$

for any possible permutation matrix $\mathbf{A}$.

This test will be satisfied by any index that satisfies a stronger property $P(\mathbf{p}_0, \mathbf{Ap_0}) = 1$ (which we may call the strong price bouncing test). This could itself be introduced as a separate test, but by testing how an index would respond to a change which doesn't match outlet prices across periods, it would be subject to the same criticism as Dalèn's price bouncing test. This property is similar to the time reversal test, and indeed it will be satisfied by any index that satisfies time reversal. The two properties are however independent as an index may satisfy the price bouncing test but not time reversal. For instance it can be shown that the index

$$\frac{\Sigma_i \Sigma_j \left(\frac{p_{1i}}{p_{0j}}\right) + \left[\Sigma_i \Sigma_j \left(\frac{p_{1i}}{p_{0j}}\right)^{-1}\right]^{-1}}{\Sigma_i \Sigma_j \left(\frac{p_{0i}}{p_{0j}}\right) + \left[\Sigma_i \Sigma_j \left(\frac{p_{0i}}{p_{0j}}\right)^{-1}\right]^{-1}} \tag{3.4}$$

satisfies the price bouncing test, as well as positivity, identity, linear homogeneity, homogeneity of degree minus one, monotonicity in current and base prices, proportionality, dimensionality, and commodity reversal - but not time reversal or circularity.

The price bouncing test is also independent of its stronger version. Again, we can demonstrate this with a example. Consider the following index, which combines two geometric means of price relatives, weighted by base and current prices

$$P(\mathbf{p_0}, \mathbf{p_1}) = \sqrt{\prod_i^N \left(\frac{p_1^i}{p_0^i}\right)^{\frac{p_0^i}{\Sigma p_0^i}} \times \prod_i^N \left(\frac{p_1^i}{p_0^i}\right)^{\frac{p_1^i}{\Sigma p_1^i}}} = \prod_i^N \left(\frac{p_1^i}{p_0^i}\right)^{\frac{p_0^i \Sigma_j p_1^j + p_1^i \Sigma_j p_0^j}{2 \Sigma_j p_0^j \Sigma_j p_1^j}}$$

This index satisfies time reversal and so the weak price bouncing test (and indeed all our other tests with the exception of commensurability and circularity), but fails the strong price bouncing test.

The Jevons and Dutot indices satisfy price bouncing as they are in any case invariant to any reordering of prices. While the CSWD index failed Dalèn's old price bouncing test, it passes this new one as it is time reversible. The Carli and Harmonic

**Table 3.3:** Price bouncing example

| | Period 0 | Period 1 | Period 2 | |
|---|---|---|---|---|
| Store A price | 1 | 1.25 | 1 | |
| Store B price | 1.25 | 1 | 1.25 | |
| | | Period (0,1) | Period (1,2) | Chained index |
| Carli | ... | $\frac{(1.25+0.8)}{2} = 1.025$ | 1.025 | 1.0506 |
| Harmonic | ... | $[\frac{1}{2} \times (\frac{1}{1.25} + \frac{1}{0.8})]^{-1} = \frac{1}{1.025}$ | $\frac{1}{1.025}$ | 0.952 |
| Dutot | ... | $\frac{1.125}{1.125} = 1$ | 1 | 1 |
| Jevons | ... | $\sqrt{1.25 \times 0.8} = 1$ | 1 | 1 |
| CSWD | ... | $\sqrt{1.025 \times (\frac{1}{1.025})} = 1$ | 1 | 1 |

indices on the other hand fail this test, as I illustrate with a simple numerical example. Table 3.3 shows how different indices respond to price bouncing in a case with two goods sold in different stores. In period 1 we swap the period 0 prices between store A and store B, and in period 2 we swap them back. In both periods 1 and 2, the Carli index increases by 2.5%, with a cumulative increase over both periods of 5.06%. Similarly, the Harmonic index decreases by 2.5% in each period for a cumulative reduction of 4.8%. This is despite prices in period 2 being no different to what they were in period 0! In fact, the Carli will always show an increase in these sorts of situations and the Harmonic index will always show a decrease (for the same reason that in general $P_C(\mathbf{p}_0, \mathbf{p}_1) P_C(\mathbf{p}_1, \mathbf{p}_0) \geq 1$ and $P_H(\mathbf{p}_0, \mathbf{p}_1) P_H(\mathbf{p}_1, \mathbf{p}_0) \leq 1$). The Jevons, Dutot, and CSWD indices on the other hand, will correctly record no price change, as they do in the example.

Given this list of requirements we can now ask, how do our elementary indices measure up? Table 3.3 summarises the results

The Jevons passes all the tests listed, while the Dutot only fails the commensurability test. The Carli and Harmonic indices fail the time reversal, circularity test and (both) price bouncing tests. The CSWD index fails the circularity test and the original price bouncing test (but not my revised test). Not all of the tests are necessarily as important as each other and, in principle, this might present us with

**Table 3.4:** Test performance of the elementary aggregate formulae

| Test | Carli | Dutot | Jevons | Harmonic | CSWD |
|---|---|---|---|---|---|
| 1. *Positivity* | ✓ | ✓ | ✓ | ✓ | ✓ |
| 2. *Identity* | ✓ | ✓ | ✓ | ✓ | ✓ |
| 3. *Monotonicity in current prices* | ✓ | ✓ | ✓ | ✓ | ✓ |
| 4. *Monotonicity in base period prices* | ✓ | ✓ | ✓ | ✓ | ✓ |
| 5. *Linear homogeneity* | ✓ | ✓ | ✓ | ✓ | ✓ |
| 6. *Homogeneity of degree minus one* | ✓ | ✓ | ✓ | ✓ | ✓ |
| 7. *Proportionality* | ✓ | ✓ | ✓ | ✓ | ✓ |
| 8. *Dimensionality* | ✓ | ✓ | ✓ | ✓ | ✓ |
| 9. *Mean value* | ✓ | ✓ | ✓ | ✓ | ✓ |
| 10. *Commodity Reversal* | ✓ | ✓ | ✓ | ✓ | ✓ |
| 11. *Commensurability* | ✓ | × | ✓ | ✓ | ✓ |
| 12. *Time reversal* | × | ✓ | ✓ | × | ✓ |
| 13. *Circularity* | × | ✓ | ✓ | × | × |
| 14. *Price bouncing test* | × | ✓ | ✓ | × | ✓ |

an aggregation problem (how to weight the different tests). However, luckily the results are definitive: whatever the weights we place on the individual tests the Jevons emerges with the strongest axiomatic backing. If we were to consider the importance of the different tests, many would point to the Carli's and Harmonic indices' failure of time reversal as being particularly serious. These indices fail this test in a biased manner, meaning that the Carli for instance will tend to give a higher rate of inflation than other indices that satisfy time reversal. This bias will then be reflected in the price changes used in calculations at higher stages of aggregation (as the indices used later satisfy monotonicity), biasing the whole index. Fisher (1922) (pg 66) was fairly unequivocal in his condemnation of the Carli index for its failure to satisfy this test, and it was on the basis that the Carli index failed the time reversal test - with an upward bias - that Diewert (2012b) recommended that the Carli index should no longer be used in the RPI. That said, there is not universal agreement on the importance of time reversal. Eichhorn and Voeller (1976) for instance introduce the time reversal and circularity tests by saying a price index needn't "necessarily" satisfy these (the only "indisputable" conditions a price index should satisfy according to these authors were their monotonicity, linear homogeneity, identity and dimensionality tests - the positivity test was included as part of the definition of a price index). Furthermore, while the Carli's failure to satisfy time reversibility (and other tests) is indeed a problem, it is important to realise that the RPI and CPI which these elementary aggregates eventually feed into are themselves not time-reversible, and nor would they be even if the elementary aggregates were time-reversible.[9] This means that fixing this particular problem associated with the RPI may not be of that great a benefit (though as the preceding discussion indicates, it is true that replacing the Carli with a time-reversible index would *reduce* any upward time reversal bias of the whole index).

## 3.3 The statistical approach

The statistical or stochastic approach to index numbers was originally associated with Jevons (1884) and Edgeworth (1925). More recent discussions of the statistical approach can be found in Selvanathan and Prasada Rao (1994), Diewert (1995b), and Clements et al. (2006). The statistical approach treats the problem of deciding on the correct price index as an estimation problem. The aim is to separate out a 'common' change in prices over two periods (a signal) from relative price changes (which can be considered noise). Different estimators indices can then be evaluated according to the standard statistical considerations of (statistical) bias, and efficiency. The bias of any index is measured against the population object of interest, but unfortunately, there seems to be little agreement in the literature for this approach on what the 'common' price change in the population should be.

The 'unweighted' stochastic approach aims to estimate the average price change from a population of price relatives $\left(p_1^i/p_0^i\right)$. Selvanathan and Prasada Rao (1994) start off by supposing that each price change is made up of a systematic part that is common to all prices and a zero-mean random component $u^i$, so

[9] After the level of the elementary aggregates, the RPI makes use of the Young and Lowe indices to aggregate further. The Young index is not time reversible, while the Lowe index is only time reversible for some comparisons. For an explanation of these see chapter 1 of International Labor Organization (2004). Diewert (2012b) also recommended that the Young index no longer be used in the RPI.

$$Dp_t^i = \alpha_t + u^i \tag{3.5}$$

where $Dp_t^i = \ln p_t^i - \ln p_{t-1}^i = \ln\left(\frac{p_t^i}{p_{t-1}^i}\right)$ (i.e the log increase in prices) and $\alpha_t$ the common trend in all prices which we aim to estimate. The International Labor Organization's consumer price manual (International Labor Organization (2004)) and Diewert (2012b) both use this statistical model to justify the Jevons index. This is because $\alpha_t = E\left[\ln\left(p_1^i/p_0^i\right)\right] = \ln P_J\left(\mathbf{p}_0, \mathbf{p}_1\right)$ gives the average log of the price changes and taking the anti-log of this gives the Jevons index. However, there are two issues with this conclusion.

Firstly, we can reasonably question whether this is the correct object of interest to consider. A longstanding criticism of the unweighted approach (associated with Keynes (1930), and Walsh (1901)) is that it treats all items equally regardless of their economic importance. Diewert (2012b) for instance refers to this (in the context of higher level indices) as a "fatal flaw". We might instead think that if for every £1 spent on item A £5 are spent on item B then we should assign B five times the weight of A. This is achieved straightforwardly if we imagine the population to be the price relatives associated with each pound spent rather than associated with each individual item regardless of its price or budget share. This would give us the object of interest $E[w^i\left(p_1^i/p_0^i\right)]$ where $w^i$ is the budget share of good $i$.[10]

A second more fundamental issue is that, even if this object of interest is the correct one, the Jevons does not necessarily give us a good statistical measure in this case. The elementary aggregate price change is one plus the percentage increase in prices and the motivation for using $Dp_t^i$ is to make use of the fact that

$$\ln p_t^i - \ln p_{t-1}^i \approx \frac{p_t^i}{p_{t-1}^i} - 1$$

(when growth rates are small). Ultimately our object of interest here is $E\left(p_1^i/p_0^i\right)$. But we know that if this is the case then the Jevons is *biased downwards* (an observation first pointed out, in this context, by Greenlees (2001)). This is because

[10] This can also be motivated by the "Expenditure Based Regression Model" in Selvanathan and Prasada Rao (1994).

Jensen's inequality tells us that $E[f(x_i)] \geq f(E[x_i])$ when $f$ is a convex function (with equality when all of the $x_i$'s are the same). When we take the anti-log of $\alpha_t$ we have $f$ as the exponential function and $x_i = \ln\left(p_1^i/p_0^i\right)$ so

$$P_J(\mathbf{p}_0,\mathbf{p}_1) = \exp\left[\ln P_J(\mathbf{p}_0,\mathbf{p}_1)\right] = \exp\left[\frac{1}{N}\sum \ln\left(p_1^i/p_0^i\right)\right] \leq E\left[\exp\left[\ln\left(p_1^i/p_0^i\right)\right]\right] = E\left[p_1^i/p_0^i\right]$$

$$\implies P_J(\mathbf{p}_0,\mathbf{p}_1) \leq E\left[p_1^i/p_0^i\right]$$

An alternative way to approach this problem is to note that the price-relatives themselves can generally (except in some extreme cases) be described by a decomposition into their mean and an additive, mean-zero, deviation

$$\frac{p_1^i}{p_0^i} = E\left[p_1^i/p_0^i\right] + e^i \tag{3.6}$$

where $E\left[e^i\right] = 0$ and the variance of $e^i$ is $\sigma^2$.

We can then estimate $E\left[p_1^i/p_0^i\right]$ in an unbiased way by taking its sample analogue

$$\frac{1}{N}\sum \frac{p_1^i}{p_0^i} = \hat{P}_C(\mathbf{p}_0,\mathbf{p}_1)$$

which is the Carli index of price relatives in the sample. Notice that this conclusion is not based on any arguments about how price relatives evolve or whether (3.5) is a more realistic model of the process generating price relatives than (3.6).[11]

This needn't mean that the use of the Jevons is ruled out by the unweighted statistical approach however, as bias is not our only consideration here. The overall performance of an estimator can be summarised by its mean squared error (MSE), which measures its expected squared deviation from the true population value of the parameter of interest (equal to the sum of its squared bias and its variance). That is, for some estimator $\hat{\theta}$ of a population parameter $\theta$

[11] The data generating process in (3.5) implies that current period prices in all outlets would be described by $p_1^i = E[p_1^i/p_0^i]p_0^i + e^i p_0^i$, which means that they would equal base period prices inflated by a common factor plus a heteroskedastic deviation. In a process described by $Dp_t^i = \alpha_t + u^i$ however, log prices in both periods would be decomposable into a mean and a homoskedastic deviation $\ln p_t^i = E[\ln p_t^i] + u_t^i$ for $t = 0,1$ where $u^i = u_1^i - u_0^i$. The latter seems the more realistic model of the way the data is generated. However, this has no bearing on the question of what the object of interest should be, or on whether the Jevons is biased as an estimator of $E[p_1^i/p_0^i]$.

$$MSE(\hat{\theta}) = E[(\theta - \hat{\theta})^2] = Var(\hat{\theta}) + Bias(\hat{\theta})^2 \tag{3.7}$$

The sample Jevons $\hat{P}_J$ may be a biased estimate of the unweighted population parameter of interest $P_C = E\left[p_1^i / p_0^i\right]$ while the Carli $\hat{P}_C$ is unbiased, but the Jevons may still perform better than the Carli in cases where it has a lower variance.

Using (3.7), we can define a measure of the performance of our indices relative to an arbitrary population object of interest $P$. Assuming we have a statistically random sample, the mean squared error of the estimator provided by the sample Carli index is simply the variance of the sample mean of price relatives $\sigma^2/N$ plus the Carli's squared bias

$$MSE(\hat{P}_C) = E[(P - \hat{P}_C)^2] = \frac{\sigma^2}{N} + (P - P_C)^2 \tag{3.8}$$

In the case of the Jevons index, the mean squared error (using a variance approximation in Dalén (1999) cited in Elliott et al. (2012) is

$$MSE(\hat{P}_J) = E[(P - \hat{P}_J)^2] \approx \frac{\sigma^2}{N}\left(1 - \frac{\sigma^2}{P_C^2}\right) + (P - P_J)^2 \tag{3.9}$$

where $\frac{\sigma^2}{N}\left(1 - \frac{\sigma^2}{P_C^2}\right)$ is the approximate variance of the Jevons. Since it can be shown that $\left(1 - \sigma^2/P_c^2\right) \leq 1$, this means that the approximate variance of the Jevons will always be smaller - by a constant factor - than the variance of the Carli.

Combining (3.8) and (3.9) tells us that the ratio of MSE's of the Carli and Jevons for our purposes will be given by

$$\frac{MSE(\hat{P}_C)}{MSE(\hat{P}_J)} \approx \frac{\sigma^2/N + (P - P_C)^2}{\frac{\sigma^2}{N}\left(1 - \frac{\sigma^2}{P_C^2}\right) + (P - P_J)^2} \tag{3.10}$$

The expression (3.10) suggests that it is possible that $MSE(\hat{P}_C)/MSE(\hat{P}_J)$ is greater or smaller than one 1 depending on the relative biases and variance of the two indices. To say anything further at this point, we need to settle on an appropriate object of interest. The preceding discussion indicates that this would likely be a weighted average of price relatives. I shall adopt period 0 weights in what follows, although one could equally well choose period 1 weights (or some combination of weights in the two periods). This decided, we can now attempt to evaluate the performance of the Carli and Jevons in different circumstances.

### 3.3.1 Empirical exercise

To investigate this, I make use of data from the Kantar Worldpanel. This is a panel run by the market research firm Kantar which surveys households throughout Great Britain (Northern Ireland is not included). In this survey, participating households are issued with barcode readers and are asked to scan all barcoded products brought home. In principle this includes groceries purchased from all retailers including online outlets, and not just supermarkets. Households also record information on the stores visited. Information on the prices is obtained from till receipts which are mailed to Kantar who then match the prices paid to the purchase record. Where no receipts are available, prices are taken from centralised databases of store- and product-specific prices, or otherwise imputed. The data also record any promotional deal attached to a purchase. The Kantar Worldpanel therefore provides us with extremely detailed data on household's expenditure on individuals products, including the date they were scanned and the shop where they were purchased, and whether or not they were on offer. The household purchases recorded in the survey give me sample of prices to use in my analysis.

To investigate the statistical performance of the Carli and Jevons indices, I first use the data to construct a 'population' of price relatives for different goods. To calculate price relatives for goods seen in different months, I first need to define what is meant by a 'good' and then how to decide on its monthly price. I take a good to be a particular product sold in a particular store (products are identified by barcode). Thus two identical loaves of bread sold in two different supermarkets would be considered different goods. The price of goods for a particular month is then taken to be the modal price consumers paid for one unit of this good in that month. The price relative for that month is calculated by deciding on a base month, and then calculating the ratio of prices in the following months relative to that month. I drop any prices that are not observed in every month (leaving us with a balanced panel of prices) which serves to exclude seasonal items that are only purchased at certain times of the year. I also drop the prices of any goods that are on any kind of promotion (such as those subject to quantity deals). The idea behind all these choices is to try and replicate what the ONS or other statistical agency would sample when calculating its elementary aggregates.

To estimate the mean squared errors of the Carli and Jevons indices, I then employ the following procedure:

1. Drawing a random sample of price relatives of size $n$ without replacement from this population

2. Calculating sample Carli and Jevons estimators

3. Drawing another sample, and so on for 30,000 iterations

I repeat this procedure for different sample sizes (for $n = 20, 50, 80, 110, 140, 170, 200$), and using different base months. In each case we will get a distribution of sample Carli and Jevons indices which I use to obtain direct estimates of their biases (the average difference between the sample Carli and Jevons and the population weighted mean of price relatives), their mean squared errors (the average squared difference between the sample Carli and Jevons and the population weighted mean of price relatives) and their variances. I look at the prices of two categories of goods: alcohol and bread. The above selection procedures leave me with 518 price quotes for alcohol in each month and 2,319 for bread. Ideally I would like to look at clothing as this is the category of spending for which the formula effect is largest (see Morgan and Gooding (2010)) and therefore the group for which the question of whether the Carli or Jevons is best statistically is most important. Unfortunately spending on this is not covered by the Kantar Worldpanel which only records grocery spending. My data cover the months of January to October 2010.

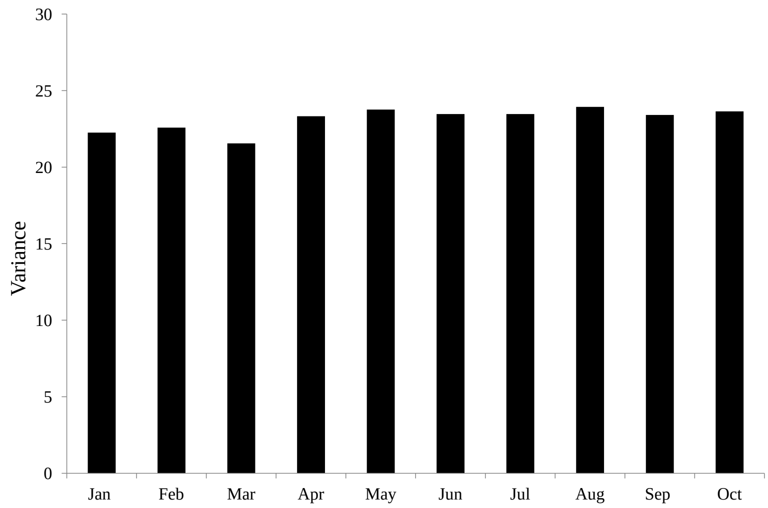
This work builds on a similar analysis by Elliott et al. (2012) who also used alcohol prices in the Kantar Worldpanel covering the years 2003-2011 to look at the statistical performance of different estimators (Elliott et al. themselves remain agnostic as to what the object of interest should be, and consider the performance of different estimators for various different target indices). They find that the MSE of the Carli tended to be smaller than that of the Jevons $E\left[w_0^i(p_1^i/p_0^i)\right]$ was the object of interest - especially as the sample size increased. In very small samples the Jevons could however perform better, suggesting that the statistical approach might favour one or the other in different circumstances. I build on the analysis of Elliot et al. by comparing the performance of the different base months and different times of the year. The interest in the impact of the choice of base month on the two estimators relative statistical performance stems from its potential to influence the size of the formula effect. Fenwick (1999) points out that January sales might increase the variation of prices (especially for goods such as clothing), and since prices in all subsequent months will then be compared with January, these would have a knock on effect on the variance of price relatives and hence the size of the difference between the Carli and Jevons throughout the rest of the year. Different budget shares between goods in different months also mean the choice of base month may affect the indices' relative biases in our case.

The population variance of modal prices in different months for alcohol is shown in Figure 3.2. There is little evidence of an impact of January sales in particular on the variance of prices, and the differences from month to month are small. The same is true for the prices of bread which I omit due to space considerations. This suggests that the channel Fenwick (1999) proposed for differences between the two indices across base months may not be too important in this data. For alcohol the variance is greatest in August (and smallest in March).

The variance of the Carli and Jevons estimators are shown in Figures 3.3 (for bread) and 3.4 (for alcohol). These plot the variance for each sample size and each month using a January base month. The expected relationships hold in that the Jevons consistently has a lower variance than the Carli and the variances of both estimators decline approximately with the square root of the sample size.

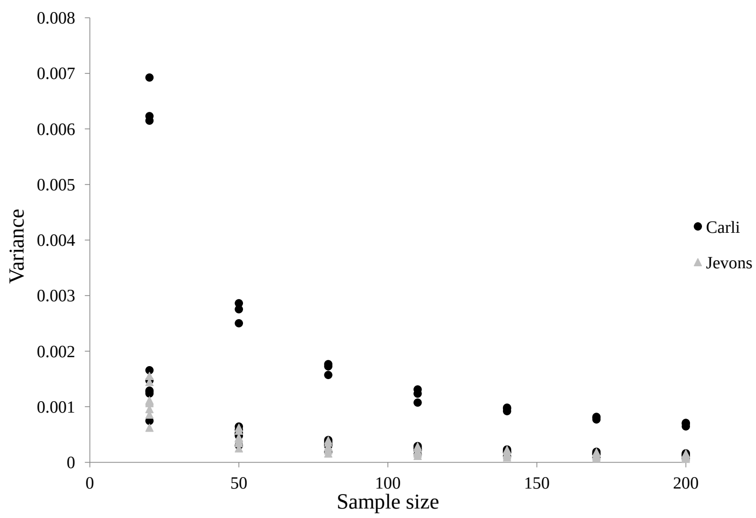
The other component of the mean squared error is the bias. Both the Carli and Jevons are biased estimators for our chosen object of interest, and it is not certain *a priori* which index will be more biased. The Jevons naturally attaches

**Figure 3.2:** Variation in population modal alcohol prices by month, 2010



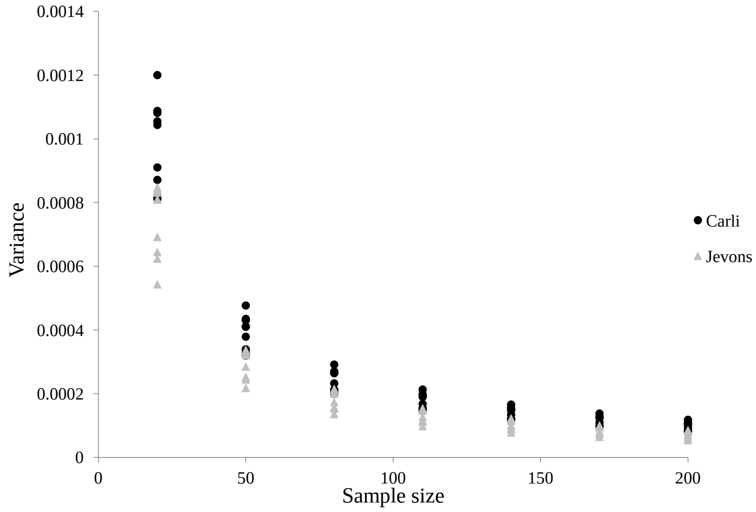

Note: Author's calculations from Kantar Worldpanel.

**Figure 3.3:** Variance of Carli and Jevons for bread prices




Note: Author's calculations from Kantar Worldpanel.

**Figure 3.4:** Variance of Carli and Jevons for alcohol prices

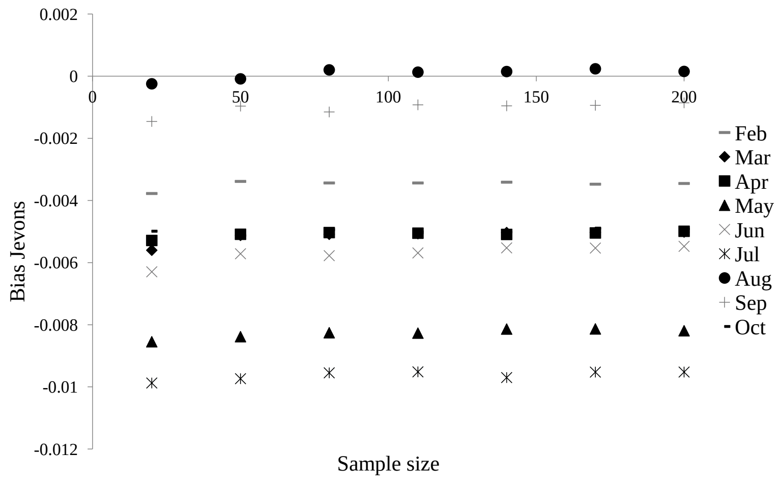



Note: Author's calculations from Kantar Worldpanel.

less importance to the largest price changes, so if these good also have the smallest weights, then the Jevons may be less biased than the Carli for example. The biases of the Jevons for different sample sizes for alcohol are shown for each month in Figure 3.5 below (again using a January base month). It is clear that unlike the variance, the bias remains roughly constant as the sample size grows. The same is true for the bias of the Carli for alcohol and the biases of both indices for bread, which I do not plot to save space. For bread, the Carli is positively biased for some months and negatively biased for others (March and April), while it always has a negative bias for alcohol prices. The Jevons has a positive bias for bread and a negative bias for alcohol prices (though the bias for alcohol prices is smaller than that of the Carli).

We have seen that the Jevons consistently has a lower variance than the Carli (which falls with sample size for both estimators) while the biases of the two estimators are essentially unaffected by the size of the sample. What does this imply for the two estimators' mean squared errors? The ratio of the MSE for the two estimators is shown below in Figures 3.5 and 3.7. For bread the ratio falls as the sample size increases, but the number of months where the Carli has a greater MSE than the Jevons is greater only after the sample size reaches exceeds 110. Even then, in

**Figure 3.5:** Bias of the Jevons index for alcohol prices



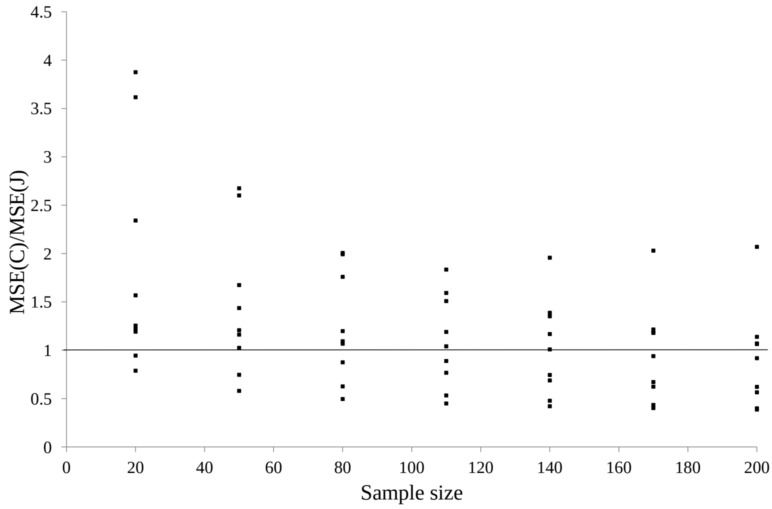

Note: Author's calculations from Kantar Worldpanel.

some months the Carli performs much worse than the Jevons (sometimes with an MSE that is more than twice as high even when the sample size reaches 200). For alcohol the Jevons seems to perform better in all samples, and the ratio of the MSE's favours the Jevons as the sample size gets larger. This is because the constant biases have a larger relative impact on the ratio as the variance of the two indices shrinks (and the Jevons is less biased than the Carli).

I now turn to the question of how the MSEs of the two estimators vary with the time of year. I plot the MSE ratios for different sample sizes in each month for bread and alcohol respectively in Figures 3.8 and 3.9 (keeping the base month at January). For alcohol, the Jevons consistently outperforms the Carli in terms of MSE in every month. For bread, this can vary. In June, the Carli has a lower MSE for all sample sizes, while in February, May and July the Jevons always has the lower MSE.

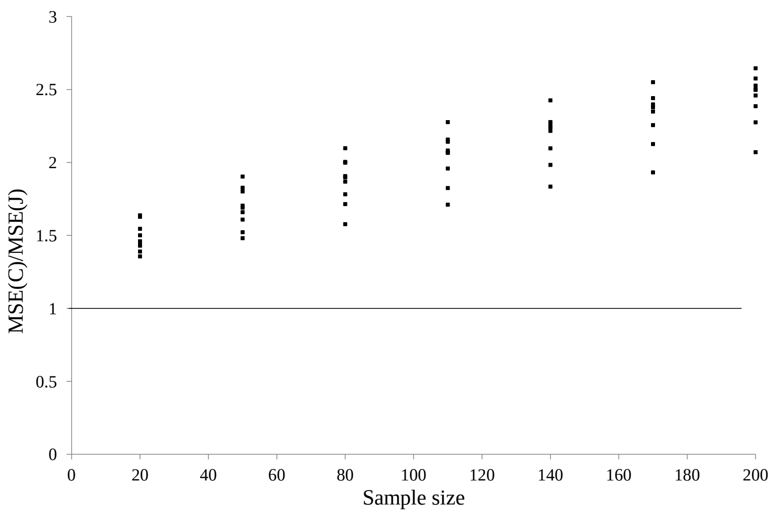
A final question concerns the impact of the choice of base month. The Carli-Jevons MSE ratio for different months given different base months is shown for bread in Figure 3.10 and for alcohol in Figure 3.11. For these plots I keep the sample size constant at 200. I calculate the ratio of MSEs for every month following the base month in the year (so when the base month is September we will only have price relatives for October). It's clear that the choice of base month matters. In my

**Figure 3.6:** Ratio of Carli and Jevons MSEs for bread by sample size



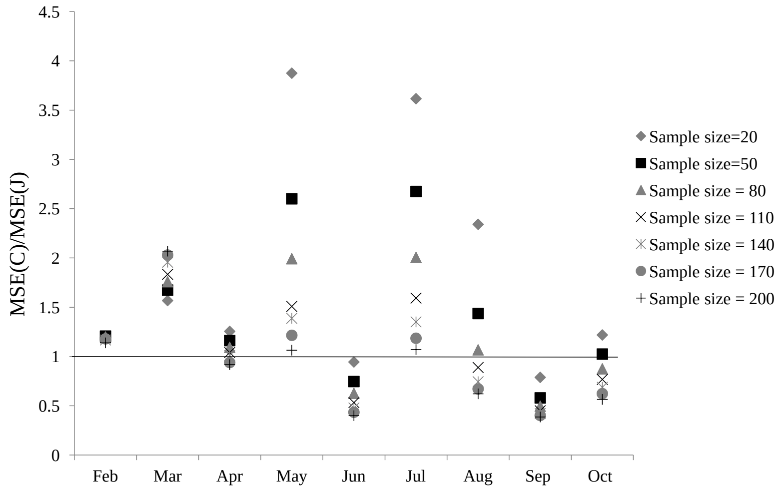

Note: Author's calculations from Kantar Worldpanel.

**Figure 3.7:** Ratio of Carli and Jevons MSEs for alcohol by sample size



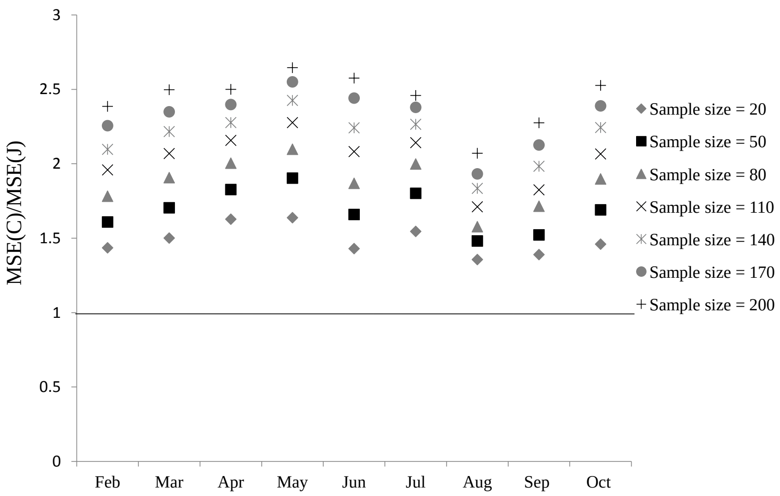

Note: Author's calculations from Kantar Worldpanel.

**Figure 3.8:** Ratio of Carli and Jevons MSEs for bread by month




Note: Author's calculations from Kantar Worldpanel.

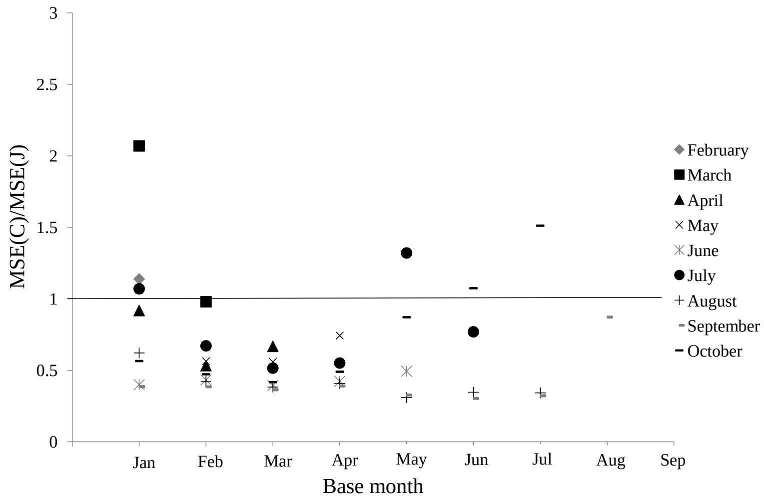
**Figure 3.9:** Ratio of Carli and Jevons MSEs for alcohol by month




Note: Author's calculations from Kantar Worldpanel.

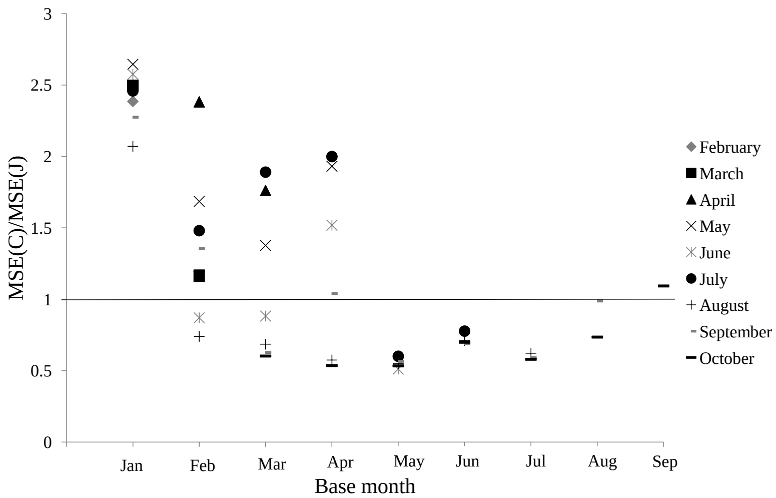
data for bread it appears that the case for using the Carli seems stronger when the base month is February for instance than in January. Similarly for alcohol, the case for using the Jevons seems much stronger when using a January base month than when we use other base months.

**Figure 3.10:** Ratio of Carli and Jevons MSEs for bread by base month (sample size =200)




Note: Author's calculations from Kantar Worldpanel.

**Figure 3.11:** Ratio of Carli and Jevons MSEs for alcohol by base month (sample size =200)




Note: Author's calculations from Kantar Worldpanel.

To summarise, the statistical approach does not decisively favour either the

Carli or the Jevons for bread in the data I use, but the Jevons consistently outperforms the Carli for alcohol. This contrasts with the findings of Elliott et al. (2012) who find, using data covering a longer time period, that the Carli tended to perform better as a measure of the population base weighted price relatives for alcohol (at least when we use a January base month). We also find that the relative performance of the two estimators varies over the year and according to the base month used. These results have implications for discussions around the size of the formula effect. Since 2010, a change to the methods used to collect clothing prices has led to a particularly large difference emerging between the Jevons index used for these goods in the CPI and the Carli used in the RPI. The changes tended to increase the sample sizes used for calculating clothing price relatives. One might have thought that this would favour the Carli over the Jevons as the variance of the Carli declines faster with sample size, and indeed this would be true if the object of interest was unweighted. However with our weighted object of interest, as the sample size increases and the variance decreases, the bias of each index becomes a more important element in determining their relative MSEs. Whether the Carli or Jevons is more biased for our object of interest depends on which goods see the fastest price increases, and how dispersed the price relatives are. These factors vary over time and across goods, making it difficult to offer any general rules for when the Carli or Jevons may be more suitable. This suggests caution when trying to generalise results such as those found in Elliott et al. (2012).

One potential pitfall with analyses such as these is that they are based on drawing a random sample of price relatives from some population (and implicitly assume that national statistical agencies would do the same). However, in practice the ONS sample for instance would not appear to be completely random. At present the ONS price sample consists of price quotes for a list of specific items judged in advance to be representative of broader categories: and the selection of these representative items is often a matter of judgement, (see Gooding (2012)). The list of representative items is updated annually based on a range of considerations. In 2012 walking/hiking boots replaced outdoor adventure boots as a representative item for footwear. A bag of branded chocolate also replaced candy coated chocolate as a representative item as its price had been becoming more difficult to collect (for more details see Gooding (2012)). It's not clear how this might affect the relative statistical performance of the sample Carli and Jevons. The fact that price relatives are not independently distributed also makes it impossible to calculate reliable standard errors for the elementary aggregates since the actual variance-covariance matrix of $e^i$ in (2) will in practice be unknown. This is unfortunate as the prospect of publishing standard errors alongside inflation rates has been mentioned as a key attraction of the statistical approach, (see for instance Clements et al. (2006)).

## 3.4 The economic approach

The economic approach to index number construction aims to answer the question: how much more income would a typical consumer require to maintain the same standard of living following a price change? We are typically only interested in maintaining the same 'economic welfare' over two periods, by which we mean that we only seek to compensate the consumer for changes in the prices they face, and not for changes in other environmental factors such as air quality, or changes in the consumer's tastes, which are held constant for the purposes of our comparison.

The question is answered conceptually by a cost of living index (COLI). The COLI is defined by means of a cost function $c(\mathbf{p}_t, u_t)$, which tells us for any given level of prices $\mathbf{p}_t$, the minimum level of expenditure needed to achieve a given level of welfare or 'utility' $u_t$. The ratio of cost functions in two periods, holding the target utility constant at some level $u$ defines the COLI or Konüs index (dating back to Konüs (1939))

$$P_K(\mathbf{p}_0, \mathbf{p}_t, u) = \frac{c(\mathbf{p}_t, \bar{u})}{c(\mathbf{p}_0, \bar{u})}$$

Every household will have its own COLI, and in order to calculate an economy-wide inflation measure it is necessary to aggregate these in some way. Various ways of doing this are discussed in Crossley and Pendakur (2010) and aggregation issues are also discussed in Diewert (2001). The COLI is distinct from a Cost of Goods

Index (COGI) which conceptually aims to compare the cost of buying a fixed basket of goods in two different periods (rather than to achieve the same utility). The ONS rejects the interpretation of both the CPI and RPI as attempts to measure changes in the cost of living (see Office for National Statistics (2011)) and instead regards these as COGIs.

Under the economic approach, the price index chosen should reflect the degree to which consumers mitigate the welfare impact of price changes by shifting their purchases away from goods and services that have become relatively more expensive and towards goods that have become relatively cheaper. This means it should explicitly take account of the dependence of prices and quantities over time given by the demand function $\mathbf{q_t}(\mathbf{p_t})$. Economists traditionally do this is by representing consumers' decision making (their preferences) with a utility function that ranks different bundles of goods and services, and more importantly is associated with particular demands and particular substitution responses. Each utility function is associated with its own cost function, and if a price index coincides with the ratio of two cost functions for a particular utility function, it can be thought of as representing the COLI for those particular preferences. Two noteworthy price indices that do just this are:

1) The Laspeyres index

$$P_L(\mathbf{p}_0, \mathbf{p}_1) = \frac{\sum p_1^i q_0^i}{\sum p_0^i q_0^i} = \sum w_0^i \left( \frac{p_1^i}{p_0^i} \right)$$

where $w_0^i$ gives the budget shares of good $i$ in period 0. This corresponds to the Leontief preferences, where the consumers utility is given by

$$u_t = \min[\alpha^1 q_t^1, \alpha^2 q_t^2 .... \alpha^N q_t^N]$$

The consumer maximises this for any given vector of prices by selecting quantities such that

$$\alpha^1 q_t^1 = \alpha^2 q_t^2 = .... = \alpha^N q_t^N$$

Thus for Leontief preferences, there are no substitution responses: the ratios of different quantities remain constant as prices change.

2) The Geometric Laspeyres

$$P_{GL}(\mathbf{p}_0, \mathbf{p}_1) = \prod_{i=1}^{N} \left(\frac{p_1^i}{p_0^i}\right)^{w_{i0}} = \exp(\sum w_0^i \ln\left(\frac{p_1^i}{p_0^i}\right))$$

This corresponds to Cobb-Douglas preferences

$$u_t = \left(q_t^1\right)^{\beta^1} \times \left(q_t^2\right)^{\beta^2} \ldots \times \left(q_t^N\right)^{\beta^N}$$

which are in turn associated with the demand functions

$$q_t^i = \frac{\beta^i}{\sum_j \beta^j} \frac{M}{p_t^i}$$

where $M$ is the consumers income. For these preferences, the budget shares will be constant, as $\frac{q_t^i p_t^i}{M} = \frac{\beta_i}{\sum_j \beta_j}$, regardless of prices. This implies that a 1% increase in the price of a good results in a 1% reduction in the quantity demanded.

If we think that substitution between products occurs *within* certain groups but not *between* those groups and others (or if preferences within the group are described by Cobb-Douglas or Leontief 'subutility' functions), then we could calculate sub-COLIs within groups using these formulae and then combine them to get an overall index in a manner similar to the process of aggregation used to construct the RPI and CPI. For instance, a Geometric Laspeyres could be used within categories of goods where we thought substitution responses were realistically described by Cobb-Douglas preferences, and a Laspeyres index could be used if we thought that it was more appropriate to assume zero substitution.

These indices differ from the unweighted Carli and Jevons and Dutot indices that are actually used in the RPI and CPI, but their resemblance is sometimes used to justify the choice of indices used in the calculation of the elementary aggregate price changes. The Jevons for example is thought to approximate a Geometric Laspeyres within an elementary stratum, and this was one reason behind the Boskin Commission's (Boskin et al. (1996)) recommendation that the US CPI should make use of the Jevons for elementary aggregates (when this change was put into effect, the BLS also argued that it would better capture consumers' substitution responses, see for instance Dalton et al. (1998)). This kind of logic has however been criticised by Diewert (2012b) who writes that "...the economic approach cannot be applied at the elementary level unless price and quantity information are both available." Since at the level of elementary aggregates such information is not available, it follows that the economic approach should have nothing to say on the subject of which index is preferable.

There are two problems with applying the economic approach when quantities are unknown. The first of these is that without knowledge of the weights which should be given to each price or price relative, we will not know if elementary indices are greater than or smaller than the Laspeyres and Geometric Laspeyres - in other words the direction and scale of their bias will be unknown. There are in fact particular assumptions about the way prices are sampled under which elementary indices will equal their COLI counterpart (set out in Chapter 20 of International Labor Organization, 2004). Most importantly for our purposes, the Carli will equal the Laspeyres index and the Jevons the Geometric Laspeyres if the price relatives of good $i$ are sampled with a probability equal to their base period expenditure shares.[12]

These conditions will be true under random sampling in the base period provided outlets stock goods in proportion to consumers' expenditures on them. However, as we saw in the last section, this assumption is unlikely to hold in practice. If these conditions are not true, and they are essentially impossible to verify, then our elementary indices may end up calculating something rather different from what we intended. Indeed, they may be upward or downward biased. Thus, unless we had some *reason* to think that the Carli or Dutot will approximate a true Laspeyres, or that the Jevons would approximate a Geometric Laspeyres, then we should be wary about economic justifications for one elementary index over another.

[12] The Dutot will equal the Laspeyres if the probability of sampling good $i$ in the base period is equal to the ratio of purchases of $i$ in the base period to the total purchases of all goods in $i$'s elementary stratum in the base period.

The second problem from a lack of quantity information at this level is that it means we will be ignorant of the nature of the interdependence of prices and quantities (consumers' substitution responses), which it is necessary to understand in order to choose whether our target COLI index should be a Laspeyres or Geometric Laspeyres.

These problems are often used to argue that the economic approach should not be used to choose index numbers at the level of the elementary aggregates, and that different approaches such as the statistical or test approach should be used instead (see for instance Diewert (2012b)). Switching to some other approach is however not a particularly satisfying solution. A statistical agency that was employing the economic approach to calculating inflation would still wish to estimate the appropriate COLIs at the elementary level: even if there was insufficient information to construct adequate approximations to these, this doesn't by itself give us justification to adopt entirely different criteria to select index numbers at this level and this level only. The elementary indices chosen for instance using the test approach could also be greater than or less than what would be suitable given consumers' spending weights and substitution behaviour, and so would be equally problematic. Employing a different approach all together does not solve the problems posed by a lack of information.

Given then that we do seem to have an alternative, how can we select index numbers to approximate consumers' COLIs when we lack all the relevant data? It turns out that, in situations such as these, there is a constructive principle that can be used to guide our choice of index number. This is the principle of maximum entropy (PME), which I now explain.

### 3.4.1 Principle of maximum entropy

Our problem is that the vectors of budget shares in the base and current periods are unknown at the elementary aggregate level. If we had grounds for selecting one particular vector of base period budget shares for each period from the infinite number of possible combinations for a set of goods, then we could use these to construct indices that approximated either the Laspeyres or Geometric Laspeyres.

If in addition we could select a vector of current period budget shares, then by considering the dependence between shares and prices over time we could also decide which of these two target indices was a more accurate reflection of the COLI for these goods. The question is: why should we choose one particular combination over another? In situations where we have limited knowledge, the PME provides a criterion which we can use to guide our choice of budget shares. This was first proposed by Jaynes in two papers (Jaynes (1957a), Jaynes (1957b)) in the context of selecting probability distributions.

To see how this approach works, consider the following example. Suppose we have a die that has been rolled many times. By 'many' we mean that a sufficient number for us to ignore any problems of sampling variation. Suppose that the only thing we are told about these dice rolls is the value of the average roll. What can we say about the probability of rolling a particular number given only this information? This problem would normally be considered insoluble, as Jaynes (1983) notes "on orthodox statistical theory, the problem is ill-posed and we have no basis for making any estimate at all."

Laplace's 'principle of insufficient reason' provides us with a first step for assigning probabilities in situations such as these. This states that in any situation where you want to assign probabilities to different outcomes, you should set them to be equal unless you have reason to do otherwise. The maximum entropy combines the principle of insufficient reason with any information we do have, and in doing so reflects the idea that we do not want to favour any outcome unless we have adequate justification to do so.

An objective function that will achieve this outcome is the entropy function proposed by Shannon (1948).

$$H(\mathbf{p}) = -\sum_i p_i \ln p_i$$

where in the dice example $p_i$ is the probability of rolling number $i$.

This function is maximised when probabilities are uniform and minimised when probabilities are degenerate on a particular outcome. In any given application, we will want to maximise entropy subject to constraints given by the knowledge we have (in the dice example, subject to knowledge of the average roll). The constrained optimum to this problem then best represents the current state of knowledge. To choose a distribution with lower entropy than the solution would be to assume information (as measured by Shannon's function) which we do not possess. To choose a distribution with higher entropy would violate the constraints provided by the information which we do possess from the data. By solving this problem, the maximum entropy approach provides us with estimates of probability distributions in cases where there is insufficient information to use standard statistical methods.

### 3.4.2 Application of maximum entropy to elementary aggregates

The PME is traditionally applied to situations where we must choose a vector of probabilities. To apply it to our case, we need only note that the budget shares $\mathbf{w}$ have all of the necessary properties of probabilities so we can apply the PME to these in the same way. In particular they conform to the Kolomogorov axioms of probability measures (so by definition we can treat them exactly like probabilities).

This suggests the following entropy measure

$$H(\mathbf{w}) = -\mathbf{w}' \ln \mathbf{w}$$

where the budget shares take the place of the probabilities. In the simplest case in which we have no other information (i.e. no constraints aside from the fact that budget shares should sum to one) the maximum entropy problem is

$$\max_{\mathbf{w}} H(\mathbf{w}) = -\mathbf{w}' \ln \mathbf{w} \text{ subject to } \sum w^i = 1 \tag{3.11}$$

This is solved by equal budget shares.

**Proposition 1.** *The solution to the maximum entropy problem (3.11) is $w^i = 1/N$ for all $i$.*

*Proof.* See Appendix to this chapter. □

The intuition behind this solution is as follows. Our problem for selecting budget shares at the level of elementary aggregates is analogous to the dice problem but in a case where we do not even know the average roll. It seems we just cannot know what the budget share of each individual good is in the same way as we couldn't know what the chances of rolling a 1 in the dice example were, which is the reason for rejecting the economic approach. However, just as we can assign some probabilities to dice rolls using the principle of insufficient reason, we can similarly assign weights using a budget share equivalent: if we do not have any reason to think that one good should have a greater or smaller budget share than any another, we will assign them all equal budget shares. [13]

This provides us with a constructive principle which can be used to address our first problem with applying the economic approach to elementary aggregates. The principle of maximum entropy justifies the Carli as an approximation for the Laspeyres index and the Jevons as an approximation of the Geometric Laspeyres. This is another way of saying that without additional knowledge, we will not assign any one good a higher base period budget share than another when calculating our Laspeyres or Geometric Laspeyres.

Our second problem concerned our lack of knowledge of the interdependence of prices and quantities over time. We can proceed in spite of this by noting that the PME can be applied to the vectors of budget shares in *both* periods $t = 0, 1$. For instance we can solve

$$\max_{\mathbf{w}_0,\mathbf{w_1}} H(\mathbf{w}_0, \mathbf{w_1}) = -\sum_t \mathbf{w_t}' \ln \mathbf{w_t} \text{ subject to } \sum_i w_t^i = 1 \text{ for } t = 0, 1 \tag{3.12}$$

We can add constraints imposed on the consumer's behaviour by economic theory to this problem. Suppose that we also have available the *total* expenditure on the sum of all of the items in the elementary stratum in each period: denoted $\{x_0, x_1\}$ where $x_0 = \mathbf{p}_0'\mathbf{q}_0$ and $x_1 = \mathbf{p}_1'\mathbf{q}_1$. This is the kind of data which may be used to weight

[13] Diewert (2012a) also suggests that an assumption of equal weighting could be used at the elementary level. He uses this to justify both the Carli and Harmonic indices as Cost of Goods Indices.

elementary aggregates in the next level up. Given this additional data the economic approach to index numbers provides constraints on the budget shares. They must satisfy certain axioms of behaviour provided by the Generalised Axiom of Revealed Preference (GARP) (for details see Afriat (1967); Diewert (1973); Varian (1982)):

$$\{\mathbf{p}_0, \mathbf{p}_1; \mathbf{w}_0, \mathbf{w}_1; x_0, x_1\} \text{ satisfies GARP}$$

GARP is a set of inequalities involving the prices, budget shares and total expenditures[14] which provide necessary and sufficient conditions for the standard economic model of consumer choice. Note that these restrictions are fully nonparametric in the sense that they do not require any knowledge of the consumer's preferences. These constraints can then be added to the maximum entropy problem which becomes:

$$\max_{\mathbf{w}_0, \mathbf{w}_1} -\sum_t \mathbf{w}_t' \ln \mathbf{w}_t \text{ subject to } \{\mathbf{p}_0, \mathbf{p}_1; \mathbf{w}_0, \mathbf{w}_1; x_0, x_1\} \text{ satisfies GARP and } \sum_i w_t^i = 1 \text{ for } t = 0, 1 \tag{3.13}$$

The result will be a set of weights which satisfy economic theory and the informational content (as measured by Shannon's index) of the data. We can show that this problem is also solved by equal budget shares in both periods (since this solves the unconstrained problem and it turns out that the restrictions from GARP are not binding).

**Proposition 2.** *The solution to the maximum entropy problem (3.13) is* $w_t^i = 1/N$ *for all* $i, t$.

*Proof.* See Appendix to this chapter. □

This means that when you have no data on quantities or budget shares, the PME provides a constructive argument for equal shares across good and periods of time. These budget shares would be chosen by consumers who had equally weighted Cobb-Douglas preferences. In terms of the choice of elementary index,

[14] Typically GARP is applied to prices and quantities but it can easily be rewritten in terms of prices and budget shares since $q_t^i = w_t^i x_t / p_t^i$.

this would justify the Geometric Laspeyres as a COLI (since this corresponds to the COLI for Cobb-Douglas preferences), and also justify the Jevons index (since when budget shares are uniform, an unweighted index will equal the COLI). To choose different vectors of budget shares would assume information which we do not have at this level, and so would not be justified without additional evidence. As was noted before, the PME can also be used to justify the Carli index as an approximation of the Laspeyres. Thus, the Carli could also be used in some elementary aggregates if one had some *a priori* grounds for believing that a Leontief utility function (i.e no substitution) better reflected consumers' preferences. For instance, the Carli might be more appropriate for elementary aggregates covering pharmaceuticals or goods sold in very different regions.

## 3.5 Conclusion

Now that I have set out my views on the different approaches to assessing elementary indices, we can now ask whether the UK's national statisticians are right to regard the Carli index as flawed and the Jevons as superior. This after all is the reason underlying both the creation of the new RPIJ index and the decision to stop classifying the old (and venerable) RPI as a national statistic. Here I will sum up my conclusions from the test, statistical and economic approaches and give an overall judgement.

Under the test approach, I noted that the Carli index fails to satisfy various properties which one would expect of a price index, including the important time reversibility test, while the Jevons satisfied all the tests considered. I also find that the Carli fails a new, revised version of the price bouncing test. It is true however, that the Carli's failure to satisfy time reversibility does not provide very strong reasons to replace it in the RPI, an index which is itself not time reversible, and which would not become time-reversible were the Carli index to be replaced (although this would serve to reduce the time-reversal bias of the index).

For the statistical approach I noted that there is no *a priori* reason to prefer either the Carli or the Jevons. Even when our object of interest is an unweighted average of price relatives, the Jevons, despite its bias, may still have a lower mean squared error than the Carli. This is because the Jevons can have a lower variance than the Carli. My view is that it is more appropriate to use the population weighted price relatives as the target to be estimated. We looked at the relative performance of the Carli and Jevons as estimators for this in different contexts for bread and alcohol. This exercise showed that the relative performance of the Carli and Jevons are not invariant to considerations such as sample size, the goods included in the elementary aggregate, the base month and the month of the year. This suggests that results found in one context needn't necessarily generalise to others.

A common view in the literature is that the economic approach cannot be applied at the level of elementary indices, where quantity information is by definition not available. However, I showed that in the absence of additional information, the principle of maximum entropy provides a constructive argument for equal shares across goods and across periods. This approach provides justification for both the Jevons as an approximation to the Geometric Laspeyres and the Carli for the Laspeyres. When applied across periods, the PME suggests use of the Geometric Laspeyres as a target index, as this is consistent with constant budget shares over time. This would favour the use of the Jevons for elementary aggregates when information on consumers' actual preferences was not available (which will in practice be true for most categories).

Thus, the test and economic approaches both seem to favour the Jevons index over the Carli, while the statistical approach doesn't provide clear, general guidance. I therefore concur with the general conclusion of the ONS and UKSA that the Jevons should be preferred to the Carli.

# Chapter 4

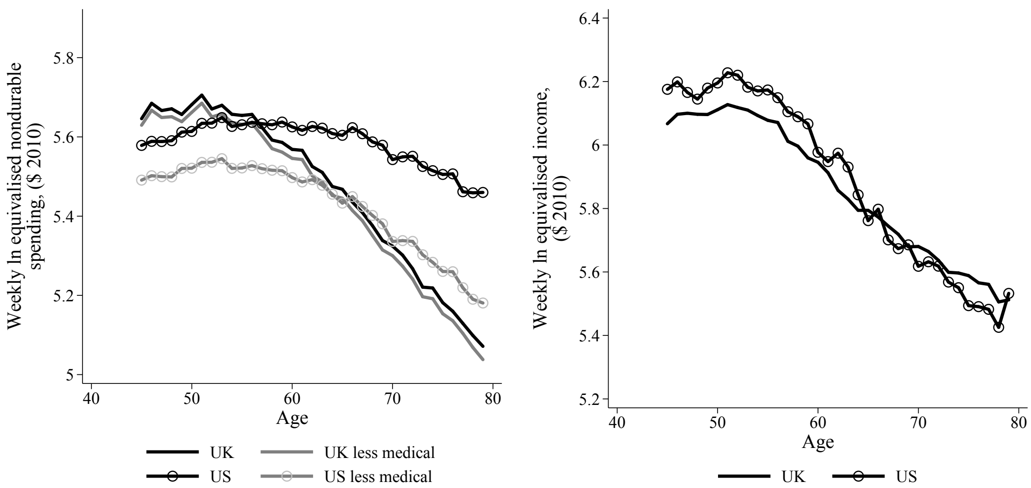
## Life-cycle Consumption Patterns at Older Ages in the US and the UK: Can Medical Expenditures Explain the Difference?

As populations in advanced countries continue to age, a key concern for policy-makers is whether individuals have saved enough to fund their consumption needs over increasingly long retirement periods. Understanding trajectories of consumption and wealth as individuals age is crucial to resolving this question. Research on life-cycle consumption patterns has typically concentrated on working ages with an emphasis on expected paths in labour income, economic wage shocks, and retirement; see for example the Review of Economic Dynamics special issue on micro facts (Krueger et al. (2010)). However, this leaves out an important and growing span of life during the post-retirement years where other factors such as health, mortality, health expenses and shifts in housing expenditures and recreation may play an increasingly central role. Moreover, these are areas where there are large cross-country institutional differences - for example in housing markets and in whether medical care is privately or government financed - that may have important implications for patterns of non-durable consumption at older ages.

In the United Kingdom, average non-durable expenditure between the ages of 45 and 79 falls by 2.2 percent each year. This compares to 1.4 percent for the United States. To illustrate, the first panel of Figure 4.1 plots non-durable expenditures in the UK and US by age averaged across birth cohorts. Its clear that spending remains roughly constant after age 50 in the US while it falls much more rapidly in the UK.

What can explain a difference of this magnitude? An obvious starting point is to examine age paths of income to assess the extent to which consumption expenditures are tracking age paths in household income. But the second panel in Figure 4.1, which plots cohort averaged paths of household income at older ages in the two countries, demonstrates that, if anything, incomes decline at a slightly faster rate in the US than the UK.[1] This therefore seems unlikely to be the major reason for a flatter spending profile in the US. In this chapter we investigate other possible reasons that may explain the dramatically different patterns of non-durable consumption of older ages in the two countries by investigating differences in inter-temporal consumption for households around and beyond retirement age.

**Figure 4.1:** Non-durable spending and incomes in the US and UK by age, 1984-2010




Notes: Authors calculations using BLS Consumer Expenditure Survey 1984-2010 and ONS Living Costs and Food Survey 1984-2010. Values are in US$ (2010). Figures equivalised using the modified OECD scale. The definition of spending includes medical expenditures.

The set of factors that we explore in this chapter include: differential cohort effects in the two countries that may distort average life-cycle age profiles, differences

[1] In both countries income is measured as the sum of salary, investment, interest, rental and transfer income and other income net of tax payments. In neither country does income include capital gains on property or other investments. UK Prices are converted to US dollars with PPP indexes.

in timing of retirement in the presence of separabilities with employment, differential paths of housing expenditures possibly driven by institutional differences in housing markets between countries, level and path differences in health status and mortality, and finally the levels, prices and volatility of medical spending, as in the US deteriorating health with age leads to higher spending there while this is not true in the UK because of the National Health Service (NHS). Figure 4.1 shows paths of non-durable spending including and excluding out-of-pocket medical spending in the two countries. It is immediately clear that excluding medical spending helps account for a significant fraction, though not all, of the difference between the two countries.[2] As we detail below, once medical expenditures are removed, the difference in the decline in spending between the two countries shrinks by around three quarters. Different papers have made different decisions about whether medical expenditures should be included in the definition of non-durable consumption. For instance, Heathcote et al. (2010) and Attanasio and Pistaferri (2014) include medical spending in their measures of expenditure while for instance Attanasio and Weber (1995), Banks et al. (1997), Blundell et al. (2008), and Attanasio et al. (2012) do not (often on the grounds that spending on healthcare is more akin to investment than consumption spending). Our results highlight the importance of giving careful consideration to such choices.

Medical spending is not the only difference between the two countries however. We therefore move on to quantify cross-country differences in three potential factors - employment, housing status and health - and look for any immediate differences that might explain the differential consumption paths observed in Figure 4.1. While there are some differences in the way these variables evolve in the two coun-

[2]Changes in medical spending at older ages could in principle be driven by changes in medical consumption in the two countries or differences in the prices paid for medical care. Purchasing Power Parities (PPP) for medical care from the OECD suggest that the level of prices (paid by both government and consumers) is consistently higher in the US than the UK (see `http://stats.oecd.org/Index.aspx?DataSetCode=PPP2014`). In 2005, for example, UK prices were estimated to be 78 percent of costs in the US. In Appendix B.2, we also consider the rate of change of medical prices in the US versus the UK for the period 1988-2010. Price movements in the two countries track each other quite closely for much of this period but US medical price inflation is higher in the latter years of the sample. If medical care is a normal good, this would tend to reduce US consumption of medical care relative to the UK.

tries, these differences do not seem large enough to account for the cross-country difference in spending patterns.

We examine this hypothesis more formally in a regression context, finding that controlling for these factors only marginally reduces the cross-country difference in the decline in non-durable consumption spending with age when medical expenditure is included. We then turn to model non-medical consumption conditional on health status and real medical expenditures. This approach allows preferences for non-medical consumption to change in a non-separable way with health and the consumption of medical goods. It also captures any substitution effects driven by the change in the relative price of medical consumption. We also consider the role medical expense uncertainty may play in explaining consumption profiles in the US, partly by exploiting differences in the institutional environments in the two countries. We find suggestive evidence that precautionary savings against medical expense risk play an important role in US consumption decisions. Controlling for both medical uncertainty and relative prices fully explains the cross-country difference in spending declines. Our regression estimates imply that medical uncertainty increases consumption growth at older ages in the US by around 0.90 percentage points per year on average for the ages we consider. Precautionary motives against medical expense risk in the UK are, by contrast, negligible.

The rest of the chapter is organised as follows. In the next section, we describe in more detail the essential features of the data we assemble to look at these issues and document cohort specific paths of non-durable spending and household income for both countries. We then move on to look at various potential explanations for the cross-country differences in turn. To illustrate, Section 4.2 provides a description for cohort specific age paths in employment in the two countries and discusses their implications for consumption profiles, Section 4.3 provides a parallel treatment for housing by describing age paths of housing ownership and Section 4.4 focuses on levels and paths of health status and differential levels and age patterns of medical expenditures. Section 4.5 presents results obtained from an inter-temporal model of growth rates in total non-durable expenditures for each country to identify factors that may account for different shaped consumption paths at older ages. The final section highlights our main conclusions.

## 4.1 The life-cycle pattern of consumption and income

We use two repeated cross-sectional surveys widely viewed as containing the highest quality measurement of household expenditure and its components in each country  the Consumer Expenditure Survey (CEX) in the US and the Living Costs and Food Survey (LCFS) in the UK. While these surveys do not cover the same individuals for long periods of time, we organise the data to create a pseudo-panel and track cohort consumption behavior by age (in the manner of Browning et al. (1985)). To do this we group individual observations by 5-year birth cohorts and take averages within each year. Cohorts are determined by the age of household head. Following this approach allows us to merge in information from other surveys at the cohort-year level where necessary.

The LCFS is an annual cross-sectional survey that has been running in one form or another since 1961. The LCFS, formerly known as the Family Expenditure Survey, is conducted by the Office for National Statistics (ONS), the UKs national statistical agency and has been the basis of a number of studies of intra- and inter-temporal spending patterns. Currently it interviews around 6,000 households throughout the UK and continuously throughout the year. The survey begins with an interview with questions about demographic characteristics, income, large purchases over the last year and regular expenditures (such as magazine subscriptions, internet subscription costs and so on). Each household member over 16 then records all spending in a diary over the next two weeks.

For the USA, we make use of the Consumer Expenditure survey (CEX). This survey has carried out by the Bureau of Labor Statistics (BLS) on a continuous basis since 1980. For some quarters prior to 1984, the survey only covered households living in urban areas. The CEX includes two separate surveys, a diary survey which works much like the LCFS, and an interview survey, where households are asked to recall their spending on a range of spending categories over the previous three months. The interview survey is also a short panel, as the same households are interviewed on up to 5 occasions. The first of these interviews collects some basic data on family characteristics. Each subsequent interview updates this information and asks questions concerning household spending over the previous 3 months. Information on incomes and labour force participation are however only collected in the 2nd and 5th interviews (except for new household members and members who have newly started work), meaning that income and spending data for the 3rd and 4th interviews need not cover the same time periods. In this chapter we only make use of the interview survey.[3] Around 5-8000 households are interviewed in each quarter.

**Table 4.1:** Spending categories

| | |
|---|---|
| Food in | Food at home |
| Food out | Food in restaurants, school dinners, catering. |
| Other non-durables | Alcohol, tobacco, clothes, books, child care, pet goods and services. |
| Medical | Health insurance premia, fees for services from health professionals, drugs, medical equipment, care in nursing homes, care of invalids. |
| Housing related | Electricity, gas and water bills, domestic services, repairs, building insurance. |
| Recreation | Sporting goods, musical instruments, CDs, entertainment, holidays |
| Transport | Motoring costs, petrol, fares for public transport, air fares. |
| Durables | Vehicles, appliances, entertainment equipment. |

In both UK and US surveys, spending data are provided for hundreds of highly disaggregated individual product codes. We allocate these goods into 8 broader categories defined to be consistent across the two countries: food in, food out,

[3] While the methodology employed in the CEX diary survey is arguably more similar to that used in the LCFS than the interview survey, the diary survey has lower sample sizes, tends to exhibit greater variability in responses, and tends to under report spending relative to the interview survey (Bee et al. (2015)). For these reasons, we make use of the interview survey instead.

other non-durables, medical, housing related, recreation and transport and durables. Some examples of what are included in these categories are given in Table 4.1. We do not include rental payments or mortgage interest in any of these definitions as we do not observe the shadow price of owned housing in the LCFS, nor can we estimate it easily (the CEX does include a self-reported imputed rental cost for owned properties). We define total non-durable expenditures to include all rows in Table 4.1 with the exception of the final row measuring durable spending.

Household income data are derived from the same surveys and cohort age profiles obtained in the same manner. Household income is defined comprehensively to include all sources of income for the head of household, the spouse/partner, and all other household members net of taxes. US expenditures and incomes are deflated to 2010 terms using the Consumer Price Index (CPI). UK variables are deflated to 2010 terms using the Retail Prices Index and then converted into dollars using PPP exchange rates for that year taken from the OECD. Both surveys contain measures of standard definitions of labour force participation. From 1994 onwards, the CEX also contains detailed questions on the nature of households health insurance policies and Medicare coverage. In both data sets we restrict our attention to households where the head is aged 45-79. This is because ages in the LCFS are top-coded at age 80 from 2002 onwards.[4]

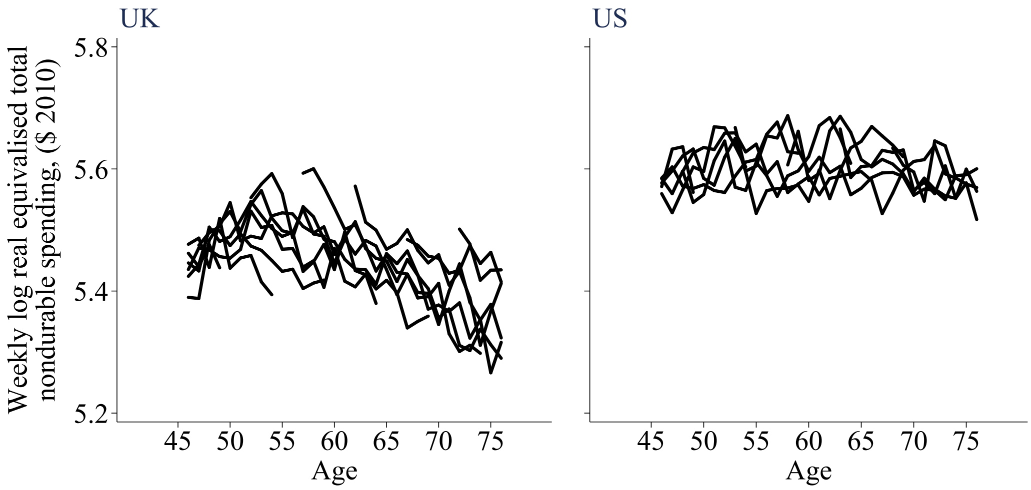
To control for measurement error and impacts of extreme values on life-cycle paths, we trim households in the top or bottom 1 percent of distribution of income and expenditure. In the CEX we take data from 1984 (to consistently include a nationwide sample) until 2010. For the LCFS we take data from 1978 until 2010. We stop in 2010 in both countries as we do not have mortality data for either country after this date.

Figure 4.1 shows spending at different ages average across different birth cohorts and different years. This means that differences between the two

countries shown there may partly be driven by differences in cohort and time

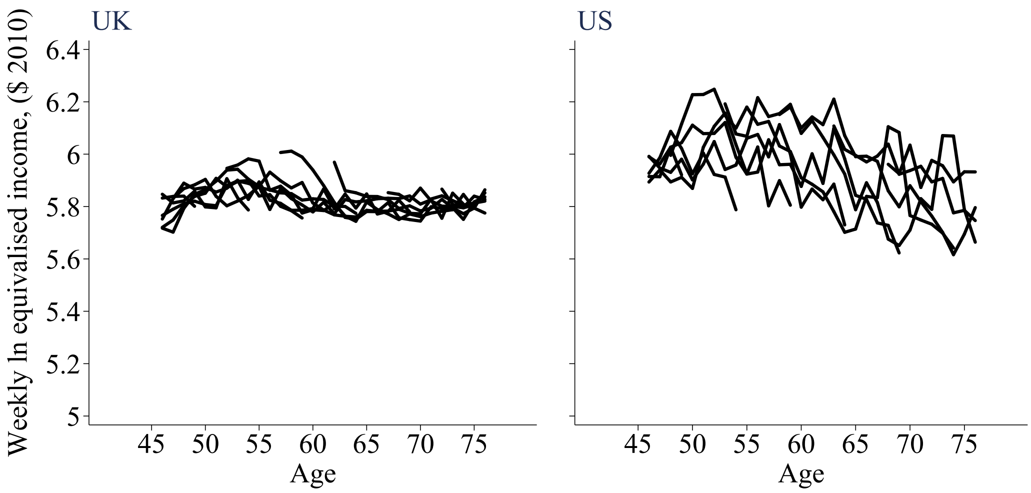
[4]We also plotted spending, income and demographics up to age 85 in the two countries using data up to 2001 only. The patterns in the two countries are very similar. Results are available on request.

**Figure 4.2:** Non-durable spending by cohort and age




Notes: Data from LCFS in the UK and CEX for the US. Each line represents average log non-durable expenditures at each age for 5-year birth cohorts over the periods they are observed between ages 45 and 79 over the period 1984-2010. Average differences across cohorts are removed by regressing spending on cohort dummies and taking the residuals. Values are in US$ (2010). UK prices are converted to dollars with PPP indexes. Figures equivalised using the modified OECD scale.

**Figure 4.3:** Log household income by cohort and age




Notes: Data from LCFS in the UK and CEX for the US. Each line represents average log incomes at each age for 5- year birth cohorts over the periods they are observed between ages 45 and 79 over the period 1984-2010. Average differences across cohorts are removed by regressing incomes on cohort dummies and taking the residuals. Values are in US$ (2010). Figures equivalised using the modified OECD scale.

effects. To understand whether the patterns in Figure 4.1 are driven by cohort effects, Figures 4.2 and 4.3 show how spending and incomes decline within cohorts in the two countries. Before plotting these, we remove average differences across cohorts by regressing spending and income on cohort dummies and taking the residuals. It is clear that cohort effects by themselves cannot account for the main puzzle with which we motivated this chapter. Although the spending decline observed in the UK is somewhat smaller when one looks within individual cohorts rather than averaging across them, the age pattern non-durable consumption at older ages in the USA remains relatively flat. Within cohort declines in incomes are also similar across the countries.[5]

## 4.2 Differences in employment and retirement

One dimension of labour force behavior at older ages that has been studied in the context of consumption age profiles involves the impact of retirement on levels and time paths of consumption. Consumption levels and paths may not be independent of the retirement decision if preferences over employment and consumption are not separable, or individuals do not fully anticipate income reductions coincident with labour market retirement (Banks et al. (1998)). The importance of this in explaining consumption trajectories at older ages is substantial. In the US, it has been estimated that work related expenditures account for the entire decline in non-durable spending from middle age to age 75 (Aguiar and Hurst (2013)). In addition to any direct costs associated with work, movements out of employment may also be associated with having more time to spend shopping for discounts or for home production of some goods (Aguiar and Hurst (2007)). This could partially explain cross-country differences if there are differences in the links between labour supply and consumption expenditures in the two countries, or if declines in employment were more rapid in one country than another (or both).

[5] In the CEX there were two changes to the way incomes were measured that matter for Figure 4.3. One occurred in 2001 and the other in 2004. The first introduced a bracketing question for those who did not report their incomes first time round. The second introduced imputation for non-responders. The income definition we employ makes use of non-bracketed responses only from 2001 and non-imputed values for income from 2006 onwards. In 2004 and 2005 it is not possible to remove non-imputed income values.

These declines in male employment by age are somewhat more rapid in the UK compared to the United States. However, in the absence of non-separabilities in employment and consumption, differences in paths of employment at older ages in the two countries do not seem large enough to be the major explanation for the substantial differences in consumption profiles. We will examine the role of non-separabilities between labour supply and consumption in explaining the cross-country difference in consumption profiles in more detail in Section 4.5 below.[6]

## 4.3 Housing ownership and downsizing

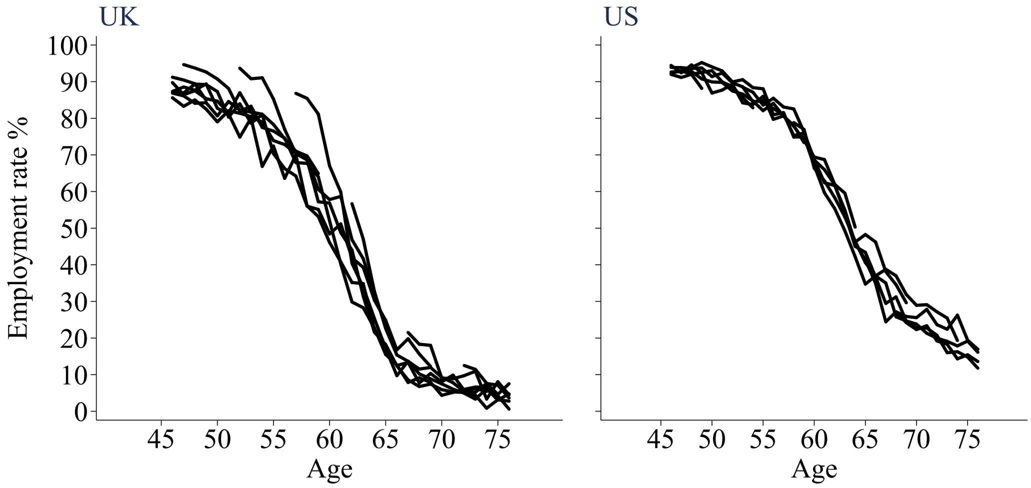
Housing related decisions and expenditures represent another spending category in which there are important institutional differences between the countries that may affect levels and age paths of expenditures at older ages. We have provided evidence in other work that there exists far less geographical mobility in Britain compared to the United States and more downsizing in the US compared to the UK as a meaningful fraction of older Americans move to smaller homes (i.e., fewer rooms) with little evidence of such downsizing in Britain (Banks et al. (2010, 2012)). While this lower rate of British mobility was characteristic of both owners and renters, the differential was particularly high among renters.

For British households over age 50, the probability of being a homeowner is about thirteen percentage points lower than for an American household, a deficit mostly offset by a higher probability of renting in highly subsidised social housing. The major secular changes in housing tenure at older ages have decidedly taken place in the UK and not the US. The fraction of older British people owning their own home increased by almost thirty percentage points (from less than half to over 80 percent) from the 1908-12 cohort to the 1943-47 cohort. In contrast over the same set of birth cohorts and age groups, the fraction of older American households who were home owners has remained relatively stable at around 80 percent.

The primary reason for this secular change in home ownership rates for older British households is due to changes in the proportion of individuals in social hous-

[6] Age paths for women (not shown) also display the same pattern of rapid declining employment rates with age as women exit the labour force in both countries.

**Figure 4.4:** Employment rates: men by cohort and age




Notes: Data from LCFS in the UK and CEX for the US. Each line represents average employment rates for men at each age for 5-year birth cohorts over the periods they are observed between ages 45 and 79 over the period 1984-2010.

ing. In the UK, there is a system of subsidised housing, often referred to as local authority, social or council housing. Those who are allocated a property pay a below-market rent, and the landlord will be either the local authority or a housing association. Individuals entitled to such a rental property are placed on a waiting list until suitable accommodation becomes available. While entitlement to live in social housing is subject to a strict means test, once allocated a property, tenants can usually stay for life irrespective of any changes in circumstance. Social renters have a severely reduced incentive and ability to move or to downsize their property, for several reasons. Even if a tenants current circumstances mean that they are still entitled to social housing, moving can be very difficult because of shortages of social housing. Existing tenants are treated the same as new applicants, so if they are not in a priority group, they may not be allocated a different property. For those whose circumstances have changed in such a way that they would no longer be entitled to social housing if they were to reapply, there is a large incentive not to move as they may not be allocated a different property at all and may have to move into the private sector and pay full market rent.

There has been a sharp across cohort decline in social rental housing in the UK that parallels the increase in home ownership across cohorts (which for space considerations we do not plot). There was an almost 30 percentage point decline in the fraction of British households in social rental housing, which is pretty much the same percentage point increase observed in home ownership. Over the same set of birth cohorts, ages, and years there was little change in the fraction of households in private rental housing. These changes reflect the introduction of a Right-to-buy in 1980 which required local authorities to sell council-owned housing at a discount to eligible tenants (the policy was later extended to other forms of social housing).

The differences in levels and trends in ownership patterns between the two countries may partially contribute to an understanding of the differences in age-consumption profiles. We examine the impact conditioning on these differences might play in Section 4.5 below.

## 4.4 Health and the divergence of medical expenditures

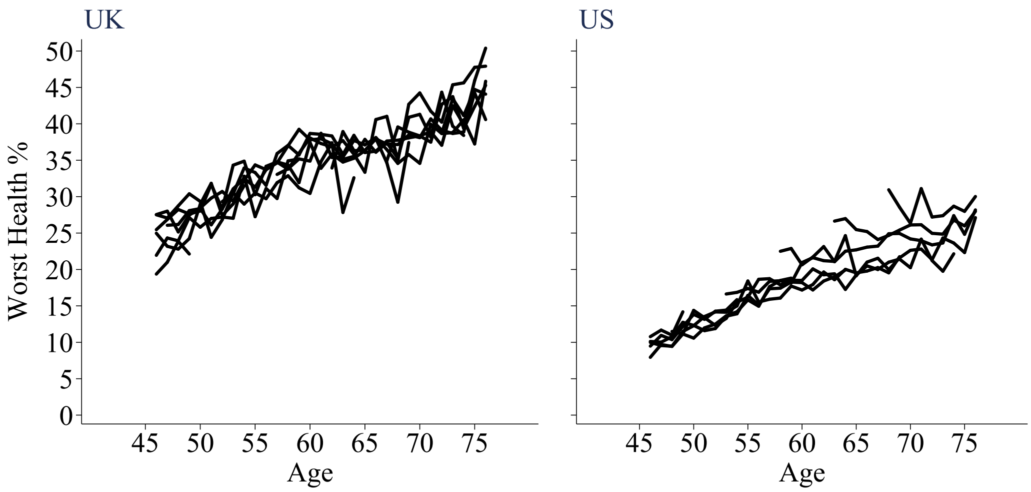
Our health measures are based on self-reported health status, age specific mortality rates, and out-of-pocket medical expenditures by cohort, age, and gender. Neither the CEX nor LCFS include information on health or mortality, so we draw these from other sources.

### 4.4.1 Health status

For the UK health status data come from two cross-sectional surveys, the Health Survey for England (HSE) and the General Household survey (GHS). These surveys contain information on households self-reported health which we average by age, sex and cohort. Two surveys are used as we do not have GHS data after 2006, and HSE data before 1991. In addition, there are two breaks in the GHS (in 1997 and 1999), due to redesigns of the survey, which interrupt the series. We make use of GHS data up to 1997 and HSE data from 1997 onwards. In the GHS, respondents are asked about their general health status over the last 12 months which they answer on a three-point scale: answers can be good, fairly good, or poor. In the HSE, households are asked to report their general health on a 5-point scale -very good, good, fair, bad, or very bad. For consistency, we group these into three categories (by putting the final three responses into a single worst health group). We then average health status by age, year, and sex and use this information to impute health of household heads in the LCFS. The switch from the GHS to the HSE surveys introduces a downward shift in the level of self-reported health statuses beginning in 1997. In what follows we remove this discontinuity by regressing health status in both surveys on a GHS dummy and taking the residuals. To our self-reported health data, we add data on mortality rates by age, sex and cohort/year from the ONS Mortality tables. For the US we use the National Health Interview Surveys (NHIS). NHIS is an ongoing nationwide survey of about 40,000 households. Since 1982, NHIS used a 5-point scale to measure respondents general health status Would you say your health in general was excellent, very good, good, fair, or poor? We create three categories for consistency with our UK measure. These three groups are excellent or very good, good, and fair or poor. We use these to impute health statuses to household heads and spouses in the CEX in the same way we do for the LCFS. We also calculate the proportion of responses that are self-reported in each cell to use as a control. Mortality data for the United States are obtained from the Berkeley life tables which also give death rates by age, gender and year (`http://www.demog.berkeley.edu/~bmd/states.html`).

Figure 4.5 plots proportions of those in worst health in both countries showing several distinct patterns in health status in both countries. First, levels of worse health are always higher in the UK than in the US. However, these different levels of subjective health status in the UK compared to the US have been shown to be due to different subjective health thresholds between the two countries. In the age groups we are considering, the British are typically healthier than the Americans with prevalence of almost all diseases higher in the US compared to the UK (Banks et al. 2006). At the same objective health levels, the British report themselves in worse health on subjective scales. The second pattern to note in Figure 4.5 is that

**Figure 4.5:** Proportion of responders in worst health by cohort and age




Notes: Data for the UK is from the HSE and GHS surveys spliced together (adjusted to remove discontinuity between the surveys). Data for the US is from the NHIS. Each line represents proportion of household heads reporting being in the worst health condition at each age for 5- year birth cohorts over the periods they are observed between ages 45 and 79 over the period 1984-2010.

the fraction of a cohort in poor health rises with age in both countries. The third pattern concerns cohort effects in these paths of health at older ages. While there is little evidence of cohort differences in the UK, cohort differences are however apparent in the US. Finally, we note that subjective health declines faster with age in the UK than the US. We attempt to account for the potential role of health status in explaining the different expenditure patterns we observe in Figure 4.1 in our regression analysis below.

The impact of declining health on consumption decisions in a life-cycle model will depend on how it affects the marginal utility of consumption. If poor health reduces the marginal utility of consumption, then we will observe that consumption declines more steeply with age as health deteriorates. Various papers have investigated the dependence of the marginal utility of consumption on health without achieving consensus on either its sign or magnitude (Finkelstein et al. (2009)) for a survey of the available literature). Lillard and Weiss (1997) find that there is substantial positive effect on marginal utility using panel data on consumption (as inferred from income flows and asset changes) and health shocks. By contrast, employing a novel approach that combines data on permanent income, utility proxies and health data, Finkelstein et al. (2013)) find a substantial negative effect. Other studies have essentially found no effect. DeNardi et al. (2010)) estimate a model allowing preferences over consumption to be health dependent. They find that the parameter governing the effect of health on the marginal utility is negative but statistically insignificant.

### 4.4.2 Life expectancies and age paths of mortality

We present information on life expectancies at different ages in two countries in Table 4.2. The top panel shows life expectancies in 1984. The bottom panel shows equivalent figures for 2010. For both men and women, life expectancies at each given age tended to be greater in the US than the UK in the early part of our sample (these differences had largely disappeared by the end of our sample period in 2010).

In the standard life-cycle model, higher age specific mortality risk acts like a decline in the interest rate encouraging current consumption and producing a steeper decline in consumption with age. Mortality risk rises steeply with age in both countries with mortality risk about ten times larger at age 70 compared to age 45. There is evidence of cohort improvements in mortality that are larger in the UK compared to the US. However, the shape of the age mortality risk function appears to be similar in the two countries suggesting once again that differential mortality risk by age, see Hurd (1989), does not appear to be the likely source of the significantly differently age shapes in consumption in the two countries documented in Figure 4.1. In any case, we account for mortality's potential role in explaining spending differences within a regression framework in what follows.

### 4.4.3 Medical expenses

On the health side of potential explanations, we have so far explored age patterns at older ages in general health status and mortality. While both health dimensions may play a role in shaping consumption profiles at older ages, their ability either alone or together to account for the much flatter non-durable consumption with age in the United States compared to the UK seems limited. The final health dimension we

**Table 4.2:** Life expectancies at different ages, 1984 and 2010

| Age | UK | US |
|---|---|---|
| | 1984 | |
| | Males | |
| 60 | 16.60 | 17.79 |
| 65 | 13.24 | 14.48 |
| 70 | 10.31 | 11.52 |
| 75 | 7.86 | 8.99 |
| 80 | 5.91 | 6.82 |
| | Females | |
| 60 | 21.06 | 22.48 |
| 65 | 17.20 | 18.63 |
| 70 | 13.63 | 15.03 |
| 75 | 10.44 | 11.77 |
| 80 | 7.70 | 8.85 |
| | 2010 | |
| | Male | |
| 60 | 22.03 | 21.64 |
| 65 | 18.03 | 17.89 |
| 70 | 14.33 | 14.39 |
| 75 | 11.00 | 11.19 |
| 80 | 8.10 | 8.37 |
| | Females | |
| 60 | 24.92 | 24.63 |
| 65 | 20.66 | 20.50 |
| 70 | 16.61 | 16.63 |
| 75 | 12.88 | 13.05 |
| 80 | 9.55 | 9.83 |

Notes: For the UK these are taken from the ONS lifetables. US figures are obtained from the Human Mortality Database (`http://www.demog.berkeley.edu/~bmd/states.html`) examine - health expenditures - appears to us to offer far more potential since there are large differences between the two countries. While consumption of medical services may increase in both countries as individuals age, differences in how the costs of these are financed will show up as differences in both the level of measured out-of-pocket expenditures and their dispersion.

How health costs are financed at older ages in the two countries are quite different. To a large extent, UK medical costs at all ages are paid by the state with very little absorbed by the individual. State provision not only includes medications and doctor visits, but hospitalizations as well. Charges are however typically levied for prescription drugs and dental care. There are also often charges for long term care costs as we discuss below.

The situation is very different in the US where government assistance for health care is incomplete and a large proportion of the costs of medical insurance are met by employers or directly by households rather than by government. Government assistance for health care in the US is mostly provided through the Medicare and Medicaid programs.[7] Figure 4.6 shows enrolment under the two schemes over the ages we consider. Medicare provides some insurance for the vast majority (over 90 percent) of households with heads over 65 but only a limited proportion of younger households. The share of households which report receiving some support from Medicaid increases somewhat from around 7 percent to around 10 percent as individuals age from 45 to 75.

While previous studies have found that Medicare eligibility reduces both the mean and variance of out-of-pocket (OOP) medical expenditures (Barcellos and Ja-

[7]Medicare is a government insurance program for the elderly. Most individuals become eligible for the scheme when they turn 65. Eligibility is automatic for those who have worked and accumulated Social Security credits for at least 10 years prior to reaching this age, but those who do not meet this requirement may also qualify on the basis of their spouses contribution history. There are however some groups who can qualify at younger ages. For example, those who have received Social Security disability benefits for at least 24 months automatically receive partial coverage. Around 12 percent of the population is already enrolled by the time they reach age 65 (Card et al. (2009)). Medicaid is general scheme that provides reduced cost or free health services for low-income and low wealth households, including those attempting to meet the costs of their long term care. Exactly who or what is eligible varies from state to state with the federal government specifying minimum standards of coverage. Over half of long term care costs are paid through Medicaid (O'Shaughnessy (2014)).

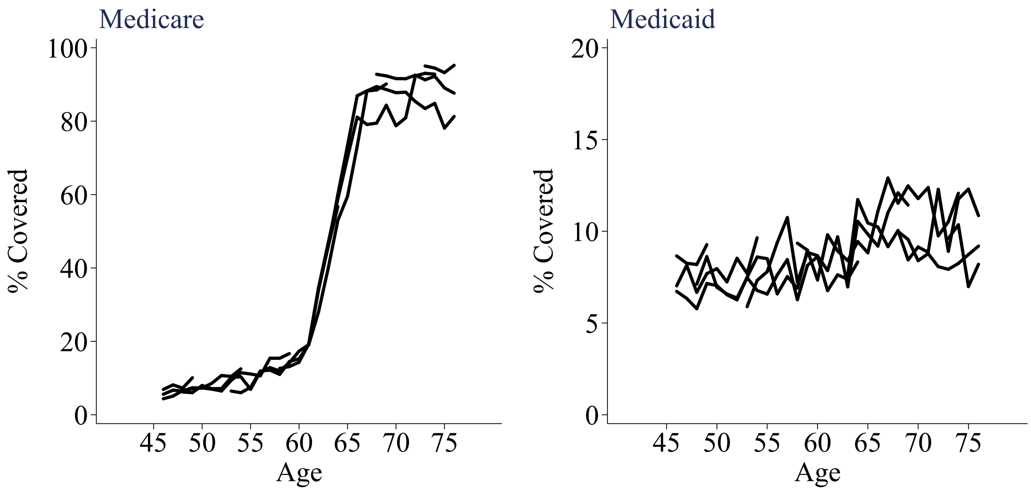
cobson (2015)), it does not eliminate the need for them entirely. Coverage is neither free nor comprehensive with various direct costs for households. While hospital insurance (Medicare Part A) is typically provided free of charge, insurance for doctors services and prescription drugs (covered under Parts B and D) involve income-contingent premia. Individuals covered under Medicare Part C (or Medicare advantage) contract with a private company to receive their part A and B coverage and may pay a higher premium for additional coverage. In addition, Medicare does not cover the costs of all treatments and even when treatments are covered patients must pay deductibles, co-payments and co-insurance from their own resources. A further institutional difference between the two countries is that, in the US, a large fraction of individuals have their private insurance costs covered by third parties (usually employers). This proportion tends to decline with age however as individuals retire and leave the labour market. Prior to age 65, a majority of American households have their insurance at least partially paid for by some third party but this falls to around 40 percent at age 70 as the left panel in Figure 4.7 shows. Similarly, the proportion of households who have insurance but pay nothing (shown in right panel of Figure 4.7) falls from 20 percent at age 45 to less than 3 percent at 75. For workers, the share of health costs paid by employers is substantial, at around 75-80 percent of the total.[8]

The institutions in the two countries naturally have consequences for paths of medical expenditures as individual's age. We plot the budget shares for medical spending in for the two countries in the two panels of Figure 4.8. Not only are medical costs in the UK lower as a share of the budget (always under 5 percent), but there are only modest increases in this share with age. In contrast, the US graph indicates much higher and sharply rising medical costs shares at older ages in the US that are not due solely to cohort effects. To illustrate, medical costs shares in the US are approximately eight percent at age 45 and rise steadily until they are around 20 percent of the total budget by age 70. The decomposition of these medical

[8] See Exhibit 4.1 in `http://meps.ahrq.gov/mepsweb/data_stats/MEPSICChartbook.pdf`

expenditures for a single cohort is shown in Figure 4.9.[9] In the UK, the majority of medical spending goes towards non-insurance costs. In the US, insurance premia are far more important. Medicare spending begins to rise when the head reaches age 65 but the trajectory of overall spending is smooth.

**Figure 4.6:** Proportions covered by government programs, US



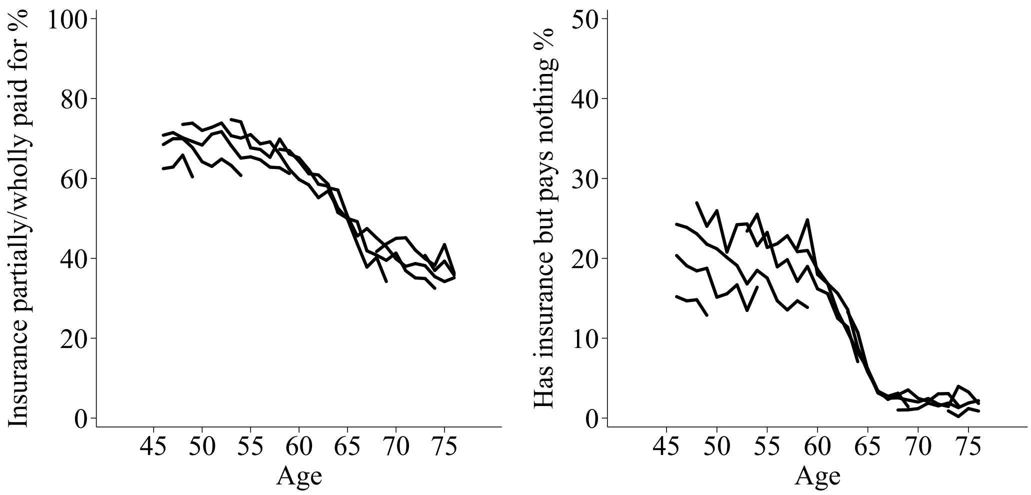

Notes: Data from CEX. Each line represents proportions of households with at least one member covered by Medicare or Medicaid at each age for 5- year birth cohorts over the periods they are observed between ages 45 and 79 over the period 1994-2010.

Information on the distribution of medical expenses, and the riskiness of such expenses, is harder to come by, particularly in the UK. Table 4.3 compares the distribution of annual OOP medical expenses by major categories in the UK and US, for all households aged 60 or over.[10] The HRS only includes medical equipment spending in later years and so these are not included in the US data. Consistent with the graphs for the 1928-1932 cohort in Figure 4.9 above, the table shows that average costs in the US are almost seven times larger in the US than they are in the UK, with a mean of over $5,201 per year compared to just $762 in the UK. Even though insurance makes up proportionately more of the US expenses, the country

[9]Results from other cohorts are very similar.

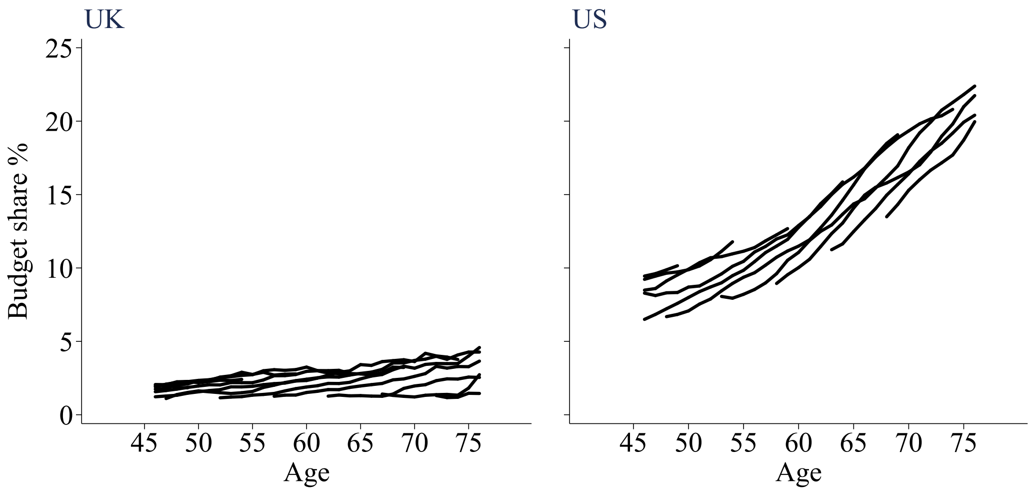
[10]The best source of information to breakdown such expenses is the longitudinal ageing surveys, and we use the US Health and Retirement Survey for this analysis. Since the level of out-of- pocket medical expenses is so low, the English equivalent of the HRS does not collect information on such spending, so we use the cross-sectional LCFS data as in rest of our analysis above.

**Figure 4.7:** Insurance paid for by others, US



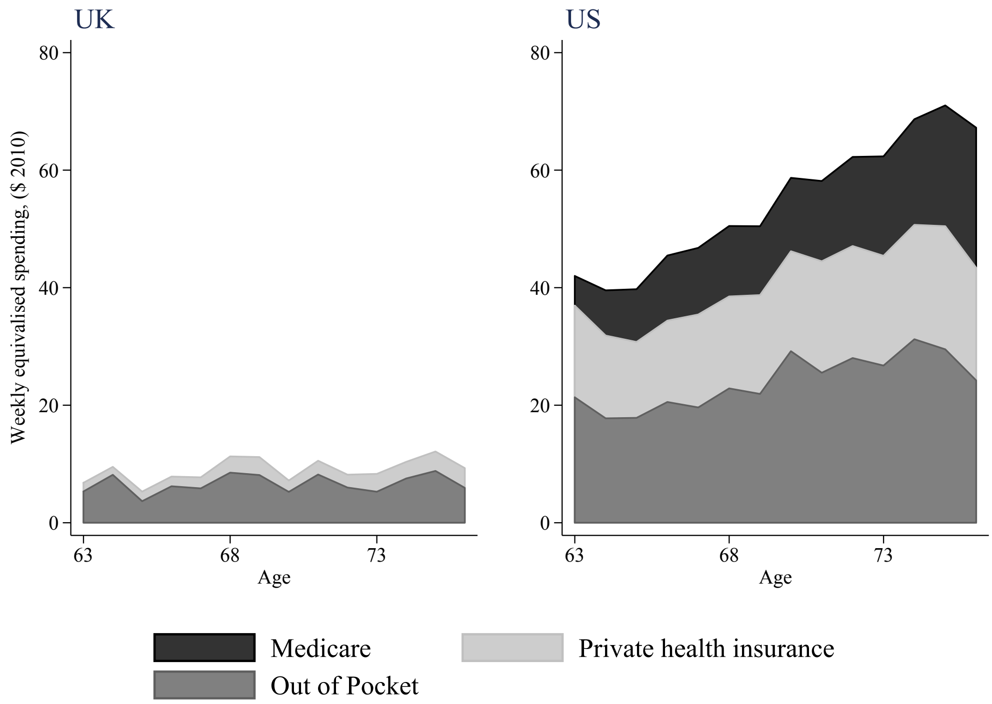

Notes: Data from CEX. Each line represents average coverage rates at each age for 5- year birth cohorts over the periods they are observed between ages 45 and 79 over the period 1994-2010. The left panel shows the proportion of households who report insurance policies wholly or partially financed by third parties. The right panel shows the proportion of households who pay no insurance costs but report being covered by insurance paid for by third parties.

**Figure 4.8:** Share of cohort spending on medical care




Notes: Data from LCFS in the UK and CEX for the US. Each line represents average budget shares out of non-durable expenditures at each age for 5- year birth cohorts over the periods they are observed between ages 45 and 79 over the period 1984-2010.

**Figure 4.9:** Composition of out-of-pocket medical spending (1928-32 birth cohort)




Notes: Data from LCFS in the UK and CEX for the US. Values shown over the period 1994-2010. Values are in US$ (2010).

differences are of the same order of magnitude if we exclude insurance payments. But the US data also exhibit considerably greater variance. To illustrate, health expenses at the 95th percentile are around 17,313 per year (compared to $3,788 in the UK), indicating a much larger risk of very large medical costs in the US.[11]

One final "institutional" difference between the two countries may be in the nature or extent of family ties and caring by family members, and this may have effects on medical expenses. A full investigation of the links between family care and other medical expenses is an important topic for future research, but it is beyond the scope of this chapter. We briefly investigated the link between health, family care and OOP medical expenses in the HRS data. For individuals reporting three

[11] Since the HRS data are a panel, we can also look at longer term spending totals, and indeed the persistence of expenses over time. As well as being highly concentrated, medical expenses are also shown to be strongly persistent over the six-year period, with the correlation between total medical expenditures in 2002 and total medical expenditures two and four years later being 0.66 and 0.6 respectively. (Full results available from authors on request).

**Table 4.3:** Yearly out-of-pocket medical expenditures by country- 2000-2006 Age 60+

| A. UK | | | | | | |
|---|---|---|---|---|---|---|
| Variable | Mean | P25 | Median | P75 | P90 | P95 |
| Total | 762 | 0 | 46 | 375 | 1,729 | 3,788 |
| Excluding insurance | 574 | 0 | 13 | 255 | 989 | 2,243 |
| Private insurance | 188 | 0 | 0 | 0 | 88 | 881 |
| Prescription drugs | 118 | 0 | 0 | 129 | 342 | 553 |
| Health services | 234 | 0 | 0 | 0 | 0 | 1,118 |
| Hospital | 41 | 0 | 0 | 0 | 0 | 0 |
| Medical equipment | 180 | 0 | 0 | 0 | 0 | 145 |
| B. US | | | | | | |
| Variable | Mean | P25 | Median | P75 | P90 | P95 |
| Total | 5,201 | 443 | 2,458 | 6,125 | 11,929 | 17,313 |
| Excluding insurance | 3,361 | 225 | 1,122 | 3,025 | 6,568 | 11,152 |
| Private insurance | 1,772 | 32 | 509 | 1,592 | 3,711 | 5,889 |
| Prescription drugs | 1,841 | 0 | 150 | 2,429 | 5,701 | 8,236 |
| Health services | 964 | 6 | 189 | 718 | 1,838 | 2,952 |
| Hospital | 301 | 0 | 0 | 0 | 365 | 1,062 |

Notes: Data from the Health and Retirement Survey in the US and LCFS in the UK. Values are annual averages for households where at least one member is aged 60 or over. Values are in US$ (2010). Figures exclude spending on nursing homes.

or more limitations in Instrumental Activities of Daily Living (IADLs), 97 percent reported receiving some assistance from family, but this had no relationship with OOP expenses. In the UK, we cannot make a similar calculation since there is no dataset with OOP expenses and health, disability or the receipt of family care, however since OOP expenses are so low for so many individuals, as discussed above, such a relationship between family caring and OOP medical expenses is unlikely to be important.

### 4.4.4 Long term care costs

One important source of medical cost uncertainty is in the cost of long term care. This tends to be most important at older ages (for instance, rising over three-fold in the US for those aged over 85 compared to those aged 75-84 (Fahle et al. (2016)). However, in so far as these expenses also generate precautionary motives, they may also affect spending behavior of households within our sample (Ameriks et al. (2015)).

In the UK, long term care costs are not typically covered by the NHS, though care costs are often paid for, wholly or partially subject to a means test of resources by local authorities. Estimates on the relative importance of private versus public spending on long term care indicates that the majority of costs in the UK are paid for by the public sector. Private spending on formal care is roughly half the value of spending by local authorities (National Audit Office (2014)) and only around a quarter of over 65s receiving formal care report paying for it themselves (Crawford and Stoye (2017)).

In the US Medicare does not directly cover the costs of long term nursing care, though it can cover related costs such as care in skilled nursing facilities and home health care. Longterm care costs are often covered under the Medicaid program, subject to a means test of resources. In 2004, the proportion of total long term care costs paid for under these two programs was nearly 60 percent (Congressional Budget Office (2004)).

Despite differences in the institutions for funding long term care costs, both the overall level and proportion of long term care financed through private spending is similar in the two countries (OECD (2005)). Census data show that the proportion of population aged 65 and over who are resident in institutions is also very similar in the two countries at around 3.6 percent in the UK and 4.1 percent in the US (Peeters et al. (2013), Fig.1). Nursing home costs are not well covered in our household expenditure surveys so to make what comparisons we can we draw on the English Longitudinal Study of Ageing (ELSA), which only includes nursing home care costs in its most recent wave (covering spending in the period 2014-2016). We then compare this to the latest wave of the HRS to which we have access (covering the period 2012-2014). Even in these two surveys, which focus specifically on the older population, the measurement of costs, and even the coverage of the survey, is not comparable for those who are resident in institutions, with the main difference being that the ELSA data does not currently include any measures of spending for those currently residing in institutions. In this respect, HRS data has 3.7 percent of households over aged 60 with at least one member resident in an institution and a mean spending over the last two years of $847 in 2010 prices. This is lower than all but one component of out-of-pocket medical expenses identified in panel B of Table 4.3 for the US. But the distribution is highly skewed for those who do incur costs (median out-of-pocket spending over the previous two years amongst those in institutions was $930, the 75th percentile was $31,157 and the 95th percentile was $104,950). The ageing surveys do, however, have comparable measures for out-of-pocket nursing home spending over the last two years for those currently residing in the household sector. Once again, mean spending is low although a minority of households pay high costs. These patterns are similar in the two countries. 98.3 percent of the US household population over aged 60 either did not use nursing home or institutional care in the previous two years or else paid nothing for their usage. The corresponding number in England is 99.4 percent. Mean annual spending was $53 in the US and $30 in England and, conditional on having to pay something the top of the distribution in each country was rather similar. Further details of the distribution of these transitory nursing home costs is in Table B.1 of Appendix B.1.

Taking all this evidence together, it is clear that nursing home costs are small on average but a significant expense but for a small minority of households as would be expected. But the risks of high nursing home expenses and the size of the out-of-pocket costs if they are incurred are both somewhat similar in the two countries.

## 4.5 Inter-temporal allocations of consumption

In the previous sections we noted possible links between trends in demographic variables and consumption at older ages. We highlighted differences in particular in the decline in employment, and the pattern of home ownership between the two countries. We also noted strikingly different patterns of medical expenditures, summarised in Figure 4.8, largely reflecting differences in the delivery of health services in the US and the UK.

To motivate our regression analysis of consumption growth, we consider the case where intertemporal preferences for non-medical consumption had the CRRA form, and where health and medical consumption is non-separable with non-medical consumption. We then write the following (approximate) *conditional* Euler equation governing intertemporal spending allocations[12]

$$\Delta \ln c_{it} = \ln r_t + \Delta X_{it} \beta + \zeta \Delta H_{i,t} + \eta \Delta \ln p_{i,t} + u_{i,t} \quad (4.1)$$

where $\Delta$ is the first difference operator (i.e. $\Delta x_t = x_t - x_{t-1}$), $r_t$ is the real interest rate, $c_{i,t}$ is non-medical consumption, $\Delta X_{it}$ the change in a variety of demographic and household characteristics which we detail below, $\Delta H_{i,t}$ is a measure of the change in health status by household members. The change in the real price of medical consumption $\Delta \ln p_{i,t}$, captures the non-separability with medical consumption.[13] For example, this price term allows for substitution away from medical consumption as the relative price of medical consumption increases.

In the application, we additionally allow for uncertainty in medical expenses that might induce precautionary saving. To do this we follow Banks et al. (2001), and incorporate an additional conditional variance term in the consumption growth equation (4.1) to reflect uncertainty over shocks to future medical expenses.[14] This is explained in more detail in Section 4.5.2 below.

### 4.5.1 Growth rates in consumer expenditures

We now turn to our analysis of inter-temporal consumption changes controlling for differences in health, labour supply, mortality and tenure, again tracking group level averages over time. In this section we split households into groups defined by education (whether or not the household head or their spouse completed high school), as well as year and 5-year birth cohorts.

Table 4.4 shows results from taking an average over the rates of decline in spending for non-durable goods, and non-durable goods not including OOP medical spending for our different cohort-education groups. Non-durable expenditures

[12]See, for example, Blundell et al. (1994).

[13]We could have conditioned directly on the change in medical consumption and used changes in medical prices as instruments. We decided instead to include the price term to directly capture the effect medical price inflation.

[14]As the preceding discussion shows, non-separabilities may be present within period (affecting relative demands for particular goods but not the level of spending) or across time (affecting the inter-temporal allocation of consumption). In the Appendix to this chapter, we examine the shares of expenditure on different goods and looking for within-period non-separabilities.

**Table 4.4:** Average percent consumption growth rates

| | UK | US | Country Difference |
|---|---|---|---|
| A. Expenditure | | | |
| Non-durable | -2.21 | -1.37 | -0.84 |
| Non-durable less medical | -2.28 | -2.05 | -0.23 |
| B. Equivalised Expenditure | | | |
| Non-durable | -0.65 | -0.05 | -0.59 |
| Non-durable less medical | -0.72 | -0.72 | 0.00 |

Notes: Observations weighted by cell size. Equivalised using the OECD scale. Equivalised using the OECD scale. The OECD scale is 1 for first adult, 0.5 for each additional adult and child 14 and over and 0.3 for each child under 14.

decline by 2.21 percent a year on average for cohort-education groups in the UK compared to 1.37 percent in the US, giving a statistically significant difference of 0.84 percent between the countries (p-value 0.034). This difference in consumption expenditures before equivalization between the two countries falls by just under three quarters when OOP medical spending is taken out. This suggests that differing healthcare financing institutions may explain a significant part of the difference between the countries.[15]

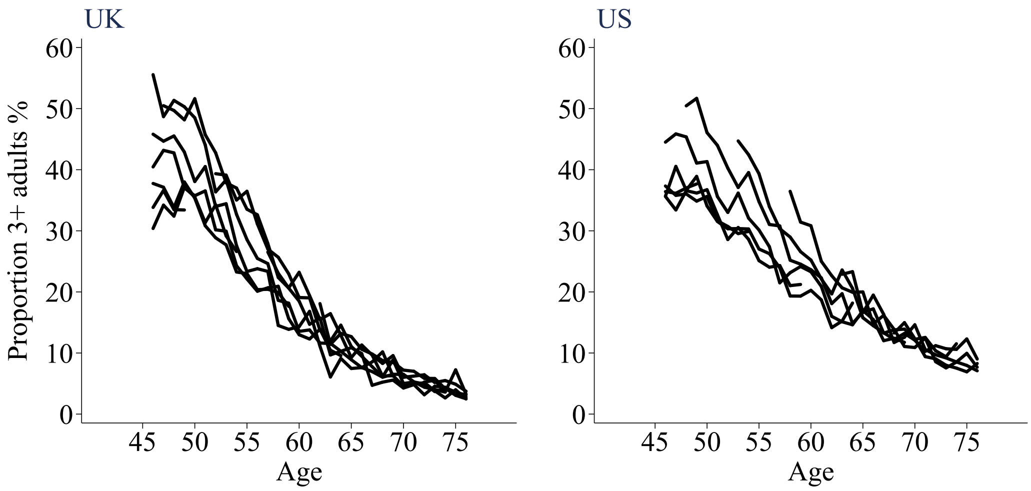
One reason consumption declines at middle and older ages is that people leave the household for several reasons which include the exit of adult children into homes of their own, divorce and the death of a spouse. This pattern is illustrated for both countries in Figure 4.10 which plots by age and cohort the fraction of households who contain three or more adults. These fractions decline significantly with age in both countries, especially between ages 45 and 60 continuing at a somewhat slower pace after age 60.

Declines in the number of adults in the household will of course play a role in producing consumption declines at older ages. When we use equivalised consumption expenditures instead in part B of Table 4.4, not surprisingly we see that rates of decline in both measures of consumption are significantly reduced in both

[15] Both surveys have seen declines in expenditure relative to aggregate measures of household spending as reported in the countries respective National Accounts. This steady decline in coverage may have implications for cross-country differences estimated here. For the definition of spending we are considering however, changes in coverage over time do not appear important for our results. We discuss this further in Appendix B.3.

countries. This indicates that reductions in the number of people in the household, primarily the exit of children and death of spouses, play an important role in the rates of decline in both measures of consumption among those ages 45 and above. However, the difference between the two countries in declines in total non-durable consumption remains large (at 0.59 percent). Once again, this difference between the countries disappears when we examine non-durable consumption less OOP medical expenses.[16]

**Figure 4.10:** Proportion of households with 3 or more adults




Notes: Data from LCFS in the UK and CEX for the US. Each line represents proportions of households with 3 or more adults (individuals over 16) for 5- year birth cohorts over the periods they are observed between ages 45 and 79.

In addition to the role of OOP medical expenses, however, the results in the previous section also highlight the potential importance of other key determinants for instance relating to housing and employment. To see the extent to which controlling for changes in these and other demographic trends can explain the steeper decline in non-durable nonmedical consumption that we see in the UK, we estimate an extended consumption growth equation of the form:

[16]In addition to considering differences in mean expenditure, we also examine growth across the 25th, 5th and 75th and 90th and 95th percentiles of the spending distribution. While the decline in spending growth in both countries is faster towards the bottom of the distribution, there is no clear evidence that cross-country difference in expenditure declines varies much across the spending distribution. This suggests that the UK-US differences are not driven by a few high spending individuals at the top of the distribution in the US.

$$\Delta \ln c_{s,k,t} = \gamma_1 US + \gamma_2 UK + \alpha \ln r_{s,t} + \theta \ln m_{s,k,t} + \Delta X_{s,k,t}\beta + \eta \Delta \ln p_{s,k,t} + u_{s,k,t} \tag{4.2}$$

where $c_{s,k,t}$ denotes non-durable consumption for a cohort-education group $k$, in country $s$ and year $t$ (initially including OOP medical expenses which we later remove). The variable US denotes a dummy for the United States and UK a dummy for the United Kingdom, $\ln r_{s,t}$ is the log real interest rate, $\ln m_{s,k,t}$ is the log mortality rate, and $X_{s,k,t}$ is a set of demographic controls including family size, employment, health status, and housing tenure. Following the discussion of non-separability between medical and non-medical consumption, for specifications where we exclude medical expenditures we include a term for the change in real medical consumption prices, $\Delta p_{s,k,t}$.

The difference between coefficients $\gamma_1$ and $\gamma_2$ in (4.2) indicates how much faster expenditures decline in the US relative to the UK once other factors are controlled, note there is no constant term. We think of this difference as the unexplained component of the cross-country difference, and report it separately in the regression results that follow (multiplied by 100 to give value in percentage point terms).[17]

Results for different versions of model (4.2) are shown in Table 4.5. Column (1) shows results using Weighted Least Squares (using cohort cell sizes as weights) with no controls and including medical spending in the consumption measure. These results are the same as those shown in Table 4.4 except that to maintain comparability across regression models, we use the same sample as we will use in subsequent regressions. The difference in the average rates of decline across the two countries is around 0.9 percentage points and significant at the 5 percent level.

Column (2) adds additional controls for employment, renter status, mortality and health, as well as the interest rate. These additional controls, capturing possible non-separabilities and macroeconomic differences between the two countries,

[17] We also run specifications including country-age interaction terms. These were not significant for either country, suggesting that the difference in the rates of decline in spending between the two countries does not change with age.

do not appear to explain the different rates of consumption growth between the two countries. Declines in rates of employment and increases in the proportion of renters within each group are both associated with lower spending growth. The faster employment declines in the UK shown in Figure 4.4 therefore help account for some of the differences between the countries. However, the effect of this on the unexplained element of the cross-country difference is offset by the larger increase in the proportion of renters in the US which other things equal imply faster spending declines there than the UK. Overall the unexplained component of the spending difference with these controls is around 0.7 percentage points.

Column (3) takes the specification used in column (2) but removes medical expenditures from the consumption variable and allows for the possibility of non-separability between medical and non-medical expenses by including the change in log relative medical prices in the regression (as implied in equation (1)). Relative medical prices are computed relative to non-medical non-durable consumption spending using a Stone price index as described in Appendix B.2. The relative price term enters significantly and indicates a negative gross substitution effect of medical consumption. Other things equal a 1 percent increase in real medical prices from one period to the next is expected to reduce consumption growth by 0.4 percentage points. Even after allowing demographics and real medical prices, there is still an unexplained gap in spending growth between the two countries of similar magnitude to what we had before medical expenditures were omitted.

We might expect some of the characteristics on the right-hand side of the consumption growth specifications in columns (2) and (3) to be endogenous. Households that move out of employment or change their tenure status may adjust their spending because these developments are responses to unexpected shocks that also lead households to reassess the value of their lifetime resources. For instance, estimating the average change in consumption when households change their employer statement may exaggerate the causal impact of employment on spending changes, if households did not already anticipate the change in job status. To account for this we run weighted instrumental variable regressions in which we instrument changes in employment, housing tenure, health and mortality with their first and second lags. Under standard rational expectations assumptions, these should be correlated with current realizations of these variables uncorrelated with unanticipated shocks that enter $u_{s,k,t}$ (we calculate lagged means excluding observations from those interviewed in the following period for CEX). However, these IV models do not produce significantly different results to those reported in Table 4.5 and Durbin-Wu-Hausman tests for endogeneity of these variables does not reject the null of exogeneity. The parameters and test statistics are reported in Appendix B.4.

### 4.5.2 Precautionary motives

One omitted factor from our consumption growth analysis so far is uncertainty over future OOP medical expenditures. As we showed in Table 4.3, older households in the US still face a high risk of large OOP medical expenses in spite of the Medicare and Medicaid programs. The important role these risks potentially play in wealth and consumption dynamics in retirement in the US have been emphasised in Palumbo (1999) and DeNardi et al. (2010). The risks of such expenses are much lower in the UK where households effectively enjoy a much greater degree of health insurance coverage. The differences in the extent of risks of incurring high OOP medical expenses are illustrated in Figure 4.11, where we plot the average differences between the 90th and 50th percentiles of the distributions of OOP medical expenses in the two countries within cohort- education cells at different ages. We plot the 90th 50th difference since, as we saw in Table 4.3, the distribution of OOP medical is highly positively skewed in both the US and the UK, and the main risk households in the US face is the relatively small but non-trivial probability of very high OOP medical expenses. Figure 4.11 shows that in the UK this measure is roughly a quarter of the size it is in the US. It also tends to increase with age and is larger for more educated households.

What implications might these differences in the dispersion of OOP medical expenses have for consumption profiles? A simple theoretical analysis, such as that in Banks et al. (2001), suggests that the effect of uncertainty over shocks to future medical expenses on consumption growth will depend on the product of three fac-

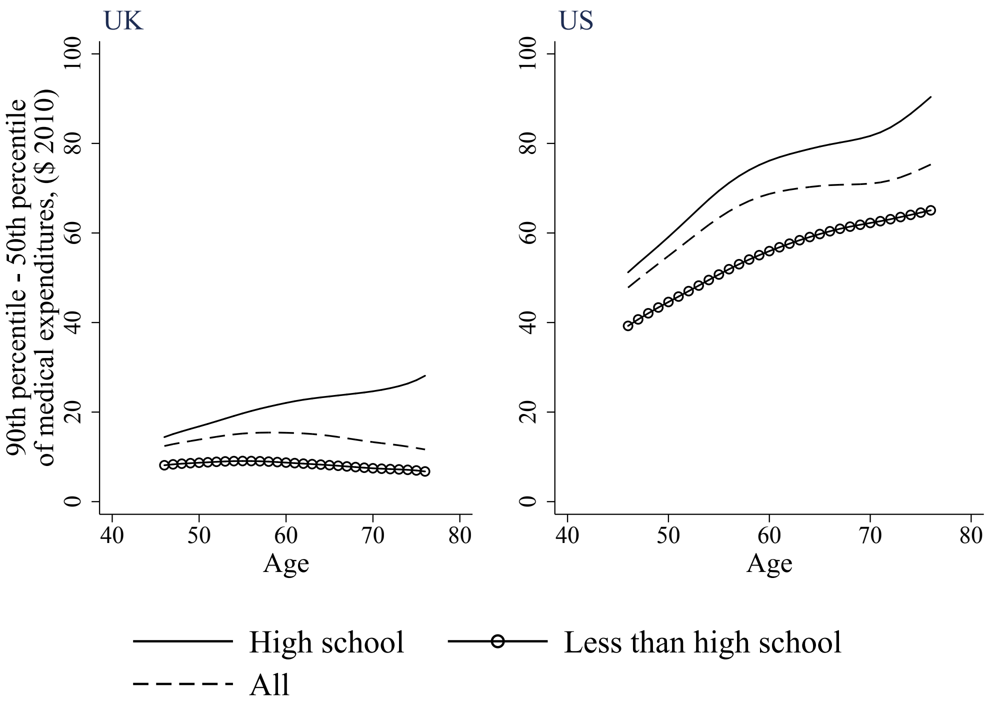
**Table 4.5:** Changes in log non-durable expenditure

| | Incl. Medical Expenditure | | Excl. Medical Expenditure | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| US | -0.013 | -0.001 | -0.004 | -0.025 |
| | (0.003) | (0.011) | (0.012) | (0.016) |
| UK | -0.022 | -0.008 | -0.011 | -0.026 |
| | (0.003) | (0.012) | (0.012) | (0.014) |
| Interest rate | | 0.040 | 0.163 | 0.206 |
| | | (0.093) | (0.096) | (0.097) |
| Log Mortality | | 0.001 | -0.000 | -0.003 |
| | | (0.003) | (0.003) | (0.003) |
| Δ Head employed | | 0.082 | 0.095 | 0.093 |
| | | (0.045) | (0.045) | (0.045) |
| Δ Renter | | -0.419 | -0.400 | -0.404 |
| | | (0.052) | (0.053) | (0.052) |
| Δ Number of kids | | -0.009 | -0.009 | -0.003 |
| | | (0.041) | (0.042) | (0.041) |
| Δ Number of adults | | 0.228 | 0.222 | 0.220 |
| | | (0.030) | (0.030) | (0.030) |
| Δ Single | | -0.249 | -0.226 | -0.228 |
| | | (0.056) | (0.057) | (0.056) |
| Δ Worst health | | -0.216 | -0.239 | -0.236 |
| | | (0.075) | (0.076) | (0.075) |
| Δ Log Medical Price | | | -0.394 | -0.388 |
| | | | (0.073) | (0.072) |
| $\pi^2_{s,k,t-1}\phi_{s,k,t}$ | | | | 0.002 |
| | | | | (0.001) |
| (US-UK)100 | 0.877 | 0.691 | 0.747 | 0.106 |
| | (0.415) | (0.390) | (0.415) | (0.543) |
| N | 616 | 616 | 616 | 616 |

Notes: Estimates presented are for weighted regressions with weights given by cell sizes in each education-year-cohort cell. The dependent variable is log nond-urable consumption (cols 1 and 2 with medical expenditure, cols 3 and 4 without). Additional controls for switch from GHS to HSE surveys in the UK, change in proportion of households reporting own health in US, and the change in proportion responding to subjective health questions. In column (4) we instrument the conditional risk term $\pi^2_{s,k,t-1}\phi_{s,k,t}$ with its lag value. Comparisons of columns (2) and (3) with fully instrumented regressions described in the text are available in Appendix B.4, differences in parameters were not found to be significant.

tors $\kappa\pi^2_{s,k,t-1}\phi_{s,k,t}$ where $\kappa$ is a constant scaling factor reflecting both the persistence of shocks and the consumers risk aversion, $\pi_{s,k,t-1}$ reflects the contribution of uncertainty in medical expenses to uncertainty in overall wealth for group $k$ in country $s$ and period $t-1$, and $\phi_{s,k,t}$ is some measure of the dispersion in OOP medical expenses conditional on information available to each individual consumer in period $t-1$.

**Figure 4.11:** Dispersion in OOP medical expenses




Notes: Data from LCFS in the UK and CEX for the US. Each line represents averages in the 90th 50th percentiles of the distribution of medical expenses within 5 year birth cohorts. Lines are smoothed using locally weighted regression.

Of the three factors, $\pi^2_{s,k,t-1}$ can be approximated by the squared ratio of OOP medical expenses to non-durable consumption excluding medical expenses in period $t-1$. This can be readily estimated from our cross-sectional data (which we do using cohort level averages by education group).[18] The patterns across cohorts and countries is very similar to the patterns shown in Figure 4.8. The choice for

[18]Specifically, the approximation to $\pi_{s,k,t-1}$ is calculated as the square of the cohort-level ratio of medical expenditures to non-medical non-durable spending in each cohort-age-education cell.

the measure of dispersion $\phi_{s,k,t}$ is less straightforward. We take $\phi_{s,k,t}$ to be the period $t$ 90th-50th range in OOP medical expenses in each cohort education group as plotted in Figure 4.11. We then add $\pi^2_{s,k,t-1}$ into the regression model in (4.2) and instrument it using its lag since the term depends on $t-1$ spending and is therefore endogenous. The coefficient on this term will then reflect the value of $\kappa$.[19] This approach identifies the scale of precautionary effects using cohort variation in the importance of medical spending uncertainty. The effects of including this term in our regression model are reported in the final column (4) of Table 4.5.[20]

The uncertainty term enters with the expected positive coefficient and is significant at the 10 percent level. The unexplained difference between the two countries falls from 0.75 to 0.11 percentage points: a remaining difference which is not statistically significant. Thus controlling for medical uncertainty eliminates the remaining gap in spending growth between the two countries. Our results also allow us to estimate the scale of precautionary motives to save against OOP medical expense risk in both countries. To calculate this, we take the predicted spending profiles using our regression results and compare them with those predicted for a counterfactual world in which there was no medical uncertainty (using results corresponding to the model in column (4) of Table 4.5 and households from the cohort born in the years 1933-1937). With medical uncertainty, the expected average annual decline in spending (excluding medical) is 2.21 percent per year in the US and 1.80 percent in the UK. Without medical uncertainty, the predicted declines are 3.10 percent in the US and 1.81 percent in the UK. We therefore estimate that precautionary motives raise consumption growth in the US by around 0.90 percentage points per year on average for the ages we consider.

[19]Ideally, $\phi_{s,k,t}$ should not include any predictable changes in medical expenses, as these do not generate precautionary motives. Calculating risk within cells defined by age, cohort and education eliminates important sources of this heterogeneity. Other sources of heterogeneity that lead to multiplicative differences between the conditional and unconditional risk (for example that might arise if lagged medical expenditures affect current spending through an autoregressive process) will be absorbed in the coefficient on $\pi^2_{s,k,t-1}\phi_{s,k,t}$.

[20]To understand whether other sources of risk may generate create precautionary motives in the US, we have also run a specification (otherwise the same as that in column (4)) where we include a term for income risk that is analogous to the term we use for medical expense risk. This enters the regression insignificantly and does not greatly affect the magnitude or sign of the medical expense risk term.

## 4.6 Conclusions

For many years, debates surrounding the question of whether individuals' have saved enough to fund their consumption needs have focused on whether documented declines in consumption spending over the retirement period could be fully accounted for by optimal behavior within the framework of the life-cycle model. For instance, early work on the "retirement savings puzzle" attributed declines in spending between pre- and post-retirement periods to a failure of consumption smoothing that indicated a lack of preparedness for retirement (Bernheim et al. (2001)). More recent work has argued that those declines that are observed can be fully accounted for through a combination of home production and non-separable preferences (Hurst (2008)).

The work we have reported in this chapter has emphasised how the interpretation of such profiles must be understood in terms of the institutional environment that individuals face, and in particular the extent to which individuals are exposed to uninsured OOP medical cost risks and uncertainties. Relatively large and uninsured risks can generate modestly declining spending profiles on average which do not necessarily indicate sufficiency of resources. We have compared consumption trajectories for older households in the UK and the USA. In the US, spending tends to remain relatively flat at older ages, while it declines quite steeply in the UK. These differences persist when we control for other variables including employment, health and so on that evolve differently in the two countries.

A key component in explaining this difference is OOP medical spending, which rises in the US much faster than in the UK where medical expenses tend to be covered by the state. Taking out OOP medical spending from our comparison reduces the gap in the average decline in consumption spending by roughly three quarters. Although other differences such as inheritance taxes, house price movements, long term care costs and risks and income risk may also play a role in explaining these differences, we find suggestive evidence that precautionary motives to save in the face of greater OOP medical risk in the US are sufficient to eliminate the remaining gap.

These findings have relevance for discussions of consumption behavior at older ages. It is often found that older households, particularly in the US, tend to continue to amass wealth as they age (see Love et al. (2009)). In this chapter, we point out and account for differences between US households and households in an environment where the risks of high medical expenses have been effectively eliminated and for whom spending declines by much more.

# Chapter 5

# Revealed Preference and Consumption Behaviour at Retirement

Simple versions of the life-cycle model predict that households' consumption should not respond to anticipated changes in economic circumstances. However, a number of studies have documented falls in consumption as workers retire (see for instance Banks et al. (1998); Bernheim et al. (2001); Luengo-Prado and Sevilla (2013) etc.). Since retirement should be largely foreseeable for most workers, and a failure of consumers to smooth their consumption around this event violates a central prediction of the standard life-cycle model, this tendency is referred to in the literature as the "retirement consumption puzzle".

A number of explanations for this phenomenon have been proposed. Falls in consumption could for example imply that households do not have adequate retirement savings, and are thus unable or unwilling to smooth their consumption. This could for example be evidence of myopia on the part of consumer or some other behavioural bias (Bernheim et al. (2001)). Alternatively, preferences over consumption may be non-separable with labour supply. For instance, consumers may spend less on work-related clothing, travel expenses and food outside of the home after they retire. They may also have more time for home production or to shop for discounts and so reduce market expenditures (Aguiar and Hurst (2005)). Finally, retirement may sometimes be driven by adverse shocks (such as job loss, or worsening health) and so not be foreseen (Smith (2006)). These explanations clearly have very different implications for what we infer about the adequacy of households' retirement savings and the optimal design of social security systems and pensions.

In this chapter I set out and apply non-parametric, "revealed preference" tests to households over the period of retirement. Such tests allow us to test whether the behaviour of retirees can be rationalised by specific variants of the life-cycle model. The tests themselves are non-parametric and avoid making specific assumptions on the form of the utility function (beyond it satisfying standard properties such as concavity, continuity, transitivity and so on). They rely solely on data on prices and the quantities of different goods that are consumed.

I test whether household's behaviour can be rationalised under the standard life-cycle model, and the extent to which this affected when we tighten assumptions on consumer's foresight, allow for non-separabilities in preferences over consumption and labour force participation, or allow for revisions in the marginal utility of wealth at the point of retirement. Tests of the life-cycle model without perfect foresight are equivalent to tests of the Generlised Axiom of Revealed Preference (GARP) and have little probative force (a fact discussed in Beatty and Crawford (2011)). The perfect foresight life-cycle model performs poorly in the sense that it is largely rejected in our data. Results for this model are not substantially improved when one allows for a revision in expectations at the point of retirement. I also find that minimising deviations from the perfect-foresight life-cycle model suggests rising marginal utility of wealth over time. Thus the smoothest possible paths of marginal utility that rationalise the data are associated with wealth decreasing faster than we would otherwise expect. I show that this particular result is unlikely to be due to changing family composition over the retirement period, aggregate shocks over the period we consider or credit constraints.

This final conclusion cannot be interpreted as strong evidence of consumer myopia. The path of marginal utility could be more stable over time if one allowed larger deviations from the perfect foresight model (due to either uncertainty or perhaps measurement error in the prices and interest rates used). Perhaps one key lesson that we can draw from the results of these exercises is the limits on what we can learn given only data on prices and quantities. The same data can often be rationalised by increasing or decreasing marginal utilities of wealth, depending on exactly what one assumes about the nature of the utility function. It is therefore not easy to conclude whether consumption patterns over retirement are consistent with smoothed marginal utility or not on the basis of expenditure data alone.

The outline of this chapter is as follows. Section 5.1 describes the life-cycle model, parametric approaches, their findings and their various limitations. Section 5.2 then sets out the form of our battery of revealed preference tests. Section 5.3 describes the data we will be using. Section 5.4 presents our results. Section 5.5 discusses further empirical implications of the life-cycle model when we allow for uncertainty. Section 5.6 concludes.

## 5.1 The life-cycle model

We begin by stipulating precisely what we mean by the "life-cycle model". Consumers are assumed to maximise a lifetime utility function that is separable across time, has a geometric discount factor and incorporates uncertainty by having the consumer maximise the *expected* sum of within-period utilities. That is the consumer will solve

$$\max_{q_1..q_T} U = E_t\left[\sum_{t=1}^{T}\beta^{t-1}u(q_t)\right] \tag{5.1}$$

subject to the sequence of flow constraints

$$\Omega_{t+1} = \Omega_t + y_t - \rho_t' q_t \tag{5.2}$$

where $\beta$ is a discount factor, $q_t$ is a vector of consumption commodities, $y_t$ is income, and $\rho_t$ is a vector of discounted prices for consumption commodities. $\Omega_t$, $\rho_t$ and $\Omega_t$ are discounted values of actual wealth $W_t$, prices $p_t$ using the discount factor

$$1/(1+r_t)(1+r_{t-1})\dots(1+r_2) \tag{5.3}$$

where $r_t$ is the (nominal) interest rate. Expectations are also assumed to be *rational* - that is they incorporate all information available at time $t$, are on average correct and are revised each period through a process of Bayesian updating. This is of course a hypothesis that can be tested like any other. We also assume that individuals can borrow or lend freely. Taken together, these assumptions form what we will call the life-cycle model.

Each period the consumer produces a new plan according to their current information set. Writing out the Lagrangian $L_t$ of the problem in (5.1) being solved with period $t$ information

$$L_t = \beta^{t-1}u(q_t) + E_t\left[\sum_{\tau=t+1}^{T}\beta^{\tau-t-1}u(q_\tau)\right] - \sum_{\tau=t}^{T}\lambda_t^\tau(\Omega_\tau + y_t + \rho_\tau' q_\tau - \Omega_{\tau+1}) - \sum_{\tau=t}^{T}\xi_\tau q_\tau \tag{5.4}$$

where the leads us the following vector of first order conditions for consumption in periods $t$ and $t+1$

$$u'(q_t) \leq \frac{\lambda_t^t}{\beta^{t-1}}\rho_t \tag{5.5}$$

$$u'(q_{t+1}) \leq E_t\left[\frac{\lambda_t^{t+1}}{\beta^t}\rho_{t+1}\right] \tag{5.6}$$

This is an inequality rather than an equality as there may be zero consumption of some commodities in some periods. Here $\lambda_t^\tau$ is the Lagrange multiplier on the budget constraint for period $\tau$ when the problem is being solved in period $t$. This term can be interpreted as the (discounted) marginal utility of wealth: the utility value of relaxing the flow constraint period $t$. For simplicity we will denote $\lambda_t^t$ (the $\lambda$ term for the within period first order condition) as $\lambda_t$ in the remainder of this chapter. Similarly, $\lambda_{t+1}$ will refer to the lagrange multiplier on the time $t+1$ constraint when the consumer is solving their lifetime optimsation problem in period $t+1$

(i.e. with period $t+1$ information). $\xi_t$ is a vector of Kuhn-Tucker multipliers on the consumer's non-negativity constraints. Further manipulation of these conditions yields Euler conditions governing growth rates of consumption over time for each good $i$

$$u'(q_t^i) \geq \beta E_t \left[ \frac{\rho_t^i}{\rho_{t+1}^i} u'(q_{t+1}^i) \right] \forall i \tag{5.7}$$

### 5.1.1 The life-cycle model and retirement

The life-cycle model makes a number of predictions about behaviour. The most important of these for individuals who are retiring is that consumer's consumption plans should only be revised from one period to the next if the consumer receives new information. Foreseeable changes to income should not affect the consumer's optimal plan. When consumption is influenced by predictable changes in income this is referred to as 'excess sensitivity'.

As we noted in the introduction, the 'retirement consumption puzzle' is the tendency of consumption expenditures to fall at the point of retirement. This phenomenon was observed in the UK by Banks et al. (1998) using household expenditure data from the UK, and it has since been observed in a number of other countries, including Italy (Battistin et al. (2009)) and the United States (Aguiar and Hurst (2007)). The retirement consumption puzzle is 'puzzling' as the retirement event should be largely foreseeable for many consumers, and so it should not result in a failure of consumption smoothing (a prediction of some variants of the life-cycle model). Some (e.g. Bernheim et al. (2001)) have argued that it represents evidence of excess sensitivity of consumption to predictable income changes, which would challenge the life-cycle view that expectations are formed rationally. This would be troubling as it raises the possibility that households may not be saving adequately to provide for their old-age out of myopia or some behavioural bias.

Various papers have however proposed or defended alternative explanations for the puzzle that are consistent with the basic life-cycle model. Smith (2006) observes that in UK data food spending only falls for those who retire involuntarily - due to poor health or job loss. This suggests that falls in consumption may be due to negative wealth shocks associated with the retirement decision. There is also evidence to suggest that the declines in expenditure are mainly in food and work-related expenses (Bernheim et al. (2001), Aguila et al. (2010) and Battistin et al. (2009)). A decline in work-related expenditures at the time of retirement would seem better explained by change in preferences over consumption goods following retirement (non-separabilites between consumption and leisure in the utility function) than by excess sensitivity. By this we do not mean that consumer has a utility function before and after retirement. Rather we mean that preferences over consumption commodities will depend on the amount of leisure the consumer enjoys, and that these will change at retirement. Aguiar and Hurst (2005, 2007) also show that in the United States there is good reason to believe that a reduction in food expenditures can be explained by an increase in home production of food and an increase in time spent shopping around (allowing retired households to pay lower prices for the same goods). Summarising these findings, Hurst (2008) deemed them sufficient to declare the "retirement of a consumption puzzle".[1]

Of special relevance to this chapter is that both Luengo-Prado and Sevilla (2013) and Christensen (2008), find no evidence of a retirement consumption puzzle in the Spanish data covering the late 80s and early 90s that we will be using. Moreover neither paper finds any evidence for change in preferences over consumption goods. Christensen (2008) points out that income remains relatively stable over retirement in Spain, and argues that this means that any changes in the allocation of spending can be attributed solely to non-separabilities between consumption and leisure. However, she finds no decline in any spending group except for medical expenses (which are heavily subsidised for Spanish retirees). Somewhat surprisingly, this includes a specially constructed category of work-related expenditures (clothing, transportation, petrol and food out). Luengo-Prado and Sevilla (2013) similarly find no decrease in expenditures on any of the sub-categories they define, though they do find evidence for a decline in food spending in a successor survey

[1]As Hurst (2008) points out these findings do not necessarily imply that households save sufficiently to provide for an adequete retirement, only that there is no evidence that this is what lies behind a fall in expenditure at the point of retirement.

covering later years. They argue with appeal to social attitude surveys that the difference between the two periods is linked to the development of more egalitarian attitudes towards women. The retirement of the (usually male) household head did not lead to an increase in home production in the earlier years as this was more often considered the role of the wife.

### 5.1.2 Parametric approaches

One way of testing the predictions of the life-cycle model is to specify a parametric model of the utility function and to then take the Euler conditions in (5.7) to data. Studies looking in consumption behaviour will typically make a few additional assumptions. First we can group individual, non-durable, commodities into an aggregate $c_t$, and then we can assume that all consumers have within-period utility functions of the power form

$$u(c_t) = \exp(X_t'\delta)\frac{c_t^{1-\gamma}}{1-\gamma} \tag{5.8}$$

where $X_t$ is a set of individual demographic and other characteristics (that may include terms such as employment status and hours worked to capture non-separabilities with leisure) and $\gamma$ is the coefficient of relative risk aversion. Consumption will never be negative so long as the consumer has access to some resources whether in the current or future periods (as this is associated with an infinite marginal utility) so in this case we get the following Euler *equation*

$$c_t^{1-\gamma} = E_t\left[\beta(1+r_t)\exp((X_{t+1}' - X_t)\delta)c_{t+1}^{1-\gamma}\right] \tag{5.9}$$

which can be 'log-linearised' to give

$$\frac{1}{\gamma}\Delta\ln c_{t+1} = \ln\beta + \ln(1+r_t) + \Delta X_t'\delta + u_t \tag{5.10}$$

where $u_t$ is a residual that captures individual heterogeneity, preference shocks and any expectational errors (see Carroll (2001) for an explanation and a discussion of some of the problems involved in doing this). Suitably instrumented for the endogeneity of the interest rate, this model can be used to test for evidence of 'excess sensitivity of consumption to predictable changes in circumstances by testing the significance of variables that should be in the consumer's information set on the right-hand side as well as to estimate important preference parameters. An equation of the form in (5.10) has been used to test hypotheses about consumption behaviour at retirement in particular in a whole range of studies (Banks et al. (1998); Bernheim et al. (2001); Smith (2006); Aguiar and Hurst (2005); Luengo-Prado and Sevilla (2013) among others).[2]

### 5.1.3 Limitations of parametric approaches

As we have seen, the standard parametric model makes a number of quite important assumptions. Firstly, the researcher must specify a functional form for within-period utility that will essentially be ad-hoc. Secondly, all consumers are assumed to have the same preferences with heterogeneity allowed in quite a restrictive manner. For instance, in power utility models, all consumers will have the same (constant) rate of relative risk aversion - a factor determining the curvature of the utility function. Thirdly, we assume that consumers' behaviour can be rationalised by an appropriate utility function at all.

The findings on consumption behaviour at retirement summarised above do not depend heavily on parametric assumptions: studies that look at the timing of changes in expenditure are not necessarily committed to a particular functional form for consumers' utility functions. However, parametric assumptions make it difficult to investigate important other, more general, questions about consumption behaviour. Firstly, we cannot use a parametric model to answer whether any individual consumer's choices can be rationalised by a stable utility function. Since this is precisely the claim that the retirement consumption puzzle is supposed to challenge, this is however certainly a question that is worthy of further investigation. Secondly, parametric models are not terribly useful for investigating questions about what is happening to the marginal utility of wealth. As we discuss below, a

[2]Sometimes consumption is separated into different components, and sometimes the interest rate is omitted.

set of observations for a given individual may be rationalised by *both* an increasing or a decreasing path for the marginal utility of wealth over time - depending on the curvature of the utility function and size of the discount factor. This means that if we want to investigate the path of the marginal utility of wealth, we need to be open-minded with respect to possible function forms. A parametric approach specifies a functional form *a priori*.

Answering questions such as these will therefore require a non-parametric method. In an ideal world we would like to estimate demands flexibly, individual-by-individual, through an equation such as

$$q_t^i = m(q_t^{-i}, \rho_t, e_t) \tag{5.11}$$

where $q_t^i$ is one element of the vector $q_t$, $q_t^{-i}$ represents all the other elements, $e_t$ represents temporal shocks, and $m$ is some (potentially) non-linear function. We could then reject the hypothesis that a stable utility function rationalised the data if the resulting demands were not integrable (an hypothesis that we can test statistically).[3] If demands were integrable, we would then have a good deal of knowledge about the consumer's utility function and the possible paths of shocks they experienced.

Unfortunately such methods are typically highly data intensive (they will be subject to the curse of dimensionality) and so cannot by applied in most datasets. An alternative, non-statistical, approach set out by Varian (1982) and in the context of the life-cycle model by Browning (1989) is the method of revealed preference. As we will show, this method can be taken to data covering only a few periods but will nonetheless allow us to contribute to the literature on consumption behaviour around retirement in a number of ways. In particular, it will allow us to directly test whether choices over time can be rationalised by *any* utility function. Secondly, as we will show the revealed preference approach will also allow us to directly investigate whether there is evidence for preference change when leisure time increases

[3]For instance, Haag et al. (2009) show how to impose and test Slustky symmetry in a nonparametric demand system.

at the point of retirement (that preferences between consumption and leisure are non-separable). Thirdly, as I describe in section 5.5 the revealed preference procedure can be used to investigate different hypotheses regarding the evolution of the marginal utility of wealth.[4]

## 5.2 Revealed preference

The revealed preference approach starts with the vector of first order conditions from the basic life-cycle model given by (5.5). A consumer maximising a lifetime utility function of the form in (5.1) must satisfy these, and so these gives us a way of telling whether the consumer's observed choices are generated from utility maximising behaviour. If so, then we say the consumer's choices can be *rationalised* by the life-cycle model (or are 'life-cycle consistent').

**Definition 1.** We say that the life-cycle model rationalises some data $(q_t, \rho_t)$ if $\exists$ a real, concave, non-satiated, real-valued, differentiable function $u(.)$ and a discount rate $\beta \in [0,1]$ such that $\beta^{t-1}u'(q_t) \leq \lambda_t \rho_t$ for all $t$.

To find a within-period utility function with the required properties, we need a condition that can be taken to data. The definition of concavity means that the utility function should satisfy the following

$$u(q_s) \leq u(q_t) + u'(q_t)(q_s - q_t) \tag{5.12}$$

$$\implies u(q_s) \leq u(q_t) + \frac{\lambda_t}{\beta^{t-1}}\rho_t'(q_s - q_t) \tag{5.13}$$

Furthermore, since $u(.)$ is continuous, there would have to be real numbers $u_s$ and $u_t$ such that

$$u_s \leq u_t + \frac{\lambda_t}{\beta^{t-1}}\rho_t'(q_s - q_t)\ \forall s,t \tag{5.14}$$

[4]One disadvantage of revealed preference methods is that is is difficult to condition on demographics or other factors which may affect spending patterns. These are relatively easier to incorporate into parametric approaches (which in this respect can be *more* flexible). In what follows I examine results where we hold key variables fixed over the period we are applying the test to attempt to control for these.

Finally, we know from monotonicity that $\lambda_t$ must be positive (and the restrictions on $\beta$ imply that $\beta^{t-1}$ must also be positive). This means that for a given $\beta$ we will have $2T$ unknowns $u_1..u_T$ and $\lambda_1..\lambda_T$ and $T(T-1)$ inequalities that may or may not be satisfied for a given set of quantities. By grid searching among possible values of $\beta$, we will be able to check if values of $u_1..u_T$ and $\lambda_1..\lambda_T$ exist such that this condition is satisfied for an individual over some period of time, and hence whether their behaviour can be rationalised by the life-cycle model.

It turns out that (5.14) is equivalent to the most basic revealed preference requirement - the generalised axiom of revealed preference or GARP - which states that if any bundle $i$ is revealed preferred to $j$ (directly or indirectly), then $i$ must be unaffordable when $j$ is chosen

$$x^i R x^j \implies p^j x^j \leq p^j x^i \ \forall i, j \tag{5.15}$$

This is due to Afriat's theorem, which implies that if GARP holds for current prices, then it will hold for a set of transformed prices (in our case prices transformed by $\frac{\lambda_t}{\beta^{t-1}} \times \frac{1}{(1+r_t)(1+r_{t-1})...(1+r_2)}$).

**Proposition 3.** *The following two statements are equivalent. 1) The data $(q_t, \rho_t)$ are rationalised by the life-cycle model. 2) The data $(q_t, \rho_t)$ satisfy GARP.*

*Proof.* See Appendix to this chapter. □

As we said GARP is the most basic revealed preference test. The reason that the life-cycle model does not make any stronger predictions about behaviour is that we are allowing for uncertainty and the possibility of the consumer replanning in every period as they process new information. In an uncertain environment, the consumer acts as though they face different lifetime constraints each period according to the news they received. Without access to this news, we will not be able impose any restrictions on the evolution of $\lambda_t$, and we are left with no more restrictions on behaviour than the simple consistency requirements of GARP. This is not to say however that a test of whether or not the consumer passes a test of GARP will exhaust the implications of the life-cycle model. In addition, in standard formulations of the life-cycle model innovations to $\lambda_t$ will be martingale (Hall (1978)): such that $E_t[\lambda_{t+1}] = \lambda_t$. We will return to the importance of this below when we look for nonrandom patterns in changes in the marginal utility of wealth across consumers.

In addition, we can employ a stricter test by imposing some restrictions on $\lambda_t$ for any given consumer. For instance we could impose the condition that $\lambda_t$ must be constant in condition (5.14). This would give us a condition of the form

$$u_s \leq u_t + \frac{\lambda}{\beta^{t-1}} \rho_t'(q_s - q_t) \ \forall s,t \tag{5.16}$$

which Browning (1989) calls cyclical monotonicity (CM). It turns out (in a result due to Browning (1989)) that this is a test of the life-cycle model with the additional assumption of perfect foresight. $\lambda_t$ is constant as consumers are able to perfectly smooth their marginal utilties. The definition of rationalisability with perfect foresight is the same as definition 1 except that $\lambda$ is not longer allowed to vary with time (and so has no $t$ subscript).

**Definition 2.** We say that the life-cycle model with perfect foresight rationalises some data $(q_t, \rho_t)$ if $\exists$ a real, concave, non-satiated, continuous, differentiable function $u(.)$ and a discount rate $\beta \in [0,1]$ such that $\beta^{t-1}u'(q_t) \leq \lambda \rho_t$.

**Proposition 4.** *The following two statements are equivalent. 1) The data $(q_t, \rho_t)$ are rationalised by the life-cycle model under the assumption of perfect foresight. 2) The data $(q_t, \rho_t)$ satisfy CM.*

*Proof.* See Browning (1989). □

Straightforwardly it can be seen that for any subset of periods (5.14) implies (5.16) and hence that $CM \implies GARP$.

The assumption of perfect foresight may well seem unreasonably strong. When we speak of perfect foresight however, we only mean over the periods when we observe the consumer. More precisely we might call this accurate foresight, since all it means is that the consumer is able to smooth their discounted marginal utility of wealth for a few periods without their plans being upset by unpredicted shocks. For short time horizons, this need not be a (completely) unreasonable assumption.

We can test GARP and CM if we have data on quantities and (discounted) prices for a sufficient number of periods. Conducting GARP and CM individual by individual for a dataset would tell us

1. What proportion of consumers' behaviour can be rationalised by a stable utility function of *any* form (GARP)?

2. What proportion of consumers' behaviour can be rationalised by the life-cycle model with perfect foresight (CM)?

A test could in principle including consumption of leisure as a commodity and the wage as a price. It may not be practical to take such a test to data however, as it is difficult to identify the price of leisure (the wage rate) in most household surveys. Information on hourly wages is often not included and calculating wages by for instance dividing earnings by hours may be misleading as wages can themselves vary with hours. Instead we can try to learn as much as possible with the limited information we have.

Without leisure, conditions (5.14) and (5.16) are necessary (but not sufficient) conditions for a version of the life-cycle model where preferences over leisure and consumption are weakly separable (Varian (1988)).[5] Since they are insufficient, if the data pass these tests this does not necessarily vindicate the model, though a failure can still be considered a violation. These are the first tests we will take to our data.

Since tests of GARP will pass if the data passes tests of perfect foresight (CM), these tests can have three possible outcomes: both GARP and perfect foresight can pass, GARP can pass but prefect foresight fails, or both hypotheses fail.

The interpretation of a high proportion of passes for the perfect foresight test is straightforward. We would say we have no reason to reject the hypothesis that the life-cycle model with perfect foresight and separable preferences successfully

[5]These conditions are not sufficient as the consumer would also have to satisfy additional restrictions on their choices over the omitted good (in this case leisure).

rationalise behaviour. Various explanations might account for a high proportion of households who pass GARP but fail the perfect foresight test. The first is that simply that there is some uncertainty leading to a failure of perfect foresight. This might be because individuals experienced a wealth shock (or news shock) at retirement or at some other point. If these shocks were coincident with the retirement decision (such as unexpected job loss), then they would be consistent with the explanation offered by Smith (2006) for the retirement consumption puzzle in the UK. One way to investigate this would be to allow one single change in $\lambda_t$ in the retirement period (and surrounding periods) and to then repeat the test. A second explanation is that preferences over consumption goods change at the point of retirement, due to the fact that preferences between consumption and leisure are non-separable and consumers now have more leisure time, but that there is not enough variation in prices for this change to be detected as a violation by the GARP test. One way to examine this hypothesis is to conduct separate tests of the perfect foresight case for the periods before and after retirement. If it turns out that while there are many failures over the whole period, there are a substantial number of passes for the two sub-periods, we could interpret this as evidence for a dependence of preferences over consumption goods on labour force participation.[6]

In the event that none of these explanations work, then we will have to consider alternatives including more general uncertainty (multiple unexpected shocks), or simple failures of the life-cycle model (behavioural biases, credit constraints and so on). We discuss these possibilities in section 5.5.

A high proportion of of individuals who fail both GARP and perfect foresight would suggest that consumption behaviour cannot be rationalised by a stable, time separable subutility function. This would point to evidence of a dependence on consumption preferences on labour force status, which we can further investigate as discussed above. If tests carried out in individual sub-periods also failed then we would have reason to question some of the fundamental assumptions of the life-

[6]This test is in the spirit of the test in Varian (1988) for cases where we do not observe prices for one of our goods: that revealed preference conditions for the remaining goods should still hold when consumption of the omitted good is constant.

cycle model.

These tests require panel data with detailed information of on expenditures and labour force status for a sufficient period of time. For this we make us of the the Spanish Continuous Household Budget Survey (*Encuesta Continua de Presupuestos Familiares*, ECPF), which we now describe.

## 5.3 Data and sample

The ECPF is a household budget survey conducted by the Spanish National Institute of Statistics (INE). The survey is a rotating quarterly panel, with individual households followed for up to 8 quarters. The survey went through two different incarnations ECPF-85, which ran from 1985-1997, and ECPF-97, which ran from 1997-2005. The latter retained many features of the old survey - including the panel component - but differs substantially in the way information is collected. We only make use of ECPF-85.

ECPF-85 covered roughly 3,200 households each quarter, and collected detailed information on a broad range of different expenditures, demographic information (including the labour force status of both husband and wife) and data on income from a variety of sources (salary, self-employment, capital, pension, transfers and other). Expenditure data is quarterly and was collected through a combination of diary and recall questions.

We group expenditures into 10 non-durable commodity groups: food at home, utilities, food out, adult clothing, child clothing, transport, communications, recreation, personal care, household goods/services. This division is intended to separate out expenditures which are likely to be substitutes or complements for leisure (such as transport or adult clothing) from those which less likely to be (such as child clothing). We exclude medical goods as these are not pure consumption goods and were subsidised for Spanish retirees during this period (Boldrin et al. (1997)).[7] Including them does not significantly affect the results.

[7] For instance public health care assistance was provided to those receiving non-contributory pensions administered by the *Instituto Nacional de Servicios Sociales.* In the later years, most regional government subsidised some medical costs for pensioners (as well as holidays and public transportation costs).

Prices are drawn from the Spanish CPI. Where no price was available for a subset of goods, a price was calculated using a stone price index. Prices are discounted using interest rates on Spanish Treasury Bills taken from the International Monetary Fund. They are converted to quarterly rates by taking them to the power 0.25 to match the period of expenditure.

Our sample consists of households where the principal earner in the first quarter the household was surveyed retires and does not return to work in the period we observe them (retirement is self-reported in the survey). We also restrict our sample to cases where households are observed at least two quarters before and after the retirement date. This gives us a sample of 312 households who we are able to follow over the retirement threshold. Some descriptive statistics for these households are presented in Table 5.1.

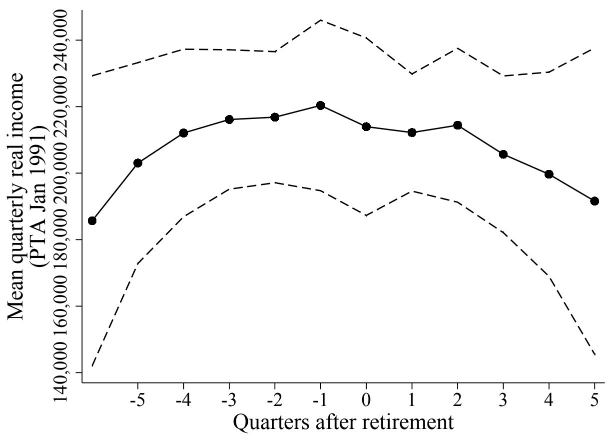
**Table 5.1:** Summary statistics

| | Mean | SD | Min | Max | Obs |
|---|---|---|---|---|---|
| Periods in Survey | 7.23 | 1.18 | 4 | 8 | 312 |
| Age at retirement | 60.02 | 6.6 | 31 | 84 | 312 |
| # HH members | 3.47 | 1.55 | 1 | 11 | 312 |
| Spouse in HH | 0.94 | 0.24 | 0 | 1 | 312 |
| Spouse always employed | 0.11 | 0.31 | 0 | 1 | 293 |
| Spouse never employed | 0.67 | 0.47 | 0 | 1 | 293 |

The average age of retirement is 60 which is quite far below the official retirement age in this period of 65, though many pension schemes allowed retirement before this (Boldrin et al. (1997)). One concern in the exercises that follow might be that the labour supply of spouses may change as the principal earner retires. The information on spouse's employment in Table 5.1 suggests that this might not be too much of a problem. In most (67%) of our households, spouses are out of the labour force for the whole period where the household is observed, and an additional 11% remain employed even as the head retires.

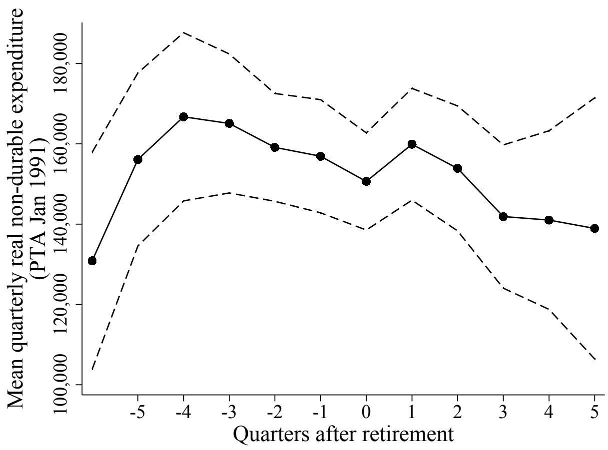
Figures 5.1 and 5.2 show what happens to income and expenditure in the periods leading up to and following the point of retirement. Both suggest that these variables are relatively constant over the retirement period in accordance with the findings of Christensen (2008).

**Figure 5.1:** Total income over retirement




Notes: Author's calculations using ECPF. Dotted lines represent 95% confidence intervals.

**Figure 5.2:** Total non-durable expenditure over retirement




Notes: Author's calculations using ECPF. Dotted lines represent 95% confidence intervals.

As an additional indication of the impact of retirement on income and spending, we run two panel regressions of the logs of these variables on an individual fixed effect and a retirement dummy with no additional controls. The results are reported in Table 5.2. There is a small statistically significant 3% decline in log expenditure and a statistically insignificant increase in log income (which becomes insignificant once household size is included as an additional control). Changing family size is not something which our revealed preference tests can easily account for, but it could plausibly explain apparent preference changes at the household level. We will return to this issue in what follows.

**Table 5.2:** Regressions: income and expenditure at retirement

| | Log non-durable expenditure | | Log HH income | |
|---|---|---|---|---|
| | | | | |
| Retired Dummy | -0.038* | -0.030 | 0.018 | 0.023 |
| | (0.016) | (0.016) | (0.022) | (0.021) |
| Household size | | 0.136** | | 0.081** |
| | | (0.020) | | (0.027) |
| # Obs | 2256 | 2256 | 2254 | 2254 |
| #Households | 312 | 312 | 312 | 312 |

Notes: * indicates significant at the 5% level ** indicates significant at the 1% level. Regressions are fixed-effects (within) models.

## 5.4 Empirical Results

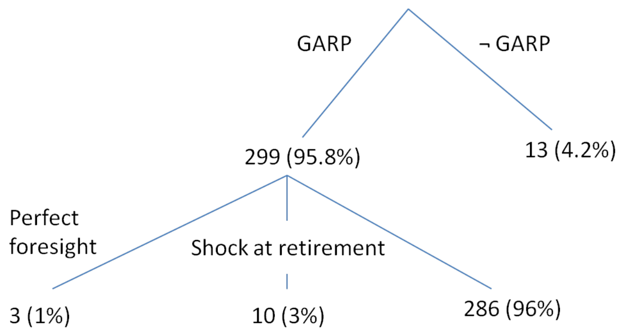
We begin by testing the life-cycle model by implementing a test of GARP, for the whole period when households are observed, followed by a test of the perfect foresight hypothesis. We then go on to consider an intermediate case - allowing a single revision to the discounted marginal utility of wealth - before carrying out a test allowing for a change in preferences over consumption goods at the point of retirement (non-separable preferences). Each of these tests is run on each household separately and independently, and so each allows for *complete* heterogeneity in the utility functions of each individual (both in terms of whether they pass or not and in terms of the form of their preferences). In tests where a discount factor is involved, we grid search for a solution over 100 possible values of $\beta$ between 0 and 1.

### 5.4.1 GARP/Life-cycle model

As we said, this is equivalent to test of the life-cycle model allowing for uncertainty in a very general way but where consumption and leisure are separable. The results of this test are summarised in the top two branches of Figure 5.3.

96% (299) of households satisfy GARP over the whole period. These pass

**Figure 5.3:** Results from different revealed preference tests




rates are similar to those found for working households in the ECPF by Beatty and Crawford (2011). This means that we have no evidence for any change in preferences over consumption goods over the period of retirement (consistent with the findings of Christensen (2008) and Luengo-Prado and Sevilla (2013)). However, the informational content of this result depends on the difficulty of passing GARP if individual tastes were *not* stable over the period, as is clear if we formulate this question in Bayesian terms. Let $U$ be the probability that an individual acts as a maximiser of a stable utility function and let $G$ be the probability that the individual passes a GARP test (or any other revealed preference test). Bayes theorem implies that

$$P(U|G) = \frac{P(G|U)P(U)}{P(G|U)P(U) + P(G|\neg U)[1 - P(U)]} \tag{5.17}$$

Since GARP is necessary and sufficient for utility maximising behaviour $P(G|U) = 1$, so

$$P(U|G) = \frac{P(U)}{P(U) + P(G|\neg U)[1 - P(U)]} \tag{5.18}$$

which depends inversely on $P(G|\neg U)$ - the probability that an individual passes GARP over the period of retirement even if their preferences did change (or if they do not in general act like utility maximisers). When $P(G|\neg U) = 1$ then we will have no reason to update our prior. Conversely, as $P(G|\neg U) \to 0$, $P(U|G) \to 1$. As this term gets smaller, the more confident we can be that a successful GARP test is evidence of utility maximising behaviour.

To compute $P(G|\neg U)$ we need to specify some alternative process that would generate our data if GARP did not hold. Unfortunately, there are many forms of irrational behaviour that could serve this purpose, and so we will have to select one among the possible contenders to utility maximisation. Following Bronars (1987) we will adopt the concept of irrational behaviour outlined in Becker (1962) - uniformly random choices across bundles on the consumer's budget constraint. Under this alternative, $P(G|\neg U)$ will be the probability of a consumer passing GARP if they had chosen which bundle to consume by rolling a die (with one side for each possible bundle). We calculate this by first observing how often a household passes GARP when their budget shares are drawn randomly from a uniform distribution, and then averaging these probabilities across individuals.[8] In our case the probability of a pass was 98.2% across the sample. This is actually greater than the pass rate of our data, and is of course very close to 100% (which would give us no reason to update any of our priors at all).[9] Thus, this exercise does essentially nothing to make the hypothesis of "no taste change" any more likely. We consider further testable implications of the life-cycle model in section 5.5 below.

### 5.4.2 Perfect foresight

A stricter test is whether or not the consumer passes a test of the life-cycle model with perfect foresight. In this test we keep the discounted marginal utility of wealth constant across periods. Passing GARP over the whole period is a necessary condition for passing this test. Since we conduct this test on all periods (before and after retirement) this test also assumes no preference change at the point of retirement.

[8] To obtain these budget shares, we first draw fractions randomly from a Dirichlet distribution and then divide them by their sum to obtain a set of budget shares that sum to one. This procedure ensures that the resulting budget shares will be uniformly distributed. We then calculate quantities as $q_t^i = w_t^i \times \frac{x_t}{p_t^i}$ where $w_t^i$ is the budget share of good i in period t and $x_t$ is the individual's expenditure and then use these to rerun the GARP test.

[9] An alternative means of demonstrating the predictive power of GARP proposed by Selten (1991) and axiomatised by Beatty and Crawford (2011) is the Selten score. This is simply the pass rate less the Selten area. This gives a value which lies between -1 and 1. A score approaching 1 represents an ideal case where the choices seen in the data all pass GARP but individuals choosing randomly would hardly ever pass. A score approaching -1 would imply the converse: the actual choices seen in the data fail GARP much more often than random choices. In our case the probability of a pass was 98.2% across the sample, giving a selten score of 0.043.

The number of passes is shown as the number of individuals under the "perfect foresight" branch in Figure 5.3. Of those households who passed GARP only 3 (1% of the sample) also pass the perfect foresight test.

### 5.4.3 Single change in $\lambda$

Now we allow for a single change to $\lambda_t$ at the time of retirement. This is an intermediate case between perfect foresight and GARP. This test allows for some uncertainty in that we allow for shocks (both positive and negative) that are coincident with the retirement decision such as job loss or health shocks. The results of this test are shown in Figure 5.3 (with those passing this test falling under the "perfect foresight" and "shock at retirement" branches). Only an additional 10 individuals pass relative to the perfect foresight case when we allow a single adjustment to $\lambda_t$ in this way.

One concern may be that the decision to retire may not be exactly coincident with any shocks the individual experiences. For instance, an individual might retire only after a shock has been realised, or else might retire in anticipation of some shock. To allow for this possibility we can simply take the whole period see if a single revision to $\lambda$ *at any time* can rationalise the data. For this exercise we need to have sufficient periods before and after a supposed shock, so we restrict our sample to those individuals observed for the maximum 8 periods. The results of the various tests are shown in Table 5.3. Allowing for a single change at any date does not give us any additional passes.

**Table 5.3:** Pass rates with single change to $\lambda$

| | No change to $\lambda$ | Single change in $\lambda$ (at retirement) | Single change in $\lambda$ (any date including retirement) |
|---|---|---|---|
| Passes | 0 | 1 | 1 |
| Total Obs | 197 | 197 | 197 |

### 5.4.4 Allowing for non-separable preferences

Here we allow for preference change over consumption goods at retirement (due to the change in labour force status) by carrying out the perfect foresight test in the periods before and then after retirement separately. The pass rate on our revealed preference tests can only increase when we reduce the number of periods we are testing (since this merely reduces the chances of encountering a violation). Nonetheless, an substantial increase in the pass rate might still give us reason to believe that a dependence on preferences over consumption goods on labour force participation is a likely explanation for the consumption behaviour we see around retirement. The results are shown in Table 5.4.

**Table 5.4:** Pass rates in periods before and after retirement

| | Passes | | | Total |
|---|---|---|---|---|
| | Before | After | Before and After | |
| GARP | 311 | 310 | 309 | 312 |
| CM | 135 | 120 | 32 | 312 |

309 out of 312 individuals pass GARP when we test separately before and after retirement. Of these, just 32 - or 10% - pass the life-cycle model both in the periods before and in the periods after retirement. Simple preference change does not do well at explaining violations of the perfect foresight model.

What can we conclude from this battery of tests? Firstly, the restrictions imposed by GARP do not provide a very strong test of the life-cycle model. Secondly, if preferences between consumption and leisure are separable then over this period something appears to happen to the marginal utility of wealth: the vast majority of households fail to pass tests of the perfect foresight model (where it is assumed to be constant). Moreover this is more systematic than a simple one-off revision whether at the time of retirement or any other time. Finally, a test of the perfect foresight model allowing for separable preferences between consumption spending and labour force participation also does poorly. In the next section we consider the nature of violations from the perfect foresight case.

## 5.5 Violations

Violations from the perfect foresight case could be the result of a failure to capture other changing aspects of the consumer's environment (durable expenditures, health, price changes that differ from CPI aggregates) or as a result of uncertainty.

As we noted previously, GARP does not exhaust the empirical implications of the life-cycle model with uncertainty. To investigate the plausibility of this model further we can consider deviations $e_t$ from the perfect foresight case

$$u_s \leq u_t + \frac{(\lambda + e_t)}{\beta^{t-1}} \rho_\mathbf{t}'(\mathbf{q_s} - \mathbf{q_t}) \tag{5.19}$$

where $(\lambda + e_t) > 0$. Under the life-cycle model, these deviations should be martingale, that is with $E_t[e_{t+1}] = e_t$. We can now ask what the paths of $\{e_1 \ldots e_T\}$ which rationalise our data look like. Multiple paths may rationalise the behaviour of any given household, and the data cannot tell us which of these is right.[10] We will select the path which minimises the sum of squared deviations $\sum e'e$, which will in turn isolate a particular path for $\lambda_t$. This is the path with the smallest deviations from the perfect foresight life-cycle model and so is as favourable as possible to the maintained hypothesis of rational, forward-looking consumers.

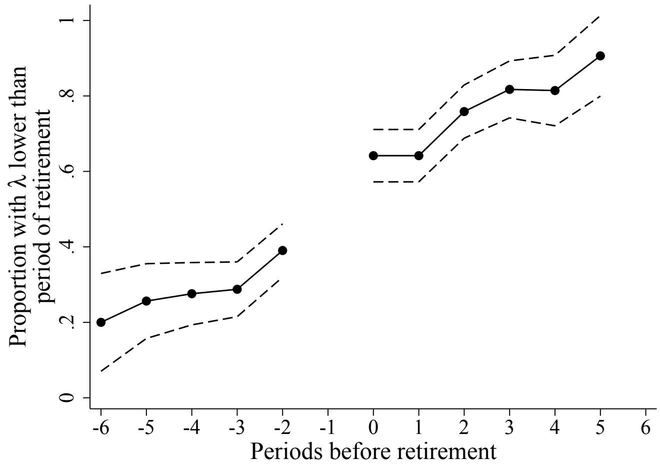
These violations are terms in a utility function and so do not have a cardinal interpretation. We cannot compare the size of violations across individuals, and there is no sense in saying that one violation is for instance 'twice as big' as another. However, the *change* in these terms from one period to another does have an interpretation. A systematic tendency to get decreasing or increasing $e_t$'s over time indicates that something systematic is happening to the discounted marginal utility of wealth. Consumption would be increasing or decreasing faster than it 'should' do under the life-cycle model, and this would point to a rejection of this model.

Figure 5.4 compares $\lambda_t$ to $\lambda$ at the point of retirement for households who we observe for all 8 quarters. It is clear that, in each period before retirement, a majority households have $\lambda_t$'s which lie below those in the first period of retirement. Moreover, in each period after retirement, a majority of households have $\lambda_t$'s which exceed it. Moreover the percentage of households with lower $\lambda_t$ than the retirement date seems to monotonically increase with time. This points to a general tendency

[10] For instance, despite finding that the in cases where we minimise the sum of squared deviations, $e_t$ is increasing, 183 profiles can also be rationalised by monotonically decreasing $e_t s$ (albeit with low discount factors).

for the marginal utility of wealth to be increasing over time when we minimise deviations from the perfect foresight case. This would suggest that consumption is falling faster than we might expect (or else deviations from the perfect foresight model are larger than these minimal values).

**Figure 5.4:** Proportion $\lambda_t$ higher than $\lambda_{retire}$

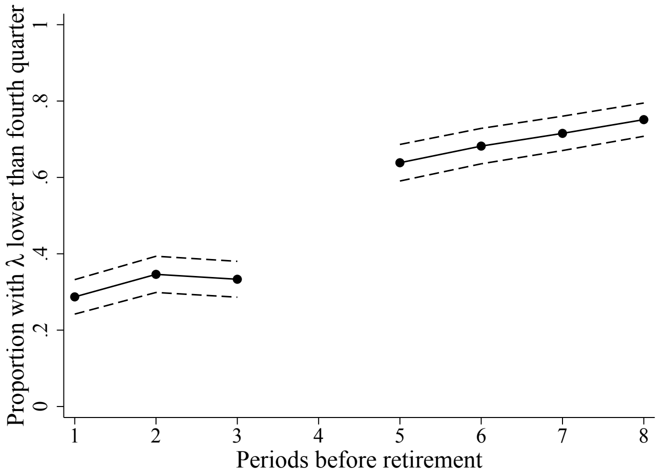



Notes: Author's calculations using ECPF. Dotted lines represent 95% confidence intervals.

This may be a phenomenon related to retirement in particular or it may be a more general phenomenon. Figure 5.5 plots the proportions of households with marginal utilities of wealth below the marginal utility seen in the 4th quarter we observe them for working age households (where the head remains employed throughout the 8 quarters we observe them and is aged 55-60). A very similar pattern is evident, pointing towards a general problem with the life-cycle model (although the extent of the increase is smaller).

The fact that it we also find increasing $\lambda$'s for households who do not change their labour supply suggests that it is not something which can be dismissed as being due to non-separable preferences between consumption and labour force participation. As we saw from Table 5.1 the proportion of spouse's who change their labour supply in our retiring households is also low, making it unlikely that this explains the result either. So what might explain these patterns? In the rest of this section we will discuss some possibilities.

**Figure 5.5:** Proportion $\lambda_t$ higher than $\lambda_{retire}$




Notes: Author's calculations using ECPF. Dotted lines represent 95% confidence intervals.

### 5.5.1 Adverse shocks

Could a series of adverse shocks account for an increasing discounted marginal utility of wealth? To see how uncertainty affects the $\lambda_t$ we need to return to the first order conditions for consumption commodities from the problem in (5.1).

Given the estimates of lifetime wealth in period $t$, the first order conditions for consumption of the goods $q$ in period $t$ (ignoring non-negativity constraints) are

$$\lambda_t = \frac{\beta^{t-1}u'(q_t)}{\rho_t} = E_t\left[\frac{\beta^t u'(q_{t+1})}{\rho_{t+1}}\right] \tag{5.20}$$

Now we know that based the assessment of lifetime wealth made in period $t+1$

$$\lambda_{t+1} = \frac{\beta^t u'(q_{t+1})}{\rho_{t+1}} \tag{5.21}$$

Substituting (5.21) into (5.20) gives

$$\lambda_t = E_t\left[\lambda_{t+1}\right] \tag{5.22}$$

or the martingale result. Given some regularity conditions we can decompose the value of $\lambda_t$ into $E_t\left[\lambda_{t+1}\right]$ (which is $\lambda_{t+1}$ if expectations are rational) and an additive, mean-zero error

$$\lambda_t = \lambda_{t+1} + \varepsilon_t \tag{5.23}$$

Since the utility function is concave, a positive value of $\varepsilon_t$ will mean that the consumer has less wealth than they had expected in the second period: the marginal utility of wealth increases when the consumer experiences a negative wealth shock, and decreases when the consumer experiences a positive wealth shock.

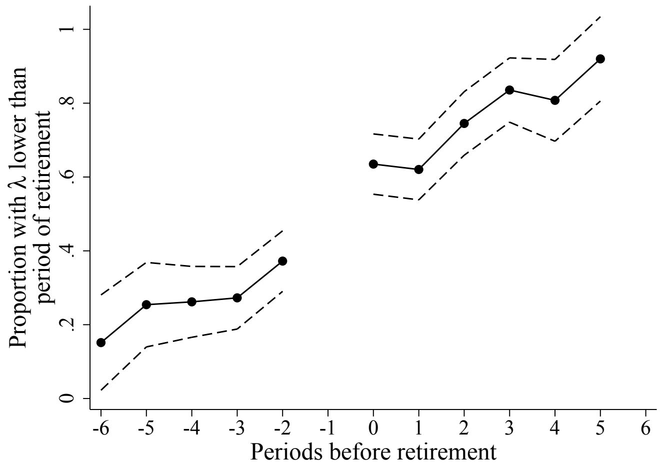
Rational expectations imply that idiosyncratic shocks should be random, not systematic across individuals, and so these would not account for the pattern of violations that we saw in Figure 5.4. A systematic tendency across a fairly large group of individuals could be the result of common, or aggregate shocks, such as unanticipated decreases in pensioner benefits or productivity shocks. But this seems an unlikely explanation for a pattern observed consistently for many individuals over a period of 12 years (apparently including those of working age). One would have to appeal to quite a long sequence of unanticipated, negative shocks to generate this.

It should be noted that increasing the martingale predicts a constant *mean* value of the discounted marginal utility of wealth, while in Figure 5.4 we saw evidence of an increasing proportion of households who had $\lambda_t$ greater than $\lambda$ at the point of retirement. This could be rationalised with the life-cycle model by a low probability possibility of a large positive wealth shock leading households to spend more than they otherwise would.

### 5.5.2 Changing household composition

One possibility suggested by the results in Table 5.4 is that the composition of households is changing over time. The evidence for a 'retirement consumption puzzle' in this data disappeared once household size was controlled for. We can check for this by looking at households where household size remains constant. Figure 5.6 is the same as Figure 5.4 except that we restrict the sample to households where there is no change in household size. It shows much the same pattern, suggesting

**Figure 5.6:** Proportion $\lambda_t$ higher than $\lambda_{retire}$, no change in HH size




Notes: Author's calculations using ECPF. Dotted lines represent 95% confidence intervals.

that this is unlikely to explain the patterns we see.

### 5.5.3 Credit/saving constraints

The life-cycle model assumes perfect capital markets, which is what allows the consumer to smooth their marginal utilities. If we relax this assumption, then the $\lambda_t$ need no longer be constant even in the perfect foresight case. To see how, we can introduce a credit constraint into the problem in (5.1): consumption in period $t$ cannot exceed some limit $b_t$. Abstracting from uncertainty (and dropping the non-negativity constraints) this gives us the Lagrangian

$$L = \sum \beta^{t-1} u(q_t) - \lambda(\sum \rho_t' q_t - W) - \mu_t(\rho_t' q_t - b_t) \tag{5.24}$$

where $W$ is *total* lifetime wealth. This is associated with the first order condition

$$u'(q_t) = \frac{\rho_t(\lambda + \mu_t)}{\beta^{t-1}} \tag{5.25}$$

and hence the cyclical monotonicity condition

$$u_t \leq u_s + \frac{(\lambda + \mu_t)}{\beta^{t-1}} \rho_t'(q_s - q_t) \tag{5.26}$$

$\mu_t$ is the Kuhn-Tucker multiplier on the credit constraint (or the marginal value of relaxing it). When the constraint is binding $\mu_t > 0$. For a saving constraint, we would have $\mu_t < 0$. A path of increasing $\lambda_t$ could therefore be rationalised if we thought consumers were unable to save in earlier periods. This would mean that while consumers knew that their future resources were going to decrease, they were not be able to smooth their consumption in they way they would like because they were unable to save. This seems unlikely. Indeed, a fair number (54%) of households saw their real incomes increase on average over the period of retirement.

### 5.5.4 Measurement error

A further possibility is that we have mismeasured the price data or interest rate we use. It is not clear why random errors in prices should tend to generate a systematic tendency for violations to increase in size over time. Proportional mismeasurements in the interest error can to an extent be offset by adjustments in the discount factor (data for which we have underestimated the interest rate by 20% can be rationalised by a utility function where $\beta$ is 20% higher). However, the extent to which this is true is limited by the fact we have bounded $\beta \in [0,1]$. Allowing $\beta$ outside of this range allows greater to minimise violations over time and so could be consistent with a flatter path.[11]

### 5.5.5 Time inconsistency

A final possibility is that consumers are non-rational. As we noted above this was the preferred explanation of Bernheim et al. (2001) for the retirement consumption puzzle. A tendency to consume too much at younger ages (due to for instance hyperbolic preferences) would explain the patterns we see in the data,and would help to explain why the same pattern is also observed among households who are

[11] Underestimating the interest rate may seem a curious explanation for why we should observe the marginal utility of wealth increasing over time (since this should mean consumers are more willing to save than we would otherwise thought). However, it arises as the feasible set of preference parameters (including $\beta$) rationalising given data is not convex.

not retiring. Further tests of these models (such as those proposed in Blow et al. (2017) using a revealed preference framework could lend further support to this hypothesis.

## 5.6 Conclusion

In this chapter I have set out revealed preference tests for different models of consumption behaviour over retirement that I then applied to a Spanish consumption panel dataset. These tests reject the perfect foresight model both with separable preferences and allowing for preference change. We do not reject a life-cycle model allowing for uncertainty, but the the first order conditions of this model do not provide very strong restrictions on possible choices. In fact, they are no stronger than those implied by the most basic revealed preference requirement: GARP. We then go on to investigate the paths of deviations from a perfectly smoothed marginal utility of wealth, in a case when these deviations have been minimised relative to the perfect foresight case. I find that the smoothest possible paths of marginal utility that rationalise the data, are associated with the marginal utility of wealth decreasing systematically over the retirement period. I show that this particular result is unlikely to be due to changing family composition, aggregate shocks over the period we consider or credit constraints. Two possible explanations are deviations from geometric discounting and mismeasurements in the interest rates facing consumers.

# Chapter 6

# Consumption Spending, Housing Investments and the Role of Leverage

Do house price booms induce households to borrow and re-leverage their balance sheets? And can leverage make households more sensitive to future house price shocks (whether good or bad)? It is now widely believed that increases in debt and leverage that accompanied the international house price boom prior to the 2008 financial crisis both deepened and prolonged the length of the slump in consumption spending that followed. As a result, recent years have seen policy-makers showing increasing interest in macro-prudential measures that limit leverage growth among households during boom periods, and relax them when economic conditions weaken.

In this chapter, we use detailed household-level data to examine the borrowing, spending and investment decisions of homeowners in response to house price increases, as well as how these decisions differ according to households' initial leverage (defined as the value of their home relative to the size of their mortgage debt). We start by documenting the extent to which households re-leverage (by taking on mortgage debt as house prices increase) using panel data from both the US and the UK. Household mortgage debt among existing home-owners, who do not move home, increases by roughly 0.3 percentage points with each 1% increase in home values in the US and by 0.2 percentage points in the UK. In this respect, households increased the size of their balance sheets in the years prior to the financial crisis in a similar way to investment banks (Adrian and Shin (2010)). Using a plausibly exogenous source of variation in households' leverage, we then go on to consider whether leverage amplifies households' spending responses to given house price changes and whether this varies across different forms of spending. To do so we link data on households balance sheets from a panel survey with spending data in a household budget survey using two-sample IV methods (Angrist and Krueger (1992)). We find strong evidence of large differences in the responses of residential investment by households according to their initial leverage. We estimate an elasticity of residential investment spending to house price increases of 0.74 among outright owners. The estimated elasticity is twice as large for those with a loan-to-value ratio of 50% (in our empirical model the effects scales linearly with each further doubling of households' debt to equity ratios). We also show that households that have greater initial leverage are more likely to make second home purchases in response to rising local prices over longer time horizons. However, we do not find evidence that more leveraged households disproportionately increase either their total, nondurable or durable consumption spending as house prices rise.

The absence of differences in the consumption behaviour of more and less leveraged households is puzzling in that the former group would be expected to see a larger increase in their net housing wealth for a given price change. It is however consistent with the view that house price increases have both positive and negative effects on the real value of households' lifetime wealth. In models with infinitely lived agents, any positive endowment effect of increased house prices on consumption is immediately offset by the negative effect of higher lifetime costs of owner-occupation. This leads to the prediction that 'housing wealth isn't wealth' and that the consumption effects of housing price changes should be small in aggregate (Buiter (2010)). To account for the disproportionately large differences we observe in responses of residential investment spending, we emphasise a novel mechanism to account for households' desire to re-leverage and make new investments in housing as prices increase. In particular, we highlight the importance of housing's dual role as a both source of immediate utility and as a source of investment returns, and the fact that there is a direct link between leverage and the share of housing in households' portfolios. Borrowing, home improvements and other forms of residential investment may be used to adjust this share.

This forms the basis of a 'portfolio-rebalancing' effect which we argue drives the residential investment decisions we observe. Standard life-cycle models of household portfolios (e.g. Merton (1969)) predict that households' desired leverage will evolve smoothly over time. In such models, house price reductions tend to leave households over-leveraged relative to their target leverage, inducing a desire to reduce consumption spending and increase savings. This behaviour provides a micro-foundation for 'debt-overhang' effects (Dynan (2012)) whereby price declines induce a desire to deleverage and depress consumption spending. These effects have an important corollary that explains household decisions to re-leverage during house price booms. In periods where house prices increase, consumers will find themselves *under*-leveraged. This will have the effect of encouraging consumers to borrow and invest in housing stock. Such portfolio adjustments increase leverage and, by increasing the size of household balance sheets, return households to their desired rate and variance of returns. Since portfolio shares are themselves nonlinear functions of leverage, we would expect these effects to be greater for households with higher loan-to-value ratios. More leveraged households will experience a larger reduction in the portfolio share of housing for a given price increase, and so we would expect that their residential investment spending should respond more strongly than other households'. This mechanism therefore accounts for our empirical findings, and in particular the larger response in the residential investment of more leveraged homeowners to house price increases.[1]

In principle household could alter the risk and return of their portfolios by adjusting holdings of other, financial assets (such as stocks). Housing wealth tends to be the most important asset that households hold and is unique in having historically offered a mix of both high and relatively low variance returns (Jordá et al.

[1] Another way leverage may affect household investment decisions is that households at greater risk of default may have reduced incentive to make investments in their property (Melzer (2017)). This is likely to be less important in our context (the UK) where mortgages are typically non-recourse.

(2017)). It is therefore perhaps not surprising that households should see housing investments as a key mechanism to adjust the riskiness of their wealth holdings.[2] Our proposed mechanism also relies on the assumption that housing investments increase the market value of homeowners' properties in addition to any private consumption benefits they provide. We argue that this is consistent with the available evidence on such investments below.

Our conclusions and empirical findings differ from previous studies that have both observed large differences in the borrowing and spending of households according to their leverage (Disney et al. (2010b); Mian and Sufi (2011); Cloyne et al. (Forthcoming); Cooper (2013); Aladangady (2017); Disney et al. (2010a)). An important difference in our empirical approach is that we attempt to account for the endogeneity of leverage. Households choose their leverage, and thus more leveraged households may respond differently to house price increases. In particular they may have different house price expectations, different financial or liquid wealth and live in differently sized homes. As a result they may be more responsive to house price changes for reasons which are not causally related to the leverage of their balance sheet. The instrument we use is a measure of credit conditions when households first moved into their current homes (the average loan-to-income ratio on new house purchases). We show using evidence from the first sample in our two-stage IV approach that this instrument is conditionally uncorrelated with a number of potential confounding variables including gross house values, financial asset income, and unsecured debt. By contrast, lagged leverage, which has been used as a source of variation in a number of previous studies is however strongly correlated with all of these factors. We find that this instrument works well for predicting leverage in the UK, which is the source of our main empirical findings. In the US, however, we find that re-leveraging in response to house price increases is so rapid that it is only very weakly related to credit conditions at the time households last moved.

Relative to previous studies, we also put much less emphasis on credit con-

[2]Other studies suggest that households adjust their holdings of other risky assets in response to house price increases - suggesting that housing and other risky assets are substitutes (Chetty et al. (2017)).

straints in explaining the different responses that we do observe between more and less leveraged households. The existing literature currently appeals to credit constraints to explain the fact that spending responses are disproportionately greater for households with the very highest leverage. The logic underpinning this conclusion is as follows. If shocks to housing wealth help to relieve credit constraints (for example by providing households with additional collateral), then they will be associated with high marginal propensities to consume. Under a credit collateral explanation for house price effects, one would expect these effects to be strongest at the very highest levels of leverage.[3] Moreover, households at greater risk of facing a binding credit constraint would be expected to accumulate precautionary savings, which they would then decumulate faster in response to house price increases (Berger et al. (2017)).

We do not believe that our findings are driven by a ‘collateral channel’ of this kind. Our instrumental variable strategy captures differences in leverage across households that result from past decisions (the timing of moves). This is unlikely to pick up differences in the propensity to be credit constrained, and we find no relationship between our instrument and measures of unsecured debt or other forms of financial wealth. Moreover, we find no evidence that leverage increases non-

[3]This is typically what is found. For example, Disney et al. (2010a) find that savings respond very little on average to house price shocks but that the response is five times greater for households emerging from a situation of negative equity than for households with initially positive equity values. Cooper (2013) finds that the spending of households with high debt service ratios, low liquid wealth and high expected future income were the most sensitive to housing wealth changes. Since all these characteristics are indicative of credit constraints, he argues that this is the most important channel at work. Looking at the UK for the years 1995-2005, Disney et al. (2010b) also find that households’ propensity to remortgage in the face of price rises does not increase with household’s LTV ratios but is higher for those with high LTV *and* high unsecured debt. In a related paper that using the US Panel Study of Income Dynamics (PSID), Disney and Gathergood (2011) again look at changes in household indebtedness by initial LTV ratios (again allowing these to be interacted with levels of unsecured debt). As before, households with high leverage - particularly those with high leverage and high unsecured debt - were seen to be much more responsive. Mian and Sufi (2011) examine the link between growth in debt and regional variation in home prices in the United States over the period 2002-2006. They find that consumers with credit scores one standard deviation above the 1997 mean have estimated responses that are twice as large as those with estimated responses one standard deviation below the mean (with an elasticity of 0.75 compared to 0.35). They find no evidence of house price effects for those in the top quartile of the credit score distribution. They also find that responses are in general larger for younger households and find similar differences between consumers with high and low credit utilisation rates. In line with the rest of the literature, they argue that these differences most likely reflect the importance of credit constraints.

durable or durable consumption responses to house prices, but rather that effects are concentrated among residential investment spending. Such behaviour would be difficult to explain through credit constraints alone. Finally, we do not see very different effects in estimates for a younger sub-sample of households than we do in our main sample.

We also do not believe that our results are driven by more traditional housing wealth effects. In contrast to what we find for residential investment spending, we see no differential effect of house price increases on spending on a basket of luxury goods (food away from home and leisure services) by more leveraged households relative to less leveraged households.

As well as explaining the lack of a consumption response, our proposed mechanism also accounts for the fact that households self-reported use of new mortgage loans is disproportionately for spending on home improvements rather than traditional forms of consumption (see Brady et al. (2000) and also results for the UK below). Our empirical analysis below confirms that this behaviour is evident in survey data on spending as well as in qualitative survey evidence.

The remainder of this chapter is structured as follows. In Section 6.1 we set-out a theoretical framework explain household's choice of leverage, motivations for re-leveraging and spending decisions. Section 6.2 describes the data we will use. In Section 6.3 we provide evidence that households do indeed re-leverage by increasing borrowing when house prices rise. In Section 6.4 we set out our empirical approach to identify the role leverage plays in the home improvement spending of existing homeowners, and present results on the impact of house prices on different components of consumption spending. Section 6.5 presents results on purchases of second homes by more and less leveraged households in response to local house price increases. Section 6.6 concludes.

## 6.1 Life-cycle portfolio choice

Consider a household with two assets available to hold in its portfolio that maximises expected lifetime utility

$$U = \max_{c,h} E_t \left[ \sum_{j=0}^{T-t} \beta^j u\left(c_{t+j}, h_{t+j}\right) \right]$$

where $c_t$ is consumption, $h_t$ is housing and $\beta$ is the household's discount factor.

Housing is a risky asset with price $p_t$, and return:

$$r_t^* = \frac{p_t}{p_{t-1}} - 1. \tag{6.1}$$

Households can also hold a risk-free asset (a bond) denoted $b_t$ with price 1 and interest rate $r$. The household can short the bond (that is, take a loan) and there are no credit constraints. The household cannot short housing. For simplicity we assume that there are no adjustment costs associated with housing.

The leverage position of the household (the loan-to-value ratio) is:

$$L_t = \frac{\text{debt}}{\text{house value}} = \frac{\text{debt}}{\text{gross housing wealth}} = \frac{-b_t}{p_t h_t} \tag{6.2}$$

and the portfolio share of housing is:

$$\omega_t = \frac{\text{gross housing wealth}}{\text{net wealth}} = \frac{p_t h_t}{p_t h_t + b_t} = \frac{1}{(1 - L_t)} \tag{6.3}$$

Notice here that leverage $0 < L_t < 1$ implies $\omega_t > 1$. For example, a household with a 95% "mortgage" ($L = 0.95$) has a housing portfolio share of $\omega_t = 20$, while for outright owners $\omega_t = 1$.

To show how leverage magnifies risk and return, denote the household's net wealth by $x_t$, and labour income by $y_t$ . Then the intertemporal budget constrant is:

$$x_t = (1 + r + \omega_{t-1}(r_t^* - r)) * (x_{t-1} - c_{t-1}) + y_t \tag{6.4}$$

or equivalently,

$$x_t = \left(1 + r + \frac{1}{1 - L_{t-1}}(r_t^* - r)\right) * (x_{t-1} - c_{t-1}) + y_t \tag{6.5}$$

Leveraged households have a greater increase in their wealth for a given house price shock, and these effects are highly nonlinear in leverage. At the same time, since housing is a risk asset, high leverage also increases the variance of portfolio returns.

To show how the consumption, saving and portfolio allocation decision of households responds to these changes in wealth, we have to specify further details household preferences, resources and markets.

Assume the household has no labour income ($y_t = 0$). Further assume that the household has CRRA preferences over consumption and that housing is just an investment good (does not yield a flow of utility). This (combined with assumptions on the stochastic process associated with house price movements) is essentially the life-cyle portfolio choice model of Merton (1969), and the policy functions are well known.[4] There is a linear consumption function:

$$c_t = \alpha_t x_t \tag{6.6}$$

and a constant target portfolio share for the risky asset:

$$\omega_t = \omega^* \tag{6.7}$$

In the Merton model, the portfolio share of the risky asset depends only on the consumer's risk aversion and moments of the return distribution. As leverage is just a transformation of the housing portfolio share, this implies there is a constant target leverage that delivers the household's desired combination of risk and return.

Now consider how this household responds to a positive house price shock:

$$x_t - E[x_t] = \omega^* \left(r_t^* - E\left[r^*\right]\right) \times \left(x_{t-1} - c_{t-1}\right) \tag{6.8}$$

$$x_t - E[x_t] = \left(p_t - E\left[p_t\right]\right) \times h_{t-1}. \tag{6.9}$$

[4]The Merton model also assumes that returns are stationary and log-normally distributed.

The shock is partly consumed:

$$c_t - E[c_t] = \alpha_t (x_t - E[x_t]) \tag{6.10}$$

and partly saved ($s_t$) according to the consumption function:

$$s_t - E[s_t] = (1 - \alpha_t)(x_t - E[x_t]) . \tag{6.11}$$

Critically, the portfolio choice rule implies that this additional saving is leveraged by $\omega^*$:

$$p_t \times h_t - E[p_t \times h_t] = \omega^* (1 - \alpha_t)(x_t - E[x_t]) . \tag{6.12}$$

Using:

$$x_t - E[x_t] = (p_t - E[p_t]) \times h_{t-1} \tag{6.13}$$

this implies extra active investment in housing of:

$$(p_t \times h_t - E[p_t \times h_t]) - (p_t * h_{t-1} - E[p_t] \times h_{t-1}) = (\omega^* \times (1 - \alpha_t) - 1)(p_t - E[p_t]) h_{t-1} \tag{6.14}$$

The first term on the left hand side of equation (6.14) is the desired extra gross housing wealth. The second term is the additional housing wealth that comes mechanically from the unexpected price increase. The difference between the two is the additional active investment in housing (funded by debt) to return the housing portfolio share to $\omega^*$. Thus, if there is an unexpected house price increase, a leveraged household will invest more in housing and borrow to do so (even if the household believes that housing returns are i.i.d). Conversely, an unexpected house price fall increasing the leverage of the portfolio and the household will want to sell housing and retire debt to return to $\omega^*$.

For example, suppose that the household owns a £600,000 house with $\alpha = 0.05$ and $\omega = 3$ (so that the household has 33% equity in the home.) If the house value unexpectedly goes up by 5% (£30,000), the consumption function implies that net wealth increases by £28,500 and the constant portfolio rule implies that the household then desires gross housing wealth of £685,500 . As the house value is now £630,000, the households makes new investment in housing of £55,500, financed by new debt. Note that the extra investment spending (£55,500) is much larger than the extra consumption spending (£1,500).

The implications of this model for housing investments are *very* stark. In practice a number of factors could moderate this mechanism. For example, if housing provides a flow of utility, this is like a dividend, but if there is diminishing marginal utility from housing (and extra housing cannot be rented efficiently), the total return to housing falls with housing wealth held, and this will temper some of the re-leveraging motive described above. Transaction costs associated with residential investment, credit constraints and the availability of other financial assets would also slow the pace of at which households return releverage through housing investments.

Nevertheless, equations (6.11) and (6.12) suggest that the re-leveraging mechanism we describe will operate so long as the policy functions for consumption and for the risky asset portfolio share are sufficiently flat in net wealth. A gently sloped consumption function implies that a significant fraction of a wealth shock is saved. A fairly flat portfolio rule implies that the household will not want to change its portfolio shares dramatically in response to a wealth shock. A wide range of life-cycle consumption and portfolio-choice models share these features.

In the remainder of this chapter, we investigate the empirical relevance of this mechanism.

## 6.2 Data

To investigate re-leveraging and the relationship between leverage and household spending, we draw on three datasets.

The first is the Living Costs and Food Survey and it's previous incarnations the Expenditure and Food Survey and Family Expenditure Survey (which we shall refer to collectively as the LCFS) (Department for Environment, Food and Rural Affairs, Office for National Statistics (2016)). The LCFS is a comprehensive, long-running survey of consumer expenditures involving between 5,000-8,000 households per year. Households are asked to record high-frequency expenditures in spending diaries over a two week period. Recall interviews are used to obtain spending on information on big ticket items (such as holidays or large durables) as well as standing costs on items such energy and water, internet bills and magazine subscriptions. The survey also collects information on incomes, demographic characteristics and, since 1992, on the value of households' mortgages (but not on other aspects of household balance sheets such as home values).

The second dataset we use is the British Household Panel Survey and its successor Understanding Society (both of which we shall refer to as the BHPS) (University of Essex. Institute for Social and Economic Research (2010); University of Essex. Institute for Social and Economic Research. (2016)). The BHPS is available in 18 waves from 1991 to 2008. Understanding Society began in 2009 and incorporated the original BHPS sample members from 2010 onwards. Both surveys include limited information on household spending on food and drink as well as self-reported house values. The BHPS contains data on total mortgage debt from 1993 onwards, while Understanding Society dropped these variables in its second wave in 2010. In the remaining years, we continue to observe whether households own their homes outright, and details on the length and type of their mortgage if they have one. We use these along with past information on mortgages values to impute mortgages in years following 2010 (see Appendix D.1 for details). Loan-to-value ratios are calculated by dividing the value of mortgages by the (self-reported) value of homes. The BHPS and Understanding also contain information on whether households own a second home.

The need to use two UK surveys comes from the fact that consumption spending is observed in the LCFS but leverage is not. At the same time the BHPS includes information on leverage but not on consumer spending. Hence our need to use two-sample methods that combine the information contained in both datasets, as we describe below.

The third dataset we use is the Panel Study of Income Dynamics (PSID). The PSID is a US-based panel of households that includes information on home ownership, household balance sheets, income and spending decisions. Since 1997, the survey has been biennial. The PSID has included questions on the value of households home equity and mortgages on an annual basis from 1999 onwards. Prior to 1999, these were only asked every 5 years. In terms of spending data, the survey only consistently included spending on food and rental payments until 1999. In that year, this was extended to cover other non-durable expenditures (including health, utilities, education and childcare). Other expenditures such as clothing and entertainment were added in 2005. Since 2001, households have been asked whether they have undertaken home improvements worth $10,000 or more since January of the year two years prior to the interview. If they answer in the affirmative, they are then asked to give the exact amount spent.[5]

For house prices we use regional/state-level data on the prices of transacted houses published by the Office for National Statistics (for the UK) and the Federal Housing Finance Agency (for the US).

In all of what follows, we drop households where the head is aged under 25 or over 65. To avoid problems of measurement error when estimating our first stage, we also drop households who have a lagged housing portfolio share in the top 1% of the distribution and those who have negative equity. We also drop households resident in Northern Ireland from both the BHPS and the LCFS samples as these were only introduced into the BHPS sample in later years. Finally, to avoid our results being partially driven by households moving at times of high or low house price growth, we drop households who have lived in their home for less than one year.

Table 6.1 provides some descriptive statistics for our three samples. We report these for 1993-2013 in the UK data, and for 2005-2013 for the PSID (i.e the years when the most comprehensive spending data was available). The proportion

[5]We annualise this figure using individuals' month of interview to determine the exact length of the period covered by this question.

of those owning their own homes and the average tenure among homeowners are similar across the two UK surveys, at around 70% of households. Ownership rates are somewhat lower in the PSID at around 55%. Focussing on home owners, the average loan-to-value ratio in our BHPS sample is 0.34, while US households tend to be more leveraged with an average loan-to-value ratio of 0.54. As we see below this is partly accounted for by the fact the age gradient in leverage is less steep in the United States, meaning that older households in the US tend to have higher loan-to-value ratios than their UK counterparts.

Figures from the LCFS show that non-durables is the largest component of expenditure (accounting for 76%). Residential investment spending, which includes extensions, renovations, household repairs, large furniture, carpets, and large household appliances accounts for roughly 7% of total spending. The remainder is accounted for by spending on non-residential durables.

The spending questions included PSID are as detailed as in the LCFS, and so we are forced to define categories differently for the US. We measure residential investment in the PSID as the sum of responses to the question "how much did your family spend altogether on household furnishings and equipment, including household textiles, furniture, floor coverings, major appliances, small appliances and miscellaneous housewares?" and responses to questions regarding home improvement spending (which are censored from below at $10,000). Since we are unable to exclude spending on small furnishings and smaller electrical appliances from this value, this definition is somewhat broader than the one used in the UK. As measured, it accounts for 6.9% of total spending. Non-residential durable spending in the PSID is essentially restricted to cars. Relative to our definition for the LCFS, it therefore excludes audio-visual equipment, as well appliances such as vacuum cleaners and microwaves (which may be included as durable household furnishings). This category accounts for 10.6% of expenditures. The remaining 82.5% of measured spending (including clothing, utilities, entertainment, vacations, motor fuel, healthcare and child care) is classified as going on non-durables. In all of these categories, non-responses to individual questions are treated as implying zero expenditures.

**Table 6.1:** Descriptive statistics, BHPS and LCFS and PSID

| | BHPS | LCFS | PSID |
|---|---|---|---|
| | (1993-2013) | (1993-2013) | (2005-2013) |
| Age | 44.5 | 44.2 | 45.3 |
| % Own home | 68.7% | 70.3% | 58.6% |
| | | | |
| *Homeowners* | | | |
| Years at address | 11.0 | 10.2 | 11.8 |
| LTV ratio | 0.34 | - | 0.44 |
| $\omega_t$ (housing share) | 3.04 | - | 3.66 |
| Total spend ($ ann.) | - | 43,764 | 65,242 |
| *Non-durable* | - | 33,430 | 53,802 |
| *Durable* | - | 6,068 | 6,917 |
| *Residential inv.* | - | 4,266 | 4,524 |
| % Res inv. $>0$ | - | 79.6% | 73.2% |

Note: UK data is for the period 1993-2013. US data is for the period 2005-2013 when more comprehensive spending measures are available in the PSID. See text for details of what is included in each spending category.

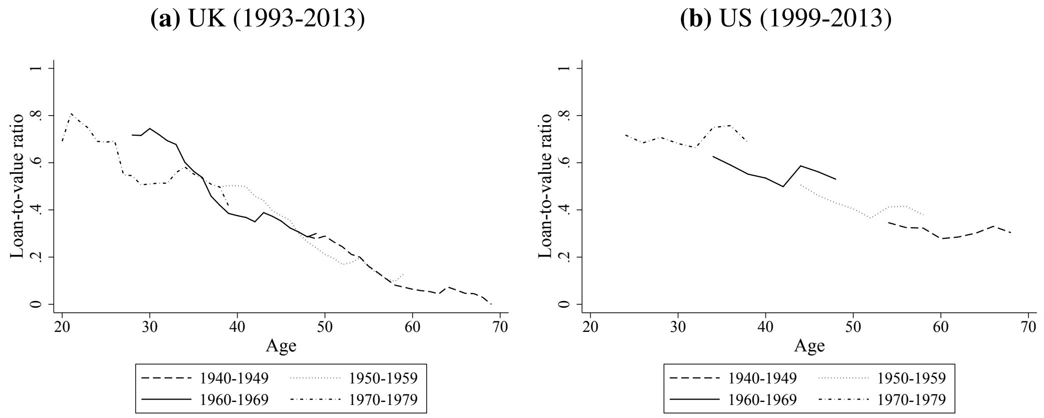
## 6.3 Household re-leveraging behaviour

We now turn to describing how the leverage of households in the UK and US has evolved over time and, in particular, how it has responded to increases in house prices in the two countries. We begin by describing aggregate trends in leverage and house prices, before moving on to household-level panel regressions that capture within household re-leveraging and de-leveraging behaviour over time.

Figure 6.2 shows how leverage evolves over time across four different 10-year birth cohorts (born between the years 1940 and 1970). In both the UK and the US there is a steady and reasonably smooth decline in leverage by age. In the UK there are pronounced differences in leverage between cohorts at younger ages. However, the different cohorts largely converge to similar leverage by around age 45. As we discuss further below, the differences in initial leverage across UK cohorts are likely to be explained by the differing credit conditions and house prices the different cohorts were exposed to at the point they became home-owners. This is a source of variation in leverage which we exploit to identify the role of leverage in spending decisions in what follows.

In the US, there is much less evidence of cohort effects, and the decline in leverage by age is much less steep than in the UK.

**Figure 6.1:** LTV ratios by age and cohort




Note: Authors' calculations using BHPS/Understanding Society and Panel Study of Income Dynamics.

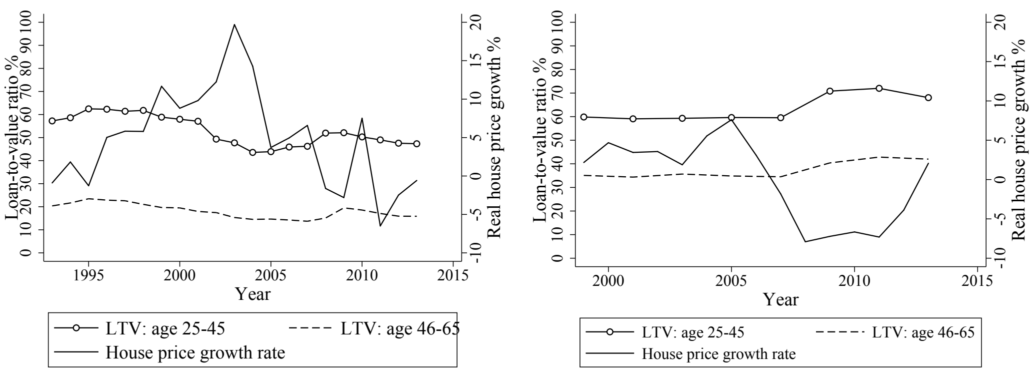
Next we consider how leverage varies over the house price cycle. The two panels of Figure 6.2 show how average real house price growth varied from year to year in the UK and US, as well as changes in households' average loan-to-value ratios among both younger (aged 25-45) and older (45-65) households in the two countries.

In the US, loan-to-value ratios among both younger and older households remained strikingly stable throughout the period of house price growth that continued until 2006. This peaked at a national rate of 7.6% in 2005, when loan-to-value ratios were essentially unchanged from the previous year (at around 60% for those aged 25-45 and 40% for those aged 46-65). When real house prices started to decline from 2007 onwards however, loan-to-value ratios rose rapidly. House prices fell by 2% in 2007 and between 7-8% in each of the years from 2008-2011. Over this period, the average LTV among younger households increased from 62 to 71%, while for older households it increased from 37 to 44%.

Such declines did not occur in the UK, where the house price slump was modest relative to both previous UK house price slumps and to the declines observed in the US (Bénétrix et al. (2012)). We plot UK house price and loan-to-value data from 1993-2013. For most of this period, UK house prices were increasing, with annual falls only observed in 1994-1995 and 2007-2009. In the period in between these years, house prices grew rapidly. Annual price increases peaked in 2003 at a rate of almost 20%. There is more evidence of a fall in average loan-to-value ratios as house prices rose in the UK than in the US. Loan-to-value ratios fell somewhat in the period of greatest house price growth among the under 45s, falling from 62% in 1995 to 43% in 2004 before climbing again as house price growth moderated (the over 45s saw smaller changes in their average leverage). However, as we now discuss, there is evidence that UK households were also engaging in re-leveraging behaviour over this period, even if the scale was not as great as it was in the US.

**Figure 6.2:** LTV ratios and house price growth rates

**(a)** UK (1993-2013) **(b)** US (1999-2013)




Note: House prices for the UK are national averages taken from the Office for National Statistics HPI deflated using the UK CPI. House prices in the US are national averages taken from the Federal Housing Finance Agency and are deflated with the US CPI. Loan-to-value ratios are taken calculated for the UK using data from BHPS and Understanding Society and for the US using the PSID.

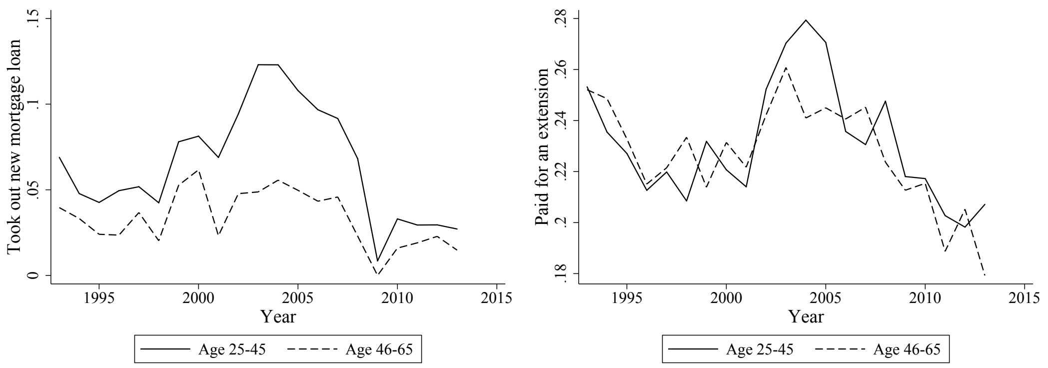
Panel (a) in Figure 6.3 shows that younger households in the UK did not passively allow leverage to fall as prices rose. Rather they responded by increasing the amount of new borrowing. The proportion of home-owning households aged 25-45 observed taking out additional mortgage debt in the BHPS increased to exceed 10%

in the period of most rapid house price growth. Panel (b) of Figure 6.3 shows that this period of new borrowing also coincided with growth in the proportion of households with positive spending on housing extensions. This activity was also focused on younger households. As we saw in Figure 6.2, older households responded much less to these developments. The importance of extensions and other home improvements as a reason for new borrowing is confirmed when we consider the uses for which households report taking out additional mortgage debt. Households in the BHPS are asked whether new loans were used for extensions, home improvements, car purchases or other household spending (households may give more than one answer). We class the first two of these responses as “residential investment’ and second two as “consumption” and plot the proportions reporting each motive for household heads aged 25-45, and 46-65 in panels a and b of in Figure 6.4. Both younger and older households are roughly four times more likely to report taking out a loan for residential investment than for consumption spending indicating that this is a key reason for households to re-leverage.

**Figure 6.3:** New mortgage loans and housing extensions by homeowners in the UK, 1993-2013

**(a)** Proportion taking out additional mortgages

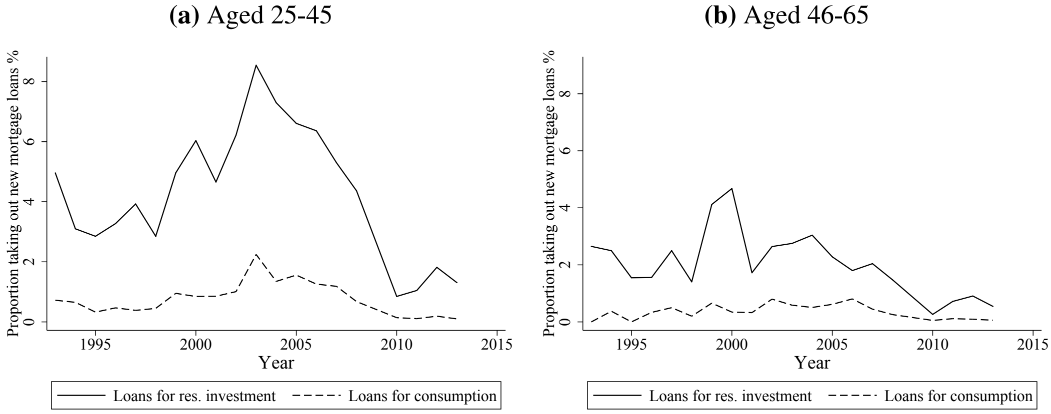
**(b)** Proportion with positive spending on housing extensions




Note: British Household Panel Survey/Understanding Society and Living Costs and Food Survey.

Figure 6.5 shows trends the proportion of home-owners engaged in new mortgage borrowing alongside changes in the proportion of households undertaking home improvement spending for the PSID (both since the previous wave - i.e in

**Figure 6.4:** Purpose of new mortgage loans, 1993-2013



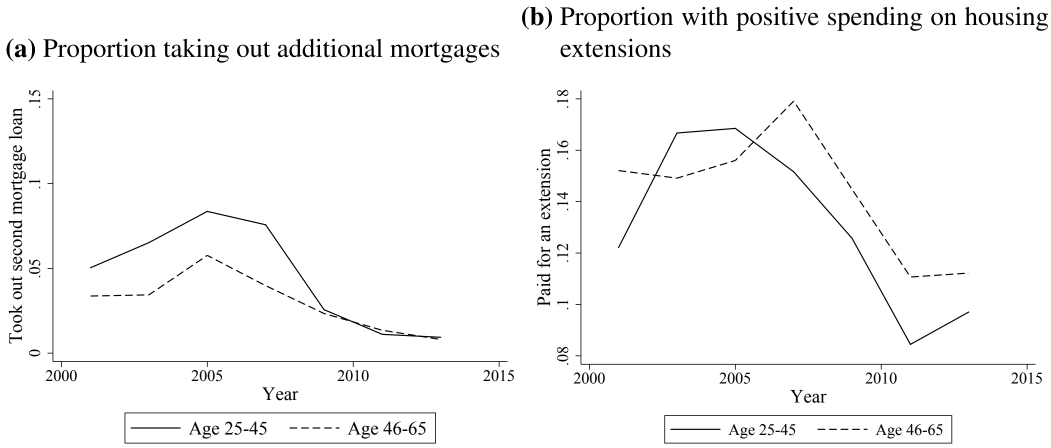

Note: Authors' calculations using British Household Panel Survey/Understanding Society.

the previous two years). As in the UK, younger households in particular were more likely to take out new mortgage loans in periods of high house price growth. There is also evidence of increases in home improvement spending when house prices growth was highest in between the 2003 and 2007 waves of the survey. Unlike in the UK, however, this is not clearly greater for younger households. While the PSID does not include questions on the motives for additional mortgage borrowing as in the BHPS, previous studies have pointed to home improvement spending as a key motive for equity withdrawal (Cooper (2010)).

Taking these trends together, Figure 6.6 shows trends in average household debt and equity (normalised by income) in the two countries. This illustrates the increase in the size of household balance sheets over the period of the house price boom, with increases in home equity (house values less mortgages) accompanied by increases in household debt.[6] The increases in home equity were much larger in the UK both relative to income and to average mortgage debt than they were in the US. In 2013, the average home equity to income ratio in the UK was five, compared to three in 1993. In the US, the average home-equity to income ratio was 1.5 in 2013 (compared to around 1 in 1999). We also note how debt burdens in the US decreased only slightly as home equity fell in the years following the financial

[6]Here home equity is calculated using data on the value of the primary residence.

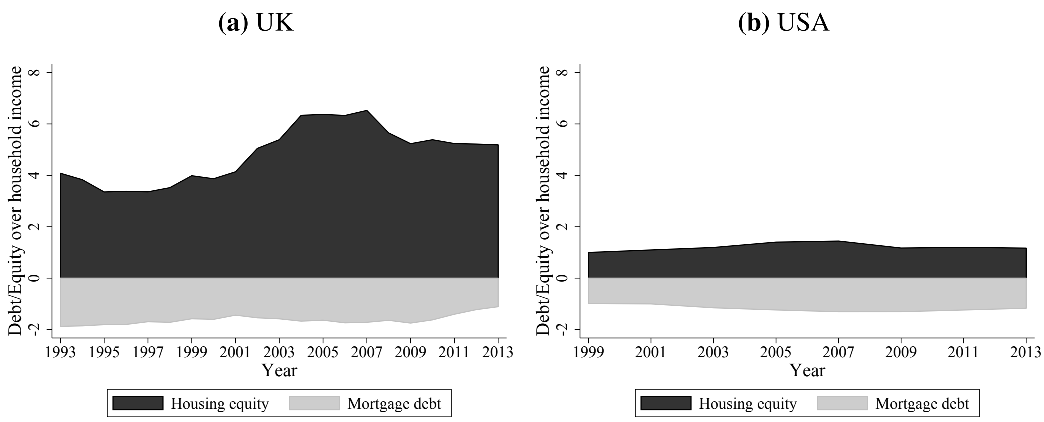
**Figure 6.5:** New mortgage loans and housing extensions by homeowners since previous PSID wave, 1999-2013




Note: Authors' calculations using Panel Study of Income Dynamics.

crisis. This is suggestive of an asymmetry in the relative ease of re-leveraging in response to house price increases and de-leveraging in response to house price falls.

**Figure 6.6:** Housing equity and debt (normalised by income)




Note: Authors' calculations using the British Household Panel Survey and Panel Study of Income Dynamics.

### 6.3.1 Panel regression

To more formally examine the extent to which households adjusted their leverage as prices rose, we report in Table 6.2 results from regressing changes in log mortgage debt on changes in households' log self-reported home values. That is, we run the regression equation

$$\Delta \log M_{it} = \delta \Delta \log HValue_{it} + \varepsilon_{it} \tag{6.15}$$

on a sample of home-owners who are never observed with an LTV of zero. Here $M_{it}$ is each household's total mortgage debt. Running this regression allows us to directly examine within household responses to changes in house values over time. We report results from both the BHPS and the PSID and since the PSID has been biennial since 1999, in both surveys we consider changes over the previous two years.

If mortgage debt did not adjust as house prices increased, leaving LTV ratios to fall passively with house prices, then we would expect the coefficient on log house values to equal 0.

Panel (a) of Table 6.2 presents these results for the US. Column (1) shows the results for a simple OLS regression, initially focussing on the sub-sample of households who do not move from one wave to the next. The results indicate that each 10% increase in house prices appears associated with a 3.4% increase in mortgage debt. Measurement error in self-reported house values has the potential to attenuate this coefficient towards zero. Column (2) therefore shows results when we instrument the change in house prices with state-level house price growth. The results are very similar, suggesting that measurement error is not a significant problem. The results also do not change much when we include a control for age to account for life-cycle changes in leverage in Column (3). As a result we report only OLS results for the subsequent columns. In Columns (4) and (5) we look for evidence in asymmetries of responses when households are re-leveraging versus de-leveraging by splitting the sample according to whether real regional house prices rose or fell relative to the previous wave. The coefficient on house price changes when prices are rising is significantly greater than at times when house prices are falling. Our results imply that a 10% increase in US house values was associated with a 3.8% increse in mortgage debt, while a 10% fall in house prices is associated with only a 2.5% reduction in mortgage debt. In the remaining three columns we consider how these results change when we include households who move address from one wave to the next. We estimate that debt responses to house price changes are nearly twice as large when movers are included. This holds true when we separately consider responses to house price falls and house price increases in Columns (5) and (6), indicating that up-sizing and down-sizing are important mechanisms by which households adjust their leverage.

Panel (b) shows equivalent responses estimated from the BHPS from 1993-2009. Here we do not include data from Understanding Society, so as to exclude the changes in mortgage debt which we have imputed from 2011 onwards. In the UK, debt responses to house price gains are not as large as they are in the US but remain substantial. The results in Column (1) indicate that each 10% increase in house prices is associated with a 2.0% increase in mortgage debt. As in the PSID, these results do not change much when we instrument for changes in house value (Column (2)) or include an age control (Column (3)). Unlike in the US, we do not find clear evidence of asymmetric responses. The estimated effect of changes in house values on mortgage debt is smaller in response to negative changes than positive ones, but it is not possible to draw clear conclusions for this as, owing to the limited number and duration of episodes of falling UK house prices over the period we consider, the coefficient estimated in Column (4) is very imprecise. When we include movers, the estimated responses of mortgage debt to house price changes rises as it does in the US, though not quite to the same levels. In the sample which includes movers, a 10% inrease in house prices is associated with a 4% increase in debt compared to 6% in the US.

**Table 6.2:** Changes in mortgage debt and changes in house prices over two-year periods

| $\Delta logDebt$ | OLS | IV | IV | OLS | OLS | OLS | OLS | OLS |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Panel a: US 1999-2013 | | | | | | | | |
| $\Delta \log HValue$ | 0.344*** | 0.326*** | 0.323*** | 0.253*** | 0.380*** | 0.602*** | 0.539*** | 0.647*** |
| | (0.058) | (0.038) | (0.038) | (0.042) | (0.097) | (0.052) | (0.038) | (0.089) |
| $R^2$ | 0.031 | 0.031 | 0.031 | 0.012 | 0.039 | 0.118 | 0.080 | 0.135 |
| $N$ | 10,811 | 10,777 | 10,777 | 5,088 | 5,723 | 11,895 | 5,489 | 6,406 |
| Clusters | 3,358 | 3,350 | 3,350 | 2,430 | 2,620 | 3,477 | 2,534 | 2753 |
| Panel b: UK 1993-2009 | | | | | | | | |
| $\Delta \log HValue$ | 0.200*** | 0.238*** | 0.238*** | 0.239 | 0.196*** | 0.413*** | 0.571*** | 0.411*** |
| | (0.024) | (0.037) | (0.037) | (0.154) | (0.024) | (0.035) | (0.179) | (0.036) |
| $R^2$ | 0.012 | 0.012 | 0.016 | 0.007 | 0.012 | 0.052 | 0.055 | 0.052 |
| $N$ | 12,928 | 12,860 | 12,860 | 541 | 12,387 | 14,131 | 586 | 13,545 |
| Clusters | 2,713 | 2,708 | 2,708 | 398 | 2,670 | 2,763 | 422 | 2,719 |
| *Controls* | | | | | | | | |
| Age | | | X | | | | | |
| *Restrictions* | | | | | | | | |
| Non-movers | X | X | X | X | X | | | |
| House price growth$< 0$ | | | | X | | | X | |
| House price growth$> 0$ | | | | | X | | | X |

Notes: * $p < 0.10$ , ** $p < 0.05$ , *** $p < 0.01$. Standard errors in parentheses. Standard errors clustered at the individual level. The log change in house values is instrumented with the log change in nominal house prices within each region.

The results show evidence of re-leveraging behaviour in the face of price increases, as we would expect in a model where consumers treated leverage as a portfolio choice. We also find evidence that de-leveraging responses in the face of price falls are slower than re-leveraging in response to house price gains.

A number of factors might explain why the re-leveraging responses we estimates in the US appear so much larger than they do in the UK. One set of reasons relates to the differing institutions in the two countries. In some US states, mortgage loans are non-recourse, meaning that lenders cannot pursue debts that are not covered through sales of foreclosed properties. This may make borrowers more comfortable with the risk of negative equity, since the costs of default in this situation are smaller. Ghent and Kudlyak (2011) for instance find that the monthly probability of default for borrowers in a state of negative equity is 32% higher in states where the is no threat of recourse. In the UK, mortgage loans are recourse loans. Secondly, in the US, mortgage interest is tax deductible, creating more of an incentive to both pay-off mortgages less quickly and to increase mortgage debt when prices rise. This was only true to a very limited extent in the UK during the period we consider. Mortgages were only tax deductible up to a fixed nominal cap of £30,000 (far below the average house price). From 1994 the size of the deduction was limited to 20% falling to 10% in 1995. The tax deduction was eliminated in 1999. Using variability in the importance of tax deductibility in the UK over time, Henderschott et al. (2003) find that it can have substantial effects on households' initial loan-to-value ratios. It may also have similar effects on incentives to re-leverage, and so partly explain some of the differences in behaviour we observe between the US and the UK.

Another important factor is the pace of house price increases in the UK. This would mean that UK households would need to make much larger adjustments to their mortgage debt in order to maintain constant loan-to-value ratios. This may have led to intolerably large increases in their loan-to-income ratios (lenders are typically only willing to lend households 3-5 times the value of their incomes). Figure D.2 in Appendix D.3 shows that the propensity for households in both countries to increase their mortgage debt in response to house price increases first rises then falls with their lagged loan-to-income ratios. It also shows that UK households were much more likely to have higher loan-to-income ratios than their US counterparts, suggesting that LTI constraints were more likely to prevent UK households from borrowing against the full value of house price increases.

## 6.4 Spending on home improvements

The following two sections discuss homeowners residential investment responses to increasing house prices. There are various ways a homeowner can invest in housing. They may purchase other residences, up-size, or make investments in their existing home through renovations and home improvements. In this section we discuss the last of these responses.

In order to properly constitute 'investment' rather than pure consumption spending, home improvements must increase the market value of the property and not just the owner's private valuation. This may be difficult however, as prospective buyers of a property may not derive the same consumption benefit to any additions or alterations in the same way as the current owner. For example, a housing extension built in a home's backyard may increase the home's value to the current owner (who desires more floor space) but discourage buyers who value a garden. Moreover, asymmetries of information may drive a wedge between the market and private valuations of a particular improvement. Necessary structural repairs may not increase the sale price of their home if buyers cannot verify that they have been done or done to an acceptable standard.

However, there is evidence that a substantial share of the costs of home improvement spending is recouped through increased home values. *Realtor* magazine conducts an annual survey of the costs and value added associated with different home improvement projects in different US housing markets. Real estate agents are asked to the expected value different projects are expected to add to a home's sale price, while professionals in the remodelling industry are asked to provide estimates of their likely cost. Taken together these two estimates provide an estimate of the proportion of costs of different projects that homeowners can expect to recoup through higher re-sale values. In 2016, the average value-cost ratio of investments made on properties sold within a year was 64%. Investments in attic insulation had the most cost effective effects on resale values, with 117% of costs recouped through higher home values.[7] Bathroom additions had the lowest returns with 56% of costs being recouped.[8]

The fact that homeowners can expect to recoup a significant fraction of the costs of home improvement means that investment motives are likely to play an important role in households' decisions to make such expenditures. Moreover, the returns to investments in one's own home appears to increase along with local home values, suggesting that this is indeed a way that households can increase the importance of housing in their overall portfolios. Gyourko and Saiz (2004) find that home improvement spending responds strongly to the ratio of local house values to construction costs, which is consistent with a rational investment motive for such projects that responds to house price growth. Choi et al. (2014) investigate the impact of local house price growth on the average ratio of costs recouped as measured by the *Realtor* survey, controlling for other factors such as local unemployment and income growth. They also find that the investment value of home improvement projects is positively associated with local house price growth.

### 6.4.1 Empirical analysis

To test the specific hypothesis that more leveraged households will disproportionately increase housing investment in response to house price increases, we estimate the equation

[7] http://www.remodeling.hw.net/cost-vs-value/2016/

[8] Similar surveys exist in the UK, for example the insurance company GoCompare provides a property investment calculator which provides estimates of the costs and returns associated with different projects. This suggests greater returns to home improvement spending in the UK, although the methodology behind the calculator has not been published. As in the US, Energy-saving investments have the highest returns, while net bathrooms have negative returns (https://www.gocompare.com/home-insurance/property-investment-calculator/).

$$\ln C_t = \beta_0 + \gamma_{crt} + \beta_1 w_{t-1} + \beta_2 \left\{ w_{t-1} \times \left( \frac{p_{rt}}{p_{rt-1}} - 1 \right) \right\} + X\beta_3 + e_t \qquad (6.16)$$

One concern about directly estimating equation (6.16) is that leverage is potentially endogenous. The conventional approach to estimating leverage effects is to use individuals' lagged leverage (uninstrumented). As we have seen however, lagged leverage is itself a choice variable that will depend on households expectations of the risk and returns associated with housing investments at the time when consumption and other spending decisions are made. As we document in what follows, lagged leverage is also correlated with gross house values and income from non-housing assets. In order for our empirical application to identify the effects of independently varying leverage, these other variables ought to be held constant. For this reason we instrument leverage using housing market conditions at the time individuals moved into their homes.

A second issue concerns data availability. Long-running surveys that contain balance sheet data on wealth and leverage rarely contain comprehensive consumption measures. A panel survey is required in order to know the consumer's lagged leverage position $\omega_{it-1}$.

Previous studies have addressed this problem by either using available proxies for consumption (such as borrowing, (Mian and Sufi (2011))), subsets of consumption that are observed (e.g. food spending as in Lehnert (2004)), or measures backed out from the consumer's budget constraint (using the difference between observed income and wealth changes, as in Cooper (2013)). Each of these approaches has drawbacks. Changes in particular categories of consumption, or variables related to consumption need not give the full spending response to shocks.[9] They also do not allow us to investigate how the composition of spending varies as house prices change, which is crucial for allowing us to test the importance of portfolio rebalancing motives for re-leveraging. In addition, the use of the budget constraint identity

[9] Credit card borrowing, which is used as proxy in Mian et al. (2013), may also be more cyclical than other forms of spending. This point was made in Aladangady (2017).

to impute consumption can in general lead to biased estimates of wealth effects in the presence of measurement error (Browning et al. (2014)). If reported wealth in the previous period is smaller than actual wealth, then leverage as observed by the researcher in that period will be too high and consumption in the current period be too large, biasing estimates upward.

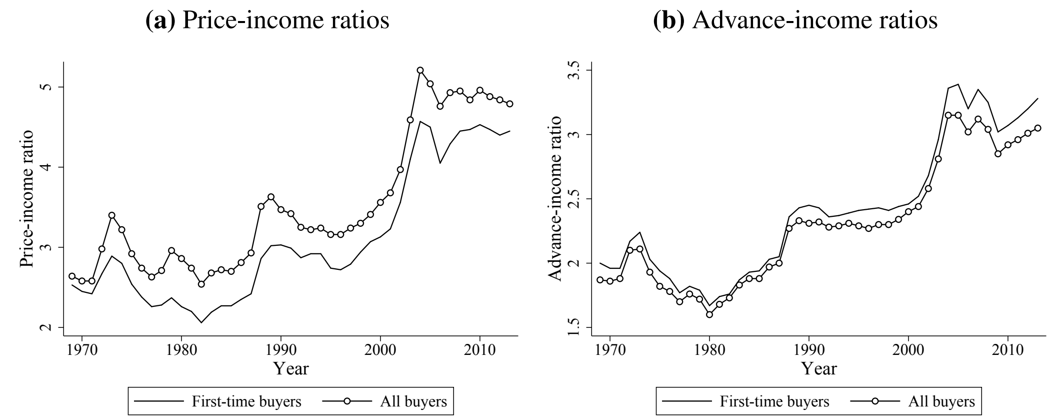
For these reasons, in the UK we use a two-sample IV approach (Angrist and Krueger (1992)) to combine spending data in the LCFS with data on leverage in the BHPS. This approach allows us to simultaneously impute and instrument for leverage in our (cross-sectional) UK expenditure dataset using balance sheet data taken from the BHPS. The instrument we use is the credit conditions households faced at the time they moved into their current residences. We discuss the strength and validity of our instrument further below. We provide additional details on the implementation of our approach in Appendix D.3.

In principle, in the US we could investigate these questions using the PSID, which in its later years contains information on both spending and leverage. However, the number of waves in which it includes comprehensive consumption data is relatively short (a problem that also applies to other panel surveys such as the HRS, used by Christelis et al. (2015) to study questions around leverage). In addition, as we saw in Figure 6.2 and in the analysis in Table 6.2, US households tend to rapidly re-leverage in response to house price increases. As a result, the leverage of US households is far less dependent on past circumstances than it is for UK households, and so our instrument has no power in the US. In what follows, we therefore focus on UK results.

### 6.4.2 Instrument and identification

For our proposed method we require a source of variation in leverage that explains why some households took out larger loans than others that is common to both the BHPS and the LCFS. For this purpose we exploit variation in the average price to income ratios for new loans at the time households moved into their current residences (denoted $P/Y_{-T}$). This variable is often used as a measure of the cost of credit (*loan* to income ratios for example included in the credit conditions index

**Figure 6.7:** Credit conditions, 1969-2013




Note: Data from Office for National Statistics.

of Fernandez-Corugedo and Muellbauer (2006)). In our case it indicates the cost of borrowing in the years house prices were made, and so the degree to which households would have been able to leverage their housing purchases at the time they moved.

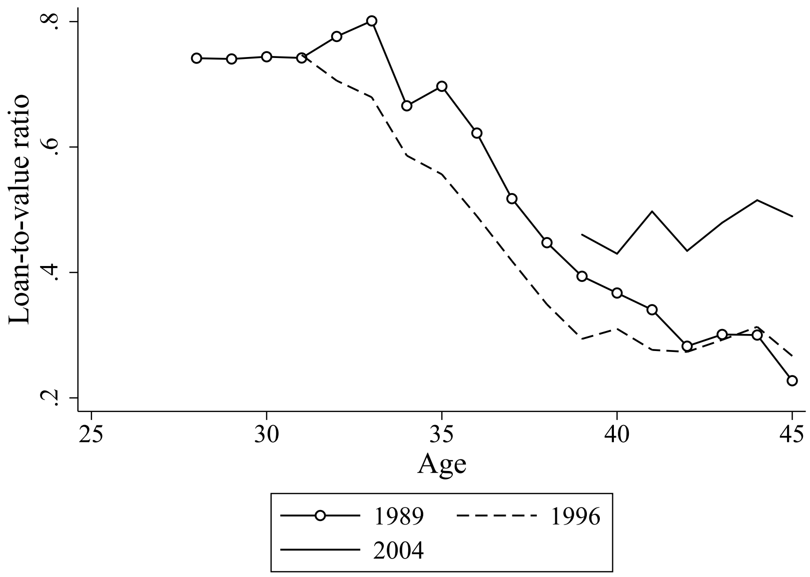
The solid line in Figure 6.7 (Panel (a)) shows how this instrument varies over time in the UK. There is a gradual upward trend in the price to income ratio suggesting that credit has become looser over time. In 2013, average loans were almost five times greater than the incomes of buyers. This compares to a ratio of 2.5 in 1969. This provides one source of identification. Importantly however, there is also cyclical variation in this variable, with for example evidence of credit tightening following the 2008 financial crisis. Movements in other measures of credit conditions such as the average deposit on new homes (Figure 6.7, Panel (b)) show similar patterns.

Our instrument is only available from 1969 onwards, and so in what follows we drop households who moved into their homes before this. This constitutes roughly 0.5% of the total number of observations in our LCFS sample.

#### 6.4.2.1 Instrument relevance and validity

For our instrument to be considered appropriate it must satisfy two requirements. The first is that the instrument must be relevant (that it is indeed correlated with the endogenous variables it is replacing). Figure 6.8 shows how our instrument relates to loan-to-value ratios for a given cohort (those born in the 1960s). This is the only ten-year birth cohort that we observe for almost our entire sample period. We plot loan-to-value ratios for households who moved into their homes in three different years: 1989,1996, and 2004. These three years represent peaks and troughs in price-to-income ratios on new housing purchases from Panel (a) in Figure 6.7. Price to income ratios reached a temporary high of 3.7 in 1989 before falling to a low of 3.2 in 1996. Thereafter they increased to a peak of 5.2 in 2004. As Figure 6.8 shows, households that moved when price-to-income ratios were relatively high in 1989 tended to have higher leverage than those in the same cohort who moved in in 1996. This is true not only at the point they moved in to their current homes but also long-afterward. Loan-to-value ratios are also persistently higher for those who moved in when credit conditions were even looser in 2004.

**Figure 6.8:** Loan-to-value ratios by age and year moved in (1960s cohort)




Note: Authors' calculations using British Household Panel Survey/Understanding Society.

This relevance of our instruments can be more formally tested by looking at the results of first stage regressions. We do this in Table 6.3.

We have two first stage regressions, one for leverage and one for leverage interacted with house prises. In both cases, the F-statistics are greater than the value of 10 suggested as a rule of thumb by Staiger and Stock (1997a) for IV estimated using a single sample.[10] Two sample IV methods may suffer less of a bias than standard 2SLS estimators, as errors in the first stage estimation will be unrelated to errors in the second stage equation. This is indeed the rationale for estimators that run first and second stages in split samples (Angrist and Krueger (1995)). Nonetheless weak instruments may still result in coefficients being biased towards zero in finite samples. The relatively strong first stage we obtain is therefore reassuring. Kleibergen-Paap statistics for the first stage also heavily reject the hypothesis of underidentification.[11]

**Table 6.3:** First stage results

| | $\omega_{it-1}$ | $\omega_{it-1} \times (\frac{p_{rt}}{p_{rt-1}} - 1)$ |
|---|---|---|
| $P/Y_{-T}$ | 0.391 | -0.011 |
| | (0.043) | (0.003) |
| $P/Y_{-T} \times (\frac{p_{rt}}{p_{rt-1}} - 1)$ | 0.451 | 0.670 |
| | (0.362) | (0.050) |
| Shea partial $R^2$ | 0.008 | 0.030 |
| F-stat (p-value) | 40.14 | 114.02 |
| | (<0.001) | (<0.001) |
| Kleibergen-Paap (p-value) | 72.46 | |
| | (<0.001) | |
| $N$ | 29,472 | |
| Clusters | 7,797 | |

The second requirement for a suitable instrument is that it is itself uncorrelated with the error term.

[10] Later we will also report results for a younger sub-sample of households (those with heads aged 25-45). In this sub-sample, we obtain an even stronger first stage (with F-stats for $\omega_{t-1}$ and $\omega_{it-1} \times (\frac{p_{rt}}{p_{rt-1}} - 1)$ of 19.75 and 80.20 respectively).

[11] A further 'first stage' check we can conduct is to look for a positive association between our instrument and total mortgage debt in the LCFS. This would us that the association between our instrument and leverage is not limited to our first sample. Regressing mortgage debt on $(P/Y_{-T})$ and our controls yields a positive coefficient with a t-statistic of 24.07.

In our case, there may be concerns that those who move home in years with higher price-income ratios will have spending patterns that are different to those who moved in other years for reasons other than the degree of their leverage. The most obvious challenge is that since price-income ratios have tended to increase over time, those households with higher values of our instrument will tend to have moved more recently. They may therefore be younger, or be more likely to furnishing a new home. We address these concerns of this nature directly by including a control for years households have spent in their current address (in addition to a dummy variable for households having moved in in the last year to account for first year 'setting up' expenses). By including a full set of cohort-region-year interactions we also control for any region or cohort specific trends in income growth that may be correlated with house price changes. These dummies can be thought of capturing shocks that are potentially correlated with house price movements but differ in their effects across young and old or across different regional labour markets. One such shock is to future income expectations which would be expected to boost the consumption of younger (and so more leveraged) cohorts by more. If effect such as these are not controlled for they could lead to spuriously large estimates of house price wealth effects for younger households (Attanasio et al. (2009)).

The use of our instrument in combination with these controls means we effectively compare the spending responses of house price changes between two households in the same region and same cohort but who moved into their homes at different times (when credit was either looser or tighter).[12]

Questions about endogeneity may remain however. For example, households may have been more likely to move when house prices were high because greater unobservable wealth made them less price sensitive. They may also have moved into larger houses. This would create a spurious association between our instrument and consumption. Households who moved at times when credit was loose may

[12]The inclusion of cohort-region-year fixed effects means that we will only identify the *relative* effects of house price changes across different households within each region-cohort-year cell. Common effects of house prices changes affecting all households (and any general equilibrium effects on either national or regional housing markets) will be absorbed by our fixed effects (Steinsson and Nakamura (2018)).

also be more likely to move in response to economic shocks and drop out of our sample introducing a selection bias. The assumption that such omitted factors do not induce a correlation between instruments and the error term is usually something which cannot be verified. Omitted variables are typically omitted because they are unobserved. However, when using a two sample approach, such tests are possible. Some variables may be observed in the sample in which we run our first stage regressions even if they are not present in our main sample.

To address additional endogeneity concerns we therefore look for an association between our instruments and gross house values, asset incomes and the probability of being a mover in the BHPS and Understanding Society panels conditional on our covariates. Table 6.4 reports results from regressions of these potential sources of endogeneity on our instruments and our other covariates. The instruments are both jointly and individually insignificant in all models suggesting that they are plausibly orthogonal to these omitted variables. In addition to the results shown in Table 6.4, we also regress unsecured debt to income ratios and an indicator for whether households have positive debts on our instruments. Debts are only observed in 3 of the 18 waves of the BHPS survey, and so these tests are necessarily conducted on a much smaller sample. The instruments are again individually and jointly insignificant in these regressions.

A further exercise we can do is test to how our instrument compares to the use of households lagged LTV ratios to sort households according to their leverage. This is the source of variation used in a number of previous studies (e.g. Disney et al. (2010), Dynan (2012)). The results of this comparison are shown in the second panel of Table 6.4. There is strong evidence that those with higher lagged leverage have fewer financial assets and tend to live in less valuable homes.[13]

### 6.4.3 Results

Table 6.5 shows results for using our estimation equation (6.16) for different forms of spending. These include total spending, non-durable spending (consumption),

[13]As an additional check we regress our instrument $P/Y_{-T}$ and lagged leverage on possible confounding variables simultaneously. We find again that these variables are not correlated with credit conditions when consumers moved into their homes, but are related to households' lagged leverage.

**Table 6.4:** Exogeneity of instruments

| Panel a | Credit conditions | | | |
|---|---|---|---|---|
| Dependent var. | $\log(HValue)$ | Invest inc. $> 1000$ | Invest inc.$= 0$ | Mover$_{t+1}$ |
| $P/Y_{-T}$ | 0.011 | -0.002 | -0.006 | -0.002 |
| | (0.012) | (0.009) | (0.014) | (0.004) |
| $P/Y_{-T} \times (\frac{p_{rt}}{p_{rt-1}} - 1)$ | -0.048 | -0.060 | 0.091 | 0.014 |
| | (0.106) | (0.078) | (0.123) | (0.037) |
| F-test (p-values) | 0.510 | 0.743 | 0.610 | 0.854 |
| N | 29,194 | 28,034 | 28,034 | 22,638 |
| Clusters | 7,681 | 7,642 | 7,642 | 6,327 |
| Panel b | Lagged leverage | | | |
| $LTV_{t-1}$ | -0.199*** | -0.168*** | 0.335*** | -0.0001 |
| | (0.0208) | (0.014) | (0.0221) | (0.008) |
| $LTV_{t-1} \times (\frac{p_{rt}}{p_{rt-1}} - 1)$ | -0.646*** | -0.133 | 0.158 | 0.050 |
| | (0.199) | (0.138) | (0.219) | (0.0860) |
| F-test (p-values) | $< 0.001$ | $< 0.001$ | $< 0.001$ | 0.841 |
| N | 29,194 | 28,034 | 28,034 | 22,638 |
| Clusters | 7,681 | 7,642 | 7,642 | 6,327 |

Notes: * $p < 0.10$ , ** $p < 0.05$ , *** $p < 0.01$. Standard errors in parentheses. Controls for education, cohort-region-year dummies, sex, house type, number of rooms, number of adults, number of children, years at address, and a dummy variable for having moved in in the previous year. Standard errors are clustered at the individual level.

durable spending (excluding residential investment), residential investment spending, spending on luxuries (food out and leisure services) and total spending less residential investment. We log each of these spending variables with the exception of durable spending, residential investment and luxuries, which we transform using the inverse hyperbolic sine (IHS) transformation as a significant fraction of households record zero spending on these categories. This is defined as

$$IHS(y) = \log(y + \sqrt{(y^2 + 1)}) \tag{6.17}$$

The IHS transformation approximates log values at high values of spending but remains defined at zero.

The regressions suggest little difference in the responses of total spending to leveraged house price gains. Our results imply that for a 10% increase in house prices, total spending would be just 0.50 % higher for outright owners and 1.00 % higher for those with a loan-to-value ratio of 50% (i.e $\omega_{it}$=2). The equivalent non-durable consumption responses fall to -0.2% and -0.4% respectively. These responses are small and not statistically significant, suggesting only limited housing wealth effects on consumption in our sample. For durable goods, we observe a negative response to house price increases, though our estimates are imprecise and not significant.

Column (4) shows however that there is strong response for residential investment spending. We estimate that a 10% increase in house prices results in a 7.4% increase in residential investment spending for outright owners, rising to 15% for those with a loan-to-value ratio of 50%.

One possible explanation for the results we obtain in column (4) is that housing is a luxury good, and that consumer spending on such goods will rise with increasing wealth independently of any investment motive. To examine this hypothesis, column (5) presents results for the effects of leveraged housing returns on a category of goods that are typically thought of as being luxuries. We do not find evidence of strong spending responses for these goods, lending additional support to our hypothesis that the increase in spending on residential investment reflects a desire to rebalance consumers' investment portfolios rather than pure consumption motives.

**Table 6.5:** Log spending responses

| | Total | Non-durables | Durables (IHS) | Res inv. (IHS) | Luxuries (IHS) | Total - Res |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $\omega_{it-1}$ | -0.020 | -0.017 | -0.041 | -0.045 | -0.049** | -0.021* |
| | (0.012) | (0.011) | (0.046) | (0.048) | (0.025) | (0.012) |
| $\omega_{it-1} \times (\frac{p_{rt}}{p_{rt-1}} - 1)$ | 0.051 | -0.025 | -0.192 | 0.741*** | -0.088 | -0.006 |
| | (0.072) | (0.065) | (0.265) | (0.279) | (0.142) | (0.070) |
| $R^2$ | 0.355 | 0.377 | 0.113 | 0.082 | 0.211 | 0.361 |
| N | 60,357 | 60,357 | 60,357 | 60,357 | 60,357 | 60,357 |

Notes: * $p < 0.10$ , ** $p < 0.05$ , *** $p < 0.01$. Standard errors in parentheses. Controls for education, cohort-region-year dummies, sex, house type, number of rooms, number of adults, number of children, years at address, and a dummy variable for having moved in in the previous year.

**Table 6.6:** Log spending responses: Ages 25-45

| | Total | Non-durables | Durables (IHS) | Res inv. (IHS) | Luxuries (IHS) | Total - Res |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $\omega_{it-1}$ | -0.013 | -0.003 | 0.002 | -0.105*** | -0.050** | -0.001 |
| | (0.010) | (0.009) | (0.037) | (0.041) | (0.020) | (0.009) |
| $\omega_{it-1} \times (\frac{p_{rt}}{p_{rt-1}} - 1)$ | 0.056 | -0.009 | -0.453** | 0.833*** | -0.007 | -0.003 |
| | (0.049) | (0.044) | (0.190) | (0.208) | (0.096) | (0.047) |
| $R^2$ | 0.319 | 0.324 | 0.106 | 0.092 | 0.184 | 0.314 |
| N | 29,557 | 29,557 | 29,557 | 29,557 | 29,557 | 29,557 |

Notes: * $p < 0.10$ , ** $p < 0.05$ , *** $p < 0.01$. Standard errors in parentheses. Controls for education, cohort-region-year dummies, sex, house type, number of rooms, number of adults, number of children, years at address, and a dummy variable for having moved in in the previous year.

In Table 6.6 we separately consider results for a younger subsample of households (those with heads aged 25-45). The results are similar to those in Table 6.5. As in the case of our full sample, there is no evidence of a differential response in total spending between leveraged and non-leveraged households.

### 6.4.4 Robustness checks

In addition to these results, we carry out a range of robustness checks. For reasons of space, we discuss the results of these briefly here, reporting the full set of results in Appendix D.4.

1. *Alternative definitions of residential investment.* The definition of residential investment we use above is relatively broad compared to what would for example be used in the national accounts. In particular, we include other fixtures and durable investments (such as for example, kitchen equipment) that we consider likely to be capitalised into the value of the property but which may be excluded in other definitions. In the Appendix, we also consider a narrow definition that is restricted to spending on changes to the structure of the property as well as household repairs and maintenance.[14] The results we obtain are very similar to our main results. We also find a positive effect when we use an indicator of whether households made investments in household extensions as our dependent variable.

2. *Alternative instruments.* We also consider results using two alternative instruments. The first is the Credit Conditions Index used in Fernandez-Corugedo and Muellbauer (2006), and the second is the average house price in each region at the time individuals moved into their current homes (as used as an instrument for mortgage debt in Chetty et al. (2017)). The use of these alternative instruments do not give very different results.[15]

[14] This definition is more in the spirit of national accounts. For example in the US National Income and Product Accounts, Private Fixed Investment includes "construction of new nonresidential and residential buildings.", "improvements (additions, alterations, and major structural replacements) to nonresidential and residential buildings." and "certain types of equipment (such as plumbing and heating systems and elevators) that are considered an integral part of the structure." (see Chapter 6 of U.S. Bureau of Economic Analysis (BEA))

[15] Differences in the total spending response of more and less leveraged households to house price

3. *Sample restrictions*. We exclude households who moved within the previous year from our analysis, but concerns may remain that our spending effects are driven by more recent movers (who are likely to be the most leveraged, and possibly at credit a constraint). We therefore consider results from an alternative sample where we exclude those who moved into their homes within the previous five years. Within this subsample we find slightly larger effects of house price changes on the investment spending of households who are predicted to be the most leveraged.

## 6.5 Other property investments

Households may also invest in housing by purchasing additional properties or by up-sizing their main residence. In this section we examine whether more leveraged households are more likely to make such investments in response to house price increases than other households, as our model would predict.

To do so we estimate the following equation using the BHPS

$$\Delta Y_{t,t+10} = X\delta_0 + \delta_1 \left( \frac{p_{rt+10}}{p_{rt}} - 1 \right) + \delta_2 \left[ \omega_{it-1} \times \left( \frac{p_{rt+10}}{p_{rt}} - 1 \right) \right] + u_t \qquad (6.18)$$

where $Y$ is some outcome of interest (second home ownership or the number of rooms in the household's main residence). We consider changes in these outcomes over a period of 10 years. This is to account for the possibility that, as a result of transaction and search costs, consumers may be slow to make new home purchases in response to increases in their housing wealth.

changes are greater when we use regional house prices at the time households move into their current residence as our instrument. However, the coefficient on the interaction between leverage and house price changes is only significant at the 10% level (and partly driven by larger estimates of effects on residential investment). Effects on nondurable and durable spending (examined separately) are not significant.

**Table 6.7:** Effects of leverage on second home ownership

| $\Delta secondhome_{t,t+10}$ | OLS (1) | OLS (2) | OLS (3) | IV (4) |
|---|---|---|---|---|
| $\left(\frac{p_{rt+10}}{p_{rt}} - 1\right)$ | -0.018 | - | - | - |
| | (0.047) | | | |
| $\omega_{it-1} \times \left(\frac{p_{rt+10}}{p_{rt}} - 1\right)$ | 0.003* | 0.003* | 0.000 | 0.037** |
| | (0.002) | (0.002) | (0.002) | (0.018) |
| Region dummies | N | Y | Y | Y |
| Other controls | N | N | Y | Y |
| N | 3,761 | 3,761 | 3,598 | 3,598 |
| Clusters | 1,465 | 1,465 | 1,393 | 1,393 |

Notes: * $p < 0.10$ , ** $p < 0.05$ , *** $p < 0.01$. Standard errors in parentheses. Controls are year dummies, dummies for 10-year birth cohorts, age, age squared, years at current address and a dummy for having just moved in. Standard errors are clustered at the individual level.

Table 6.7 shows results for the change in second home ownership. Column (1) presents OLS results for the effect of regional house price changes and regional house price changes interacted with leverage with no additional controls. This suggests that more leveraged households are more likely to purchase a second home following increases in local house prices than less leveraged households. Column (2) reports results for a model where local house price increases are replaced by region dummies. The estimates are very similar. In column (3) we include other controls for year, 10-year birth cohort, a quadratic in age and the years the household head has been living at the current address. The latter control accounts for the fact that households who have moved recently will likely be closer to their desired leverage and so less likely to need to rebalance their portfolios. Once these other factors have been accounted for, leveraged households do not appear any more likely to make second home purchases than other households. However, this may reflect the fact that more leveraged households are less wealthy than other households (as implied by the results in Table 6.4), rather than the effects of leverage itself. To account for this possibility, as well as to address problems of potential measurement error in the leverage variable, we instrument leverage with the price to income ratio at the time households moved into their current residence as we did in Section 6.4. Once we do this we find that the second home purchases of more leveraged households are indeed more responsive to house price increases than the purchases of other households.

In Table 6.8 whether more leveraged households are more likely to up-size (as measured by changes in the number of rooms in their primary residence). The pattern of results is similar to that shown in Table 6.7. However, in this case the results we obtain when we instrument are not statistically significant.

**Table 6.8:** Effects of leverage on home size

| $\Delta nrooms_{t,t+10}$ | OLS<br>(1) | OLS<br>(2) | OLS<br>(3) | IV<br>(4) |
|---|---|---|---|---|
| $\left(\frac{p_{rt+10}}{p_{rt}} - 1\right)$ | -0.293** | - | - | - |
| | (0.144) | | | |
| $\omega_{it-1} \times \left(\frac{p_{rt+10}}{p_{rt}} - 1\right)$ | 0.050*** | 0.049*** | 0.009 | 0.073 |
| | (0.007) | (0.007) | (0.007) | (0.068) |
| Region dummies | N | Y | Y | Y |
| Other controls | N | N | Y | Y |
| N | 4,834 | 4,834 | 4,625 | 4,625 |
| Clusters | 1,513 | 1,513 | 1,440 | 1,440 |

Notes: * $p < 0.10$ , ** $p < 0.05$ , *** $p < 0.01$. Standard errors in parentheses. Controls are year dummies, dummies for 10-year birth cohorts, age, age squared, years at current address and a dummy for having just moved in. Standard errors are clustered at the individual level.

## 6.6 Conclusion

In this chapter, we have shown how households re-leveraging and investments in housing can be driven by a desire to maintain the risk and return of their portfolios. We set out an empirical approach which accounts for the endogeneity of households' leverage, and that combines information from two samples in way which allows us to see how house price changes affect different forms of household spending (as opposed to looking at total spending or borrowing).

We show using panel data that homeowners re-leveraged as house prices rose and that the self-reported use of new mortgage loans was for home improvements and other residential investments rather than consumption. In addition we find that households that were initially more leverged are more likely to both purchase other residential properties and invest in their own homes in response to local house price increases than other households. However, these households do not disproportionately increase their consumption spending as house prices rise, as would be expected if re-leveraging behaviour were driven by traditional housing wealth effects.

Our findings have relevance for the designers of macro-prudential policy interventions. Previous studies have examined the total spending or borrowing response of households to house price shocks, and concluded that access to credit drives differences in responses across more and less leveraged households. Our results suggest that leverage has an important influence on the *composition* of household spending response to house price shocks even when households are not at their credit constraints. Our findings point to potentially important feedback mechanisms. As house prices rise, households desire to re-leverage may lead to greater demand for housing and further price increases - increasing households' exposure to future house price changes and amplifying housing booms.

# Chapter 7

# Aggregating Elasticities: Intensive and Extensive Margins of Women's Labour Supply

The size of the elasticity of labour supply to changes in wages has been studied for a long time. Recent debates have focused on the perceived discrepancy between estimates coming from micro studies, which with a few exceptions, point to relatively small values of such an elasticity, and the assumptions made in macro models, which seem to need relatively large values. Keane and Rogerson (2015) and Keane and Rogerson (2012) survey some of these issues and the papers by Blundell et al. (2011), Ljungqvist and Sargent (2011) and Rogerson and Wallenius (2009) contain some alternative views on the debate. To reconcile the micro evidence and the assumptions made in macroeconomics, much attention has been given to the distinction between the extensive and intensive margins of labour supply, see, in particular, Chetty et al. (2011). Perhaps surprisingly, in this debate, aggregation issues and the pervasive and complex heterogeneity that characterise labour supply behaviour have not been given much attention.[1] This chapter aims to fill this gap, while making some original methodological contributions and presenting new empirical evidence.

[1] One exception is Keane and Wasi (2016) who show men's labour supply responses vary substantially with age, education and the tax structure. Aggregation issues are also discussed in Erosa et al. (2016).

Preferences for consumption and leisure are likely to be affected in fundamental ways by family composition, fertility and wealth, as well as by unobserved taste 'shocks', and so heterogeneity in labour supply elasticities in these dimensions is something to be expected. Labour supply elasticities will vary in the cross-section and over the business cycle. The key issue, however, is how significant this heterogeneity is and whether it is important at the aggregate level: does it make any sense to talk about *the* elasticity of labour supply as a *structural* parameter? Aggregation issues are likely to be relevant both for the intensive and extensive margin, as we show.

In this chapter, we address these issues focusing on women's labour supply. Our approach consists in taking a relatively standard life-cycle model of labour supply to the data. Whilst the essence of the model is relatively simple, we stress two elements that are important for our analysis and that make our contribution novel. First, we consider all the relevant intertemporal and intratemporal margins and choices simultaneously; in particular, consumption and saving as well as participation and hours of work. This allows for interaction between different decisions. Second, we specify a flexible utility function that allows for substantial heterogeneity, fits the data well and, at the same time, allows us to make precise quantitative statements. These elements are important because they allow us to address directly the interaction between extensive and intensive margins and to evaluate empirically the importance of aggregation issues and to calculate both micro and macro elasticities.

In evaluating aggregate labour supply elasticities, it is necessary to specify the whole economic environment because, as noted by Chang and Kim (2006), the aggregate response depends on the distribution of reservation wages. On the other hand, an important methodological contribution of this chapter is to stress that key components of the model can be estimated using weaker assumptions which closely approximate the overall model structure. We separate our estimation into three steps and specify what assumptions are needed at each step and what variation in the data is used for identification. The first step identifies the within-period preferences over consumption and labour supply at the intensive margin. We use group level variability driven by group or aggregate shocks such as policy reforms, similar to Blundell et al. (1998). These estimates are used to compute within-period Marshallian and Hicksian elasticities, which hold intertemporal allocations constant and are conditional on participation. The second step estimates intertemporal preferences that generate Frisch labour supply elasticities. We estimate these parameters by using the Euler equation for consumption, using synthetic cohorts, similar to Blundell et al. (1993) and Attanasio and Weber (1995), and without taking a stance on the determinants of participation and a variety of other issues, such as retirement or the cost of children. Finally, to characterise behaviour at the extensive margin, we specify the model fully. In this step, we calibrate key parameters to a number of life-cycle moments, and explicitly aggregate individual behaviour, similar in spirit to Erosa et al. (2016). Labour supply responses to wages in a life-cycle model may change beyond the static response if savings decisions are affected by wages. Our life-cycle elasticities account for these effects and we discuss the circumstances in which static elasticities provide a good approximation to the overall life-cycle response.

We use a flexible specification for utility to allow for observed and unobserved heterogeneity in tastes at both intratemporal and intertemporal margins, and at the same time allowing for possible non-separability of consumption and leisure. Our specification of preferences is much more flexible than generally allowed for in the literature and we show this is important. Classic papers in the micro literature (such as Heckman and Macurdy (1980)) imply a strong relationship between the Frisch intertemporal elasticity and the intratemporal Marginal Rate of Substitution conditions, which, in turn, forces a strict relationship between within-period and intertemporal conditions. Our approach avoids this restriction. In the macro literature, most papers impose additive separability between consumption and leisure, and isoleastic, homothetic preferences that conform to the restrictions for balanced growth, as in Erosa et al. (2016).[2] Here, we show that the isoelastic specification for

[2]This assumption is predicated on the perceived need to work with models that match historical trends showing steady secular increases in real wages with little change in aggregate hours. Brown-

consumption and hours is strongly rejected by the data. The challenge, therefore, is to work with specifications that allow much more heterogeneity and changes over time.

Estimates of the size of the elasticity of labour supply for women vary considerably. Our estimates, at the median, are not too different from some estimates in the literature. In particular, on the intensive margin, we obtain a median static Marshallian elasticity of 0.18, with the corresponding Hicksian elasticity considerably larger at 0.54, indicating a sizeable income effect. For the same median household, the Frisch elasticity for hours is 0.87. At the same time, we document considerable variation in estimated elasticities in the cross-section: the Marshallian, for instance, has an inter-decile range of -0.14 to 0.79. As we show, these static Marshallian elasticities are smaller than the responses when we allow savings to adjust.

In comparing our estimates to the literature, we investigated what drives, in our data, differences in results. A key factor is that the size of the estimates depends on the specific estimator and normalisation used. When using standard IV or GMM methods, we typically obtain very large estimates when we put wages on the left-hand side of the MRS equation. Instead, we get much smaller estimates when put consumption or hours worked on the left-hand side. In our baseline estimation, we use methods robust to the normalization, using a method proposed by Fuller (1977), which is a generalization of a LIML approach.

We use the fully specified model to run two experiments: in the first, we evaluate the labour supply response to temporary changes in wages; in the second, we evaluate the response to a change in the entire life-cycle wage profile. The first experiment captures the impact of a temporary tax cut, which has little effect on the marginal utility of wealth, even if the cut is unanticipated. Without an extensive margin, the response would be captured by the Frisch elasticity. Introducing the extensive margin doubles the size of the response, and is particularly important at younger ages when non-participation because of children is prevalent. The second experiment captures the impact of a permanent tax cut which will change

ing et al. (1999) already noted that the fact that the historical trend for aggregate hours is roughly constant hides a large decrease for men and an increase for women.

the marginal utility of wealth. The response to the second experiment would be approximated by the static Marshallian elasticity if there was no change in savings behaviour. Allowing intertemporal allocations to adjust gives what we call *life-cycle Marshallian and Hicksian elasticities*. These life-cycle elasticities are greater than the static approximations because not all income is spent on non-durable consumption in the period it is earned. However, these life-cycle elasticities are lower than the Frisch responses to temporary changes.

Using the entire model, we can aggregate explicitly individual behaviour and study aggregate elasticities that correspond to the concept used in the macro literature. We find an important role for the extensive margin in generating aggregate movements in labour supply. Importantly in linking the micro and macro analysis of labour supply, we show that what we call the 'aggregate' elasticity changes considerably over the business cycle, and is typically larger in recessions. Moreover, it gets larger in longer recessions. To the best of our knowledge, changes in the elasticity over business cycles have never been discussed.

The closest macroeconomic paper to this chapter is Erosa et al. (2016), who have similar aims of building aggregate elasticities from men's labour supply behaviour over the life-cycle, and of distinguishing the intensive and extensive margins using a fully specified life-cycle model. The focus of this chapter is on women's labour supply responses. A second related paper is Guner et al. (2012), who model heterogeneous married and single households with an extensive margin for women and an intensive margin for both men and women. Their focus is on evaluating different reforms of the US tax system and they abstract from wage uncertainty. Both papers operate with very specific preference specifications. We discuss the extent to which our results differ from these papers in the conclusions. Among papers using microeconometrics, this chapter builds on a long literature starting from MaCurdy (1983) and Altonji (1986), and on Blundell et al. (1993), who condition on the extensive margin, and estimate jointly the within period decision and the intertemporal decision.

Our exercise is not without important caveats. In much of our analysis, we do not consider the effect of tenure and experience on wages. Such effects can obviously be important, as labour supply choices may change future wages and, therefore, future labour supply behaviour, as stressed by Imai and Keane (2004). Keane and Wasi (2016) model human capital and find that labour supply elasticities are highly heterogenous and vary substantially with age, education and the tax structure. In Appendix E.5, we extend our analysis to introduce returns to experience on the extensive margin. Introducing returns only on the extensive margin means within-period allocations at the intensive margin are unaffected. By contrast, if the return to experience operates on the number of hours (rather than only on participation), we would need to change our analysis substantially.

The rest of the chapter is organized as follows. In Section 7.1, we outline the life-cycle framework. We show how the preference parameters can be mapped into static, intertemporal and life-cycle elasticities, and discuss the meaning of the different elasticities. In Section 7.2 we explain the three steps of our empirical strategy to identify the preference parameters and opportunity set. Section 7.3 describes the data. Section 7.4 presents the parameter estimates. Section 7.5 contains the key results of the chapter: the implications of our estimates for labour supply elasticities, distinguishing between Marshallian, Hicksian and Frisch elasticities, and distinguishing static from life-cycle responses. We also report responses on the extensive margin, aggregate responses and, more generally, the aggregation issues that are central to this chapter. Section 7.6 concludes. The Appendix to this chapter provides supporting evidence.

## 7.1 A life-cycle model of women's labour supply

To study both the intensive and the extensive margins of women's labour supply, we use a rich model of labour supply and saving choices embedded in a unitary household, life-cycle framework. Both the intensive and extensive margins are meaningful because of fixed costs of going to work related to family composition and because of preference costs specifically related to participation. The intensive choice is over the typical number of hours work per week, the extensive margin is over whether to work at all in each quarter. Changes at different margins interact and heterogeneity in these responses is important to understand aggregate labour supply responses to changes in wages.

We consider married couples, who maximise the lifetime expected utility of the household, $h$, and choose consumption and women's labour supply within each period.

$$\max_{c,l} E_t \sum_{j=0}^{T} \beta^j u\left(c_{h,t+j}, l_{h,t+j}, P_{h,t+j}; z_{h,t+j}, \chi_{h,t+j}, \zeta_{h,t+j}\right) \tag{7.1}$$

where $c$ is consumption, $l$ is hours of leisure for women, and $P$ is an indicator of the woman's labour force participation which can affect utility over and above the effect of hours worked. $z_{h,t}$ is a vector of demographic variables (which include education, age and family composition), $\chi_{h,t}$ and $\zeta_{h,t}$ represent taste shifters. We assume that demographics, $z_{h,t}$, are observable, whereas $\chi_{h,t}$ and $\zeta_{h,t}$ are unobservable to us, but are known to the individual. Leisure for men does not enter the utility function.

The period utility function is given by:

$$u\left(c_{h,t}, l_{h,t}, P_{h,t}\right) = \frac{M_{h,t}^{1-\gamma}}{1-\gamma} \exp(\varphi P_{h,t} + \pi z_{h,t} + \zeta_{h,t}) \tag{7.2}$$

The preference aggregator for hours of lesuire and consumption, $M_{h,t}$ is:

$$M_{h,t}\left(c_{h,t}, l_{h,t}; z_{h,t}, \chi_{h,t}\right) = \left(\frac{(c_{h,t}^{1-\phi} - 1)}{1-\phi} + \left(\alpha_{h,t}(z_{h,t}, \chi_{h,t})\right) \frac{(l_{h,t}^{1-\theta} - 1)}{1-\theta}\right) \tag{7.3}$$

The function $\alpha_{h,t}$ that determines the weight on leisure as a function of demographics is specified as:

$$\alpha_{h,t} = \exp(\psi_0 + \psi_z z_{h,t} + \chi_{h,t}) \tag{7.4}$$

The unknown parameters governing within period utility over consumption and leisure are $\phi$, $\theta$, $\psi_0$ and $\psi_z$, with additional parameters governing the full utility specification $\gamma$, $\varphi$ and $\pi$. Our specification allows for non-separability between consumption and leisure both at the intensive and extensive margin. The taste shifter

$\chi_{h,t}$ affects within period utility over consumption and leisure, and the taste shifter $\zeta_{h,t}$ affects intertemporal choices. These are specific to the cohort-education group, known to the individual and may be correlated. Non-separability between consumption and leisure depends on the value of $\gamma$ and so cannot be identified from within-period choices alone.

The general specification of utility allows substantial heterogeneity across individuals in intratemporal and intertemporal preferences, across the intensive and extensive margins, and does not impose that the elasticities of intertemporal substitution for leisure and consumption are constant. Heterogeneity arises partly because elasticities will differ by observable characteristics, $z$, such as education and the presence of children, and partly because elasticities differ at different levels of consumption and hours of work. Our parametric specification gives a log linear Marginal Rate of Substitution (MRS) and guarantees integrability. Further, our approach is more flexible than alternatives which have less scope for heterogeneity at the intensive margin, and so heterogeneity has to come through the extensive margin and the distribution of reservation wages.

Maximisation is subject to the intertemporal budget constraint:

$$A_{h,t+1} = (1+r_{t+1})\left(A_{h,t} + \left(w^f_{h,t}\left(L-l_{h,t}\right) - F\left(a_{h,t}\right)\right)P_{h,t} + y^m_{h,t} - c_{h,t}\right) \tag{7.5}$$

where $A_{h,t}$ is the beginning of period asset holding, $r_t$ is the risk-free interest rate, $F$ the fixed cost of work, dependent on the age of the youngest child $a_{h,t}$, and $L$ is maximum hours available. Wages for the woman are given by $w^f_{h,t}$, and earnings for the man by $y^m_{h,t}$.

There are no explicit borrowing constraints but households cannot go bankrupt. Therefore, in each period, households are able to borrow against the minimum income they can guarantee for the rest of their lives. This minimum income is a positive amount because we bound men's income away from zero. Households have no insurance markets to smooth aggregate or idiosyncratic shocks.

We assume that the cost of work has a fixed component and a component that depends on the child care cost needed for the youngest child, whose age is $a_{h,t}$.

Denoting with $G(a_{h,t})$ child care services and $p$ their price, we have:

$$F(a_{h,t}) = pG(a_{h,t}) + \bar{F} \tag{7.6}$$

Women differ in their age at childbirth, but this is assumed to be deterministic and fully anticipated.[3] The fixed cost of work is deterministic and known. The presence of fixed costs and discrete utility costs of participating mean some women decide not to work at all, especially at low levels of productivity. If a woman does not work, she does so by choice, given the offered wage, demographics, taste shifters and unearned income. By the same token, it is unlikely that if a woman does work, that she will work only very few hours.

Women's wages are given by the following process:

$$\ln w^f_{h,t} = \ln w^f_{h,0} + \ln e^f_{h,t} + v^f_{h,t} \tag{7.7}$$

where $e^f_{h,t}$ is the level of human capital at the start of the period. We assume that wage rates do not depend on the number of hours worked in that period, ruling out part-time penalties. This assumption, for women, is consistent with what we observe in our data and with other US-based studies (Hirsch (2005); Aaronson and French (2004)).

In our baseline specification, human capital does not depend on the history of labour supply and is assumed to evolve exogenously according to:

$$\ln e^f_t = \iota^f_1 t + \iota^f_2 t^2 \tag{7.8}$$

Equation (7.8) implies that decisions on current labour supply do not have a direct effect on continuation values.[4] Therefore, the only linkage across periods is through the decision about total within-period spending. This assumption, com-

[3] In reality, there is of course some degree of uncertainty in the realisation of households fertility decisions. We do not consider fertility as a stochastic outcome, as that would increase the numerical complexity of the problem substantively.

[4] In Appendix E.5, we relax the assumption that there are no returns to experience. We distinguish the cases where returns to experience depend on participation and where returns depend on hours worked. The first two steps of our estimation approach go through in former case but not in the latter.

bined with the intertemporally additive structure of preferences, implies that standard two-stage budgeting holds so that we can focus on the within-period problem without considering explicitly the intertemporal allocation.

Men always work and their earnings are given by:

$$\ln y^m_{h,t} \quad = \quad \ln y^m_{h,0} + \iota^m_1 t + \iota^m_2 t^2 + v^m_{h,t} \tag{7.9}$$

There are initial distributions of wages for women, $w^f_{h,0}$, and earnings for men $y^m_{h,0}$. Both women's wages and men's earnings are subject to permanent shocks that are positively correlated, as in MaCurdy (1983) and Abowd and Card (1989):

$$\begin{aligned} v_{h,t} &= v_{h,t-1} + \xi_{h,t} \qquad (7.10)\\ \xi_{h,t} &= (\xi^f_{h,t}, \xi^m_{h,t}) \sim N\left(\mu_\xi, \sigma^2_\xi\right) \qquad (7.11)\\ \mu_\xi &= (-\frac{\sigma^2_{\xi^f}}{2}, -\frac{\sigma^2_{\xi^m}}{2}) \text{ and } \sigma^2_\xi = \left(\begin{array}{cc} \sigma^2_{\xi^f} & \rho_{\xi^f,\xi^m} \\ \rho_{\xi^f,\xi^m} & \sigma^2_{\xi^m} \end{array}\right) \end{aligned}$$

One period in the model is one quarter. Households choose typical hours of work each week (the *intensive margin*) and this is kept constant across weeks within the quarter, to give within-period hours of work. The *extensive margin* is the decision whether or not to work that quarter. We do not allow individuals to choose how many weeks to work in a quarter.[5] We provide empirical support for this approach in Section 7.3.2.

Within the dynamic problem just described, households make decisions taking the stochastic processes as given. When considering aggregation, we need to take a stand on the degree of correlations in the shocks different households receive. We assume that households are subject to both idiosyncratic and aggregate shocks, by allowing the shocks that affect individual households at a point in time to be correlated. However, from an individual perspective, households do not distinguish

[5]This restriction is driven by data limitations. In our data, we observe typical hours per week, whether an individual is working at that point in time, and the number of weeks per year but we do not observe the number of weeks per quarter that an individual works. We also cannot distinguish the number of days per week, from the number of hours per day, as in Castex and Dechter (2016).

aggregate and idiosyncratic shocks and condition their future expectations only on their own observed wage realisations. Our framework is not a general equilibrium one: we do not construct the equilibrium level of wages (and interest rates). Rather, we study women's aggregate labour supply and its elasticity to wages by simulating a large number of households and aggregating explicitly their behaviour.

### 7.1.1 Marginal Rate of Substitution, Marshallian and Hicksian Elasticities

We use a two-stage budgeting approach and consider the allocation of resources between consumption and hours of leisure within each period. We define within-period resources that are not earned by women as:

$$y_{h,t} = \left(A_{h,t} + y^{m}_{h,t} - F\left(a_{h,t}\right) P_{h,t}\right) - \frac{A_{h,t+1}}{1+r_{t+1}} \tag{7.12}$$

As in Blundell and MaCurdy (1999), $y_{h,t}$ accounts for resources saved into the next period. When taken to the data, this measure of unearned resources implicitly also includes (with a negative sign) durable and other spending not included in consumption $c_t$, giving the within period budget constraint:

$$c_{h,t} + w_{h,t} l_{h,t} = y_{h,t} + w^{f}_{h,t} L \tag{7.13}$$

For an interior solution with a strictly positive number of hours of work, the first order condition for within-period optimality implies that the ratio of the marginal utility of leisure to that of consumption, that is the Marginal Rate of Substitution, equals the after tax real wage:

$$w_{h,t} = \frac{u_{l_{h,t}}}{u_{c_{h,t}}} = \alpha_{h,t} \frac{l^{-\theta}_{h,t}}{c^{-\phi}_{h,t}} \tag{7.14}$$

These equations can be used to compute Marshallian and Hicksian labour supply elasticities. The Marshallian and Hicksian elasticities are fundamentally static concepts, as both hold constant the intertemporal allocation of resources.[6] The Marshallian response captures the change in behaviour due to a change in the price of leisure and the related change in resources available to spend. This latter income effect arises even if the intertemporal allocation of resources $y_{h,t}$ is held constant, because total resources within the period change with the wage.

In the full dynamic model, when the realised wage is permanently higher than expected, lifetime resources increase, and these extra resources are allocated across periods. The static Marshallian elasticity is a good approximation to the full response if extra resources are spent on non-durable consumption in the period they are earned. To the extent that resources are reallocated, the static Marshallian elasticity only captures part of the labour supply response. If within period spending is homothetic, and wages have gone up by the same amount in every period, then there may be little change in saving patterns following the wage increase. In this case, the Marshallian elasticity gives a good approximation of the complete life-cycle response. On the other hand, if all extra income from the wage increase is saved to spend in retirement, then there would be no within period income effect and the response will be closer to a Hicksian compensated response. More generally, how well the static Hicksian and Marshallian elasticities approximate the complete life-cycle responses to compensated and uncompensated wage changes is an open question. In Section 7.5, we use the full structural model to evaluate how closely the static elasticities approximate the full life-cycle ones.

We differentiate the within period budget constraint (7.13) and the MRS equation (7.14) with respect to wages to get an expression for Marshallian elasticities for hours of work and consumption (see Appendix E for details on the derivations):

[6] Blundell and MaCurdy (1999) and Keane (2011) discuss how the static concepts of Marshallian and Hicksian elasticities can be put within the framework of a dynamic life-cycle model through two-stage budgeting, as developed by Gorman (1959) and applied to labour supply by MaCurdy (1981), MaCurdy (1983) and Blundell and Walker (1986).

$$\varepsilon_h^M = \frac{\partial \ln h}{\partial \ln w} = -\left(\frac{\phi w (L-l) - c}{\theta c + \phi wl}\right) \frac{l}{L-l} \tag{7.15}$$

$$\varepsilon_c^M = \frac{\partial \ln c}{\partial \ln w} = \frac{\theta w (L-l) + wl}{\theta c + \phi wl}$$

If preferences were Cobb-Douglas, $\theta$ and $\phi$ would both equal 1; and the Marshallian wage elasticities for consumption and for hours of work would be equal to 1 and 0, respectively, if there were no unearned income or savings. For balanced growth (in women's labour supply) we would require $\phi = 1$. If preferences were a standard CES, $\theta = \phi$. If this value were greater than 1, $\varepsilon_c^M < 1$, and $\varepsilon_h^M < 0$. In Section 7.5, we show how much heterogeneity is introduced through our more general specification in equations (7.15) and through allowing for unearned income.

The static Hicksian response nets off the increase in within-period resources due to the wage increase, again holding constant the intertermporal allocation, $y_{h,t}$. We calculate the Hicksian response from the Marshallian elasticities by using the Slustky equation and income elasticities, as would be done in a static labour supply model:

$$\begin{aligned} \varepsilon_h^H &= \left(\varepsilon_l^M - \frac{\partial \ln l}{\partial \ln(c+wl)} \frac{w(L-l)}{(c+wl)}\right) \frac{-l}{L-l} &= \frac{-wl^2}{(\theta c + \phi wl)(L-l)} \\ \varepsilon_c^H &= \varepsilon_c^M + \frac{\partial \ln c}{\partial \ln(c+wl)} \frac{wl}{(c+wl)} &= \frac{-c}{\theta c + \phi wl} \end{aligned} \tag{7.16}$$

The Marshallian and Hicksian elasticities are the relevant concepts to think about the labour supply responses to permanent changes in wages or taxes. However, as we discuss in Section 7.5, estimates based on the within period problem might miss potential intertemporal reallocations that might occur in response to wage changes.

Two additional points are worth noting. First, despite their simplicity, the Marshallian and Hicksian elasticities are non-linear in $c$ and $l$: they have the potential of varying greatly across consumers and not aggregating in a straightforward way. Second, for the specification we use, the Marshallian and Hicksian elasticities depend only on $\phi$ and $\theta$ (and on the values of earnings, leisure and consumption). In particular, they do not depend on intertemporal parameters or on whether the utility function is separable in consumption and leisure, which depends on $\gamma$.

### 7.1.2 Frisch Elasticities

A change in the structure of wages (possibly induced by changes in taxes) may induce a reallocation of resources over time through changes to the time path of hours of work or of the marginal utility of wealth, or both. The Frisch elasticity captures the change over time in hours worked in response to the anticipated evolution of wages, with the marginal utility of wealth unchanged, as the wage change conveys no new information.[7] The Frisch elasticity is therefore the right concept to think about the implications of changes in wages over the business cycle or about temporary changes to taxation.

The expression for the Frisch elasticity for hours of work, derived in Appendix E, is given by:[8]

$$\varepsilon_h^F = -\frac{u_c u_{cc}}{u_{cc} u_{ll} - u_{cl}^2} \frac{w}{h} \tag{7.17}$$

As is well known, Frisch intertemporal elasticities must be at least as large as Hicks elasticities. Thus, the static elasticities discussed above provide a bound on the intertemporal elasticity, which is particularly useful if data are limited or direct estimation of Frisch elasticities difficult.[9]

In addition to changes in hours, anticipated changes in wages might also change participation. While, an elasticity is easily defined when thinking of the intensive margin, the same concept is somewhat vaguer at the extensive margin,

[7] When wages change stochastically, the response of hours worked is affected by the change in the marginal utility of wealth due to a particular wage realisation, whose size depends on how permanent the wage shock is. If the wage shock is temporary, lifetime wealth and the marginal utility of wealth will be approximately unchanged.

[8] Analogous expressions for the consumption Frisch wage elasticities, as well as the interest rate elasticities can be found in Appendix E.

[9] In the context of quasi-linear utility as used by Chetty (2012), the Frisch elasticity equals the Hicks elasticity (and the Marshallian) because there are no wealth effects on hours of work.

especially in the case of the Frisch elasticity, which keeps the marginal utility of wealth constant. We define the extensive-margin Frisch elasticity as the impact of a change in wages on the fraction of individuals that participate, given the distribution of state variables. The extensive margin brings to the forefront aggregation issues: aggregate participation responses to an aggregate shock are bound to depend on the distribution of state variables in the cross-section.

## 7.2 Empirical strategy

In this section, we discuss our empirical approach, identification assumptions, and the variability we use in the data. We proceed in three steps, with each successive step identifying a set of structural parameters. In the first step, we consider only the static first-order (MRS) condition that determines within-period optimal allocations, conditional on participation. This first set of parameters can be identified while being agnostic about intertemporal conditions and on life-cycle prospects. In the second step, we identify the parameters that govern the intertemporal allocation of resources using the Euler equation for consumption, making use of additional assumptions. However, we can still identify these parameters without specifying the entire life-cycle environment faced by households. For instance, we can be silent about pension arrangements or the specifics of the wage and earning processes. When estimating the parameters that determine the MRS or those that enter the Euler equation we use an estimator proposed by Fuller (1977). This choice of estimator turns out to matter for the results we obtain and has advantages over standard methods, as we discuss in Appendix E.1. Finally, in the third step, we characterise behaviour at the extensive margin. This step requires solving the entire model and, therefore, specifying completely the environment in which households operate. We identify the final set of parameters by calibration, matching a set of life-cycle statistics.

### 7.2.1 Intratemporal margins

In the first step, we estimate the parameters of the within-period utility function: $\theta, \phi$ and $\alpha$. Taking logs of the MRS equation (7.14), and noticing from equation

(7.4) that $\log \alpha_{h,t} = \psi_0 + \psi_z z_{h,t} + \chi_{h,t}$, we obtain:

$$\ln w_{h,t} = \phi \ln c_{h,t} - \theta \ln l_{h,t} + \psi_z z_{h,t} + \psi_0 + \chi_{h,t} \tag{7.18}$$

where the vector $z_{h,t}$ includes observable demographic variables.

The econometric estimation of this MRS equation poses two problems. First, the subset of households in which the woman works and the MRS condition holds as an equality is not random. For this selected group, the unobserved heterogeneity term $\chi_{h,t}$ would not average out to zero and might be correlated with the variables that enter equation (7.18). Second, even in the absence of participation issues, individual wages (and consumption and leisure) are likely to correlate with $\chi_{h,t}$, so that the OLS estimation of equation (7.18) would result in biased estimates of the structural parameters $\phi$ and $\theta$. We discuss these two issues in turn.

For participation, we specify a reduced form equation for the extensive margin. Given this participation equation, we use a Heckman-type selection correction approach to estimate the MRS equation (7.18) only on the households in which the woman works. In particular, we augment the MRS equation with a polynomial in the estimated residuals of the participation equation.[10] Identification requires that some variables that enter the participation equation do not enter the MRS specification: these variables are men's earnings and employment status, and we assume these are uncorrelated with $\chi_{h,t}$.

The fully-specified participation decision depends on a large set of state variables, some of which are not observable. In our 'reduced form', participation depends only on a subset of these variables. Therefore, our reduced form participation equation is not fully consistent with the complete model we use to characterise participation and, at best, could be considered an approximation of the 'true' participation equation. In Appendix E.6, we investigate how well this approximation

[10] One issue to worry about is the intrinsic non-linearity of the participation equation. The omission of some state variables could change the properties of the residuals of such a non-linear equation and, therefore, the shape of the appropriate control function to enter the MRS equation. For this reason, we use a polynomial to model the dependence between the residuals of the participation equation and those of the MRS equation. We assume that $\chi_{h,t} = \beta_0 + \beta_1 e_{h,t} + \beta_2 e_{h,t}^2 + \beta_3 e_{h,t}^3$ and then compute $E[e_{h,t}^s | e_{h,t} > -\Pi Z_{h,t}], s = 1,2,3$ where $e_{h,t}$ is the normally distributed residual from the participation equation and $Z_{h,t}$ are the determinants of participation.

to the full model performs: we estimate MRS parameters using our reduced form empirical strategy on simulated data from the full model. We are able to recover the true parameter estimates and our conclusion is that our reduced form provides an accurate approximation in this context.

The second issue in the estimation of equation (7.18) is that consumption and hours, as well as our measures of individual wages, obtained dividing earnings by hours, might be correlated with the residual term $\chi_{h,t}$, either because of the possible correlation between tastes for leisure and heterogeneity in productivity or because of measurement error in hours or earnings. To avoid these problems, following the literature on labour supply (such as Blundell et al. (1998)), we do not use variation in individual wages to identify the parameters of our equation. Instead, we exploit variation induced by changes in taxation and/or aggregate demand for labour and use changes in cohort-education groups' average wages over time.[11] The Monte Carlo evidence on our MRS estimation in Appendix E.6 shows that both this endogeneity issue and the selection issue have to be taken into account in our context to obtain sensible estimates.

We use as instruments the interaction of ten-year birth cohort and education dummies with a quintic time trend. Our use of a quintic time trend rather than fully interacted time dummies helps smooth intertemporal movements in wages, consumption and hours for each of our cohort-education groups.[12]

In our estimating equation, we allow many variables to shift the taste for leisure through an effect on the term $\alpha_{h,t}$ in the CES utility function. The $z$ vector includes:

[11] Various papers have used variation across education groups; for example MaCurdy (1983) and Ziliak and Kniesner (1999) both use age-education interactions as instruments for wages and hours in their MRS/labour supply conditions. Similarly, Kimmel and Kniesner (1998) use education interacted with a quadratic time trend. One concern with this approach is that individuals with different levels of education might have different preferences for leisure and consumption. Moreover, the composition of education groups has changed substantially over time, particularly for women. In 1980, 19.4% of married women had not attained a high school diploma, and only 18.4% had obtained a college degree in our data. By 2012, these proportions had changed to 9.7% and 36.5% respectively. These compositional changes may lead to changes in the mix of ability and preferences of workers within each education group over time - making education an invalid instrument.

[12] Using fully interacted cohort-education and year dummies would be equivalent to taking averages within cells defined by year, education and cohort groups, to use group level rather than individual level variability. Given our sample size, this would result in averages over relatively small cells and, therefore, in very noisy estimates. Using very finely defined and small groups can introduce the very biases grouping is meant to avoid.

log family size, woman's race, a quartic in woman age, an indicator for the presence of any child, the numbers of children aged 0-2, 3-15, and 16-17, the number of individuals in the household 65 or older, region and season dummies, and, most importantly, cohort-education dummies. A corollary of putting variables such as cohort and education dummies in the vector $z$ is that we do not exploit the variation in wages (and leisure and consumption) over these dimensions to identify the structural parameters $\phi$ and $\theta$. In our estimation, we also control for year dummies, therefore removing year to year fluctuations from the variability we use to identify the parameters of interest. The inclusion of year dummies, as in Blundell et al. (1998), is needed because aggregate fluctuations change the selection rule year to year in ways that are not fully captured by the selection model we use.[13]

### 7.2.2 Euler equation estimation

The second step of our approach estimates the preference parameters that govern the intertemporal substitutability and non-separability between consumption and leisure, $\gamma$, and the non-separability with participation, $\varphi$. While in principle we could use either the Euler equation for hours or that for consumption, only one is relevant when coupled with the intratemporal condition (7.14). If we were to use the Euler equation for labour supply, we would need to consider corner solutions at different points in time (and the dynamic selection problems these involve). Instead, we focus on the Euler equation for consumption, as in Blundell et al. (1993). In the absence of binding borrowing constraints, the following intertemporal condition holds:

$$E\left[\beta\left(1+r_{t+1}\right)\frac{u_{c_{h,t+1}}(\cdot)}{u_{c_{h,t}}(\cdot)}\middle| I_{h,t}\right]=1 \tag{7.19}$$

The term $I_{h,t}$ denotes the information available to the household at time $t$.

A natural approach to the estimation of equation (7.19) is non-linear GMM. However, as discussed in Attanasio and Low (2004), the small sample properties of non-linear GMM estimators can be poor in contexts similar to ours. Moreover, given the specification of the utility function and nature of the data, we can only

[13] We have also run specifications where we do not control for time dummies in the MRS and checked that our results are not affected much by the introduction of the time dummies.

estimate its log-linearised version.

The evolution of the marginal utility of consumption can then be written as:

$$\beta\left(1+r_{t+1}\right)u_{c_{h,t+1}}\left(\cdot\right)=u_{c_{h,t}}\left(\cdot\right)\varepsilon_{h,t+1} \tag{7.20}$$

where $\varepsilon_{h,t+1}$, whose conditional expectation is 1, is the innovation to the expected discounted marginal utility of consumption. Equation (7.20) uses the variability in $r_t$ to identify the parameters of $u_c(.)$. Taking the log of equation (7.20), given utility is given by equation (7.2):

$$\eta_{h,t+1}=\kappa_{h,t}+\ln\beta+\ln(1+r_{t+1})-\phi\Delta\ln c_{h,t+1}-\gamma\Delta\ln(M_{h,t+1})+\varphi\Delta P_{h,t+1}+\pi\Delta z_{h,t+1} \tag{7.21}$$

where $\eta_{h,t+1}\equiv\ln\varepsilon_{h,t+1}-E\left[\ln\varepsilon_{h,t+1}\left|I_{h,t}\right.\right]+\Delta\zeta_{h,t+1}$ and $\kappa_{h,t}\equiv E\left[\ln\varepsilon_{h,t+1}\left|I_{h,t}\right.\right]$.

The identification and estimation of the parameters of this equation depends, obviously, on the nature of the 'residual' term $\eta_{h,t+1}$, which contains expectations errors ($\varepsilon_{h,t+1}$), higher order moments and taste shifters unobservable to the econometrician ($\zeta_{h,t+1}$). Aggregate shocks mean expectation errors may be correlated in the cross-section, and average to zero only in the time dimension. Consistency then requires time series variation, as discussed in Attanasio and Low (2004). We construct a long time dimension using a synthetic cohort approach (see Browning et al. (1985)), defining groups using married couples in ten year birth-cohorts. We assume that the lagged variables used as instruments are uncorrelated with the innovations to the taste shifters $\Delta\zeta_{h,t+1}$. This is trivially true if taste shifters are constant over time or if they are random walks. We maintain one of these two assumptions, a hypothesis that we test in part by checking over-identifying restrictions.

We aggregate equation (7.21) to be estimated across group $g$ households. For this approach to work, it is necessary that the equation to be estimated is linear in parameters, which would be the case if $M_{h,t}$ were observable. However, $M_{h,t}$ is a non-linear function of data and unobserved parameters, so that, in principle it cannot be aggregated within groups to obtain $M_{g,t}$. On the other hand, the parameters that determine $M_{h,t}$ can be consistently estimated using the MRS conditions as discussed in Section 7.2.1.[14] These estimates can be used to construct consistent estimates of $M_{h,t}$, which can be aggregated across households to give $M_{g,t}$. This gives an equation analogous to equation 7.21, but where variables are group averages and where all variables on the right hand side are now observable. We use this procedure to recover the intertemporal preference parameter $\gamma$ and the participation preference parameter $\varphi$. We cannot identify any additional effect of participation that is separable in the utility function. Nor, at this stage, do we know the fixed costs of work and so we cannot identify the extensive margin response to wage changes.

Using group averages on repeated cross-sections introduces a number of other econometric problems, linked to the presence of estimation errors in small samples. These issues, as discussed in Deaton (1985), have implications for the choice of instruments and computation of standard errors. Further details of this procedure are discussed in Appendix E.1.

In principle, the first two steps of our estimation could be followed without making parametric assumptions about the utility function and, instead, estimating leisure and consumption demands directly. However, such an approach would require that the demand functions satisfy integrability conditions. Furthermore, the actual underlying utility function would still need to be recovered to study participation and the extensive margins.

### 7.2.3 Extensive margins

The last step of our approach obtains estimates of the remaining model parameters, including the fixed costs of work and childcare costs, which drive the extensive margin decision. When considering the extensive margin, it is necessary to solve explicitly the whole dynamic problem. This involves making assumptions on the entire economic environment faced by households over the life-cycle, including both present and future conditions. We solve the model numerically and use the solution to estimate and calibrate the model parameters. To reduce the numerical burden, when simulating the model, we assume a fixed interest rate. As the MRS

[14] $M_{h,t}$ includes $\chi_{h,t}$ which is unobserved. However, since it is the residual from the MRS equation, it can be included in the calculation of $\alpha_{h,t}$ that is needed to calculate $M_{h,t}$.

conditions do not change, this assumption will not change within period elasticities, but the life-cycle solution of the model and life-cycle elasticities will be affected to the extent that uncertainty about interest rates affects saving. We provide the value functions of the household's problem and details about the numerical solution in Appendix E.1.

We take as given the estimates of the parameters we obtained from the MRS and the Euler Equation, and obtain some parameters from the literature and from direct regressions. We estimate the remaining parameters so that data generated from simulations match key life-cycle aspects of the extensive margin: the participation rate, the participation rate of mothers and average wage growth of participants (which is endogenous because of selection). Finally, we simulate the model for a large number of individuals to study the properties of individual and 'aggregate' labour supply.

We then assess the model's goodness of fit by exploring the life-cycle profiles of several variables as well as participation rates conditioning on individual characteristics and the distribution of hours worked and wages.

## 7.3 Data and descriptive statistics

We take our data from the Consume Expenditure Survey (CEX) for the years 1980-2012. In the CEX, households are interviewed up to four times, answering detailed recall questions on expenditures as well as on the demographics, incomes and labour supply of household members.

We calculate gross hourly wages for individuals using information on the value of each individual's last pay cheque, the number of weeks it covered and the typical number of hours worked per week. Net wages are then calculated by subtracting marginal federal income tax rates generated using the NBER TAXSIM model (Feenberg and Coutts, 1993).[15] We deflate all expenditures, wages and incomes using the Consumer Price Index. Weekly leisure is calculated by subtracting weekly hours worked from the maximum number an individual has to divide be-

[15] We are grateful to Lorenz Kueng for making his mapping of the CEX to TAXSIM publically available.

tween leisure and labour supply per week (which we set to 100). Participation is defined by employment status at the time of the interview. Consumption covers non-durable goods excluding medical and education spending. We divide quarterly consumption spending by 13 to put it in weekly terms.

Our sample consists of couples with women aged between 25 and 60 and men aged between 25 and 65. We drop those in rural areas; those in the top 1% of the consumption and net wages distribution; those earning less than three-quarters of the national minimum wage in any given year; and those who are employed but who report working less than 5 hours a week. Since labour supply and income questions are (almost always) only asked in the first and last interviews, we drop responses from interviews apart from these two. Our sampling choices leaves just under 79,000 households (50,895 where the woman is working). Appendix E.2 presents descriptive statistics on individual characteristics over time.

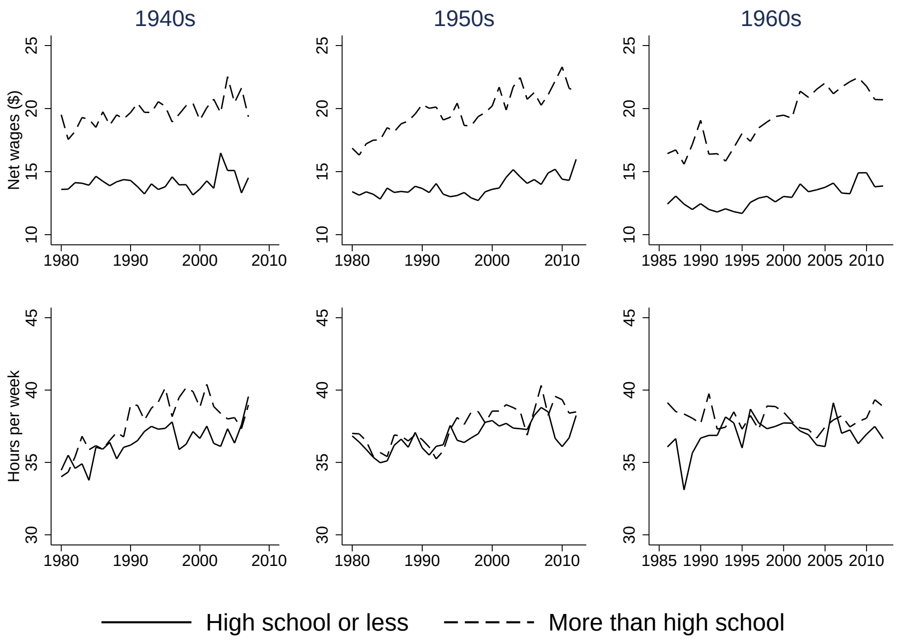
### 7.3.1 Cohort averages

We separate households into birth cohorts and examine the evolution of wages and hours by education within each cohort group. In Figure 7.1, we report patterns for the cohorts born in the 40's, 50's and 60s and for females with high school or less and with more than high school.[16]

Within the 1950s cohort, the net wages of those with more than high school education increased from an average of \$16.90 per hour in 1980 to \$21.40 in 2012 (an increase of 27%), while the wages of those with less than high school education only increased by 19% from \$13.40 to \$16.00. Despite this, the bottom row of Figure 7.1 shows average weekly hours of less educated worked actually increased by more than those from the more educated group (increasing from 36.8 hours per week to 38.2 compared to an increase from 37.4 to 38.5 for those with more than high school education).

[16] The advantage of considering the variability over time of a given cohort, is that composition is unlikely to change, as it is rare for workers to increase their educational qualifications after age 25.

**Figure 7.1:** Wages and hours by education group and cohort




Note: Data from Consumer Expenditure Survey.

### 7.3.2 Individual Variation in hours and wages

In addition to changes in average hours and wages over our sample period, there are two important issues at the individual level: what is the relative importance of the intensive and extensive margins in the raw data and what fraction of individuals are experiencing changes in hours or wages over time.

The individual extensive margin decision is whether to incur a fixed cost $F(a_{h,t})$ and participate in the current quarter. We measure this by the stated current employment status. The intensive margin decision is over how many hours to work per week (when working). An additional labour supply response may be through changing weeks worked per quarter. However, we are not able to estimate this margin of adjustment because the CEX asks current workers about the number of weeks they worked over the previous year rather than the previous quarter.

Whether ignoring the margin of the number of weeks worked within a quarter matters, depends on how much of the variance of workers' quarterly hours is driven by differences in weeks worked within a quarter rather than hours per week. Table

7.1 decomposes the variance of log annual hours into variation in log annual weeks, variation in log workers' typical weekly hours, and their covariance. The first panel shows this breakdown for the entire sample of workers. The variance in annual weeks worked is around two thirds of the total variance in hours worked. Much of this is likely to be workers not participating for entire quarters: our extensive margin. In the second panel, we restrict the sample to workers who work for more than 39 weeks (and thus could not have been unemployed for a complete quarter). These workers account for 84% of the total and for them, almost all of the variance in annual hours is a result of differences in hours worked per week, with differences in weeks worked making a negligible contribution. In the third panel, we restrict our sample further to those working exactly 52 weeks per year and notice that even among workers who do not differ in the number of weeks worked, the variance in log hours per week remains substantial (at 0.08).

These results suggest that hours worked per week is the key margin by which workers adjust their quarterly hours. We thus use this measure when estimating our MRS and Euler equations. In Appendix E.4, we check the robustness of this strategy by showing that our estimates and results are little affected by replacing our current measure with a measure of annual hours worked.

A further question is whether individual workers are able to adjust their weekly hours in response to wage changes, or whether there are market frictions that prevent this. Table 7.2 shows the proportion of workers who changed their typical hours from the first to the last CEX interview (a period of nine months). While it is true that most women do not change their hours within this period, a substantial fraction (46%) do. Around a quarter of workers change their weekly hours by 1-5 hours, and 2% change their hours by more than 20 hours.

## 7.4 Results: Parameter Estimates and Calibration

In this section, we report estimates of the structural parameters of our model. In subsections 7.4.1 and 7.4.2 we report the estimation results obtained using the MRS conditions and the Euler equation. In subsection 7.4.3, we report the calibration of

**Table 7.1:** Variances of labour supply measures, 2012

| | Less than high school | High school | Some college | Degree or higher | All |
|---|---|---|---|---|---|
| *All workers* | | | | | |
| Variance (ln hours per week) | 0.148 | 0.117 | 0.128 | 0.126 | 0.126 |
| Variance (ln weeks per year) | 0.550 | 0.271 | 0.231 | 0.482 | 0.367 |
| Covariance (ln hours, ln weeks) | 0.031 | 0.046 | 0.010 | 0.028 | 0.027 |
| Variance (ln annual hours) | 0.761 | 0.479 | 0.380 | 0.665 | 0.546 |
| *Working at least 39 weeks (84% of workers)* | | | | | |
| Variance (ln hours per week) | 0.061 | 0.040 | 0.086 | 0.110 | 0.086 |
| Variance (ln weeks per year) | 0.001 | 0.003 | 0.003 | 0.005 | 0.004 |
| Covariance (ln hours, ln weeks) | -0.001 | 0.001 | 0.002 | 0.000 | 0.001 |
| Variance (ln annual hours) | 0.059 | 0.045 | 0.094 | 0.115 | 0.092 |
| *Working 52 weeks (69% of workers)* | | | | | |
| Variance (ln hours per week) | 0.064 | 0.031 | 0.068 | 0.117 | 0.080 |

**Table 7.2:** Changes in weekly hours among the employed

| Change in Weekly Hours | No change | 1-5 hrs | 6-10 hrs | 11-20 hrs | >20 hrs |
|---|---|---|---|---|---|
| All Workers | 53.8% | 25.2% | 11.9% | 6.9% | 2.2% |
| Extent of Change in wages: | | | | | |
| $<$ 5% wage change | 74.9% | 17.5% | 4.7% | 2.3% | 0.71% |
| $>$ 5% wage change | 47.5% | 27.5% | 14.0% | 8.2% | 2.7 % |

Notes: Changes in hours are measured between the 2nd and 5th interviews for individuals who are employed at each interview.

the remaining parameters that govern choices at the extensive margin. In the last subsection, we show how well the complete model fits a number of features of the data that were not used explicitly to obtain the parameter estimates.

### 7.4.1 MRS estimates

In Table 7.3, we report the estimates of key parameters for the MRS equation and tests on the quality of our instruments, with results for the participation model reported in Appendix E.3. We estimate values for $\theta$ and $\phi$ at 1.75 and 0.76 respectively: there is much more curvature in utility on leisure than on consumption. We test the restrictions implied by Cobb-Douglas and standard CES specifications using a wild-cluster residual bootstrap. The Cobb-Douglas specification for preferences, $\phi = \theta = 1$, is rejected at the 5% level (p-value 0.01), while the standard CES specification, $\phi = \theta$, is rejected with a p-value of 0.06.

Table 7.3 also reports the coefficients, $\psi$, on demographic variables in $z_{h,t}$. A larger (positive) value for $\psi$ means, other things equal, a higher marginal utility of leisure and so women will supply fewer hours of work in the market. The positive and significant coefficient on the dummy for having children indicates that the presence of children tends to reduce hours worked, but the effect of children depends on their age. The coefficient on the number of children aged 0-2 is positive and highly significant, on children aged 3-15 the coefficient is positive, but smaller; for older children, the coefficient is negative.

We include three Heckman selection terms corresponding to the first, second and third moments of the truncated normal distribution (as described in footnote 10). We test the joint significance of these in both our first and second stage regressions. These terms are highly significant in each of the first stages, where we are predicting individual consumption, hours and wages. On the other hand, the selection terms are insignificant in the second stage of the MRS. The Cragg-Donald statistic for weak instruments in our MRS equation takes a value of 2.00 for 138 instruments, well above the relevant Stock and Yogo (2005) critical level of 1.69, and therefore suggesting that weak instruments are not a problem.[17] The Sargan test does not reject the null of no violation of the overidentifying restrictions.

[17] The value of 1.69 is given for two endogenous variables and 100 instruments, and given that the critical values for a maximum 5% relative bias for the Fuller estimator are decreasing in the number of instruments, the use of this test statistic is conservative.

**Table 7.3:** Estimation of MRS equation

| Parameter | Estimate | (Standard Error) | [95% C.I] |
|---|---|---|---|
| $\theta$ | 1.75** | (1.230) | [0.34,5.12] |
| $\phi$ | 0.76*** | (0.103) | [0.55,0.95] |
| $\Psi$ | | | |
| $\ln(famsize)$ | -0.32*** | (0.037) | [-0.38,-0.23] |
| Has kids | 0.07*** | (0.021) | [0.04, 0.10] |
| No. of kids 0-2 | 0.15*** | (0.030) | [0.10, 0.22] |
| No. of kids 3-15 | 0.06*** | (0.017) | [0.04, 0.10] |
| No. of kids 16-17 | -0.02** | (0.011) | [-0.05,0.00] |
| Joint tests of selection terms (p-value) | | | |
| First stage: ln wage | | 166.47 ($< 0.001$) | |
| First stage: ln consumption | | 311.75 ($< 0.001$) | |
| First stage: ln leisure | | 40.83 ($< 0.001$) | |
| Main equation | | 0.72 (0.87) | |
| Cragg-Donald statistic | | 2.00 | |
| Sargan statistic (p-value) | | 127.8 (0.66) | |

Notes: N = 50,895. *p<0.10, ** p<0.05, *** p<0.01. Additional controls: number of elderly (aged over 65) in the household, a quadratic in age, race, region, season, cohort-education interactions and year dummies. Consumption and leisure are instrumented with the interaction of cohort and education groups and a fifth-order polynomial time trend. Confidence intervals are bootstrapped with 1000 replications allowing for clustering at the individual level.

### 7.4.2 Euler equation estimates

Table 7.4 reports estimates of the Euler equation (7.21) using group averages. We estimate $\gamma$ to be 2.07, significantly different from zero at the 10% level, providing evidence that preferences are non-separable and that consumption and leisure are substitutes ($\gamma = 0$ would imply additively separable preferences over consumption and leisure). Since $\phi$, $\theta$ and $\gamma$ are all positive, the concavity requirements of the utility function are satisfied. The coefficients on the control variables included in the vector $z_t$ are not significant, implying demographics have no role over and above their impact on the relative weight on leisure within-period. The specification in

Table 7.4 imposes that $\varphi$, the parameter on participation in equation (7.2), is zero. When we include this term (instrumented with its own lags), the coefficient estimate is negative but not significantly different from zero.

**Table 7.4:** Estimation of Euler equation

| Parameter | Estimate | (Standard Error) | [95% C.I] |
|---|---|---|---|
| hline $\gamma$ | 2.07* | (0.656) | [-0.11,2.60] |
| $\bar{\kappa}+\ln(\beta)$ | 0.03 | (0.040) | [-0.08, 0.10] |
| $\pi$ | | | |
| $\ln(famsize)$ | -0.47 | (0.244) | [-0.69, 0.31] |
| Has kids | 0.05 | (0.069) | [-0.09, 0.19] |
| No. of kids aged 0-2 | 0.22 | (0.099) | [-0.05, 0.35] |
| No. of kids aged 3-15 | 0.03 | (0.038) | [-0.06, 0.09] |
| No. of kids aged 16-17 | 0.03 | (0.071) | [-0.11, 0.18] |
| First Stage F-stats (p-values) | | | |
| $-\phi(\Delta \ln c_{g,t} + \ln(1+r_{t+1})$ | | 7.95 (<0.001) | |
| $\Delta \ln M_{g,t}$ | | 2.08 (0.08) | |
| Sargan statistic (p-value) | | 5.70 (0.13) | |

Notes: N = 1,519. *p<0.10, ** p<0.05, *** p<0.01. Additional controls: season dummies, a quartic in age, the change in the proportion of households in each of four education groups, the change in proportion who are white, and the change in the average number of elderly individuals per household. Instruments are second, third and fourth lags of $\ln M_{g,t}$, as well as the lagged real interest rate. Confidence intervals are bootstrapped with 1000 replications.

Our instruments are second, third and fourth lags of $\ln M_{g,t}$ and the lagged real interest rate (defined as the 3 month Treasury Bill rate minus the inflation rate), and we have two endogenous variables $\phi(\Delta \ln c_{g,t} + \ln(1+r_{t+1}))$ and $\Delta \ln M_{g,t}$. We place the second of these on the left-hand side of the equation. With only one left-hand side endogenous variable, the Cragg-Donald test for weak instruments is equivalent to a standard F-test of the instruments' joint significance in the first stage regression. The critical values of these F-tests suggest that the instruments are highly correlated with the dependent variable (with an F-statistic of 7.95), but less strongly correlated with our choice of left-hand side variable (with an F-statistic of 2.08). The relevant Stock and Yogo test statistic for having less than a 5% relative bias in our parameter estimates when there are four instruments and one left-hand side endogenous variable is 7.63. When we carry out a Sargan test for the Euler equation, we fail to reject the null of over-identification (p-value 0.13) as we do for the MRS.

### 7.4.3 Calibration of the remaining parameters

There are three sets of parameters used in the calibration of the full model: those estimated via the MRS conditions and the Euler equation, those coming from external sources and those that we calibrate using simulations of the full model.

We focus on the cohort of women born in the 1950s, using moments from women age 25-55. We assume that $\chi$ and $\zeta$ are homogeneous within a cohort. Attanasio et al. (2008) show that women's labour supply behaviour differs substantially across cohorts. The main cause in that paper is differences in costs of childcare, but there are also differences in wage processes across cohorts. These differences will lead to different responses across cohorts on the extensive margin and could also lead to differences in the intensive margin because of different levels of consumption and leisure.

Within the 1950s cohort, we assume there are nine different groups of women: one group of women who remain childless for the whole of their lifetime, and eight groups of women who differ by maternity experience. These women exogenously receive two kids but differ in the age at which the first child arrives. To determine when these children are born, we draw on Rendall et al. (2010) who use population and survey data sources to calculate the distribution of maternity age at arrival of the first child for different cohorts of women in various countries.[18] We assume that the second child arrives 2 years after the first.

**External Parameters.** The complete set of external parameters is reported in Table 19 in Appendix E.3. We fix the annualized interest rate to equal the average real

[18]Consistent with the distribution for the 1950s cohort, we assume 16% of women are childless, 27% have their first child at the age of 19, 12% at the age of 22, 11% at the age of 24, 5% at the ages of 26, 28, 30 and 32 and, finally, 14% at the age of 34.

**Table 7.5:** Calibrated parameters and targets

| **Parameters** | | **Value** |
|---|---|---|
| Constant term weight of leisure | $\psi_0$ | 4.20 |
| Childcare Cost | $p$ | 967 |
| Fixed Cost of Working | $\bar{F}$ | 468 |
| Offered Wage Gender Gap at age 22 | $y_0^f/y_0^m$ | 0.72 |
| Std Dev. of Permanent Shock (Women) | $\sigma_{\xi^f}$ | 0.063 |
| Std Dev. of Initial Wage (Women) | $\sigma_{\xi_0^f}$ | 0.50 |
| Exogenous growth in offered wage | $\iota_1^f$ | 0.052 |
| Exogenous growth in offered wage | $\iota_2^f$ | - 0.0006 |
| Discount Factor (annualized) | $\beta$ | 0.99 |
| **Targets** | **Data** | **Model** |
| Weekly hours worked | 37.2 | 37.2 |
| Participation Rate | 0.684 | 0.679 |
| Participation Rate of Mothers 0-2 | 0.538 | 0.546 |
| Observed Wage Gender Gap | 0.720 | 0.727 |
| Observed Var. Wage Growth (Women) | 0.004 | 0.004 |
| Observed Initial Var. of Wages (Women) | 0.14 | 0.15 |
| Wage Growth (if younger than 40) | 0.012 | 0.010 |
| Wage Growth (if older than 40) | 0.001 | 0.004 |
| Median wealth to income ratio | 1.84 | 1.80 |

Notes: Statistics for women born in the 1950s and aged 25 to 55. Wage growth is annual.

return on three monthly T-bill at 0.015. The deterministic component of the male earnings process is estimated from the CEX: we take the two parameters of a regression of husband log earnings on age and age squared. The standard deviation of the innovation for husband's earnings, $\sigma_{\xi^m}$, is set to be 0.077, consistent with Huggett et al. (2011). Further, we estimate an initial standard deviation of husband earnings $\sigma_{\xi_0^m}$ of 0.54. There is limited evidence on the variability of women's wages and/or earnings, and further since this statistic is highly affected by non-random self-selection into the labour market, we calibrate the parameters that characterise the women's wage process within the model as explained below. Finally, we assume that the correlation coefficient between the two shocks (for husband and wife) $\rho$ is equal to 0.25 as estimated by Hyslop (2001).

As in Attanasio et al. (2008), there are two components to child care costs: the function $G(a_{h,t})$ and the price $p$. We estimate the function $G(a_{h,t})$ directly from data. For households where the mother is working, we regress total childcare expenditure on the age of the youngest child, the age of the oldest child, the number of children and a dummy equal to one if the youngest child is 0. The shape $G(a_{h,t})$ can be derived from the coefficients of this regression function, using the assumption that in our model all women have two children at an interval of two years.[19]

Finally, we assume individuals in this cohort live for 50 years from age 22, with the last 10 in retirement, and that the household receives a pension equal to 70% of the husband's earnings in the final working period.

**Calibrated parameters.** There are nine parameters that we calibrate within our decision model: the fixed cost of working, $\bar{F}$; the price of child care, $p$; the wage gender gap in offered wages, expressed as $y_0^f/y_0^m$; the standard deviation of the permanent shock to women, $\sigma_{\xi^f}$; the standard deviation of the initial wage for women, $\sigma_{\xi_0^f}$; two parameters that determine exogenous wage growth, $\iota_1^f$ and $\iota_2^f$; and the base weighting on leisure in the CES utility function, $\psi_0$, which, together

[19]Our estimate of $G\left(a_{h,t}\right)$ combines the cost of the first born child along with any subsequent costs associated with additional children who are born later. In this way, any economies of scale in child costs will be captured by $G\left(a_{h,t}\right)$, but we do not identify separately the marginal cost of extra children.

with demographics $z$ and the estimates of $\psi_z$, determine the total weight on leisure in the utility function. Finally, we calibrate the discount rate $\beta$.

The calibration targets are the participation rate of all women, the participation rate of mothers, average hours worked, the observed wage gender gap, the observed variance of wage growth, the observed initial variance in wages and the observed wage growth at two different stages of the life-cycle. Finally, we target the median wealth to median household income ratio as in Low (2005).

In Table 7.5, we report the calibrated parameters, and the targeted moments in the data and in the simulations. The monetary fixed cost of working is about 6% of median earnings of women aged 25 to 55. The additional monetary fixed childcare cost is up to 13% of median earnings for a child age 0-2. The wage ratio between women and men at age 22 is calibrated to be 0.72. This is needed to match the average observed ratio over the lifetime of 0.72. In addition to the initial wage gender gap, there is a further, exogenous wage gap that opens up through differential wage growth for men and women over the life-cycle. Exogenous wage growth implies that men's wages are on average 77% higher by the age of 45 than at the moment of entering the labour market. By contrast, for women the figure is only 31%. We calibrate the standard deviation of wage innovations for women to be 0.063 and the standard deviation of the initial wages to 0.50.

### 7.4.4 Goodness of fit

Our next step is to show whether the model can account for some observed features of women's labour supply behaviour that were not explicitly targeted in the calibration. The calibration focused on averages taken over the life-cycle. Our focus here is on life-cycle paths and on the distribution of hours and wages.

Figure 7.2 shows the life-cycle paths of women's labour supply in the model and in the data, which match well at both the extensive and intensive margins. Table 7 reports additional moments on heterogeneity. The model matches the participation of different demographic groups, such as women who have no dependent children, and mothers of children aged 3 to 17. Goldin and Mitchell (2017) show that, for women born in 1957-58, the fraction who had worked more than 80% of the years

**Figure 7.2:** Life-cycle profiles: baseline model (solid line) versus data (dashed line)

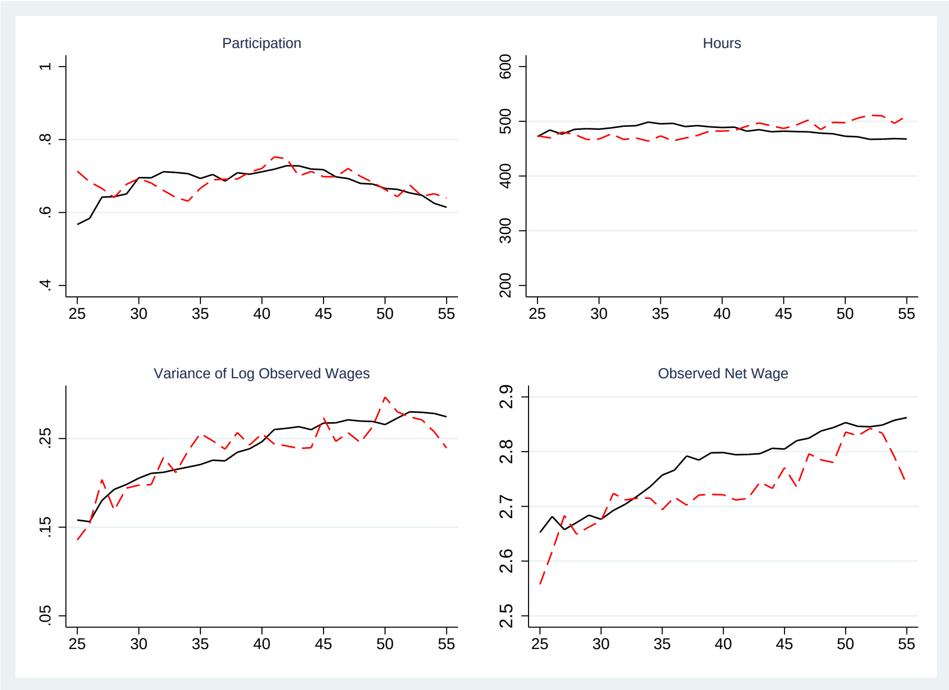



**Table 7.6:** Statistics on hetereogeneity

| | Data | Model |
|---|---|---|
| Participation Rate: Mothers with Children Aged 3-17 | 0.682 | 0.688 |
| Participation Rate: Women without Dependent Children | 0.755 | 0.692 |
| Average Hours Worked 10th Percentile | 20 | 25 |
| Average Hours Worked 25th Percentile | 35 | 31 |
| Average Hours Worked 50th Percentile | 40 | 38 |
| Average Hours Worked 75th Percentile | 40 | 44 |
| Average Hours Worked 90th Percentile | 48 | 48 |
| Wage 10th Percentile | 8.16 | 8.11 |
| Wage 50th Percentile | 15.05 | 16.20 |
| Wage 90th Percentile | 29.23 | 31.12 |
| Correlation of wages and hours | 0.33 | 0.54 |

Notes: Women without dependent children are women who have never had children and those whose children are over 17.

between age 25 and 54 was 0.53, while the fraction who had worked less than 20% was 0.09. In our benchmark economy the comparable fractions are 0.57 and 0.21. The distribution of observed wages in the model is similar to that in the data, as is the distribution of hours worked, although the fraction of women working an average of 40 hours a week is higher in the data than in the model. Observed wages and the variance of wages are increasing with age in our simulations, consistent with the data. The correlation of wages and hours worked for those employed is 0.33 in the data, compared to 0.54 in the simulations.

Finally, as discussed in Subsection 7.2.1 and in Appendix E.6, using the approximate selection correction to estimate the MRS equation could introduce a bias. To assess the importance of this bias, we take simulated data generated from the complete life-cycle model with taste heterogeneity and estimate the MRS equation using our reduced form procedure which approximates the full model. The estimates of the MRS parameters $\theta$ and $\phi$ used to generate the simulated data are almost identical to those we recover using our reduced form estimation. Given the complexity of the model and of the full-selection process, this is an important validation of the approximation used in the reduced form selection model.

## 7.5 Labour supply elasticities

This section provides the key results of the chapter. We use the estimates of the model to show implications for various wage elasticities. We start with the static Marshallian and Hicksian elasticities obtained from the MRS parameters. We then move to the Frisch elasticities at the intensive margin using estimates from the Euler equation. Finally, we simulate the full model to obtain elasticities at the extensive margin and the aggregate response of labour supply to changes in wages. When using the full model, first we analyse responses to transitory changes to wages, which do not have wealth effects and so are analogous to the Frisch elasticities; and second, we analyse the effect of shifts in the entire wage profile allowing savings and wealth to change, generating life-cycle Marshallian and Hicksian elasticities.

### 7.5.1 Marshallian and Hicksian hours elasticities

The first two columns in Table 7.7 show how the MRS parameters translate into within-period Marshallian and Hicksian wage elasticities separately for hours of work and for consumption. These elasticities vary according to family characteristics, wages and the levels of consumption and leisure. We report elasticities at different points of the distribution of Marshallian elasticities to highlight the heterogeneity across individuals.

The median Marshallian hours elasticity is estimated to be 0.18, implying an upward sloping labour supply function. Hicksian elasticities are greater than Marshallian elasticities: for the household with the median Marshallian elasticity, the Hicksian hours elasticity is three times larger at 0.54, indicating large income effects.

The Marshallian and Hicksian elasticities show substantial heterogeneity. The 90-10 range of the Marshallian hours elasticity is 0.93 (from -0.14 to 0.79) while for the Hicksian one it is 0.78 (from 0.38 to 1.16). Differences in hours worked are an important source of variation in both the Hicksian and Marshallian elasticities. Figure 7.3 plots average elasticities by wages and by hours worked. Those working the fewest hours and those with the lowest wages have the largest proportional response to a wage increase.

Our median estimates of the Marshallian and Hicksian elasticities are quite small (see Keane (2011) for a survey), and are similar to estimates in the literature obtained using a similar methodology to ours. Blundell et al. (1998) estimate values of the Marshallian elasticity ranging from 0.13 to 0.37 and of the Hicksian from 0.14 to 0.44 (depending on the age of the youngest child). The meta-study by Chetty et al. (2011) reports an average Hicksian elasticity (for men and women) of 0.33. Some results in the literature, however, report much larger estimates. MaCurdy (1983), for instance, estimates elasticities ranging from 0.74 to 1.43 (for men).

Different studies take different approaches and use different sources of variation to estimate elasticities. We investigated extensively the main reasons for different estimates of labour supply elasticities. Our hypotheses ranged from the type

**Table 7.7:** Elasticities at percentiles of Marshallian distribution

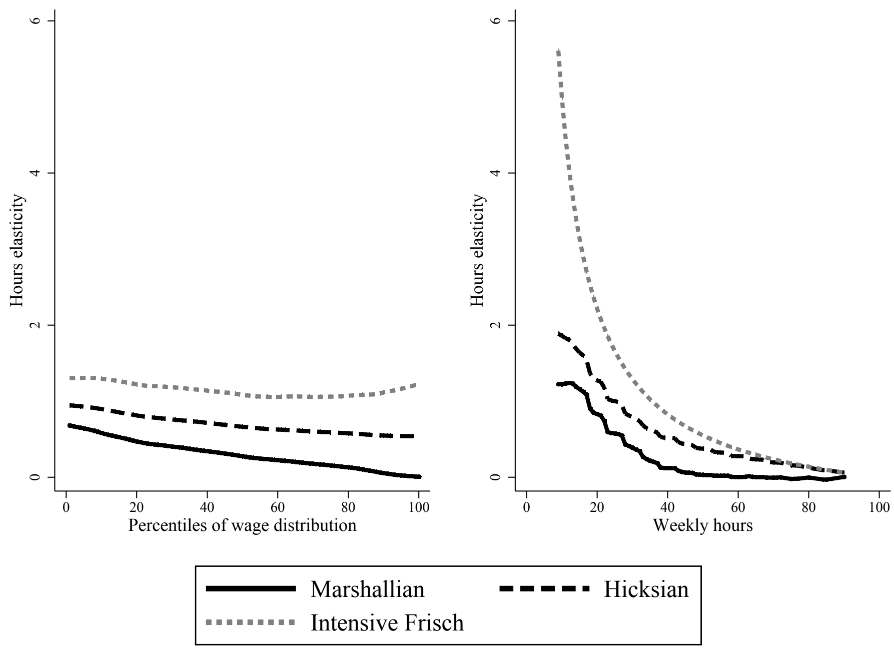
| | Wage | | | Interest rate |
|---|---|---|---|---|
| | Marshallian | Hicksian | Frisch | Frisch |
| | *Hours worked* | | | *Hours worked* |
| 10th | −0.14<br>[−0.31,0.00] | 0.38<br>[0.21,0.62] | 0.80<br>[0.25,1.85] | 0.78<br>[0.25,1.61] |
| 25th | 0.01<br>[−0.11,0.13] | 0.44<br>[0.22,0.79] | 0.80<br>[0.24,1.99] | 0.76<br>[0.24,1.68] |
| 50th | 0.18<br>[0.05,0.38] | 0.54<br>[0.24,1.07] | 0.87<br>[0.26,2.29] | 0.81<br>[0.24,1.90] |
| 75th | 0.39<br>[0.16,0.86] | 0.69<br>[0.28,1.49] | 1.00<br>[0.31,2.85] | 0.93<br>[0.31,2.34] |
| 90th | 0.79<br>[0.36,1.65] | 1.16<br>[0.51,2.30] | 1.92<br>[0.57,4.96] | 1.82<br>[0.57,4.07] |
| | *Consumption* | | | *Consumption* |
| 25th | 0.82<br>[[0.68,1.08] | 0.43<br>[0.18,0.87] | 0.04<br>[−0.02,0.50] | −1.17<br>[−1.83,−0.56] |
| 50th | 1.05<br>[0.94,1.23] | 0.52<br>[0.24,0.98] | 0.05<br>[−0.02,0.57] | −1.19<br>[−1.84,−0.52] |
| 75th | 1.30<br>[1.14,1.46] | 0.61<br>[0.31,1.06] | 0.05<br>[−0.02,0.63] | −1.20<br>[−1.84,−0.50] |

Notes: Elasticities are calculated as averages within five percentage point bands around the 10th, 25th, 50th, 75th and 90th percentiles of the Marshallian distribution. 95% confidence intervals in square brackets. Confidence intervals are bootstrapped with 1000 replications.

of specification used,[20] to the type of variation in wages that is used to identify the elasticity (that is what type of instruments are used), to sample selection rules. To estimate equilibrium conditions such as the MRS equation, researchers often use methods, such as 2SLS and GMM, that are sensitive to the normalization used. Therefore, we also investigated whether the results we obtain depend on which variable is used as a dependent variable. It turns out that the normalization used drives the result in a fundamental fashion, while results are robust to the other hypotheses considered. In particular, we find that IV or GMM estimates obtained using wages

[20] That is whether one uses consumption to proxy for the marginal utility of wealth or other indicators.

**Figure 7.3:** Intensive elasticities




Note: Lines show the distributions of Marshallian, Hicksian and intensive Frisch elasticities smoothed using a local polynomial.

as the left hand side variable (as in MaCurdy (1983)) result in very large elasticities while putting hours of leisure on the left hand side (similar to Blundell et al. (1998), who use hours worked) yields much smaller elasticities. As noted above, we use the Fuller estimator, which is less sensitive to the normalisation of the estimating equation than alternative methods. In Appendix E.3, we report results from GMM estimation with different normalisations.

### 7.5.2 Frisch hours elasticity

We compute Frisch elasticities with respect to wages at the intensive margin using equation (7.17) and estimates of the Euler equation parameters reported in Section 7.4.2. We report these elasticities in the third column of Table 7.7 and plot them alongside Hicksian and Marshallian elasticities in Figure 7.3.

The Frisch elasticity for hours of work is larger than the Hicksian elasticity, as theory would predict. The elasticity also varies in the cross-section rising from 0.8 at the 10th percentile of the Marshallian elasticity to 1.92 at the 90th percentile. The median value is 0.87. It is quite common to find large estimates of the Frisch hours elasticity among married women, and our findings are broadly in line with those of previous studies. Blundell et al. (2016a) find a Frisch elasticity for married women of 0.96; Kimmel and Kniesner (1998) estimate a Frisch elasticity of 0.67. Part of the heterogeneity we observe in the Frisch elasticities is due to differences across the life-cycle and in demographics, but, once again, much of it is also due to differences in the level of hours of work. As with the Hicksian and Marshallian elasticities, Figure 7.3 shows that Frisch hours elasticities are largest for those working the fewest hours.

The elasticity of consumption with respect to anticipated wage changes is small but positive (owing to the fact consumption and leisure are substitutes). The Frisch elasticity of consumption with respect to the interest rate at the median level of consumption is -1.19.

We compare these results with those obtained when we impose additive separability for preferences over consumption and leisure, as well as when we use a standard CES utility specification in Appendix E.4. This exercise highlights the importance of adopting a flexible utility specification. A standard CES specification, which is shown to be rejected by the estimation in Section 7.4.1, leads to similar estimates of Marshallian hours elasticities, but much larger Hicksian and Frisch elasticities. The median Frisch hours elasticity estimated using the more restrictive standard CES specification is 1.33, which is roughly 50% larger than our baseline result. The corollary of this result is that the Frisch elasticity of consumption with respect to the interest rate is much lower: imposing a standard CES forces consumption and leisure to have the same substitution parameters, making consumption less elastic and hours of work more elastic than in our baseline. In addition, the standard CES utility implies much greater non-separability between consumption and leisure: implying a Frisch wage elasticity of consumption of 0.4 compared to 0.05 under our more general utility specification. On the other hand, when we impose additive separability with our general CES specification, the Frisch hours elasticity is very similar to the one we estimate allowing for non-separability.

### 7.5.3 The extensive margin, aggregate elasticities and life-cycle responses

This section discusses labour supply responses at the extensive margin, life-cycle responses and aggregation issues at different margins and across households. We define the extensive margin elasticity as referring to the change in the percentage of women participating as the wage changes. We also calculate how total hours worked by women change as a result of both the extensive and intensive margin responses. This is what we call the "aggregate response" to a wage change. We also calculate aggregate changes in efficiency units, because women with different levels of productivity may respond differently, as suggested by Figure 7.3.

We explore responses to two different types of wage changes. First, in Section 7.5.3.1, we focus on the response to temporary changes in wages, which is relevant for temporary tax changes.[21] We report heterogeneity by age, across the wealth distribution, across demographic groups and over the business cycle. Then, in Section 7.5.3.2, we report labour supply responses to changes in the entire life-cycle wage profile, which we call *life-cycle Marshallian* and *life-cycle Hicksian* elasticities. These are interesting for two reasons. First, for thinking about the implications of permanent tax changes or differences in taxes across countries; and, second, for comparing these life-cycle Marshallian and Hicksian elasticities with the static elasticities from the MRS to assess the accuracy of the static approximation.

#### 7.5.3.1 Response to temporary wage changes

Frisch responses are calculated by comparing labour supply at a given age between the baseline economy and a counterfactual economy in which wages are anticipated to be higher at that particular age. The wage difference generates differences in participation rates, differences in hours worked for participants and, therefore, differences in aggregate labour supply. In Table 7.8, we report the average response for different age groups. The third column reports the 'extensive response', calculated

[21] We compute responses to both anticipated and unanticipated the temporary changes. The results are almost identical because there is very little effect on the marginal utility of wealth, $\lambda$, of a temporary change.

as the percentage point change in participation following a one percent increase in the wage. The fourth to sixth columns report different percentiles of the distribution of the intensive margin elasticity at each age, computed by considering only those individuals who participate both in the baseline economy and in the counterfactual economy. Changes in participation also induce changes in the distribution of hours worked that would be reflected in the aggregate response of labour supply. Finally, therefore, the last two columns report the 'aggregate' elasticity: the change in the total number of hours worked and the change in efficiency units of labour, considering both intensive and extensive margins.

**Table 7.8:** Frisch responses by age

| Age Band | Participation Rate (Percent) | Extensive Response (Percent Pt) | Intensive Elasticity | | | Aggregate Elasticity | |
|---|---|---|---|---|---|---|---|
| | | | 25th | 50th | 75th | Hours | Eff units |
| 25-29 | 61.61 | 0.82 | 0.69 | 0.85 | 1.09 | 1.93 | 1.44 |
| 30-34 | 70.07 | 0.63 | 0.66 | 0.82 | 1.11 | 1.51 | 1.12 |
| 35-39 | 70.00 | 0.63 | 0.64 | 0.82 | 1.14 | 1.49 | 1.08 |
| 40-44 | 72.05 | 0.56 | 0.64 | 0.85 | 1.21 | 1.37 | 1.01 |
| 45-49 | 69.53 | 0.59 | 0.65 | 0.88 | 1.25 | 1.42 | 1.04 |
| 50-55 | 65.37 | 0.59 | 0.67 | 0.91 | 1.30 | 1.49 | 1.06 |

Notes: The extensive response is the percentage point change in participation in response to a 1% increase in the wage. The aggregate elasticity reports the percentage change in hours corresponding to a percentage change in the wage, accounting for changes at both the extensive and intensive margins.

A first point to notice is the variation in the size of the extensive margin elasticity over the life-cycle. As a consequence, the age composition of the population may have important implications for the aggregate response of labour supply to changes in wages. Early in life, the percentage point response is about 0.82, falling to 0.63 between 30 and 35 and to a minimum of 0.56 for the 40-45 group. The median of the intensive margin elasticity is stable over the life-cycle, at around 0.85,[22]

[22]The comparable value calculated directly from step 2 of the estimation process is 0.86. The

however the elasticity at the 75th percentile increases substantially with age.

The aggregate elasticity for hours is about 1.45 on average, but again is larger at the start of the life-cycle. The relative importance of the extensive and intensive margins to explaining the macro elasticity varies with age. Before age 30, the intensive margin response contributes approximately 46% of the response in the aggregate. However, by age 50-55, the contribution of the intensive response has increased to 63%. The contribution of the intensive margin is somewhat larger than Erosa et al. (2016), who find that the response through the intensive margin contributes about 38% to the aggregate response. This difference is not surprising since the Erosa et al. (2016) calculation is for men, where we see less variability in hours worked but it highlights the difficulty of aggregating behaviour to create a single labour supply elasticity. The aggregate elasticity for efficiency units is smaller than that for hours, but also declines with age.

**Household Wealth.** In Table 7.9, we report household responses across the wealth distribution. We calculate the percentiles of household's wealth at each age and classify households into quartiles. We find a clear pattern of a decreasing response of the extensive margin with increasing wealth. This is the case at all ages. There is also heterogeneity in the intensive margin elasticity by wealth, with the wealthy being less responsive, but the differences are more moderate than with the extensive margin response. The message from these results is that the distribution of wealth is crucial to understanding the response of aggregate labour supply to changes in wages.

**Macroeconomic Conditions.** Labour supply responses may change across the business cycle. Differences in the economic environment will lead to differences in the estimated elasticity for the same underlying preference parameters, as also discussed by Keane and Rogerson (2012). This issue is likely to be relevant particularly for the extensive margin, which is driven by non-convexities in the dynamic problem, such as fixed costs of going to work. If these non-convexities are important, it is likely that a certain sequence of aggregate shocks will tend to bunch

similarity of estimates from step 2 and step 3 of the estimation provides further validation of our multi-step approach.

**Table 7.9:** Frisch responses by household wealth

| Wealth Quartile | Participation Rate (Percent) | Extensive Response (Percent Pt) | Intensive Elasticity (Median) | Aggregate Elasticity |
|---|---|---|---|---|
| Below $p25$ | 45.42 | 1.20 | 1.20 | 3.53 |
| $p25 - p50$ | 59.25 | 0.77 | 1.03 | 2.07 |
| $p50 - p75$ | 76.80 | 0.39 | 0.81 | 1.23 |
| Above $p75$ | 90.10 | 0.16 | 0.66 | 0.81 |

Notes: The extensive response is the percentage point change in participation in response to a 1% increase in the wage. The aggregate elasticity reports the percentage change in hours corresponding to a percentage change in the wage, accounting for changes at both the extensive and intensive margins.

(or further disperse) households around the kinks that determine the extensive margin response. As a consequence, different distributions of the state variables will trigger different responses in the aggregate. In particular, whether an economy is in a recession or not may well affect how much individuals are willing to respond to wage growth.

**Table 7.10:** Frisch responses across the business cycle

| Business Cycle | Extensive Response (Percent) | Intensive Elasticity (Median) | Aggregate Elasticity | |
|---|---|---|---|---|
| | | | Hours | Eff units |
| Baseline | 0.63 | 0.86 | 1.53 | 1.12 |
| Recession | | | | |
| First quarter | 0.67 | 0.87 | 1.61 | 1.15 |
| Fourth quarter | 0.73 | 0.86 | 1.71 | 1.20 |

Notes: The extensive response is the percentage point change in participation in response to a 1% increase in the wage. The aggregate elasticity reports the percentage change in hours corresponding to a percentage change in the wage, accounting for changes at both the extensive and intensive margins.

In Table 7.10, we report responses to temporary changes in wages that occur at different points of the business cycle.[23] We report the labour supply response in the first and fourth quarters of the recession. The key finding is that responses are higher in recessions than in the baseline, and further, responses increase with the duration of the recession. From the results in Table 7.9, the decrease in wealth that households suffer over a recession could be behind the increasing responsiveness of the extensive margin to anticipated changes in wages. Effects may persist beyond the end of the recession, especially if wages or wealth are permanently lower. Both lower wages and lower wealth lead to higher elasticities: households who have been hit by recessions earlier in their life are more responsive throughout the remainder of their lives.[24]

**Demographics.** Finally, we explore the effect of children on the size of the elasticities. Mothers of children aged 0 to 2 are more elastic at the extensive margin (0.82) than mothers of older children (0.68) and childless women show the lowest elasticity (0.57). In contrast, differences in intensive margin elasticities are less pronounced, with mothers of young children being slightly less elastic.

#### 7.5.3.2 Life-cycle responses to wage changes

In this section, we use our model to compute the response to a change in the entire wage profile, so to measure the response to a permanent tax change. The life-cycle Marshallian elasticity captures the response of labour supply to changes in wages when savings are allowed to change, that is when the extra income that arises in period $t$ due to the increased wage, does not have to be spent in that period.

The life-cycle Hicksian elasticity arises after netting off the extra lifetime resources from the lifetime budget constraint, in contrast with a static Hicksian response, which would net off the extra resources within period. Life-cycle compensation is calculated as the change in income needed to keep the original bundle of

[23] We define a recession as a situation in which all men and women receive an unexpected negative earnings shock for four consecutive quarters. These wage changes are to the permanent wage and will affect the marginal utility of wealth as well as changing intertemporal incentives. We consider responses to temporary changes in wages at different points in such a recession.

[24] We show this by using our simulations to compare women hit by a recession at age 25 with those not hit by recession. Differences persist throughout their lifetimes. The detail of these results are not reported here.

life-time consumption and hours worked exactly affordable. The change in income from each period that needs to be compensated for is $\Delta w^{f}_{h,t} * \left(L - l_{h,t}\right)$. Summing across all periods would give the extra resources from a wage increase that need to be subtracted in a life-cycle context. This compensation can be implemented either by imposing a person-specific lump-sum tax that is equal across periods, or a person-specific lump-sum tax at a given point in time. The choice will matter because uncertainty means the timing of income is important.[25] The alternative to this exact compensation is to do the compensation within a group, or indeed within the whole population as discussed by Keane (2011), which would mean calculating the extra income for all individuals as with the exact calculation, but then redistributing through a common per period lump-sum payment. This approach does not give exactly the life-cycle Hicksian response because some households will be over-compensated and some under-compensated relative to their individual change in lifetime resources. On the other hand, it may be the right way to calculate the response to a funded tax change. If preferences are quasi-linear then there are no income effects and so there is no effect on labour supply of any redistribution associated with the lump sum compensation.

In Table 7.11, we report the life-cycle Marshallian and Hicksian responses. The first panel shows responses when the Hicksian compensation is common across all individuals. The second panel shows responses when compensation is common within quartiles of the initial wage distribution for women. We compare these life-cycle elasticities with the static elasticities estimated from the MRS. As we argued in Section 7.1.1 and emphasized by Meghir and Phillips (2008), life-cycle labour supply responses may be approximated by the static elasticities computed from the MRS.

The median life-cycle Marshallian elasticity for the intensive margin is 0.43, substantially above the 0.18 static Marshallian elasticity. The static elasticities are calculated from the MRS using non-durable consumption, holding constant saving and also, implicitly, durable spending. In a full life-cycle model, however, following

[25] In a model with substantial ex-ante and ex-post heterogeneity, either from of compensation is computationally costly to calculate.

a wage increase, savings adjust and individuals reallocate resources across periods. Furthermore, all life-cycle resources are spent, so that we have a broader consumption measure in these calculations. In other words, the extra income from the wage increase is not all spent on non-durables in the period it is earned. Spreading these resources across periods and other goods reduces the amount of extra income and hence the income effect in the period it is earned. This means the life-cycle Marshallian elasticity is more like the static Hicksian elasticity. However, the life-cycle Hicksian elasticity is close to the Hicksian elasticity we estimate with the MRS.

Looking at the responses by quartile in the bottom panel, there is substantial heterogeneity in the size of the life-cycle Marhsallian intensive margin response depending on initial conditions, particularly in the extensive margin response. On the other hand, the life-cycle Hicksian elasticity when there is within quartile compensation, does not vary much with the quartile of the initial conditions.[26] The substitution effect is very similar across groups, and it is the income effect which matters more for the heterogeneity in the Marshallian labour supply responses across groups.

### 7.5.4 Elasticities with returns to experience

An important maintained assumption to this point has been the absence of any returns to experience. Imai and Keane (2004) argue that assuming wages are exogenous may introduce a downward bias in estimates of the willingness to substitute intertemporally. Indeed, they present estimates of such a parameter as high as 3.8 in a model that accounts for returns to labour market experience. We consider as a robustness exercise an alternative framework in which returns to experience accrue to individuals who are participating, but in which returns to experience are not affected by the number of hours worked conditional on participation. Appendix E.5 details the estimation results allowing for returns to experience. Intensive elasticities are similar to our baseline, but the extensive margin response differs: with returns to experience, the current wage is only part of the return to work and so changes to the

[26] We experiment with more finely targetted compensation, in particular making the individual transfer contingent on initial husband earnings and the maternity group, but this does not alter the overall intensive margin response.

**Table 7.11:** Life-cycle responses

| | Extensive Response (Percent Pt) | Intensive Elasticity 25th | Intensive Elasticity 50th | Intensive Elasticity 75th | Aggregate Elasticity Hours | Aggregate Elasticity Eff Units |
|---|---|---|---|---|---|---|
| | | | **Whole Sample** | | | |
| **Marshallian** | | | | | | |
| Life-cycle Response | 0.51 | 0.28 | 0.42 | 0.67 | 0.91 | 0.63 |
| Static (MRS) | | 0.01 | 0.18 | 0.39 | | |
| **Hicksian** | | | | | | |
| Life-cycle Response | 0.65 | 0.42 | 0.63 | 0.96 | 1.26 | 0.84 |
| Static (MRS) | | 0.44 | 0.54 | 0.69 | | |
| | | | **By Quartile of Initial Wage** | | | |
| **Life-cycle Marshallian** | | | | | | |
| 1st quartile | 0.62 | 0.40 | 0.57 | 0.80 | 2.25 | 1.88 |
| 2nd quartile | 0.70 | 0.34 | 0.48 | 0.78 | 1.44 | 1.21 |
| 3rd quartile | 0.55 | 0.32 | 0.48 | 0.75 | 0.97 | 0.83 |
| 4th quartile | 0.17 | 0.22 | 0.33 | 0.52 | 0.46 | 0.40 |
| **Life-cycle Hicksian** | | | | | | |
| 1st quartile | 0.66 | 0.46 | 0.65 | 0.87 | 2.47 | 2.05 |
| 2nd quartile | 0.81 | 0.45 | 0.62 | 0.94 | 1.71 | 1.43 |
| 3rd quartile | 0.64 | 0.48 | 0.67 | 0.97 | 1.25 | 1.04 |
| 4th quartile | 0.21 | 0.41 | 0.56 | 0.81 | 0.71 | 0.60 |

Notes: The extensive response is the percentage point change in participation in response to a 1% increase in the wage. The baseline participation rate is 67.8%. Within quartiles, the baseline participation rates are 29,56,77 and 95% respectively. The aggregate elasticity reports the percentage change in hours corresponding to a percentage change in the wage, accounting for changes at both the extensive and intensive margins.

current wage make little difference to participation. The extensive margin response becomes very small.

## 7.6 Conclusion

This chapter shows that to understand labour supply behaviour and to calculate aggregate labour supply elasticities, it is crucial to account for heterogeneity across individuals. To make this point precisely and show its quantitative importance, we estimate a life-cycle model of intratemporal and intertemporal choices over consumption, saving and work and characterise the response of women's labour supply to different types of wage changes. In estimating such a model, we use a flexible specification of preferences that allows us to test some of the assumptions commonly used both in the macro and labour literature on labour supply.

We find substantial heterogeneity in labour supply responses, and this heterogeneity is prevalent at both the intensive and extensive margins. The median static Marshallian elasticity is 0.18, but has a *90-10* range of -0.14 to 0.79. The corresponding Hicksian elasticity is 0.54, with *90-10* range of 0.38 to 1.16; and the corresponding Frisch wage elasticity is 0.87, with *90-10* range of 0.8 to 1.92. The static Marshallian and Hicksian concepts assume there is no intertemporal reallocation of resources in response to a wage change. We use the full life-cycle model to show that these static concepts underestimate the full life-cycle responses, especially for the life-cycle Marshallian response. Finally, over the business cycle we find that the aggregate hours elasticity increases in recessions and more so in longer recessions.

In terms of heterogeneity in the intensive margin responses, the Marshallian, Hicksian and Frisch elasticities are greatest for those working the least number of hours, those with the lowest wages and those with the least wealth. For the extensive margin, the response to anticipated wage growth is large for women under 30 and can explain 54% of their labour supply response. This sizable contribution of the extensive margin declines with age. We find some evidence of non-separability between consumption and leisure, but assuming there is separability does not substantially change the distribution of estimates of the Frisch elasticity.

Our preference parameter estimates reject the restrictions required for balanced growth, which are widely used in the macro literature. The curvature on consumption in utility is less than log, and the curvature on hours worked is much greater than the curvature on consumption. This implies individuals are less willing to substitute hours of work over time than they are willing to substitute consumption. Further, the heterogeneity we observe means it is not sensible to talk about a single elasticity to measure how aggregate labour supply responds to wage changes. Instead, we aggregate explicitly from individual behaviour to the aggregate in order to understand how economy wide hours of work change given the demographic and age structure of the economy, the wealth distribution and the state of the business cycle.

Our results on the importance of the extensive margin in explaining macro elasticities can be compared to others in the literature, especially Erosa et al. (2016) and Guner et al. (2012). Our estimates put a greater importance on intensive margin changes in hours worked per week than those papers, but we do find that a substantial fraction of the changes in total hours is due to changes in participation, ranging from 54% to 37%. Erosa et al. (2016) find that the extensive margin is the dominant labour supply response, explaining 62% of the aggregate response. Their model has a similar life-cycle structure to ours, but is focused on men's labour supply and the conclusion on the importance of the extensive margin is for men where hours of work are less variable. Guner et al. (2012) analyse the importance of the extensive margin for the aggregate response of labour supply to changes in taxes in a model with heterogeneous married and single households, and with an extensive margin for women as well as an intensive margin for men and women. As with Erosa et al. (2016), they find that the extensive margin for women is a key contributor to the aggregate response to tax reform. The key difference from our framework is their assumption that there is no uncertainty in wages and this assumption of certainty tends to lead to greater labour supply responses, as shown in Low (2005).

One key point that emerges from our exercise is that aggregate responses of labour supply to changes in wages (both at the intensive and the extensive margin) is not constant: it changes with the structure of the population as well as with the state of the economy. This finding is similar to Keane and Rogerson (2012), who argue that there is no contradiction between macro and micro elasticities of labour supply and that they are simply measuring different concepts. Our conclusion is however stronger: the macro elasticity is not a structural parameter, it is simply the result of highly non-linear aggregation which depends on demographic structure as well as the distribution of wealth and the particular point in the business cycle.

# Chapter 8

# General Conclusions

The chapters in this thesis all address questions relating to households' consumption and labour supply decisions over the life-cycle.

Chapters 2 and 3 addressed issues around how we measure expenditure and price changes. Chapter 2 documented the declining share of expenditure surveys in the US and UK relative to the national accounts and explored possible reasons for this. Chapter 3 discussed how we measure inflation and in particular the appropriate choice of formula for price changes at the elementary aggregate level of price indices.

The next two chapters discussed consumption patterns at older ages. Chapter 4 discusses spending declines in two countries - the US and the UK - and the role of medical expenses in accounting for these differences. Chapter 5 applied non-parametric, 'revealed preference' tests of different variants of the life-cycle model to retiring households in a Spanish consumption panel dataset.

The final two chapters looked at how consumers' spending and labour supply choices are affected by changes in their economic environment. Chapter 6 considered how households responses to house price changes are affected by their initial leverage. It showed how more leveraged households tend to disproportionately increase their residential investment spending as house prices rise, and set out a portfolio-rebalancing motive to explain this behaviour. Chapter 7 looked at how womens' labour supply responds to changes in wages along both intensive and extensive margins.

A common theme across these chapters is the importance of accounting for heterogeneity in consumer's behaviours and responses to shocks. Chapter 4 highlighted how risks and precautionary savings motives change as individuals age. Chapter 5 discussed the importance of allowing for heterogeneity in preferences when conducting tests of the life-cycle model. Chapter 6 and Chapter 7 both showed how consumer responses to shocks depend non-linearly on their individual wealth and indebtedness. Overall responses to shocks will depend heavily on how these characteristics vary across households in the economy. Individual responses will also not easily aggregate to give a single 'macro' response to changes in the economic environment.

# Appendix A

# Appendix for Chapter 3

## A.1 Proofs of chapter results

*Proof.* Proof of fact 1.

Using $x_k = \frac{p_1^i}{p_0^i}$, the classical geometric-arithmetic inequality is

$$\prod \left(x^i\right)^{1/N} \leq \sum \frac{1}{N} x^i$$

Using the change of variables

$$x^i = \left(y^i\right)^s$$

and substituting

$$\prod \left[\left(y^i\right)^s\right]^{1/N} \leq \sum \frac{1}{N} \left(y^i\right)^s$$

then taking the $s$th root gives

$$\prod \left(y^i\right)^{1/N} \leq \left(\sum \frac{1}{N} \left(y^i\right)^s\right)^{1/s}$$

For $s \in (0,1)$ we can use Jensen's inequality again to give

$$\left(\sum \frac{1}{N} \left(y^i\right)^s\right)^{1/s} \leq \sum \frac{1}{N} y^i$$

since $z^s$ is concave with $s \in (0,1)$

$$\sum \frac{1}{N}\left(y^i\right)^s \le \left(\sum \frac{1}{N} y^i\right)^s$$

Now set $s = 1/2$. Then since $Var(X) = E\left(X^2\right) - \left[E(X)\right]^2$

$$\begin{aligned} E(X) &= \sum \frac{1}{N}\left(y^i\right)^{1/2} \\ \left[E(X)\right]^2 &= \left(\sum \frac{1}{N}\left(y^i\right)^{1/2}\right)^2 \\ E\left(X^2\right) &= \sum \frac{1}{N} y^i \end{aligned}$$

$$Var(X) = \sum \frac{1}{N} y^i - \left(\sum \frac{1}{N}\left(y^i\right)^{1/2}\right)^2$$

Using the fact that $\prod (y_i)^{1/N} \le \left(\sum \frac{1}{N}\left(y^i\right)^s\right)^{1/s}$(which we have already established) with $s = 1/2$ gives

$$\prod \left(y^i\right)^{1/N} \le \left(\sum \frac{1}{N}\left(y^i\right)^{1/2}\right)^2$$

so

$$Var(X) \le \sum \frac{1}{N} y^i - \prod \left(y^i\right)^{1/N}$$

hence

$$P_C(\mathbf{p}_0, \mathbf{p}_1) - P_J(\mathbf{p}_0, \mathbf{p}_1) \ge Var\left(\frac{p_1^i}{p_0^i}\right)$$

□

*Proof.* Proof of fact 2.

First notice that we can rewrite the Dutot as

$$P_D\left(\mathbf{p}_0,\mathbf{p}_1\right)=\frac{E\left[p_1^i\right]}{E\left[p_0^i\right]}=\frac{E\left[\left(\frac{p_1^i}{p_0^i}\right)p_0^i\right]}{E\left[p_0^i\right]}$$

Then notice that the definition of the covariance between $p_0^i$ and $\left(\frac{p_{i1}}{p_{i0}}\right)$ is

$$Cov\left(p_0^i,\left(\frac{p_1^i}{p_0^i}\right)\right)=E\left[p_0^i\left(\frac{p_1^i}{p_0^i}\right)\right]-E\left[p_0^i\right].E\left[\left(\frac{p_1^i}{p_0^i}\right)\right]$$

$$\Rightarrow Cov\left(p_{i0},\left(\frac{p_1^i}{p_0^i}\right)\right)/E\left[p_0^i\right]=E\left[p_0^i\left(\frac{p_1^i}{p_0^i}\right)\right]/E\left[p_0^i\right]-E\left[\left(\frac{p_1^i}{p_0^i}\right)\right]$$

This is just the difference between the Dutot and the Carli, so we have that

$$\Rightarrow P_D(\mathbf{p}_0,\mathbf{p}_1)-P_C(\mathbf{p}_0,\mathbf{p}_1)=\frac{Cov\left(p_0^i,\left(\frac{p_1^i}{p_0^i}\right)\right)}{E\left[p_0^i\right]}$$

□

*Proof.* Proof of proposition 3.

Writing prices as in equation (3), we can think of the Dutot as the empirical counterpart of

$$P_D\left(\mathbf{p}_0,\mathbf{p}_1\right)=\frac{E\left(p_1^i\right)}{E\left(p_0^i\right)}$$

and the Jevons as the counterpart of

$$P_J\left(\mathbf{p}_0,\mathbf{p}_1\right)=\prod\left(\frac{E\left(p_1^i\right)\left(1+e_1^i\right)}{E\left(p_0^i\right)\left(1+e_0^i\right)}\right)^{1/N}=\frac{E\left(p_1^i\right)}{E\left(p_0^i\right)}\prod\left(\frac{1+e_1^i}{1+e_0^i}\right)^{1/N}$$

Rearranging gives

$$P_J(\mathbf{p}_0, \mathbf{p}_1) = P_D(\mathbf{p}_0, \mathbf{p}_1) \prod \left( \frac{1+e_1^i}{1+e_0^i} \right)^{1/N}$$

The Jevons is equal to the Dutot multiplied by a function of the deviations in each period. We can approximate the value of $\prod \left( \frac{1+e_1^i}{1+e_0^i} \right)^{1/N}$ by taking a second order Maclaurin expansion.

Let

$$\prod \left( \frac{1+e_1^i}{1+e_0^i} \right)^{1/N} = f(\mathbf{e}_1, \mathbf{e}_0)$$

then our approximation is

$$f(\mathbf{e}_1, \mathbf{e}_0) \approx f(\mathbf{0}, \mathbf{0}) + \left[ \begin{array}{cc} \left. \frac{\partial f(\mathbf{e}_1, \mathbf{e}_0)}{\partial \mathbf{e}_1} \right|_{\mathbf{e}_1, \mathbf{e}_0 = 0} & \left. \frac{\partial f(\mathbf{e}_1, \mathbf{e}_0)}{\partial \mathbf{e}_0} \right|_{\mathbf{e}_1, \mathbf{e}_0 = 0} \end{array} \right] \left[ \begin{array}{c} \mathbf{e}_1 \\ \mathbf{e}_0 \end{array} \right]$$
$$+ \frac{1}{2} \left[ \begin{array}{cc} \mathbf{e}_1' & \mathbf{e}_0' \end{array} \right] \left[ \begin{array}{cc} \frac{\partial^2 f(\mathbf{e}_1, \mathbf{e}_0)}{\partial \mathbf{e}_1 \partial \mathbf{e}_1'} & \frac{\partial^2 f(\mathbf{e}_1, \mathbf{e}_0)}{\partial \mathbf{e}_0 \mathbf{e}_1'} \\ \frac{\partial^2 f(\mathbf{e}_1, \mathbf{e}_0)}{\partial \mathbf{e}_1 \partial \mathbf{e}_0'} & \frac{\partial^2 f(\mathbf{e}_1, \mathbf{e}_0)}{\partial \mathbf{e}_1 \partial \mathbf{e}_1'} \end{array} \right] \left[ \begin{array}{c} \mathbf{e}_1 \\ \mathbf{e}_0 \end{array} \right]$$

The derivatives of $f(\mathbf{e}_1, \mathbf{e}_0)$ are the following

$$\left. \frac{\partial f(\mathbf{e}_1, \mathbf{e}_0)}{\partial e_1^i} \right|_{\mathbf{e}_1, \mathbf{e}_0 = 0} = 1/N, \forall i$$

$$\left. \frac{\partial f(\mathbf{e}_1, \mathbf{e}_0)}{\partial e_0^i} \right|_{\mathbf{e}_1, \mathbf{e}_0 = 0} = -1/N, \forall i$$

$$\left. \frac{\partial^2 f(\mathbf{e}_1, \mathbf{e}_0)}{\partial e_1^i \partial e_0^j} \right|_{\mathbf{e}_1, \mathbf{e}_0 = 0} = -(1/N)^2, \forall i, j$$

$$\left. \frac{\partial^2 f(\mathbf{e}_1, \mathbf{e}_0)}{\partial e_1^i \partial e_1^j} \right|_{\mathbf{e}_1, \mathbf{e}_0 = 0} = \left. \frac{\partial^2 f(\mathbf{e}_1, \mathbf{e}_0)}{\partial e_0^i \partial e_0^j} \right|_{\mathbf{e}_1, \mathbf{e}_0 = 0} = (1/N)^2, \forall i \neq j$$

$$\left. \frac{\partial^2 f(\mathbf{e}_1, \mathbf{e}_0)}{\partial (e_1^i)^2} \right|_{\mathbf{e}_1, \mathbf{e}_0 = 0} = \left. \frac{\partial^2 f(\mathbf{e}_1, \mathbf{e}_0)}{\partial (e_0^i)^2} \right|_{\mathbf{e}_1, \mathbf{e}_0 = 0} = \frac{1}{N} (\frac{1}{N} - 1), \forall i$$

So our approximation evaluates to

$$= 1 + \frac{1}{2}\left[\left(\frac{1}{N}\sum e_1^i\right)^2 - \left(\frac{1}{N}\sum (e_1^i)^2\right) - 2\left(\frac{1}{N}\sum e_1^i\right)\left(\frac{1}{N}\sum e_0^i\right) - \left(\frac{1}{N}\sum e_0^i\right)^2 + \left(\frac{1}{N}\sum (e_0^i)^2\right)\right]$$

$$= 1 + \frac{1}{2}\left[\left(\frac{1}{N}\sum (e_0^i)^2\right) - \left(\frac{1}{N}\sum (e_1^i)^2\right)\right] \tag{A.1}$$

so we have that

$$P_J(\mathbf{p}_0, \mathbf{p}_1) \approx P_D(\mathbf{p}_0, \mathbf{p}_1)\left(1 + \frac{1}{2}\left[Var\left(e_0^i\right) - Var\left(e_1^i\right)\right]\right) \tag{A.2}$$

□

*Proof.* Proof of proposition 1.

To find the solution to $\max_{\mathbf{w}} -\mathbf{w}' \ln \mathbf{w}$ subject to $\sum w^i = 1$, we first set up the Lagrangian

$$\mathcal{L} = -\mathbf{w}' \ln \mathbf{w} - \lambda(\sum_i w^i - 1) \tag{A.3}$$

where $\lambda$ is the Lagrange multiplier. Taking first order conditions gives

$$w^i \frac{1}{w^i} + \ln w^i - \lambda = 0, \forall i \tag{A.4}$$

$$\implies \ln w^i = \lambda - 1, \forall i \tag{A.5}$$

$$\implies w^i = \exp(\lambda - 1), \forall i \tag{A.6}$$

which implies that at the point of maximum entropy, budget shares are constant across $i$. Combining this with constraint tells us that the entropy maximising budget shares will be $1/N$.

□

*Proof.* Proof of proposition 2.

This proof consists of two stages. First we show that the solution to the maximum entropy problem (3.12) is $w_t^i = 1/N$ for all $i,t$. Then we show that constant and equal budget shares are consistent with GARP, and so the additional constraint in this problem is not binding.

The first stage is to solve

$$\max_{\mathbf{w}_0,\mathbf{w}_1} H(\mathbf{w}_0,\mathbf{w}_1) = -\sum_t \mathbf{w_t}' \ln \mathbf{w_t} \text{ subject to } \sum_i w_t^i = 1 \text{ for } t = 0,1$$

which is associated with the Lagrangian

$$-\sum_t \mathbf{w_t}' \ln \mathbf{w_t} - \lambda_0 (\sum w_0^i - 1) - \lambda_1 (\sum w_1^i - 1)$$

with $\lambda_0$ and $\lambda_1$ Lagrange multipliers. This has first order conditions

$$1 + \ln w_t^i - \lambda_t = 0, \forall i,t$$

which as we demonstrate in the proof of proposition 1 implies that $w_t^i$ is constant across $i$ for both $t$. Combining this with the constraints implies that budget shares will be $1/N$ (constant across both $i$ and $t$).

To prove the second stage, we will show that a violation of GARP is impossible with equal budget shares. A violation of GARP implies that there exist two periods $t$ and $s$, when the consumer chooses quantities $q_t$ and $q_s$ such that $\mathbf{p}_t'\mathbf{q}_s \leq x_t$ and $\mathbf{p}_s'\mathbf{q}_t < x_s$ or equivalently

$$\sum p_t^i \left[ \frac{w_s^i x_s}{p_s^i} \right] \leq x_t$$

$$\Rightarrow \sum \left( \frac{p_t^i}{p_s^i} \right) w_s^i \leq \frac{x_t}{x_s}$$

and

$$\sum \left(\frac{p_s^i}{p_t^i}\right) w_t^i < \frac{x_s}{x_t}$$

Our working assumption is that budget shares are constant $w_t^i = w_s^i = \frac{1}{N}$ for all $i$,$t$. So these conditions imply that

$$\frac{1}{N}\sum \left(\frac{p_t^i}{p_s^i}\right) \leq \frac{x_t}{x_s}$$

and

$$\frac{1}{N}\sum \left(\frac{p_s^i}{p_t^i}\right) < \frac{x_s}{x_t}$$

Now we know from Jensen's inequality that

$$\frac{1}{\frac{1}{N}\sum \left(\frac{p_t^i}{p_s^i}\right)} \leq \frac{1}{N}\sum \left(\frac{p_s^i}{p_t^i}\right)$$

since the reciprocal is a convex function. However, since for positive prices and expenditures

$$\frac{1}{\frac{1}{N}\sum \left(\frac{p_t^i}{p_s^i}\right)} \geq \frac{x_s}{x_t}$$

then this implies that

$$\frac{x_s}{x_t} < \frac{x_s}{x_t}$$

which is a contradiction. It follows that equal budget shares must satisfy GARP. □

# Appendix B

# Appendix for Chapter 4

## B.1 Long term care costs

**Table B.1:** Distribution of out-of-pocket long term care costs, non-institutional population

| **HRS: United States** | percent | Mean | P25 | P50 | P75 | P95 | P99 |
|---|---|---|---|---|---|---|---|
| All household population 60+ | 100 | 53 | 0 | 0 | 0 | 0 | 526 |
| Any stays in institutions (past 2 yrs) | 5.3 | 991 | 0 | 0 | 246 | 3768 | 29075 |
| Any paid-for stays in institutions (past 2 yrs) | 1.7 | 3085 | 263 | 645 | 2010 | 12975 | 38189 |
| **ELSA: England** | | | | | | | |
| All household population 60+ | 100 | 30 | 0 | 0 | 0 | 0 | 0 |
| Any stays in institutions (past 2 yrs) | 1.1 | 3236 | 0 | 0 | 1257 | 15699 | 37704 |
| Any paid-for stays in institutions (past 2 yrs) | 0.6 | 6619 | 628 | 1782 | 10055 | 28278 | 37704 |

Notes: Data from English Longitudinal Study of Ageing for England and the Health and Retirement Survey in the US for those not in institutional residences at the time of interview. Values are annual averages over the previous two years in US$ (2010). US spending is for 2012-2014. UK spending is for 2014-2016.

## B.2 Within period demand system

In this appendix, we estimate an extension of the Almost Ideal specification of Deaton and Muellbauer (1980) that includes an additional quadratic term in income (Banks et al. (1997)). Our interest is in establishing the nature of within-period non-separabilities between consumption and housing, health and employment in the two countries through the effect of these variables on household budget shares. By including total expenditure and prices, we control for differences in trends in relative prices and wealth across different birth cohorts in the two countries, which may otherwise confound our estimates.

We run the following consumer demand model in each of the two countries:

$$w_{ik} = \alpha_{ik} + \sum_{k}^{G} + \gamma_k \ln p_k \beta_k + \ln\left(\frac{x_i}{a_i(p)}\right) + \theta_k + \ln\left(\frac{x_i}{(a_i(p))}\right)^2 \tag{B.1}$$

where $w_{ik}$ is the budget share of individual $i$ for each of the $G$ goods $k$, $p_k$ is the price of good $k$ and $x_i$ is total expenditure on the goods included in the demands system by individual $i$. There are $M$ demographic variables $z_{mi}$ for each individual $i$ including housing, employment, health and mortality are included in $\alpha_{ik}$

$$\alpha_{ik} = \alpha_{k0} + \sum_{k}^{M} \alpha_{mk} z_{mi} \tag{B.2}$$

Expenditures are deflated using the price index

$$\ln \alpha_i(p) = \alpha_O + \sum_{k}^{G} \alpha_{ik} \ln p_k + 1/2 \sum_{l}^{G} \sum_{k}^{G} \gamma_{lk} \ln p_l \ln p_k \tag{B.3}$$

This model differs slightly from the Almost Ideal specification ofDeaton and Muellbauer (1980) in that it includes an additional quadratic term on income (although it is still only an approximation to the fully integrable QUAIDS model (Banks et al. (1997)). Our interest is in establishing the nature of within-period non-separabilities between consumption and housing, health and employment in the two countries through the effect of these variables on household budget shares. By including total expenditure and prices, we control for differences in trends in relative prices and wealth across different birth cohorts in the two countries which may otherwise confound our estimates. The use of the household specific price index $\alpha_i(p)$ means that income deflators can vary across groups according to their differing consumption patterns.

Prices for each of our categories are computed from the individual components and sub-indices of the UK Retail Price Index and the US CPI, which go back to 1978 and 1988 respectively.[1] Typically, sub-indices are not available for the particular category grouping we use (defined in Table 4.1). For instance, in the UK RPI

[1] The authors are grateful to Brendan Williams of BLS for constructing price indices that go back to this date.

**Figure B.1:** Real price of medical spending


Notes: UK prices are a weighted geometric mean of the RPI categories that include medical spending. US prices are a weighted geometric mean of prices for medical care and hospital services. In both calculations the weights are shares of total medical expenditure. Prices are made real with a Stone price index for total non-durable spending.

medical costs are split between "personal services," "chemists goods," "personal articles" and other categories. So in order to calculate price indices for these goods we calculate a Stone price index for a given category $k$

$$p_k = \exp(\sum_{j=1}^{G_k} w_{jk} \ln p_j) \tag{B.4}$$

where $w_{jk}$ is the cohort-year budget share of good $j$ within some spending category $k$ for which there are $G_k$ goods in category $k$ for which we want a price (e.g. "other non-durables"). We plot the estimated series for medical costs in Figure B.1. This shows that real medical prices tended to increase faster in the US than they did in the UK over the period we are considering. This implies that the growth in medical consumption in the US may not have been as large relative to the UK as the growth in medical expenditures.

Our demand system includes sex, number of children, number of adults, and linear and quadratic time trends as controls in all models reported below. We also include dummies for being over state pension age in the UK (60 for women, 65 for men) and for being over 65 in the US. These are included to control for the effects of Medicare (to which US households become eligible at 65) and benefits such as free-prescriptions, the Winter Fuel Payment, and transport subsidies which UK households become eligible for at state pension age. We do not otherwise control for age - our view is that age is usually included as a proxy for health and mortality effects, and these are affects that we are directly interested in (and include separately).

The health and mortality variables are cell averages for the population (by age, year and sex) based on the data we described in Section 4.4 above. We instrument the expenditure and expenditure squared variables using income and income squared (dummying out changes in the income question in the CEX that occurred from the 2nd quarter of 2001 - introducing a bracketing question for those who failed to report their incomes - and income imputation which was introduced in 2004).

The coefficients on the taste shifters, $\alpha_{mk}$ are shown in Table B.2. The particular specification of the demographic variables, $z$, includes: (1) housing tenure with dummy variables for being a renter and housing owners with no mortgage so that the reference group are owners with remaining mortgages; (2) marital status represented a dummy variables for being single; (3) employment proxied by two dummies - household head employed and both partners working; (4) the log of mortality of the head (5) the health of head captured by the proportion of individuals in their cohort who have the worst health status.

In both countries, the demand system results show that those who rent not surprisingly spend a much smaller share of the budget on housing related expenditures. In the US the share spent on housing related expenses is 10 percentage points lower for renters than those who own. In the UK the equivalent number is 4 percentage points. The estimates in Table B.2 indicate renters consequently devote higher shares to all other goods (except food at home in the US), with a particularly large effect for other non-durable spending. Owning a home outright (compared to owners who still have a mortgage to pay off) leads to small reduction in housing related expenses in both countries (though the effect is not significant in the UK).

**Table B.2:** Estimated demand system coefficients $\alpha_{mk}$

| | Food in | Food out | OthND | Medical | Housing | Recrea | Transport |
|---|---|---|---|---|---|---|---|
| | | | | **UK (1978-2010)** | | | |
| *Mean Budget Shares (percent):* | | | | | | | |
| | 24.36 | 4.97 | 25.10 | 1.88 | 23.82 | 7.29 | 12.58 |
| Single | -6.63 | 2.62 | 3.34 | -0.26 | -2.39 | -0.08 | 3.40 |
| | (0.12) | (0.07) | (0.16) | (0.06) | (0.14) | (0.15) | (0.13) |
| Renter | 0.85 | 0.38 | 3.26 | 0.01 | -4.17 | 0.43 | -0.77 |
| | (0.09) | (0.05) | (0.12) | (0.05) | (0.11) | (0.11) | (0.10) |
| Own-outright | 0.14 | -0.11 | -0.75 | 0.16 | -0.04 | 0.92 | -0.31 |
| | (0.08) | (0.05) | (0.11) | (0.04) | (0.10) | (0.10) | (0.09) |
| Head-empl. | -0.03 | 0.61 | -0.42 | -0.05 | -0.81 | -0.40 | 1.11 |
| | (0.09) | (0.05) | (0.12) | (0.05) | (0.10) | (0.11) | (0.10) |
| Both work | -0.67 | 0.23 | 0.71 | -0.14 | -0.47 | 0.43 | -0.10 |
| | (0.09) | (0.06) | (0.13) | (0.05) | (0.11) | (0.12) | (0.11) |
| ln(mortality) | 0.85 | -0.00 | -1.81 | 0.36 | 0.78 | 0.17 | -0.35 |
| | (0.06) | (0.04) | (0.09) | (0.04) | (0.07) | (0.08) | (0.07) |
| Worst health | -0.53 | -0.29 | -0.16 | 0.28 | -1.12 | 0.98 | 0.84 |
| | (0.46) | (0.28) | (0.63) | (0.26) | (0.55) | (0.55) | (0.48) |
| | | | | **US (1988-2010)** | | | |
| *Mean Budget Shares (percent):* | | | | | | | |
| | 22.00 | 6.50 | 17.82 | 12.61 | 19.99 | 4.26 | 16.81 |
| Single | -4.74 | 3.02 | 2.17 | -3.57 | -0.54 | 1.33 | 2.34 |
| | (0.24) | (0.15) | (0.24) | (0.35) | (0.24) | (0.13) | (0.22) |
| Renter | -0.28 | 2.14 | 4.86 | 0.83 | -10.32 | 1.45 | 1.34 |
| | (0.26) | (0.17) | (0.26) | (0.37) | (0.26) | (0.14) | (0.24) |
| Own-outright | 0.08 | 0.53 | -0.71 | 0.57 | -0.33 | 0.09 | -0.25 |
| | (0.13) | (0.08) | (0.13) | (0.20) | (0.13) | (0.07) | (0.12) |
| Head-empl. | 0.96 | 0.46 | -0.60 | -1.68 | 0.10 | -0.32 | 1.05 |
| | (0.18) | (0.11) | (0.18) | (0.26) | (0.18) | (0.09) | (0.16) |
| Both work | -2.18 | 0.45 | 1.33 | -0.09 | -0.90 | 0.31 | 1.09 |
| | (0.18) | (0.11) | (0.18) | (0.26) | (0.18) | (0.09) | (0.17) |
| ln(mortality) | -0.53 | -0.45 | -1.80 | 2.72 | 0.99 | -0.12 | -0.81 |
| | (0.14) | (0.08) | (0.13) | (0.20) | (0.14) | (0.07) | (0.12) |
| Worst health | -0.19 | -0.30 | 2.28 | -0.30 | -1.90 | -0.89 | 1.32 |
| | (0.65) | (0.39) | (0.63) | (0.94) | (0.63) | (0.33) | (0.58) |

Notes: UK N=99,425; US N= 50,796, standard errors in parentheses. We take only data from the first interview in the CEX. Additional controls for log expenditure, log expenditure squared, number of children, number of adults, dummy for whether head or spouse has compulsory education, a quadratic time trend, being over state pension age and self-reported health missing. Expenditure is instrumented using income (with additional dummies in US model for year greater than 2001 and year greater than 2004, when changes to the survey income questions were introduced).

Employment effects look as expected in both countries when the head is employed less is spent on recreation and more is spent on food out and on transport, which is most likely associated with transport to work. Employment in the United States is associated with more food consumption both in and out of the home, but in the UK there is a substitution of food consumption to out of the home. When both head and spouse are working, there is a reduction in spending on food at home in the US.

Important differences emerge in the relationship between employment and health costs, however. In the United States where people bear more of the responsibility for paying their medical costs, the heads employment reduces out-of-pocket medical expenses, a much larger effect than in the UK where the effect is essentially zero. Although this could partly be explained by incomplete controls for health in the model, the key difference is the association between medical insurance and being in a job in the United States (as reflected in Figure 4.7). In the US, the head being employed reduces the proportion spent on medical spending by 1.7 percentage points but there is no similar effect in the UK. This could reflect employers meeting some healthcare costs for their employees in the US (which in the UK would be met by the state). Whether the spouse works or not, does not appear to contribute to this effect.[2]

Due to the data limitations described above, our mortality and subjective health measures capture variations in health status that occur on average at the cohort level rather than individual level variation. A higher risk of mortality among the cohort increases medical spending in both the US and UK with, perhaps unsurprisingly in light of the differential financing of medical care in the two countries, a much larger effect in the US. In the UK, reductions in subjective health controlling for mortality have little effect on the composition of total household consumption (except for a reduction in spending away from home). In contrast, a worsening of

[2]When we exclude health insurance spending from medical expenditures the estimated impact of employment on the share of spending on medical in the US goes from -1.7 to -0.9. The value of this coefficient may seem surprising given discounts for employer-sponsored insurance, However, those leaving employment may still be covered by third parties as Figure 4.7 suggests (e.g. through retiree benefit plans).

the cohorts subjective health status in the United States leads to an apparent (but statistically insignificant) reduction in medical expenses once the effects of mortality are controlled. This likely reflects some difference in health spending among cohorts that we have not been able to control for (for instance, those caused by institutional changes in Medicare coverage or changes in the availability of expensive, technology-intensive health services over time).

Comparing the positive impact of mortality probabilities on medical spending with the zero or negative effects for self-reported health suggests that subjective measures of health may not improve even when objective measures of health do. One possible explanation for this is that people assess their health relative to others in their cohort (so self-reported health status would tend to vary within but not between cohorts), weakening its association with actual health conditions and so medical expenditures.

In Table B.3 we show coefficients from demand system excluding medical spending, additionally controlling for the quantity of medical consumption (defined as the volume of medical spending or expenditure divided by price). This is a model of conditional demands (using the language of Pollak (1969)), allowing us to test for the presence of non-separabilties in medical consumption over and above those associated with ill-health. In both countries, much medical consumption is publicly provided and can only be obtained in rationed quantities. As a result, we instrument medical consumption with its price. To make our results easier to interpret, we also scale medical expenses by their standard deviation in both countries. We use a test of the significance of the medical quantity term to test the hypothesis of separability between medical and other demands. In both countries, we find evidence of non-separability. In the US, higher medical quantities are associated with significantly lower spending on other non-durables and recreation. Our results imply a one standard deviation increase in medical quantities in the US is associated with an increase in the US budget share on housing related goods by around 7 percentage points. The direction of effects for the UK are similar to those for the US, except that higher medical consumption in the UK is associated with lower (rather than higher) spending on housing and with higher (rather than lower) recreation spending. However, the latter of these effects is not significant in the UK.

## B.3 Coverage of household surveys

Comparisons of both the LCFS and the CEX to the aggregate National Income and Product Accounts (NIPA) in the respective countries have highlighted the possibility of increasing measurement error over time in the two surveys. It is now well-documented that coverage rates (the proportion of consumer expenditure in the national accounts that is accounted for by the household surveys) have been declining in both the US and UK (see for example Passero et al. (2015) and Attanasio et al. (2006)). This potentially has consequences for our estimates of consumption growth. In this appendix we compare trends in coverage rates for the two countries to understand better what the implications of this might be.

Any comparison of national account and survey data must take into account the fact that the two sources of information measure different spending concepts. For example, the two sources cover different populations. Both the LCFS and the CEX surveys exclude foreign residents and those in institutional residences whose spending is included in NIPA. In addition, some items of spending that may be thought of as taxes are included as expenditures in surveys but are counted as transfers rather than expenditures in the NIPA. Finally, there are items for which the definitions of spending differ. For example, the NIPA impute rental costs to owner-occupiers while not including the outgoings on for example mortgage interest payments. In the US spending on healthcare made on behalf of households by employers and the government (including the Medicare and Medicaid programs) are also counted as household spending in the NIPA but are not counted in the CEX. Many of these measurement differences might plausibly be thought to have been increasing over time, perhaps differentially so in our two countries. In what follows, we calculate coverage rates after first making adjustments to both our survey data and to the NIPA to make them more comparable. We start by removing spending by non-profit institutions on households behalf from the personal consumption expenditures in both the

**Table B.3:** Estimated conditional demand system coefficients $\alpha_{mk}$ (no medical)

| | Food in | Food out | OthND | Housing | Recrea | Transport |
|---|---|---|---|---|---|---|
| | | | **UK (1978-2010)** | | | |
| *Mean Budget Shares (percent):* | | | | | | |
| | 24.81 | 5.07 | 25.55 | 24.29 | 7.44 | 12.83 |
| Single | -8.50 | 6.71 | 1.21 | -3.37 | 0.01 | 3.79 |
| | (0.41) | (0.68) | (0.51) | (0.37) | (0.23) | (0.21) |
| Renter | 1.45 | -0.89 | 3.91 | -3.89 | 0.36 | -0.88 |
| | (0.22) | (0.34) | (0.28) | (0.15) | (0.13) | (0.12) |
| Own-outright | 1.09 | -2.09 | 0.25 | 0.50 | 0.84 | -0.51 |
| | (0.24) | (0.38) | (0.30) | (0.19) | (0.13) | (0.13) |
| Head-empl. | -0.08 | 0.74 | -0.49 | -0.89 | -0.39 | 1.10 |
| | (0.19) | (0.28) | (0.25) | (0.11) | (0.11) | (0.11) |
| Both work | -2.11 | 3.41 | -0.88 | -1.31 | 0.57 | 0.21 |
| | (0.32) | (0.54) | (0.41) | (0.29) | (0.18) | (0.17) |
| ln(mortality) | 2.30 | -2.99 | -0.08 | 1.51 | 0.01 | -0.64 |
| | (0.30) | (0.50) | (0.37) | (0.28) | (0.17) | (0.15) |
| Worst health | -0.68 | -0.23 | -1.27 | -0.64 | 1.46 | 1.36 |
| | (1.07) | (1.55) | (1.40) | (0.61) | (0.62) | (0.62) |
| Medical quantity | -24.66 | 59.62 | -30.00 | -14.94 | 1.66 | 5.98 |
| | (4.73) | (8.44) | (5.71) | (5.09) | (2.52) | (2.16) |
| | | | **US (1988-2010)** | | | |
| *Mean Budget Shares (percent):* | | | | | | |
| | 25.28 | 7.38 | 20.31 | 23.07 | 4.80 | 19.15 |
| Single | -6.33 | 3.12 | 0.68 | 0.38 | -0.41 | 2.50 |
| | (0.52) | (0.31) | (0.42) | (0.47) | (0.30) | (0.38) |
| Renter | 0.10 | 2.46 | 5.66 | -11.75 | 1.67 | 1.91 |
| | (0.30) | (0.19) | (0.27) | (0.31) | (0.19) | (0.27) |
| Own-outright | 0.18 | 0.65 | -0.51 | -0.55 | 0.54 | -0.30 |
| | (0.18) | (0.11) | (0.15) | (0.18) | (0.11) | (0.15) |
| Head-empl. | 0.60 | 0.41 | -1.31 | 0.17 | -0.67 | 0.76 |
| | (0.22) | (0.14) | (0.20) | (0.23) | (0.14) | (0.19) |
| Both work | -2.78 | 0.57 | 1.30 | -0.35 | -0.39 | 1.64 |
| | (0.29) | (0.17) | (0.24) | (0.28) | (0.17) | (0.23) |
| ln(mortality) | 0.13 | -0.34 | -0.62 | 0.42 | 1.03 | -0.56 |
| | (0.34) | (0.21) | (0.27) | (0.30) | (0.21) | (0.23) |
| Worst health | -0.43 | -0.23 | 2.23 | -1.79 | -1.61 | 1.84 |
| | (0.71) | (0.43) | (0.66) | (0.77) | (0.45) | (0.66) |
| Medical quantity | -2.18 | 1.10 | -2.99 | 6.98 | -5.28 | 2.03 |
| | (1.75) | (1.08) | (1.31) | (1.47) | (1.04) | (1.08) |

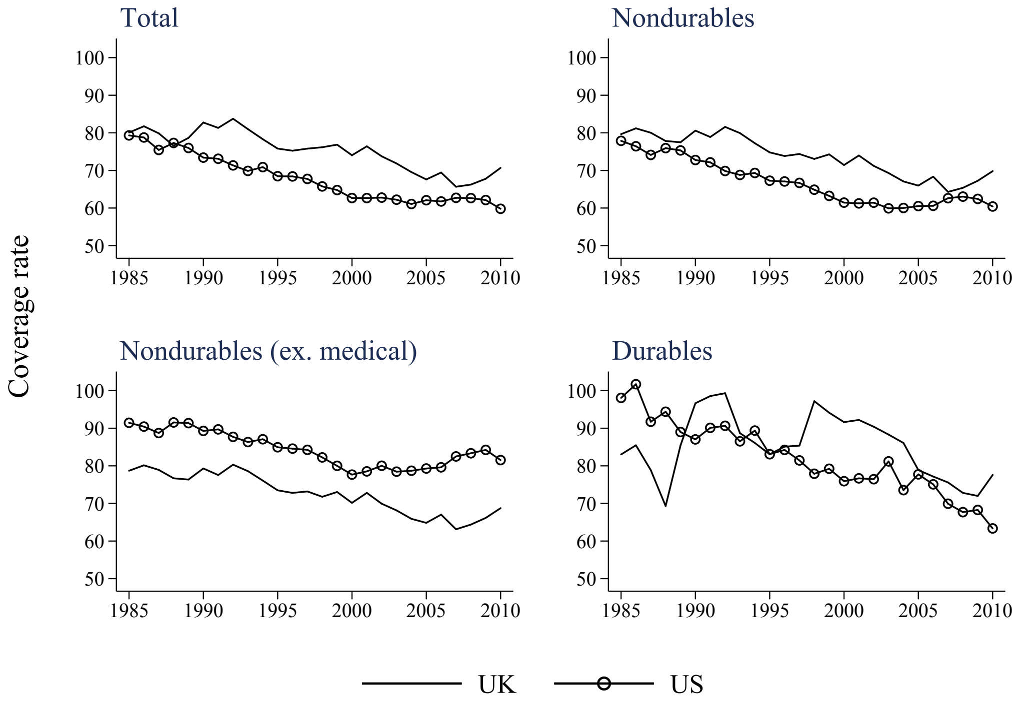
Notes: UK N=99,425; US N= 50,796, standard errors in parentheses. We take only data from the first interview in the CEX. Additional controls for log expenditure, log expenditure squared, number of children, number of adults, dummy for whether head or spouse has compulsory education, a quadratic time trend, being over state pension age and self-reported health missing. Expenditure is instrumented using income (with additional dummies in US model for year greater than 2001 and year greater than 2004, when changes to the survey income questions were introduced). Medical quantity is instrumented with its log price and scaled by its standard deviation in both countries.

UK and the US. We then exclude spending on imputed rent to owner-occupiers in the NIPA. In our surveys we remove the costs of mortgage interest, vehicle licensing costs, property taxes and (in the UK) TV licenses. We also show the consequences of removing health spending from both sources.

Figure B.2 plots the coverage rates for total expenditure, non-durables, non-durables excluding medical expenditures and durables. The first thing to notice is that there is still evidence of a steady decline in coverage in both countries. The top left panel shows coverage rates for total spending (including medical) which decline faster in the UK than the US. These fall from 80 percent to 71 percent in the UK over the period 1985-2010 compared to a fall from 80 percent to 60 percent in the US. A decline in coverage of this magnitude would reduce annual spending growth as measured in surveys by around 0.5 percentage points in the UK compared to 1.2 percentage points in the US.

The coverage rates of non-durable spending, which is the definition of spending examined in this paper, decline at similar rates. However, when we remove health spending in the bottom left panel, the picture is very different. Coverage rates are now higher in the US (where they fall from 91 percent to 82 percent) than the UK (where the fall is from 79 percent to 69 percent). The implied falls would now suggest a slightly larger understatement of spending growth in the UK (by 0.54 percentage points compared to 0.46 percentage points in the US). The difference that arises from excluding healthcare reflects the rapid growth of medical spending on US households by government and employers. As mentioned above, these expenditures are not included as household spending in the CEX survey but are included in the US NIPA. Passero et al. (2015) estimate that spending by government on behalf of households in the US increased by 271 percent from 1992 to 2010 and that this accounts for one fourth of the growth in the gap between the coverage of the CEX survey and NIPA consumption spending. An additional proportion is likely to be explained by growth in the proportion of health costs paid by employers. In the UK spending on the NHS is not attributed to households in the national accounts in the same way, and employer coverage is much less widespread. As a result, excluding

**Figure B.2:** Coverage rates, 1985-2010




Notes: Coverage rate is the proportion of consumer expenditure in the national accounts that is accounted for in the household surveys. Household survey data comes from the LCFS in the UK and the CEX in the US. National Income and Product Account (NIPA) data comes from the UK Office for National Statistics and the US Bureau of Economic Analysis.

health spending has a much smaller effect on coverage rates in the UK. Durable spending in our household surveys has higher rates of coverage in both countries. The CEX accounted for roughly 100 percent of the durable spending in the national accounts by our measure in the US in 1985. This fell to just 63 percent in 2010. In the UK the decline was from 83 percent to 77 percent over the same period.

## B.4 Additional consumption growth regressions

**Table B.4:** Changes in log non-durable expenditure

| | Including Medical Expenditure | Excluding Medical Expenditure | |
|---|---|---|---|
| | (1) | (2) | (3) |
| US | -0.009 | -0.010 | -0.028 |
| | (0.016) | (0.016) | (0.019) |
| UK | -0.018 | -0.023 | -0.034 |
| | (0.017) | (0.017) | (0.017) |
| Interest rate | 0.199 | 0.318 | 0.313 |
| | (0.155) | (0.160) | (0.140) |
| Log Mortality | -0.001 | -0.001 | -0.003 |
| | (0.004) | (0.003) | (0.004) |
| ΔHead employed | -0.068 | -0.156 | -0.117 |
| | (0.181) | (0.180) | (0.154) |
| ΔRenter | -0.029 | -0.073 | -0.163 |
| | (0.205) | (0.201) | (0.167) |
| ΔNumber of kids | -0.043 | -0.057 | -0.042 |
| | (0.059) | (0.059) | (0.051) |
| ΔNumber of adults | 0.248 | 0.237 | 0.233 |
| | (0.041) | (0.039) | (0.034) |
| ΔSingle | -0.303 | -0.275 | -0.275 |
| | (0.080) | (0.078) | (0.068) |
| ΔWorst health | -0.781 | -0.847 | -0.585 |
| | (0.532) | (0.530) | (0.430) |
| ΔLog Medical Price | | -0.510 | -0.470 |
| | | (0.112) | (0.096) |
| $\pi^2_{s,k,t-1}\phi_{s,k,t}$ | | | 0.002 |
| | | | (0.002) |
| (US-UK)100 | 0.905 | 1.304 | 0.558 |
| | (0.551) | (0.611) | (0.658) |
| Hausman endogeneity test (p-value) | 0.204 | 0.336 | 0.629 |
| N | 616 | 616 | 616 |

Notes: Estimates presented are for weighted regressions with weights given by cell sizes in each education-year-cohort cell. The dependent variable is log non-durable consumption (col 1 with medical expenditure, cols 2 and 3 without). Additional controls for switch from GHS to HSE surveys in the UK, change in proportion of households reporting own health in US, and the change in proportion responding to subjective health questions. We instrument employment, renter, health and mortality (and GHS, self-report dummies) with their first and second lags. In column (4) we instrument the conditional risk term $\pi^2_{s,k,t-1}\phi_{s,k,t}$ with its lag value.

# Appendix C

# Appendix for Chapter 5

## C.1 Proofs of chapter results

Proof of proposition 1.

*Proof.* To prove this proposition we need to show that firstly the test in (5.14) is equivalent to GARP. Secondly, we need to show that data that satisfy the test in (5.14) then this implies that they will be rationalised by the life-cycle model in the sense given in definition 1.

*Equivalence of (5.14) and GARP:* Afriat's theorem states that the following two statements are equivalent (see Afriat (1967) or Varian (1982) for a proof):

(i) The data satisfy GARP: if $q_s R q_t$, then $p_t q_t \leq p_t q_s$ for all $s,t$

(ii) There exist 'Afriat numbers' $U_s, U_t, \lambda_t > 0$ such that $U_s \leq U_t + \lambda_t p_t'(q_s - q_t)$ for all $s,t$

Now if condition (ii) is satisfied for some data $p_t, q_t$, then it will also be satisfied for the data $\sigma_t p_t, q_t$ where $\sigma_1, .., \sigma_T$ are any positive numbers. This is because if $U_1 ... U_T$ and $\{\lambda_1 .. \lambda_T\}$ satisfied (ii) in the old data, then $U_1 ... U_T$ and $\{\hat{\lambda}_1 .. \hat{\lambda}_T\}$, where $\hat{\lambda}_t = \lambda_t / \sigma_t$, will satisfy it in the new.

Conversely, if (ii) is satisfied for some $\sigma_t p_t, q_t$, then it will also be satisfied for $p_t, q_t$. Hence, by Afriat's theorem $p_t, q_t$ will satisfy GARP if $\sigma_t p_t, q_t$ satisfy (ii).

If we set $\sigma_t = \beta^{t-1}(1+r_t)(1+r_{t-1})\ldots(1+r_2)$ , then (ii) becomes the test of the life-cycle model in (5.14). Since this both implies and is implied by GARP, we have that *GARP* $\Longleftrightarrow$(5.14).

*Equivalence of the data satisfying (5.14) and the data being rationalised by the life-cycle model (LCM):* That LCM implies (5.14) is demonstrated in the text. Thus, (5.14) is a necessary condition for LCM. It remains to be demonstrated that (5.14) is sufficient for LCM. We can rearrange our test to get

$$\frac{\lambda_t}{\beta^{t-1}}\rho_t'(q_s - q_t) \geq u(q_s) - u(q_t)$$

Let's call $\frac{\lambda_t}{\beta^{t-1}}\rho_t' = \Xi_t$. Across periods this gives us

$$\Xi_t(q_s - q_t) \geq u(q_s) - u(q_t)$$

$$\Xi_s(q_u - q_t) \geq u(q_u) - u(q_s)$$

$$\vdots$$

$$\Xi_u(q_z - q_u) \geq u(q_z) - u(q_u)$$

Adding these up gives

$$\Xi_t(q_s - q_t) + \Xi_s(q_u - q_s) + \ldots + \Xi_u(q_z - q_u) \geq 0$$

This matches what Rockafellar (1997) and Browning (1989) call "cyclical monotonicty". Following Browning (1989) we can then apply Theorem 24.8 in Rockafellar to demonstrate the existence of a concave utility function $u(.)$ such that $u'(q_t) = \Xi_t$ for all $t$. This satisfies the requirement in definition 1, completing the proof. $\square$

# Appendix D

# Appendix for Chapter 6

## D.1 Mortgage imputation in Understanding Society

The BHPS contains data on mortgage values from 1993 (wave 3) onwards, while Understanding Society dropped these variables in its second wave in 2010 except for households who had newly moved. However, in all years of the BHPS and Understanding Society the data contains a great deal of information on household mortgages, including whether households are outright owners, the mortgage type, the value of any additional loans, and the years left to pay on the mortgage. So as to avoid throwing data out unnecessarily, we use this information to impute mortgages for the remaining three waves of Understanding Society.

For those with interest only or 'endowment' mortgages, we assume no principal repayments. In this case, we take the current value of the mortgage to be its lagged value plus any additional loans the household may have taken out since its previous interview. For those with standard repayment mortgages we assume the loan is amortised with annual payments (which consist of both interest and principal) determined by

$$\text{Ann. Payment} = M_{t-1} \times i/(1-(1+i)^{-(\ell+1)}) \tag{D.1}$$

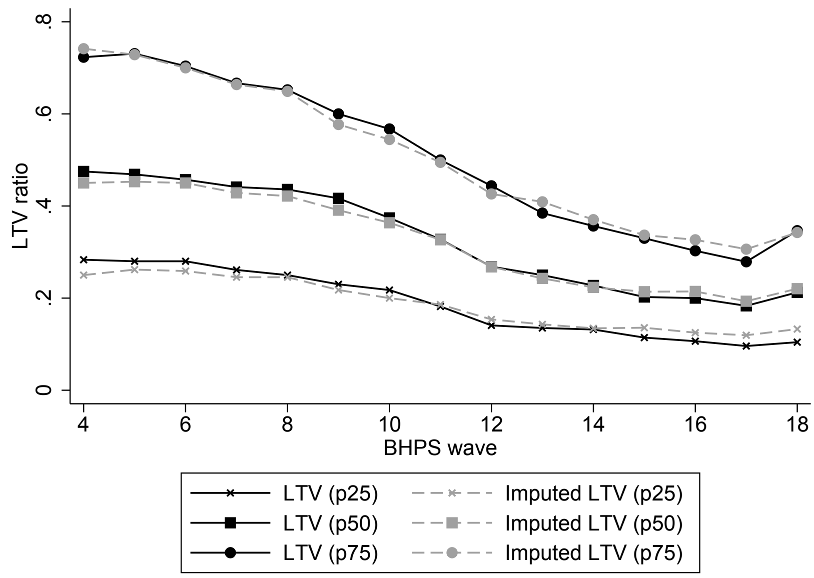
where $M_{t-1}$ is the value of the households mortgage in the previous year, $i$ is the interst rate and $\ell$ is the remaining life of the mortgage. This means that the mortgage in any given period is given by

$$M_t = M_{t-1} - \text{Ann. Payment} + iM_{t-1} + M_t^{new} \tag{D.2}$$

where $M_t^{new}$ is the amount of additional mortgage we observe the household borrowing between periods $t$ and $t-1$.

To assess the accuracy of our imputation procedure, we implemented it on waves of the BHPS for which we observe the true value of households' mortgages. That is we took a set of households observed in the 3rd wave of the BHPS, and imputed their mortgage values for all subsequent waves. We then plot the LTV ratios implied by our imputation procedure against actual values calculated from the survey for different percentiles of the LTV distribution (25th, 50th and 75th). The results of this exercise are shown in Figure D.1. Our imputation procedure appears to work extremely well - accurately predicting households' LTV ratios even after 15 waves.

**Figure D.1:** Imputed and actual LTV values, BHPS




Note: Data from British Household Panel Survey and Understanding Society.

## D.2 LTI constraints and re-leveraging behaviour

To assess how re-leveraging responses vary according to households loan-to-income and loan-to-value ratios, we run regressions of the following form

$$I[\Delta logM_t > 0] = \delta \Delta \log HValue_{it} + \varepsilon_{it} \quad \text{(D.3)}$$

separately for households with LTI ratios falling within different categories, and separately in the BHPS and the PSID. As in Section 6.3, we instrument changes in house values with changes in regional house prices. We use an indicator for whether households *increase* their mortgage debt rather than the actual change in mortgage debt to mitigate problems of endogeneity that come from splitting the sample according to households' debt values. We plot the coefficients on house value changes for different LTI ratios in the left-hand panel of Figure D.2. The right-hand panel shows the proportion of households who fall in each of the LTI categories in the two countries.

The graph shows that in both countries, re-leveraging responses first rise and the fall as households' LTI ratios increase, suggesting that borrowing constraints bind at high LTI ratios. US households appear more likely to re-leverage for a given LTI value than their UK counterparts. UK households are more likely to have high LTI ratios.

## D.3 Two-sample IV

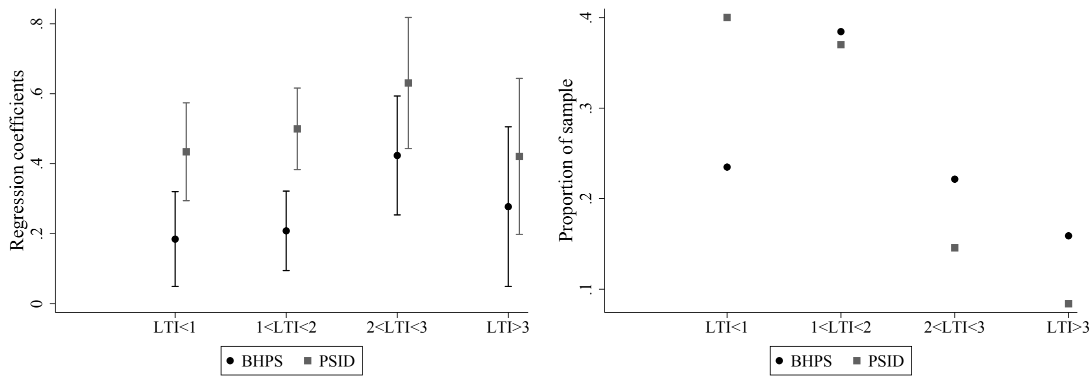
In this paper we make use of Two Sample Two Stage Least Squares (TS2SLS). Inoue and Solon (2010) show that this approach is more efficient than the TSIV estimator of Angrist and Krueger (1992).

TS2SLS is best explained by first considering a standard two-stage least squares (2SLS) approach.

Let $\mathbf{M} = [\ X \quad \omega_{t-1} \quad \omega_{t-1} \times (\frac{p_{rt}}{p_{rt-1}} - 1)\ ]$ denote the $n \times (k+p)$ matrix of right-hand side variables ($p$ of which are endogenous). Suppose we face the problem of consistently estimating the $1 \times (k+p)$ vector of coefficients $\delta$ in the model

**Figure D.2:** Re-leveraging behaviour by LTI ratios

**(a)** Regressions coefficients by LTI category **(b)** Proportion of households in each LTI category




Notes: Data from British Household Panel Survey and Panel Study of Income Dynamics. The left-hand panel shows regression coefficients from a regression of a indicator for increases in mortgage debt over a two-year period on the change in log nominal house values (instrumented with the log change in regional/state house prices).

$$c = \mathbf{M}\delta + e$$

where $\omega_{t-1}$ and $e$ are correlated. It is well known that the coefficients estimated using a naive OLS regression of $c$ on $\mathbf{M}$ will be biased. To solve this problem, instrumental variable methods make use of an $n \times (k+q)$ matrix of instruments $\mathbf{Z}$ where the $p$ endogenous variables in $\mathbf{M}$ are replaced with $q \geq p$ variables that are assumed to be exogenous. This assumption implies that $E[e|\mathbf{Z}] = 0$ and means that $\delta$ can be consistently estimated using the 2SLS estimator

$$\hat{\delta}_{2SLS} = (\hat{\mathbf{M}}'\hat{\mathbf{M}})^{-1}\hat{\mathbf{M}}'c \tag{D.4}$$

where $\hat{\mathbf{M}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{M}$, or the fitted values from the set of reduced form regressions of the columns of $\mathbf{M}$ on $\mathbf{Z}$

$$\mathbf{M} = \mathbf{Z}\Pi + v$$

Notice here that while this estimator requires knowledge of both the cross-products $\mathbf{Z}'\mathbf{M}$ and $\mathbf{Z}'c$ we do not require the cross product $\mathbf{M}'c$. This insight was the basis for two sample IV proposed in Angrist and Krueger (1992).[1] They show that under certain conditions, it is possible to estimate $\delta$ even if no sample can be found that contains data on $\mathbf{M}$, $c$ and $\mathbf{Z}$ simultaneously. All that is required is a sample that includes both $c$ and $\mathbf{Z}$ (but not necessarily the endogenous components of $\mathbf{M}$) and another which includes $\mathbf{Z}$ and $\mathbf{M}$ (but not necessarily $c$) . This allows us to calculate a two sample 2SLS estimator (TS2SLS) that is analagous to (D.4)

$$\hat{\delta}_{TS2SLS} = (\hat{\mathbf{M}}_1{}'\hat{\mathbf{M}}_1)^{-1}\hat{\mathbf{M}}_1{}'c_1 \tag{D.5}$$

where $\hat{\mathbf{M}}_\mathbf{1} = \mathbf{Z}_\mathbf{1}(\mathbf{Z}_\mathbf{2}'\mathbf{Z}_\mathbf{2})^{-1}\mathbf{Z}_\mathbf{2}'\mathbf{M}_\mathbf{2} = \mathbf{Z}_1\hat{\Pi}_2$. Here $c_1$ and $\mathbf{M}_1$ contain $n_1$ observations from the first sample while $\mathbf{M}_2$ and $\mathbf{Z}_2$ contain $n_2$ observations from the second. $\hat{\Pi}_2$ is the coefficient matrix formed from a regression of $\mathbf{M}_2$ on $\mathbf{Z}_2$.

This estimator can be implemented using a simple two step procedure:

1. Run a first stage regression in sample 2 and using the recovered coefficients to impute $\mathbf{M}$ in sample 1.

2. In sample 1, regress $c_1$ on the imputed values of $\mathbf{M}$ to recover $\hat{\delta}_{TS2SLS}$.

We adjust standard errors from our second stage regression to account for the two-step nature of the procedure. Because we cluster observations from the same household in our first stage regression, we use the robust standard error correction for TS2SLS derived in Pacini and Windmeijer (2016).

# D.4 Alternative estimation approaches

## D.4.1 Alternative definitions of residential investment

First we investigate the extent to which our results depend on our chosen measure of residential investment. The measure of residential investment that we use for our main results includes certain white goods such as cookers, refrigerators and washing machines which are often capitalised into property values but which would not

[1] In their orginal article, Angrist and Krueger (1992) in fact proposed orginally an alternative GMM estimator $\hat{\delta}_{IV} = (\mathbf{Z}_2'\mathbf{M}_2/n_2)^{-1}(\mathbf{Z}_1'c_1/n_1)$. Asymptotically this gives identical results to the TS2SLS estimator. However, Inoue and Solon (2010) show these two approaches will in general give different answers in finite samples, and that the TS2SLS is more efficient. This gain in efficiency arises because the latter estimator corrects for differences in the two samples in the distribution of $\mathbf{Z}$

necessarily be considered residential investment spending in for instance a national accounting framework. Here we examine the extent to which our results are robust to the removal of these items by restricting out definition to goods such as electric tools, floor coverings and the costs of installing or repairing heating and air conditioning units (along with spending on household extensions).

We show results using these alternative measures in Table D.1. Column (1) shows results using the inverse hyperbolic sine transformation of our narrower residential investment measure. The effects of increases in prices for more leveraged households are still large and statistically significant (and indeed very similar to those obtained in our main results). In Column (2) we show results from a linear probability model in which the dependent variable takes a value of 1 if the household is observed spending a positive amount on household extensions. This is probably the purest measure of residential investment in that it only includes structural modifications to the home. Again we find that the investment spending of more leveraged households is significantly more responsive to house price changes than the spending of other home-owners. A 10% increase in local house prices is associated with a 2 percentage point increase in the probability that a household with a 50% LTV ratio builds an extension in a given region, compared to a 1 percentage point increase for outright owners.

**Table D.1:** Results with alternative definitions of residential investment

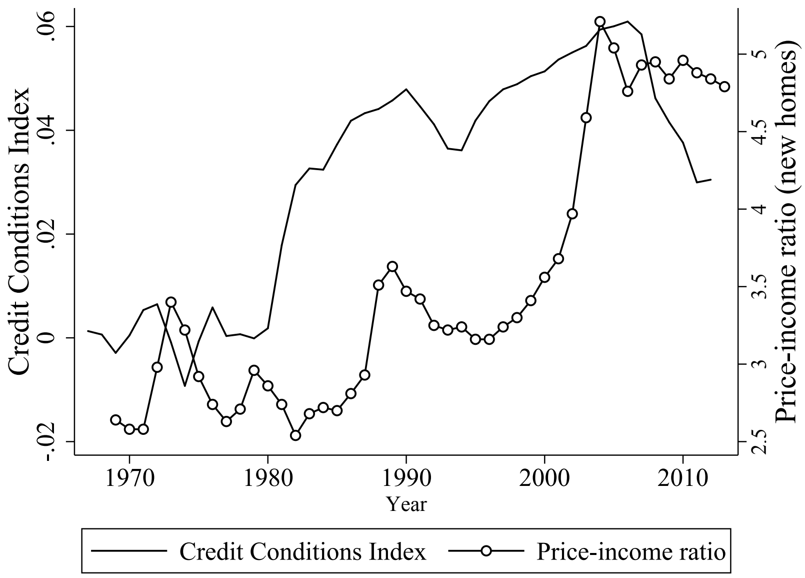
| | Narrow Res inv. | Extensions>0 |
|---|---|---|
| | (1) | (2) |
| $\omega_{it-1}$ | -0.045 | -0.010 |
| | (0.047) | (0.010) |
| $\omega_{it-1} \times (\frac{p_{rt}}{p_{rt-1}} - 1)$ | 0.725*** | 0.100* |
| | (0.278) | (0.056) |
| $R^2$ | 0.081 | 0.039 |
| N | 60,357 | 60,357 |

Notes: * $p < 0.10$ , ** $p < 0.05$ , *** $p < 0.01$. Standard errors in parentheses. See table 6.5 for list of controls.

## D.4.2 Alternative instruments

In this section, we consider how our results are affected when we use two alternative instruments in place of the price-to-income ratio at the time individuals moved into their current residences.

**Figure D.3:** Credit Conditions Index vs price-income ratio




Note: The construction of the Credit Conditions Index is described in Fernandez-Corugedo and Muellbauer (2006).

The first of these is the Credit Conditions Index (CCI) assembled in Fernandez-Corugedo and Muellbauer (2006). This index contains 10 indicators of credit conditions. Two are aggregate measures of unsecured and mortgage debts. The remaining 8 are fractions of mortgages for first time buyers that are above given loan-to-value and loan-to-income ratios for different age groups and regions. The index is constructed controlling for various determinants of credit demand to ensure the index reflects credit supply conditions.[2] The series is plotted alongside our instrument in Figure D.3. The CCI shows a discontinuous increase in 1981. Because this is not

[2]These controls are: nominal and real interest rates, a measure of interest rate expectations and of inflation and interest rate volatility, mortgage and housing return, 36 risk indicators, house prices, income, a proxy for expected income growth, the change in the unemployment rate, demography, consumer confidence, portfolio wealth components, proxies for sample selection bias and institutional features.

matched by a similarly discontinuous increase in leverage for those moving in these years in our sample, when we include households who moved before this date we find the instrument to be weak and our results imprecise. The first two columns of Table D.2 present results for log total spending and residential investment (conditional on moving in 1981 or after). The results are very similar to what we obtain in our main specification, with the implied elasticity much greater for residential than other forms of spending.

The second alternative instrument we consider is the average regional price at the point homeowners moved into their homes. This makes use of interregional variation as well as intertemporal variation in house prices. We report results for this approach in Table D.2. We find that they are again very similar to our main results.

**Table D.2:** Results with alternative instruments

| | CCI | | Reg. house prices | |
|---|---|---|---|---|
| | Total | Res. inv | Total | Res. inv |
| | (1) | (2) | (3) | (4) |
| $\omega_{it-1}$ | -0.027 | -0.135 | 0.034** | 0.104* |
| | (0.036) | (0.144) | (0.015) | (0.058) |
| $\omega_{it-1} \times (\frac{p_{rt}}{p_{rt-1}} - 1)$ | 0.097 | 0.847* | 0.156* | 0.829*** |
| | (0.112) | (0.436) | (0.081) | (0.315) |
| **Instruments:** | | | | |
| $CCI_{-T}$ and $CCI_{-T} \times (\frac{p_{rt}}{p_{rt-1}} - 1)$ | x | x | | |
| $P_{-rT}$ and $P_{-rT} \times (\frac{p_{rt}}{p_{rt-1}} - 1)$ | | | x | x |
| $R^2$ | 0.353 | 0.087 | 0.355 | 0.082 |
| N | 52,155 | 52,155 | 60,357 | 60,357 |

Notes: * $p < 0.10$ , ** $p < 0.05$ , *** $p < 0.01$. Standard errors in parentheses.
See table 6.5 for list of controls.

### D.4.3 Alternative sample

A further concern might that our results for more leveraged households are driven entirely by households who have just moved into their homes (and are thus more likely to be at a credit constraint). Since price-to-income ratios have tended to increase over time, our first stage regressions will tend to predict higher rates of leverage for more recent movers.

In our main results, we exclude households who moved into their homes in the previous year only. In Table D.3 we consider how our results are affected when we exclude households who moved into their homes within the previous five years. The results from this exercise are remarkably similar to our main set of results.

**Table D.3:** Log spending responses

| | Total | Non-durables | Durables (IHS) | Res inv. (IHS) | Luxuries (IHS) | Total - Res |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $\omega_{it-1}$ | -0.026** | -0.016 | -0.053 | -0.063 | -0.032 | -0.023* |
| | (0.013) | (0.012) | (0.049) | (0.050) | (0.026) | (0.013) |
| $\omega_{it-1} \times (\frac{p_{rt}}{p_{rt-1}} - 1)$ | -0.006 | -0.001 | -0.320 | 0.898** | 0.128 | -0.014 |
| | (0.120) | (0.107) | (0.444) | (0.458) | (0.233) | (0.116) |
| $R^2$ | 0.378 | 0.405 | 0.122 | 0.079 | 0.235 | 0.387 |
| N | 42,285 | 42,285 | 42,285 | 42,285 | 42,285 | 42,285 |

Notes: * $p < 0.10$ , ** $p < 0.05$ , *** $p < 0.01$. Standard errors in parentheses. See table 6.5 for list of controls.

# Appendix E

# Appendix for Chapter 7

## Appendix to Section 7.1: Derivation of elasticities

### Marshallian and Hicksian elasticities

For convenience we reproduce the consumer's budget constraints and MRS condition:

$$y_t = \left(A_{h,t} + y^m_{h,t} - F\left(a_{h,t}\right) P_{h,t}\right) - \frac{A_{h,t+1}}{1+r_{t+1}} \tag{E.1}$$

$$c_{h,t} + w^f_{h,t} l_{h,t} = y_{h,t} + w^f_{h,t} L \tag{E.2}$$

$$w^f_{h,t} = \frac{u_{l_{h,t}}}{u_{c_{h,t}}} = \alpha_{h,t} \frac{l^{-\theta}_{h,t}}{c^{-\phi}_{h,t}} \tag{E.3}$$

Taking the derivative of the budget constraint and the MRS equation and stacking them gives a matrix equation (omitting subscripts and superscripts for convenience):

$$\begin{bmatrix} 1 & \frac{wl}{c} \\ \phi & -\theta \end{bmatrix} \begin{bmatrix} \frac{\partial \ln c}{\partial \ln w} \\ \frac{\partial \ln l}{\partial \ln w} \end{bmatrix} = \begin{bmatrix} \frac{w(L-l)}{c} \\ 1 \end{bmatrix}$$

This can be inverted to give the Marshallian elasticities in the main text (equation 7.15):

$$\varepsilon_c^M = \frac{\partial \ln c}{\partial \ln w} = \frac{\theta w(L-l)+wl}{\theta c+\phi wl} \tag{E.4}$$

$$\varepsilon_l^M = \frac{\partial \ln l}{\partial \ln w} = \frac{\phi w(L-l)-c}{\theta c+\phi wl} \tag{E.5}$$

$$\varepsilon_h^M = \frac{\partial \ln h}{\partial \ln w} = -\left(\frac{\phi w(L-l)-c}{\theta c+\phi wl}\right)\frac{l}{L-l} \tag{E.6}$$

To calculate the Hicksian elasticities, we first calculate the income elasticities by differentiating the MRS equation and the budget constraint with respect to income:

$$\varepsilon_c^y = \frac{\partial \ln c}{\partial \ln y} = \frac{\theta y}{\theta c+\phi wl} \tag{E.7}$$

$$\varepsilon_l^y = \frac{\partial \ln l}{\partial \ln y} = \frac{\phi y}{\theta c+\phi wl} \tag{E.8}$$

The income elasticity and the marshallian elasticity are then used to calculate the Hicksian elasticity using the Slutsky equation:

$$\begin{aligned}
\varepsilon_c^H &= \varepsilon_c^M + \frac{\partial \ln c}{\partial \ln y}\frac{wl}{(c+wl)} \\
\varepsilon_l^H &= \varepsilon_l^M - \frac{\partial \ln l}{\partial \ln y}\frac{w(L-l)}{(c+wl)}
\end{aligned}$$

## Frisch Elasticities

In this section we provide the formulae for the first and second derivatives that are used to calculate the different elasticities. We define $D = \exp(\pi z + \xi P + \zeta)$ (omitting subscripts for convenience). Then it is easy to show that:

$$u_c(c,l) = DM^{-\gamma}c^{-\phi} \tag{E.9}$$

$$u_l(c,l) = D\alpha M^{-\gamma}l^{-\theta} \tag{E.10}$$

$$u_{cl}(c,l) = (-\gamma) D M^{-\gamma-1} \alpha c^{-\phi} l^{-\theta} \tag{E.11}$$

$$u_{ll}(c,l) = (-\gamma) \frac{u_l(c,l)}{\alpha M} l^{-\theta} - u_l(c,l)\, \theta l^{-1} \tag{E.12}$$

$$u_{cc}(c,l) = (-\gamma) \frac{u_c(c,l)}{M} c^{-\phi} - u_c(c,l)\, \phi c^{-1} \tag{E.13}$$

Finally, note that:

$$u_{cl}(c,l) = (-\gamma) u_c(c,l)\, l^{-\theta} \frac{\alpha}{M} = (-\gamma) u_l(c,l)\, c^{-\phi} \frac{1}{M} \tag{E.14}$$

These expressions can be used to calculate the Frisch elasticities in the main body of the chapter. The formula for the wage Frisch for intensive margin choices can be derived as follows:

$$\begin{bmatrix} u_{cc} & u_{cl} \\ u_{cl} & u_{ll} \end{bmatrix} \begin{bmatrix} \frac{\partial c}{\partial w} \\ \frac{\partial l}{\partial w} \end{bmatrix} = \begin{bmatrix} 0 \\ u_c \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial c}{\partial w} \\ \frac{\partial l}{\partial w} \end{bmatrix} = \begin{bmatrix} u_{cc} & u_{cl} \\ u_{cl} & u_{ll} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ u_c \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial c}{\partial w} \\ \frac{\partial l}{\partial w} \end{bmatrix} = \frac{1}{u_{cc}u_{ll} - u_{cl}^2} \begin{bmatrix} u_{ll} & -u_{cl} \\ -u_{cl} & u_{cc} \end{bmatrix} \begin{bmatrix} 0 \\ u_c \end{bmatrix}$$

$$\varepsilon_c^F = \frac{w}{c} \frac{\partial c}{\partial w} = -\frac{u_c u_{cl}}{u_{cc}u_{ll} - u_{cl}^2} \frac{w}{c} \tag{E.15}$$

$$\varepsilon_l^F = \frac{w}{l} \frac{\partial l}{\partial w} = \frac{u_c u_{cc}}{u_{cc}u_{ll} - u_{cl}^2} \frac{w}{l} \tag{E.16}$$

$$\varepsilon_h^F = \frac{w}{L-l} \frac{\partial (L-l)}{\partial l} \frac{\partial l}{\partial w} = -\frac{u_c u_{cc}}{u_{cc}u_{ll} - u_{cl}^2} \frac{w}{L-l} = -\varepsilon_l \frac{l}{L-l} \tag{E.17}$$

The formula for the interest-rate Frisch can similarly be derived as follows:

$$\begin{bmatrix} u_{cc} & u_{cl} \\ u_{cl} & u_{ll} \end{bmatrix} \begin{bmatrix} \frac{\partial c}{\partial(1+R_{t+1})} \\ \frac{\partial l}{\partial(1+R_{t+1})} \end{bmatrix} = \begin{bmatrix} u_c \\ u_l \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial c}{\partial(1+R_{t+1})} \\ \frac{\partial l}{\partial(1+R_{t+1})} \end{bmatrix} = \begin{bmatrix} u_{cc} & u_{cl} \\ u_{cl} & u_{ll} \end{bmatrix}^{-1} \begin{bmatrix} u_c \\ u_l \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial c}{\partial(1+R_{t+1})} \\ \frac{\partial l}{\partial(1+R_{t+1})} \end{bmatrix} = \frac{1}{u_{cc}u_{ll} - u_{cl}^2} \begin{bmatrix} u_{ll} & -u_{cl} \\ -u_{cl} & u_{cc} \end{bmatrix} \begin{bmatrix} u_c \\ u_l \end{bmatrix}$$

$$\varepsilon_c^{F_R} = \frac{(1+R_{t+1})}{c} \frac{\partial c}{\partial(1+R_{t+1})} = \frac{u_c u_{ll} - u_l u_{cl}}{(1+R_{t+1})(u_{cc}u_{ll} - u_{cl}^2)} \frac{1+R_{t+1}}{c} = \frac{u_c u_{ll} - u_l u_{cl}}{c(u_{cc}u_{ll} - u_{cl}^2)} \tag{E.18}$$

$$\varepsilon_l^{F_R} = \frac{(1+R_{t+1})}{l} \frac{\partial l}{\partial(1+R_{t+1})} = \frac{u_l u_{cc} - u_c u_{cl}}{(1+R_{t+1})(u_{cc}u_{ll} - u_{cl}^2)} \frac{1+R_{t+1}}{c} = \frac{u_l u_{cc} - u_c u_{cl}}{c(u_{cc}u_{ll} - u_{cl}^2)} \tag{E.19}$$

$$\varepsilon_h^{F_R} = \frac{(1+R_{t+1})}{L-l} \frac{\partial(L-l)}{\partial l} \frac{\partial l}{\partial(1+R_{t+1})} = -\frac{u_l u_{cc} - u_c u_{cl}}{(1+R_{t+1})(u_{cc}u_{ll} - u_{cl}^2)} \frac{1+R_{t+1}}{L-l} = -\varepsilon_l^{F_R} \frac{l}{L-l} \tag{E.20}$$

## E.1 Appendix to Section 7.2: estimation strategy and solution method

### Fuller's estimator

When estimating the parameters that determine the MRS or those that enter the Euler equation, we use first order conditions to derive restrictions on the data to identify structural parameters. Although these sets of conditions are different, as one set is static in nature and one set is dynamic, they are of a similar nature, in that they can be reduced to an expression of the type

$$E\left[h(X;\theta)\mathcal{Z}\right] = 0 \tag{E.21}$$

where $h(\cdot)$ is a function of data $X$ and parameters, $\theta$, and is linear in the vector of parameters. The vector $\mathcal{Z}$ contains observable variables that will be assumed to be orthogonal to $h$. The nature of the instruments that deliver identification depends on the nature of the residual $h$ and, as we discuss below, is different when we estimate the MRS conditions or the Euler equations. However, in both cases, we exploit a condition such as (E.21).

In equation (E.21), one needs to normalise one of the parameters to 1. In the context of the MRS equation (7.18), for example, we set the coefficient on $\ln w_{h,t}$ to 1, but we could have set the coefficient on $\ln l_{h,t}$, or that on $\ln c_{h,t}$ to be 1. A well-known issue with many estimators in this class is that in small samples they are not necessarily robust to the normalisation used. A number of alternative estimators that avoid this issue are available, ranging from LIML-type estimators, to the estimator discussed in Alonso-Borrego and Arellano (1999), to the iterated GMM proposed by Hansen et al. (1996). We use the estimator proposed by Fuller (1977) to estimate both our MRS and Euler equations. This estimator is a modified version of LIML with an adjustment that is designed to ensure that it has finite moments. Roughly speaking, it can be thought of as a compromise between LIML and 2SLS (being closer to LIML when the sample size is large relative to the number of instruments). While this estimator is not completely normalisation free, it is much less sensitive to the choice of normalisation than estimators such as 2SLS and GMM.

An additional advantage of the Fuller estimator is that it is known to have better bias properties than estimators such as 2SLS, when instruments are relatively weak. In Section 7.2, we test the strength of our instruments comparing the values of the Cragg-Donald test statistic to the relevant entries of the table supplied in Stock and Yogo (2005).[1] For the Fuller estimator that we employ, these critical values are typically lower than those for 2SLS, and, unlike 2SLS, they are decreasing in the number of instruments used.

## Euler equation estimation with repeated cross-sections

We estimate the intertemporal parameters using the Euler equation. We need a long time series because, even under rational expectations, expectations errors do not necessarily average out to zero (or are uncorrelated with available information) in the cross-section, but only in the time series: expectation errors may be correlated with available information in the cross-section in the presence of aggregate shocks. See the discussion in Hayashi (1987), Attanasio (1999), or Attanasio and Weber (2010). We also need to assume that the lagged variables used as instruments are uncorrelated with the innovations to the taste shifters $\Delta\zeta_{h,t+1}$. This is trivially true if taste shifters are constant over time or if they are random walks. We maintain one of these two assumptions, a hypothesis that we can in part test by checking over-identifying restrictions.

We estimate equation (7.21) using the Consumer Expenditure Survey (CEX). Although the CEX covers many years, each household is only observed for a few quarters and so we use a synthetic cohort approach (see Browning et al. (1985)): we aggregate equation (7.21) over groups with constant membership and follow the average behaviour of the variables of interest (or their non-linear transformation) for such groups. A time series of quarterly cross sections can be used to construct consistent estimates of these aggregates and, in this fashion, use a long time period to estimate the parameters of the Euler equation and test its validity.

[1] The Cragg-Donald statistic is usually used to provide a test of underidentification. Stock and Yogo (2005) propose using it as a test of instrument relevance as well.

We define groups using married couples in ten year birth-cohorts. The assumption of constant membership of these groups might be questioned at the beginning and at the end of the life-cycle for a variety of reasons, including differential rates of family formation, differential mortality and so on. To avoid these and other issues, we limit our sample to households whose husband is aged between 25 and 65 and where wives are aged between 25 and 60.[2]

Having identified groups, we aggregate equation (7.21) to be estimated across group $g$ households. For this approach to work, however, it is necessary that the equation to be estimated is linear in parameters, which would be the case if $M_{h,t}$ were observable. However, $M_{h,t}$ is a non-linear function of data *and* unobserved parameters, so that, in principle it cannot be aggregated within groups to obtain $M_{g,t}$. On the other hand, the parameters that determine $M_{h,t}$ can be consistently estimated using the MRS conditions as discussed in Section 7.2.1.[3] These estimates can be used to construct consistent estimates of $M_{h,t}$, which can be aggregated across households to give $M_{g,t}$.

We can obtain consistent estimates of the grouped variables from the time series of cross-sections, giving the group average log-linear Euler equation:

$$\widetilde{\eta}_{g,t+1} = \overline{\kappa} + \ln\beta + \ln(1+r_{t+1}) - \phi\Delta\overline{\ln c_{g,t+1}} + -\gamma\Delta\ln(\overline{\widehat{M}_{g,t+1}}) + \varphi\Delta\overline{P_{g,t+1}} + \pi\Delta\overline{z_{g,t+1}} \tag{E.22}$$

The residual term $\widetilde{\eta}_{g,t+1}$ now includes, in addition to the average of the expectation errors and of the changes in taste shifters, several other terms: (i) a linear combination of the difference between the population and sample averages at time $t$ and $t+1$ for all the relevant variables (induced by the fact that we are considering sample means rather than population means for group $g$); (ii) the difference between the (consistently) estimated $M_{g,t}$ and its actual value (induced by estimation error in

[2] If credit constraints are binding, the Euler equation will not be holding as an equality. The youngest consumers are excluded because they are more likely to be affected by this issue. For older consumers, in addition to changes in labour force participation and family composition, health status also changes in complex ways that may be difficult to capture with the taste shifters that we have been considering.

[3] $M_{h,t}$ includes $\chi_{h,t}$ which is unobserved. However, since it is the residual from the MRS equation, it can be included in the calculation of $\alpha_{h,t}$ that is needed to calculate $M_{h,t}$.

the parameters of the MRS); (iii) the difference between the innovation over time to the average value of $\kappa_{g,t}$, which we have denoted with the constant $\overline{\kappa}$.

All the variables on the right hand side of equation (E.22) are observable. We can therefore use this equation to estimate the parameters of interest. However, the instruments need to be uncorrelated with $\widetilde{\eta}_{g,t+1}$.[4] The covariance structure of the $\widetilde{\eta}_{g,t+1}$ is quite complex: the contemporaneous covariance of $\widetilde{\eta}_{g_i,t+1}$ and $\widetilde{\eta}_{g_j,t+1}$ is not, in general, zero, as aggregate shocks have effects that correlate across different groups. When computing the variance-covariance matrix of the estimates, this structure should be taken into account. Whilst it is in principle possible, given our assumptions, to estimate the variance-covariance matrix of $\widetilde{\eta}_{g,t+1}$ from estimated parameters, in practice it turns out to be cumbersome, as there is no guarantee that, in small samples, these estimates are positive-definite. Given these difficulties, we follow a different and, as far as we know, novel approach, based on bootstrapping our sample, with a structure consistent with the basic assumptions of our model. We describe the bootstrapping procedure in detail in the next subsection.

## Bootstrap procedure

We bootstrap standard errors and confidence intervals for both our MRS and Euler equations.

The two step Heckman-selection procedure for estimating the MRS coefficients can be bootstrapped in the standard way. Bootstrapping results for our Euler equation requires a slightly more complicated procedure however. This is because we aggregate our data into cohort groups and then implement an IV procedure. Taking $Z_t$ as a vector of exogenous variables, and $X_t$ and $Y_t$ as endogenous variables (with $Y_t$ as our dependent variable) we can reformulate our approach as estimating the equations

---

[4]As noted by Deaton (1985) and discussed extensively in the context of the CEX by Attanasio and Weber (1995), the use of sample rather than population averages for all the 'group' variables induces an MA(1) in the residuals, because of the sampling variation in the rotating panel structure. We need to assume that the instruments are not correlated with the (average) estimation error of the $M_{h,t}$ or with the innovations to the higher moments of the expectation errors $(\kappa_{g,t} - \overline{\kappa})$. This last assumption is discussed in Attanasio and Low (2004).

$$
\begin{aligned}
X_t &= \Pi Z_t + v_t \\
Y_t &= X_t \beta + u_t
\end{aligned}
$$

where $v_t$ is a vector of errors in our first stage. These can be thought of as economic shocks which may have a complicated structure. For instance they may be correlated across time for a given cohort, or may have an aggregate component which is correlated across cohorts for a given time period. Errors may also be correlated across the equations for different exogenous variables $Z_t$. We will wish to preserve these correlations when we implement our bootstrap procedure. In order to do this, we attempt to construct the variance-covariance matrix of the residuals $v$. Rather than filling in all possible cross-correlations in this matrix, we calculate the following moments for each cohort $c$, and equation $i$

$$
\begin{aligned}
& var(v^{i,c}) \\
& cov(v_t^{i,c}, v_{t-1}^{i,c}) \\
& cov(v_t^{i,c}, v_t^{j,c}) \\
& cov(v_t^{i,c}, v_t^{i,k})
\end{aligned}
$$

Setting all other correlations to zero. Thus we impose for instance that there is zero correlation between $v_t^{i,c}$ and $v_{t-1}^{i,k}$. Unfortunately, there is no guarantee that this matrix will be positive definite. In our procedure we therefore apply weights to the non-zero elements of our ‘off-diagonal’ matrices - which give the covariances across different cohorts for the same equation - and to our 1st autocovariances for residuals for the same cohort and same equation. The weights we apply to these are the maximum that ensure the resulting matrix is positive definite: in our case they are both set at 0.23.

Once we have this matrix we can Cholesky decompose it to obtain a vector of orthogonalised residuals

$$\Omega = \nu\nu' = \varepsilon CC'\varepsilon'$$

We then draw from the orthogonalised residuals, premultiply them by $C$ and then add them to $\Pi Z_t$ to reconstruct the endogenous variables (including $Y$). We then reestimate our second stage equation to obtain a new set of estimates for $\beta$.

The values of $Z_t$ in our case will depend on the results we obtain from our MRS equation, so in each iteration of our bootstrap we resample with replacement from from our disaggregated data, re-run the MRS equation, reaggregate to obtain the cohort averages which make up $Z_t$ and then make a draw from our residuals.

## Solution method

Households have a finite horizon and so the model is solved numerically by backward recursion from the terminal period. At each age we solve the value function and optimal policy rule, given the current state variables and the solution to the value function in the next period. This approach is standard. The complication in our model arises from the combination of a discrete choice (to participate or not) and a continuous choice (over saving). This combination means that the value function will not necessarily be concave. We briefly describe in this appendix how we deal with this potential non-concavity. An alternative would be to follow the method in Iskhakov et al. (2017).

In addition to age, there are four state variables in this problem: the asset stock, the permanent component of earnings of the husband, $v^m_{h,t}$, the permanent component of wife's wage, $v^f_{h,t}$,and the experience level of the wife. We discretise both earning and wage variables and the experience level, leaving the asset stock as the only continuous state variable. Since both permanent components of earnings are non-stationary, we are able to approximate this by a stationary, discrete process only because of the finite horizon of the process. We select the nodes to match the paths of the mean shock and the unconditional variance over the life-cycle. In particular, the unconditional variance of the permanent component must increase linearly with age, with the slope given by the conditional variance of the permanent shock. Our estimates of the wage variance are for annual shocks, but the model period is one quarter. We reconcile this difference by imposing that each quarter an individual receives a productivity shock with probability 0.25, and this implies that productivity shocks occur on average once a year.

Value functions are increasing in assets $A_t$ but they are not necessarily concave, even if we condition on labour market status in $t$. The non-concavity arises because of changes in labour market status in future periods: the slope of the value function is given by the marginal utility of consumption, but this is not monotonic in the asset stock because consumption can decline as assets increase and expected labour market status in future periods changes. By contrast, in Danforth (1979) employment is an absorbing state and so the conditional value function will be concave. Under certainty, the number of kinks in the conditional value function is given by the number of periods of life remaining. If there is enough uncertainty, then changes in work status in the future will be smoothed out leaving the expected value function concave: whether or not an individual will work in $t+1$ at a given $A_t$ depends on the realisation of shocks in $t+1$. Using uncertainty to avoid non-concavities is analogous to the use of lotteries elsewhere in the literature.

The choice of participation status in $t$ is determined by the maximum of the conditional value functions in $t$. In our solution, we impose and check restrictions on this participation choice. In particular, we use the restriction that the participation decision switches only once as assets increase, conditional on permanent earnings and experience. When this restriction holds, it allows us to interpolate behaviour across the asset grid without losing our ability to determine participation status. We therefore define a reservation asset stock to separate the value function and the choice of consumption made when participating from the value function and choice of consumption made when not participating. There are some regions of the state space where individuals are numerically indifferent between working and not working. Since we solve the model by value function iteration, it does not matter which conditional value function we use in these regions

In solving the maximisation problem at a given point in the state space, we use a simple golden search method. Note that in addition to the optimal total expenditure, the optimal amount of leisure is computed in each period by solving the MRS condition. We solve the model and do the calibration assuming this process is appropriate and assuming there is a unique reservation asset stock for each point in the state space, and then check ex-post.

There are no non-concavities due to borrowing constraints in our model because the only borrowing constraint is generated by the no-bankruptcy condition which is in effect enforced by having infinite marginal utility of consumption at zero consumption.

Finally, we include here the value functions of the household problem. In each period, if the woman chooses to participate, the value function is given by

$$V^1_{h,t}\left(A_{h,t}, v_{h,t}\right) =$$
$$\max_{c_{h,t}, l_{h,t}} \left\{ u\left(c_{h,t}, l_{h,t}, P_{h,t}=1\right) + \beta E_t \left[ \max \left\{ \begin{array}{c} V^0_{h,t+1}\left(A_{h,t+1}, v_{h,t+1}\right) \\ V^1_{h,t+1}\left(A_{h,t+1}, v_{h,t+1}\right) \end{array} \right\} \right] \right\} \tag{E.23}$$

Note that the state variable $v_{h,t}$ is a vector containing the woman and the man's productivity type. If she chooses not to participate, the value function is given by,

$$V^0_{h,t}\left(A_{h,t}, v_{h,t}\right) =$$
$$\max_{c_{h,t}} \left\{ u\left(c_{h,t}, P_{h,t}=0\right) + \beta E_t \left[ \max \left\{ \begin{array}{c} V^0_{h,t+1}\left(A_{h,t+1}, v_{h,t+1}\right) \\ V^1_{h,t+1}\left(A_{h,t+1}, v_{h,t+1}\right) \end{array} \right\} \right] \right\} \tag{E.24}$$

The decision of whether or not to participate in period $t$ is determined by comparing $V^0_{h,t}\left(A_{h,t}, v_{h,t}\right)$ and $V^1_{h,t}\left(A_{h,t}, v_{h,t}\right)$. The participation choice, the hours choice and the consumption choice in $t$ determines the endogenous state variable (assets) at the start of the next period.

## E.2 Appendix to Section 7.3: data sources and descriptive statistics

As discussed in the the main chapter, most of the data are from the CEX. One important exception are the data on the real interest rate. We define this variable as the 3 month T-Bill rate (on a quarterly basis) minus the rate of growth in the CPI. The source for the T-Bill rate is from the St Louis Fed (https://fred.stlouisfed.org/series/TB3MS).

In for Table E.1 presents descriptive statistics at the individual level using data from three particular years (1980, 1995 and 2012). Married women have seen large changes in their wages, hours and patterns of employment over our sample period. Employment rates increased from 60% in 1980 to 69.8% in 1995 before falling back to 61.9% in 2012.

Table E.1 also shows wage levels over the three years. Average real wages increased over this period, though with marked differences across different education groups. The wages of those with less than high school education actually fell slightly from $12.16 in 1980 to $11.33 in 2012. By contrast, married women with a college degree or higher saw a 20% increase in their wages between 1980 and 2012 (from $19.30 to $23.20). This increase in the education premium has been attributed to skill-biased technological change which outstripped the supply of educated workers (Goldin and Katz, 2007).

Changes in hours worked across education groups appear to mirror these patterns. While all education groups worked very similar hours in 1980, by 2012 those with a college degree were working on average five hours more per week than those with less than high school education, although the fraction with a college degree has markedly increased over the period.

**Table E.1:** Descriptive statistics for married women, 1980, 1995 and 2012

| | | 1980 | 1995 | 2012 |
|---|---|---|---|---|
| *Demographics* | No. of children | 1.25 | 1.15 | 1.17 |
| *Education* | % Less than high school | 19.4 | 12.3 | 9.7 |
| | % High school | 44.1 | 36.8 | 25.3 |
| | % Some college | 18.1 | 25.3 | 28.5 |
| | % Degree or higher | 18.4 | 25.5 | 36.5 |
| *Hours (workers)* | All | 35.2 | 37.5 | 38.4 |
| | Less than high school | 34.9 | 37.4 | 34.2 |
| | High school | 35.2 | 36.2 | 38.6 |
| | Some college | 35.0 | 36.7 | 37.1 |
| | Degree or higher | 35.5 | 39.7 | 39.5 |
| *Hourly net wages ($ 2016)* | All | 15.58 | 16.63 | 18.95 |
| | Less than high school | 12.16 | 11.23 | 11.33 |
| | High school | 14.22 | 13.41 | 14.62 |
| | Some college | 16.62 | 16.41 | 17.28 |
| | Degree or higher | 19.30 | 22.26 | 23.20 |
| % Employed | All workers | 60.0 | 69.8 | 61.9 |
| | % Workers part-time | 28.4 | 23.7 | 20.6 |
| *Sample sizes* | All | 2,199 | 2,064 | 2,026 |
| | Workers | 1,318 | 1,441 | 1,254 |

Note: Part-time is defined as working less than 35 hours per week.

## E.3 Appendix to Section 7.4

### Alternative methods of estimating the MRS

In this appendix we discuss results from alternative MRS specifications. For comparison with later results, we present a more complete set of parameter estimates from our baseline MRS specification in Table E.3. First, we present results for the selection probit we run prior to estimating our MRS equation. Husband's earnings are strongly negatively correlated with participation.

**Table E.2:** Selection probit results

| | | |
|---|---|---|
| Log earnings of husband | -0.164*** | (0.007) |
| Husband employed | -1.929*** | (0.064) |
| No. of Elderly HH members | 0.023 | (0.026) |
| Log family size | -0.110*** | (0.022) |
| Wife: White | -0.015 | (0.014) |
| Age | -0.056 | (0.042) |
| $\text{Age}^2$ | 0.001 | (0.001) |
| $\text{Age}^3$ /1000 | 0.003 | (0.018) |
| $\text{Age}^4$ /10000 | -0.003* | (0.001) |
| Has kids | -0.034 | (0.018) |
| No. of kids aged 0-2 | -0.515*** | (0.014) |
| No. of kids aged 3-15 | -0.167*** | (0.008) |
| No. of kids aged 16-17 | 0.071*** | (0.017) |
| North East | -0.004 | (0.015) |
| Mid-West | 0.119*** | (0.014) |
| South | 0.035** | (0.013) |

Notes: N= 78,674. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ Standard errors in parentheses. Additional controls for season and year dummies and cohort-education interactions.

## Estimation method and normalisation

We start by considering the issue of how the MRS is normalised. Recall that our MRS relationship is

$$\ln w^f_{h,t} = \psi_0 + \psi z_{h,t} - \theta \ln l_{h,t} + \phi \ln c_{h,t} + \upsilon_{h,t} \tag{E.25}$$

As Keane (2011) notes, this is not a labour supply equation but an equilibrium condition in which wages, leisure and consumption are all endogenous. All three variables are potentially correlated with the error term $\upsilon_{h,t}$ and so there is no natural choice of the dependent variable.

Despite this, we find that, when conventional methods are used, results can be highly sensitive to whether wages, leisure or consumption are placed on the left hand side of the MRS equation. Table E.4 shows results from estimating $\phi$ and $\theta$ using GMM under the three different possible normalisations. We include results

**Table E.3:** Baseline MRS estimates

| Parameter | Estimate | (Standard Error) | [95% C.I] |
|---|---|---|---|
| $\theta$ | 1.75** | (1.230) | [0.34,5.12] |
| $\phi$ | 0.76*** | (0.103) | [0.55,0.95] |
| $\Psi$ | | | |
| Age | 0.05** | (0.02) | [0.01,0.09] |
| Age$^2$ | -0.0005 | (0.0007) | [-0.002,0.001] |
| Age$^3$/1000 | -0.01 | (0.01) | [-0.03,0.01] |
| Age$^4$/10000 | 0.002** | 0.0007 | [0.0002,0.003] |
| North East | 0.01 | (0.03) | [-0.02,0.08] |
| Mid West | -0.05** | (0.01) | [-0.07,-0.02] |
| South | -0.11*** | (0.02) | [-0.18,-0.09] |
| White | -0.04 | (0.03) | [-0.09,0.04] |
| No. elderly HH members | 0.02 | (0.02) | [-0.02,0.05] |
| $\ln(famsize)$ | -0.32*** | (0.037) | [-0.38,-0.23] |
| Has kids | 0.07*** | (0.021) | [0.04, 0.12] |
| No. of kids 0-2 | 0.15*** | (0.030) | [0.10, 0.22] |
| No. of kids 3-15 | 0.06*** | (0.017) | [0.04, 0.10] |
| No. of kids 16-17 | -0.02* | (0.011) | [-0.05,0.00] |
| Constant ($\Psi_0$) | 4.70 | (4.94) | [-1.19,18.52] |
| *Heckman selection terms* | | | |
| $e_1$ | 0.07 | (0.167) | [-0.18, 0.48] |
| $e_2$ | 0.05 | (0.172) | [-0.21, 0.51] |
| $e_3$ | 0.01 | (0.052) | [-0.08, 0.13] |

Notes: N = 50,895. *p<0.10, ** p<0.05, *** p<0.01. Additional controls for season and year dummies and cohort-education interactions. Confidence intervals are bootstrapped with 1000 replications allowing for clustering at the individual level.

form our baseline specification in the first column. The implied parameter estimates and elasticities vary a great deal across these different approaches. When wages are selected as the left-hand side variable, elasticites are relatively large. When leisure is the dependent variable, they are much smaller. Very similar considerations apply to the estimation of our Euler equation.

Differences of this kind can emerge in IV estimation in 2SLS and GMM esti-

**Table E.4:** MRS estimates using GMM

| | Fuller | GMM | | |
|---|---|---|---|---|
| *Dependent variable:* | Wages | Wages | Leisure | Consumption |
| *Parameters* | | | | |
| $\theta$ | 1.75** | 0.46* | 13.8 | 0.13 |
| | [0.34,5.12] | [−0.04,0.61] | [−120.13,186.11] | [−0.54,0.58] |
| $\phi$ | 0.76*** | 0.61*** | 0.17 | 1.38*** |
| | [0.55,0.95] | [0.48,0.66] | [−3.44,2.78] | [1.24,1.74] |
| *Wage elasticities at median* | | | | |
| Marshallian | 0.18 | 0.55 | 0.09 | −0.17 |
| | [0.05,0.38] | [0.50,1.16] | [0.00,0.12] | [−0.41,−0.08] |
| Hicksian | 0.54 | 1.19 | 0.11 | 0.77 |
| | [0.27,1.29] | [1.10,2.25] | [−0.01,0.14] | [0.59,1.10] |

Notes: N = 50,895. *p<0.10, ** p<0.05, *** p<0.01. Controls as in Table 7.3. Elasticities are calculated as averages within a 5 percent band of the 50th percentile of the Marshallian distribution. 95% confidence intervals in square brackets. Confidence intervals are bootstrapped with 1000 replications.

mation when the instruments chosen are relatively weak. Indeed, Hahn and Hausman (2003) propose using the differences in parameters implied by 2SLS estimates run under different normalisations as a test of instruments' strength.

Various papers have discussed possible remedies for cases when strong instruments are not available (Hahn and Hausman, 2003; Hausman et al., 2012). One possible solution is the use estimators such as Limited Information Maximum Likelihood (LIML) rather than 2SLS, which is known to have poor bias properties in such circumstances (Staiger and Stock, 1997b; Nelson and Startz, 1990). Using the notation from Davidson and MacKinnon (2004), for the case where

$$\begin{aligned} y &= Z\beta_1 + Y\beta_2 + u = X\beta + u \\ Y &= \Pi W + v \end{aligned}$$

where $Z$ is a matrix of exogenous variables, $Y$ a matrix of endogenous variables, and $W = [Z, W_1]$ (with $W_1$ being a matrix of instruments). Matrices $X$ and $W$ are $n \times k$ and $n \times l$ respectively (with $l \geq k$). In general, so-called $k$-class estimators such as OLS, 2SLS, and LIML can be written in the form

$$\hat{\beta}^{\text{LIML}} = (X'(I - kM_W)X)^{-1}X'(I - kM_W)y \tag{E.26}$$

where $M_W = I - W(W'W)W'$. In the case of OLS $k = 0$, and in the case of 2SLS $k = 1$. In the case of LIML we use

$$k = k_{\text{LIML}} = \frac{(y - Y\beta_2)'M_Z(y - Y\beta 2)}{(y - Y\beta_2)'M_W(y - Y\beta 2)} \tag{E.27}$$

While LIML is often found to have better bias properties than 2SLS, it has long been recognised that conventional normalisations of LIML do not have finite moments (Mariano and Sawa, 1972; Sawa, 1972), and simulation exercises have shown that this can add considerable volatility to empirical estimates (Hahn et al., 2004). As a result Hahn et al. (2004) recommend the use of either jack-knifed 2SLS or the modification of LIML proposed by Fuller (1977). For this latter estimator, we replace $k$ in equation (E.26) with

$$k_{\text{Fuller}} = k_{\text{LIML}} - \frac{\lambda}{(n - k)} \tag{E.28}$$

where $\lambda$ here is a parameter chosen by the researcher, to obtain a value for $\hat{\beta}^{\text{Fuller}}$. We choose a value of one for this as suggested by Davidson and MacKinnon (2004) as it yields estimates that are approximately unbiased. The resulting estimator is guaranteed to have bounded moments in finite samples Fuller (1977). Since the adjustment to LIML is smaller when $(n - k)$ is large, the Fuller estimator will be closer to LIML when sample sizes are large relative to the number of instruments. In our case, the Fuller estimator can be thought of as a compromise between 2SLS and LIML, as it adjusts the value of $k$ we use downwards slightly towards one.

As well as its superior bias properties, the Fuller estimator has the advantage that is much less sensitive than GMM or 2SLS to the choice of the dependent variable, as Table E.5 shows. Both the elasticity and parameter estimates obtained using alternative normalisations of the Fuller estimator are very similar to our baseline results.

**Table E.5:** MRS estimates with different dependent variables

| | Dependent variable | | |
|---|---|---|---|
| | Wages | Leisure | Consumption |
| *Parameters* | | | |
| $\theta$ | $1.75^{**}$ [0.34,5.12] | $1.84^{*}$ [−0.43,5.38] | $1.75^{*}$ [−0.00,4.60] |
| $\phi$ | $0.76^{***}$ [0.55,0.95] | $0.76^{***}$ [0.53,0.95] | $0.77^{***}$ [0.58,0.95] |
| *Wage elasticities at median* | | | |
| Marshallian | 0.18 [0.05,0.38] | 0.17 [0.06,0.37] | 0.18 [0.07,0.42] |
| Hicksian | 0.54 [0.24,1.07] | 0.53 [0.23,0.95] | 0.54 [0.27,1.29] |

Notes: N = 50,895. *p<0.10, ** p<0.05, *** p<0.01. Controls as in Table 7.3. Elasticities are calculated as averages within a 5 percent band of the 50th percentile of the Marshallian distribution. 95% confidence intervals in square brackets. Confidence intervals are bootstrapped with 1000 replications.

## Alternative instruments

In Table E.6 we show results using alternative choices of instruments. We show results using GMM (with wages as the dependent variable) and the Fuller estimator described above, in both cases using a *full* set of cohort-education-year interactions as used in Blundell et al. (1998). This approach is similar to the approach we adopt for our main results but interacts cohort-eduction dummies full set of year effects rather than a polynomial in time trends. The estimates we obtain from fully adopting the Blundell et al. (1998) approach are very similar to our main results, though somewhat less precise.

The sensitivity of our results to the choice of instruments is on the whole quite small when we compare it to the differences that can arise from the choice of estimation method. Just as we find for our main set of results, the hours elasticities estimated using the GMM estimator with wages as the dependent variable are substantially larger than those using the Fuller estimator when using the alternative instrument set.

**Table E.6:** MRS estimates using alternative instruments

| | Fuller | GMM |
|---|---|---|
| *Parameters* | | |
| $\theta$ | 1.93<br>[−9.09,11.58] | 0.08<br>[−0.20,0.19] |
| $\phi$ | 0.76***<br>[0.42,1.03] | 0.52***<br>[0.41,0.52] |
| *Wage elasticities at median* | | |
| Marshallian | 0.17<br>[−0.84,1.13] | 1.08<br>[1.04,2.29] |
| Hicksian | 0.51<br>[−1.91,2.65] | 1.97<br>[1.82,3.85] |

Notes: N = 50,895. *p<0.10, ** p<0.05, *** p<0.01. We use a full set of cohort-education-year dummies as instruments, following Blundell, Duncan and Meghir (1998). Controls as in Table 7.3. Elasticities are calculated as averages within a 5 percent band of the 50th percentile of the Marshallian distribution. 95% confidence intervals in square brackets. Confidence intervals are bootstrapped with 1000 replications.

## Alternative samples

Table E.7 shows how our MRS results are affected by alternative sample selection choices. Column (1) presents results when we exclude those individuals who report working exactly 40 hours a week. The justification of this experiment is that these individuals may be affected by some kind of friction that does not allow them to adjust their hours worked as desired. Such frictions would mean that the MRS condition that we exploit to recover $\phi$ and $\theta$ need not hold. Excluding these observations, we obtain greater estimates of our Marshallian and Hicksian hours elasticities (at 0.45 and 0.72 respectively). These values are however somewhat imprecisely estimated and the confidence bands that surround them include our baseline estimates.

In Column (2) we show results when we exclude individuals working less than 20 hours per week (with an appropriate adjustment to our selection correction). We consider results from this specification because there may be certain frictions that prevent individuals working fewer hours than this, which would again lead to potential violations of the MRS condition. Excluding these observations delivers somewhat lower elasticity estimates, but again the estimates are imprecise.

Finally, in column (3) we consider only those ten-year birth cohorts with the most similar labour-supply behaviour over the life-cycle. In particular we exclude those born before 1925 as they tend to work fewer hours at older ages than other cohorts, and those born after 1975, as less-educated individuals born after this date tend to have lower employment rates than other earlier cohorts at the same ages. Using this sample, we obtain a Marshallian elasticity of 0.27 and a Hicksian 0.53. While the Marshallian elasticity estimated from this sample is slightly higher than our baseline estimates, the Hicksian elasticity is essentially unchanged.

## Alternative definitions of hours

In column (4) of Table E.7 we consider how elasticity estimates are affected when we use an alternative measure of hours of leisure. The measure we use here is

$$\text{leisure} = \frac{5200 - \text{hours per week} \times \text{weeks worked per year}}{52} \tag{E.29}$$

This measure accounts for the observed variation in weeks worked per year in addition to variation in hours worked per week across workers.

The elasticities resulting from this exercise are in general lower but on the

**Table E.7:** MRS estimates using alternative samples/hours measures

| | Exc. 40 hours (1) | Exc. <20 hours (2) | Born 1925-1965 (3) | Ann. hours (4) |
|---|---|---|---|---|
| *Parameters* | | | | |
| $\theta$ | 1.52 [−3.16,5.69] | 2.81 [−2.48,9.81] | 2.08*** [0.68,4.66] | 2.30* [−1.30,7.00] |
| $\phi$ | 0.42* [−0.05,0.92] | 0.76*** [0.33,1.03] | 0.56*** [0.42,0.82] | 0.78*** [0.53,1.01] |
| *Wage elasticities at median* | | | | |
| Marshallian | 0.45 [−0.45,1.41] | 0.13 [0.02,0.30] | 0.27 [0.09,0.62] | 0.13 [−0.09,0.41] |
| Hicksian | 0.72 [−1.29,2.57] | 0.39 [−0.14,1.43] | 0.53 [0.27,1.09] | 0.42 [−0.25,1.24] |
| $N$ | 26,060 | 47,743 | 39,057 | 50,895 |

Notes: *p<0.10, ** p<0.05, *** p<0.01. Specification (1) excludes individuals who work exactly 40 hours. Specification (2) excludes those working less than 20 hours (part-time workers). Specification (3) only includes individuals from cohorts with the most similar labour supply choices over the life-cycle. Elasticities are calculated as averages within a 5 percent band of the 50th percentile of the Marshallian distribution. 95% confidence intervals in square brackets. Confidence intervals are bootstrapped with 1000 replications.

whole similar to than those in our baseline specification, with a Marshallian elasticity of 0.13 and a Hicksian elasticity of 0.42. The value of $\theta$ is larger than in our main results (at 2.30), and much less precisely estimated. The value of $\phi$ is essentially unchanged.

## External parameters for the calibration

Table E.8 reports the complete set of estimated and external parameters used in the calibration. The first panel reports the estimated parameters from Tables 7.3 and 7.4 above. The second panel reports parameters which come from external sources.

**Table E.8:** External parameters

| | | |
|---|---|---|
| Estimated Parameters (from first-order conditions) | | |
| Curvature on leisure | $\theta$ | 1.75 |
| Curvature on consumption | $\phi$ | 0.76 |
| Curvature on utility | $\gamma$ | 2.07 |
| Exogenous Parameters | | |
| Interest Rate (annual) | $r$ | 0.015 |
| Regression Log Wage on Age and Age$^2$ (Men) | $\iota_1^m, \iota_2^m$ | 0.0684, -0.00065 |
| Husband and Wife Wage Correlation | $\rho$ | 0.25 |
| Std Dev. of Permanent Shock (Men) | $\sigma_{\xi^m}$ | 0.077 |
| Std Dev. of Initial Wage (Men) | $\sigma_{\xi_0^m}$ | 0.54 |
| Length of Life (in years) | $T$ | 50 |
| Length of Working Life (in years) | $R$ | 40 |

## E.4 Appendix to Section 7.5: Results for CES and additive separability

In this Appendix we discuss results for alternative specifications of our utility function. In particular we consider results from a standard CES utility function (where we impose that $\theta$=$\phi$), and one where we impose additive separability between consumption and leisure (i.e $\gamma = 0$).

Table E.9 presents parameter estimates when we impose the restrictions implied by CES utility. Under this functional form for utility, we get a slightly larger value of $\phi$ and a much lower value of $\theta$ than we obtain from our preference specification (at 0.83 compared to 0.76 and 1.75 that we obtain for $\phi$ and $\theta$ respectively in Table 7.3). We also obtain a slightly larger value of $\gamma$ however (at 3.04 compared to 2.07 for our less restrictive utility function).

Taken together, the CES parameter estimates imply that utility is less concave in leisure, and hence that labour supply elasticities are greater. We show the elasticities implied by these estimates in Table E.10. While Marshallian hours elasticities for the CES specification are only greater at the upper end of the distribution, the estimated Hicksian and Frisch hour elasticities are roughly 50% larger. The CES es-

**Table E.9:** Parameter values

| | CES |
|---|---|
| $\phi$ | 0.83 [0.66,0.97] |
| $\theta$ | 0.83 [0.66,0.97] |
| $\gamma$ | 3.04 [0.64,4.27] |

timates also imply a more substantial degree of non-separability between consumption and leisure. The Frisch elasticity of consumption with respect to predictable wage increases has a median of around 0.4 compared to 0.05 from our main estimates. This reflects both a greater sensitivity of the marginal utility of consumption to changes in leisure and the fact that leisure responses to given wage changes will in general be greater under these preferences. Finally we note that, the interest rate Frisch elasticity at the median is much lower than in our baseline specification.

Table E.10 also shows Frisch elasticities for our preference specification in the case where we impose additive separability for preferences over consumption and leisure (that is we impose that $\gamma = 0$). This necessarily sets the Frisch consumption responses to wage changes to zero. It turns out that Frisch hours elasticities are very similar to those estimated when we allow for non-separability in our main specification. This reflects the fact that when, as we find, the parameters $\theta$ and $\phi$ are small and $\alpha$ large, then the numerator and denominator in formulae for Frisch elasticities given in equations (E.15) and (E.16) will be dominated by the term $M_t$. Consequently, the impact of small changes in $\gamma$ will be limited.

When additive separability is imposed, the Frisch elasticity is identical - a direct result of setting $u_{cl} = 0$ in expressions (E.18) and (E.19). The estimated Frisch elasticity of consumption with respect to the interest rate (now simply given by $-1/\phi$ ) also falls relative to our baseline results, from a median value of -1.19 in our baseline results to -1.31.

**Table E.10:** Elasticities at percentiles of Marshallian distribution: CES

| | $\gamma = 0$ | | CES | | | |
|---|---|---|---|---|---|---|
| | Wage | Interest rate | Wage | | | Interest rate |
| | Frisch | Frisch | Marshallian | Hicksian | Frisch | Frisch |
| | *Hours worked* | | *Hours worked* | | | |
| 10th | 0.84 [0.23,3.17] | 0.84 [0.23,3.17] | −0.24 [−0.30,−0.11] | 0.48 [0.41,0.60] | 1.08 [0.97,1.47] | 0.83 [0.63,1.32] |
| 25th | 0.83 [0.23,3.15] | 0.83 [0.23,3.15] | −0.04 [−0.13,0.12] | 0.60 [0.51,0.76] | 1.16 [1.06,1.56] | 0.85 [0.65,1.35] |
| 50th | 0.90 [0.25,3.40] | 0.90 [0.25,3.40] | 0.21 [0.10,0.42] | 0.77 [0.67,0.98] | 1.33 [1.23,1.75] | 0.93 [0.71,1.47] |
| 75th | 1.04 [0.29,3.93] | 1.04 [0.29,3.93] | 0.54 [0.39,0.82] | 1.04 [0.89,1.32] | 1.66 [1.53,2.17] | 1.13 [0.86,1.78] |
| 90th | 1.98 [0.55,7.50] | 1.98 [0.55,7.50] | 1.11 [0.88,1.55] | 1.62 [1.39,2.06] | 2.71 [2.51,3.57] | 1.89 [1.44,2.99] |
| | *Consumption* | | *Consumption* | | | |
| 25th | 0.00 [−,−] | −1.31 [−1.81,−1.05] | 0.91 [0.82,1.08] | 0.63 [0.54,0.80] | 0.32 [0.12,0.53] | −0.59 [−0.93,−0.45] |
| 50th | 0.00 [−,−] | −1.31 [−1.81,−1.05] | 1.07 [0.97,1.27] | 0.72 [0.62,0.91] | 0.37 [0.15,0.62] | −0.58 −0.91,−0.44] |
| 75th | 0.00 [−,−] | −1.31 [−1.81,−1.05] | 1.23 [1.12,1.44] | 0.80 [0.68,1.01] | 0.42 [0.17,0.69*s*] | −0.57 [−0.90,−0.43] |

Notes: Elasticities are calculated as averages within 5 percent bands of the 10th, 25th, 50th and 75th and 90th percentiles of the Marshallian distribution. 95% confidence intervals in square brackets. Confidence intervals are bootstrapped with 1000 replications.

## E.5 Returns to experience

We recalibrate parameter values: the fixed cost of working, $\bar{F}$, child care price, $p$, the offered wage gender gap and $\psi_0$. In addition to these parameters, we also need to calibrate the parameter that characterises human capital accumulation function and its depreciation rate.[5] As in the baseline, we identify these parameters by targeting participation rate of women, the participation rate of mothers, the average hours worked, the observed wage gender gap, the observed wage growth at early ages,

[5] Note this is only one parameter in contrast to the two parameters $\iota_1^f$ and $\iota_2^f$ for the exogenous wage growth that were used in the baseline economy.

and the observed depreciation of wages during non-participation (we take this figure from Attanasio et al. (2008)). We report the calibrated parameters in Table E.11 and compare them to the baseline. In the context of returns to experience, where there is a strong incentive to work to reap future returns, a much larger childcare cost is required in order to reduce participation and match participation statistics.

Analogously to Figure 7.2, Figure E.1 shows life-cycle profiles in the simulations and in the data; and Table E.12 reports additional statistics on the distribution of hours and of wages. There are some differences between the model with returns to experience and the baseline. First, there is a decline in the participation profiles at ages beyond 35. These patterns are not observed either in the data or in the baseline model. Second, very few women change their participation decisions. For example, the fraction of women who worked in all previous periods at the age of 52 is 57%, which compares to 40% in the economy without returns to experience. Third, the childcare cost that is needed here to keep women out of the labour market during childbearing is substantially higher because of the incentive to accumulate labour market experience. In particular the monetary fix childcare cost is up to 76% of median earnings of a women aged 25 to 55.

### E.5.1 Response to temporary wage changes

In Table E.13, we report the labour supply responses in the economy with returns to experience. The key finding is that, in contrast to the economy without returns to experience, the extensive margin response is close to zero and, as a result, the aggregate elasticity is about half of the one in the baseline economy (reproduced in the final column). In the return to experience economy, there is a strong incentive to participate to obtain the return to experience. The larger childcare cost of participating that is estimated in this economy alongside the strong incentive to participate implies that changes in the current wage makes little difference to the incentive to participate. As expected, the size of the intensive margin response is similar to the one in the economy without returns to experience. Our results here are in line with Imai and Keane (2004) who argue that the response of labour supply to transitory changes in wages may be mitigated when there are returns to experience. Our re-

**Table E.11:** Baseline economy: calibrated parameters and targets

| Parameter Name | | Values | |
|---|---|---|---|
| | | Ret to Exp | Baseline |
| Constant term weight of leisure | $\psi_0$ | 4.13 | 4.20 |
| Childcare Cost | $p$ | 5820 | 967 |
| Fixed Cost of Working | $\bar{F}$ | 315 | 468 |
| Offered Wage Gender Gap at age 22 | $y_0^f/y_0^m$ | 0.78 | 0.74 |
| Std Dev. of Permanent Shock (Women) | $\sigma_{\xi^f}$ | 0.063 | 0.063 |
| Std Dev. of Initial Wage (Women) | $\sigma_{\xi^f,0}$ | 0.50 | 0.50 |
| Exogenous growth in offered wage | $\iota_1^f$ | - | 0.052 |
| Exogenous growth in offered wage | $\iota_2^f$ | - | -0.0006 |
| Women's Human Capital Tech | $\nu$ | 0.003 | - |
| Discount Factor (annualized) | $\beta$ | 0.99 | 0.99 |
| Depreciation rate | $\delta$ | 0.017 | - |

| Targets | Data | Ret to Exp | Baseline |
|---|---|---|---|
| Weekly hours worked | 37.2 | 37.5 | 37.2 |
| Participation Rate | 0.684 | 0.690 | 0.679 |
| Participation Rate of Mothers | 0.538 | 0.544 | 0.546 |
| Observed Wage Gender Gap | 0.720 | 0.716 | 0.727 |
| Observed Var. Wage Growth (Women) | 0.004 | 0.005 | 0.004 |
| Observed Initial Var. of Wages (Women) | 0.14 | 0.15 | 0.15 |
| Wage Growth (if younger than 40) | 0.012 | 0.013 | 0.010 |
| Wage Growth (if older than 40) | 0.001 | 0.013 | 0.004 |
| Median wealth to income ratio | 1.84 | 1.82 | 1.80 |
| Observed Depreciation Rate | -0.050 | -0.040 | 0.02 |

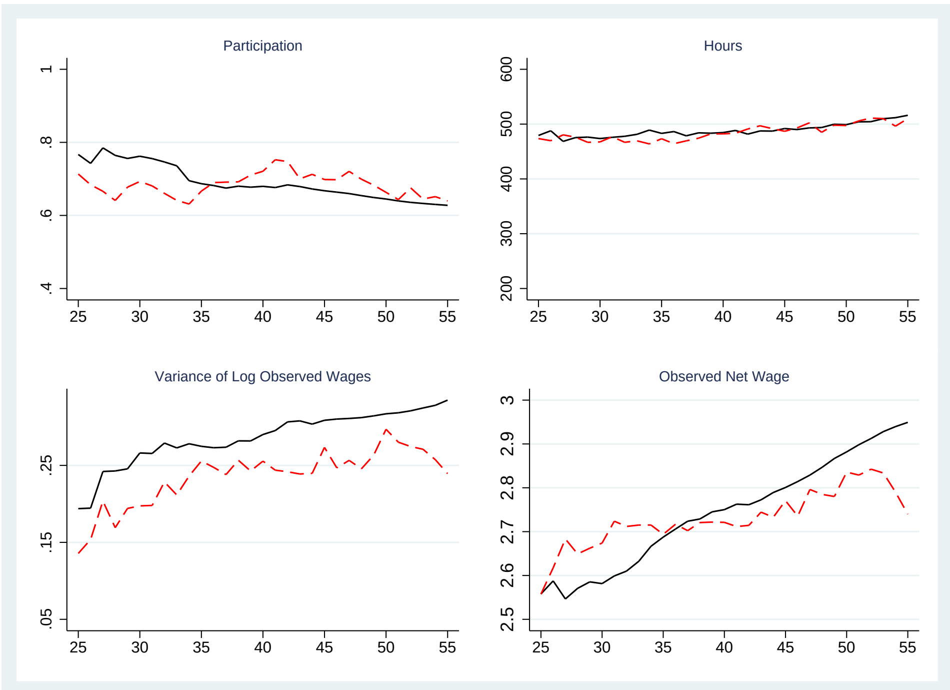
Note: Statistics for women born in the 1950s and aged 25 to 55. Wage growth and depreciation rate are annual.

**Table E.12:** Returns to experience: statistics on hetereogeneity

| | Data | Model |
|---|---|---|
| Participation Rate Mothers with Children Aged 3-17 | 0.682 | 0.672 |
| Participation Rate Childless Women | 0.755 | 0.724 |
| Average Hours Worked 10th Percentile | 20 | 21 |
| Average Hours Worked 25th Percentile | 35 | 31 |
| Average Hours Worked 50th Percentile | 40 | 40 |
| Average Hours Worked 75th Percentile | 40 | 46 |
| Average Hours Worked 90th Percentile | 48 | 50 |
| Wage 10th Percentile | 8.16 | 7.43 |
| Wage 50th Percentile | 15.05 | 15.58 |
| Wage 90th Percentile | 29.23 | 31.71 |

Note: Women without dependent children are women who have never had children and those whose children are over 17.

**Figure E.1:** Life-cycle profiles: baseline model (solid line) versus data (dashed line)



sults show that the response of the participation margin to a transitory anticipated change in the wage (for given preferences on the intensive margin) may be very different depending on the assumption that is made about the nature of the wage growth over the life-cycle (exogenous or endogenous). The extra wage that is provided by an anticipated increase in the wage in a particular period is a small fraction of the total return to participate in that period (in particular at early ages) and then it has a small impact on the participation decision.

**Table E.13:** Returns to experience: Frisch changes

| | Extensive Response | Intensive Elasticity 25th | 50th | 75th | Agg Hours Elasticity | Baseline |
|---|---|---|---|---|---|---|
| 25-29 | 0.02 | 0.65 | 0.81 | 1.15 | 0.91 | 1.85 |
| 30-34 | 0.04 | 0.63 | 0.79 | 1.17 | 0.91 | 1.48 |
| 35-39 | 0.03 | 0.63 | 0.78 | 1.17 | 0.90 | 1.45 |
| 40-44 | 0.03 | 0.61 | 0.79 | 1.19 | 0.89 | 1.35 |
| 45-49 | 0.04 | 0.60 | 0.77 | 1.19 | 0.88 | 1.39 |
| 50-55 | 0.07 | 0.58 | 0.75 | 1.09 | 0.86 | 1.45 |

Notes: The extensive response is the percentage point change in participation in response to a 1% increase in the wage. The aggregate hours elasticity reports the percentage change in hours corresponding to a percentage change in the wage, accounting for changes at both the extensive and intensive margins.

It may well be that the small response of the extensive margin labour supply that we find is related to the simple model of return to experience we have considered. Whether returns to experience operate in a more subtle manner through intensive margins and the number of hours is a question we leave for future research. If that is the case, we would need to change substantially the estimation methods we used in the chapter.

One possibility, of course, is that returns to tenure are important for some occupations and/or skill levels and not for others. In such a case, it would be necessary to introduce an additional dimension of heterogeneity that would make the aggregation issues we have repeatedly stressed even more salient.[6]

## Life-Cycle responses to changes in wage profiles

Finally, in Table E.14 we report the extensive, intensive margin and the macro responses to an increase in the entire wage profile of 10% for both husband and wife. In this case the response both at the extensive and the intensive margin is very similar in the economy with and without returns to experience.

**Table E.14:** Labour supply changes, Marshallian

| | Extensive Response | Intensive Elasticity 25th | 50th | 75th | Agg Hours Elasticity |
|---|---|---|---|---|---|
| Ret to experience | 0.53 | 0.25 | 0.40 | 0.77 | 0.99 |
| Baseline | 0.51 | 0.28 | 0.42 | 0.67 | 0.91 |

Notes: The extensive response is the percentage point change in participation in response to a 1% increase in the wage. The aggregate hours elasticity reports the percentage change in hours corresponding to a percentage change in the wage, accounting for changes at both the extensive and intensive margins.

[6] Alternatively it could be that returns to experience depend on hours worked. Blundell et al. (2016b) show that these returns are close to zero for part-time work.

## E.6 Appendix to Section 7.4: selection correction

In this Appendix we consider an extension of the full model that we calibrate in Section 7.4.3. We allow for taste shocks $\chi$ and $\zeta$ in the utility function. We solve and simulate this economy and explore the ability of our empirical strategy in Section 7.2 to recover the preference parameter values that are assumed in the simulations ($\phi = 0.76$ and $\theta = 1.75$). We discretise both $\chi$ and $\zeta$.

**Table E.15:** MRS estimates with simulated data

| | OLS | IV | IV+Selection Correction |
|---|---|---|---|
| $\phi$ | 0.52*** | 0.61*** | 0.76*** |
| | (0.000130) | (0.00229) | (0.0666) |
| $\theta$ | 1.97*** | 2.20*** | 1.71* |
| | (0.000561) | (0.00528) | (0.902) |

Notes: Standard errors in parentheses.$*p < 0.10$, $**p < 0.05$,$***p < 0.01$

In the first column of Table E.15 we report the OLS estimates of the MRS equation using simulated data. We estimate $\phi = 0.52$ and $\theta = 1.97$. These are clearly biased with respect to the assumed parameter values. As discussed in Section 7.2, there are two reasons for the bias. First, regressors are endogenous since consumption and leisure are correlated with the error term in the MRS equation, and, second, there is non-random self-selection of women into the labour market. In order to address the first issue we solve and simulate the economy for several cohorts of women that differ in the average wage they face. We use the variation in wages across cohorts to instrument consumption and leisure in the MRS equation and we provide the estimates in the second column of Table E.15. Estimated parameter values are still biased with respect to the assumed parameter values. In order to address the consequences of non-random selection we estimate a probit of the participation decision in which we include as exclusion restrictions a cohort dummy and the log of husband earnings (the same as what we used in the data). We then include the selection correction terms as regressors of the MRS equation and report the results in the third column of the Table. In this case estimated parameter values are very close to the assumed parameter values.

# Bibliography


**Aaronson, Daniel and Eric French**, "The Effect of Part-Time Work on Wages: Evidence from the Social Security Rules," *Journal of Labor Economics*, 2004, *22* (2), 329–352.

**Abowd, John M. and David Card**, "On The Covariance Structure of Earnings and Hours Changes," *Econometrica*, March 1989, *57* (2), 411–445.

**Adler, Hans J. and Michael Wolfson**, "A Prototype Micro-Macro Link for the Canadian Household Sector," *Review of Income and Wealth*, 1988, *34* (4), 371–392.

**Adrian, Tobias and Hyun Song Shin**, "Liquidiy and Leverage," *Journal of Financial Intermediation*, 2010, *19*, 418–437.

**Afriat, S. N.**, "The Construction of Utility Functions from Expenditure Data," *International Economic Review*, February 1967, *8* (1), 67–77.

**Aguiar, Mark and Erik Hurst**, "Consumption versus Expenditure," *Journal of Political Economy*, October 2005, *113* (5), 919–948.

**_ and _** , "Life-Cycle Prices and Production," *American Economic Review*, December 2007, *97* (5), 1533–1559.

**_ and _** , "Deconstructing Life Cycle Expenditure," *Journal of Political Economy*, June 2013, *121* (3), 437–492.

**Aguila, Emma, Orazio Attanasio, and Costas Meghir**, "Changes in Consumption at Retirement: Evidence from Panel Data," *Review of Economics and Statistics*, October 2010, *93* (3), 1094–1099.

**Aladangady, Aditya**, "Housing Wealth and Consumption: Evidence from Geographically Linked Microdata," *American Economic Review*, 2017, *107* (11), 3415–3446.

**Alonso-Borrego, César and Manuel Arellano**, "Symmetrically Normalized Instrumental-variable Estimation Using Panel Data," *Journal of Business & Economic Statistics*, 1999, *17* (1), 36–49.

**Altonji, Joseph**, "Intertemporal Substitution in Labor Supply: Evidence from Micro Data," *Journal of Political Economy*, 1986, *94* (3), S176–S215.

**Alvaredo, Facundo, Anthony B. Atkinson, Thomas Piketty, and Emmanuel Saez**, "The World Top Incomes Database," 2012.

**Ameriks, John, Joseph S Briggs, Andrew Caplin, Matthew D Shapiro, and Christopher Tonetti**, "Long-Term-Care Utility and Late-in-Life Saving," Working Paper 20973, National Bureau of Economic Research February 2015.

**Angrist, Joshua D and Alan B Krueger**, "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples," *Journal of the American Statistical Association*, 1992, *87*, 328–336.

**_ and _** , "Split Sample Instrumental Variables," *Journal of Business and Economic Statistics*, 1995, *13* (2), 225–235.

**Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez**, "Top Incomes in the Long Run of History," *Journal of Economic Literature*, 2011, *49* (1), 3–71.

**Attanasio, Orazio and Luigi Pistaferri**, "Consumption Inequality over the Last Half Century: Some Evidence Using the New PSID Consumption Measure," *American Economic Review*, May 2014, *104* (5), 122–126.

— , **Erich Battistin, and Andrew Leicester**, "From Micro to Macro, from Poor to Rich: Consumption and Income in the UK and US," Working Paper, National Poverty Center, Gerald R. Ford School of Public Policy, University of Michigan. 2006.

— , **Erik Hurst, and Luigi Pistaferri**, "The Evolution of Income, Consumption, and Leisure Inequality in The US, 1980-2010," Working Paper 17982, National Bureau of Economic Research April 2012.

**Attanasio, Orazio P.**, *Handbook of Macroeconomics*, Vol. 1, Elsevier,

— **and Guglielmo Weber**, "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from The Consumer Expenditure Survey," *Journal of Political Economy*, 1995, *103* (6), 1121–1157.

— **and** — , "Consumption and Saving: Models of Intertemporal Allocation and their Implications for Public Policy," *Journal of Economic Literature*, 2010, *48* (3), 693–751.

— **and Hamish Low**, "Estimating Euler Equations," *Review of Economic Dynamics*, 2004, *7* (2), 406–435.

— , — **, and Virginia Sánchez-Marcos**, "Explaining Changes in Female Labor Supply in a Life-cycle Model," *American Economic Review*, 2008, *98* (4), 1517–1552.

— , **Laura Blow, Robert Hamilton, and Andrew Leicester**, "Booms and busts: Consumption, House Prices and Expectations," *Economica*, 2009, *76*, 20–50.

**Banks, James, Richard Blundell, and Agar Brugiavini**, "Risk Pooling, Precautionary Saving and Consumption Growth," *The Review of Economic Studies*, October 2001, *68* (4), 757–779.

— , — **, and Arthur Lewbel**, "Quadratic Engel Curves and Consumer Demand," *The Review of Economics and Statistics*, November 1997, *79* (4), 527–539.

__ , __ , **and Sarah Tanner**, "Is There a Retirement-Savings Puzzle?," *The American Economic Review*, 1998, *88* (4), 769–788.

__ , __ , **Zoë Oldfield, and James P. Smith**, "Housing Price Volatility and Downsizing in Later Life," *Research Findings in the Economics of Aging*, February 2010, pp. 337–379.

__ , __ , __ , **and** __ , "Housing Mobility and Downsizing at Older Ages in Britain and the USA," *Economica*, 2012, *79* (313), 1–26.

**Barcellos, Silvia Helena and Mireille Jacobson**, "The Effects of Medicare on Medical Expenditure Risk and Financial Strain," *American Economic Journal: Economic Policy*, November 2015, *7* (4), 41–70.

**Battistin, Erich, Agar Brugiavini, Enrico Rettore, and Guglielmo Weber**, "The Retirement Consumption Puzzle: Evidence from a Regression Discontinuity Approach," *American Economic Review*, December 2009, *99* (5), 2209–2226.

**Beatty, Timothy K. M and Ian A Crawford**, "How Demanding Is the Revealed Preference Approach to Demand?," *American Economic Review*, October 2011, *101* (6), 2782–2795.

**Becker, Gary S.**, "Irrational Behavior and Economic Theory," *Journal of Political Economy*, February 1962, *70* (1), 1–13.

**Bee, Adam, Bruce D. Meyer, and James X. Sullivan**, *Improving the Measurement of Consumption Expenditures*, University of Chicago Press,

**Bénétrix, Agustín, Barry Eichengreen, and Kevin O'Rourke**, "How Housing Slumps End," *Economic Policy*, 2012, *72*, 647–692.

**Berger, David, Veronica Guerrieri, Guido Lorenzoni, and Joseph Vavra**, "House Prices and Consumer Spending," *Review of Economic Studies*, 2017, *3* (1), 15021542.

**Bernheim, B. Douglas, Jonathan Skinner, and Steven Weinberg**, "What Accounts for the Variation in Retirement Wealth among U.S. Households?," *American Economic Review*, September 2001, *91* (4), 832–857.

**Blow, Laura, Martin Browning, and Ian Crawford**, "Nonparametric Analysis of Time-Inconsistent Preferences," Economics Series Working Papers 835, University of Oxford, Department of Economics 2017.

**Blundell, Richard, Alan Duncan, and Costas Meghir**, "Estimating Labor Supply Responses Using Tax Reforms," *Econometrica*, 1998, *66* (4), 827–861.

_ **and Ian Walker**, "A Life-Cycle Consistent Empirical Model of Family Labour Supply Using Cross-Section Data," *Review of Economic Studies*, 1986, *53* (4), 539–558.

_ **and Thomas MaCurdy**, *Handbook of Labor Economics*, Vol. 3, Elsevier,

_ **, Antoine Bozio, and Guy Laroque**, "Labor Supply and the Extensive Margin," *American Economic Review*, 2011, *101* (3), 482–486.

_ **, Costas Meghir, and Pedro Neves**, "Labour Supply and Intertemporal Substitution," *Journal of Econometrics*, 1993, *59* (1-2), 137–160.

_ **, Luigi Pistaferri, and Ian Preston**, "Consumption Inequality and Partial Insurance," *American Economic Review*, December 2008, *98* (5), 1887–1921.

_ **, _ , and Itay Saporta-Eksten**, "Consumption Inequality and Family Labor Supply," *American Economic Review*, 2016, *106* (2), 387–435.

_ **, Martin Browning, and Costas Meghir**, "Consumer Demand and the Life-Cycle Allocation of Household Expenditures," *The Review of Economic Studies*, 1994, *61* (1), 57–80.

_ **, Monica Costa, Costas Meghir, and Jonathan Shaw**, "Female Labor Supply, Human Capital and Welfare Reform," *Econometrica*, 2016, *84* (5), 17051753.

**Boldrin, Michele, Sergi Jimenez-Martni, and Franco Peracchi**, "Social Security and Retirement in Spain," Working Paper 6136, National Bureau of Economic Research August 1997.

**Boskin, M., E. Dulberger, R. Gordon, Z. Griliches, and D. Jorgeon**, *Toward a more accurate measure of the cost of living: final report to the Senate Finance Committee*, Washington DC: US Senate Finance Committee, 1996.

**Brady, Peter J, Glenn B Canner, and Dean M Maki**, "The Effects of Recent Mortgage Refinancing," *Federal Reserve Bulletin*, 2000, *86* (7), 441–450.

**Bronars, Stephen G.**, "The Power of Nonparametric Tests of Preference Maximization," *Econometrica*, May 1987, *55* (3), 693–698. ArticleType: research-article / Full publication date: May, 1987 / Copyright  1987 The Econometric Society.

**Browning, Martin**, "A Nonparametric Test of the Life-Cycle Rational Expections Hypothesis," *International Economic Review*, November 1989, *30* (4), 979–992.

_ , **Angus Deaton, and Margaret Irish**, "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle," *Econometrica*, 1985, *53* (3), 503–544.

_ , **Lars Peter Hansen, and James J. Heckman**, "Micro Data and General Equilibrium Models," in J. B. Taylor and M. Woodford, eds., *Handbook of Macroeconomics*, Vol. 1 of *Handbook of Macroeconomics*, Elsevier, 1999, chapter 8, pp. 543–633.

_ , **Thomas F Crossley, and Joachim Winter**, "The Measurement of Household Consumption Expenditures," *Annual Review of Economics*, 2014, *6*, 475–501.

**Brzozowski, Matthew and Thomas F. Crossley**, "Viewpoint: Measuring the well-being of the poor with income or consumption: a Canadian perspective," *Canadian Journal of Economics/Revue canadienne d'conomique*, 2011, *44* (1), 88–106.

**Buiter, Willem**, "Housing Wealth isn't Wealth," *Economics: The Open-Access, Open Assessment E-Journal*, 2010, *3*, 2010–22.

**Bureau of Labor Statistics**, *Bureau of Labor Statistics Handbook of Methods*, US Bureau of Labor Statistics, November 2011.

_ , *Consumer Expenditure Survey Anthology, 2011* Statistics Report 1030, US Bureau of Labor Statistics, 2011.

**Card, David, Carlos Dobkin, and Nicole Maestas**, "Does Medicare Save Lives?," *The Quarterly Journal of Economics*, May 2009, *124* (2), 597–636.

**Carroll, Christopher Dixon**, "Death to the Log-Linearized Consumption Euler Equation! (And Very Poor Health to the Second-Order Approximation)," *Advances in Macroeconomics*, January 2001, *1* (1).

**Carruthers, A. G., D. J. Sellwood, and P. W. Ward**, "Recent Developments in the Retail Prices Index," *Journal of the Royal Statistical Society. Series D (The Statistician)*, 1980, *29* (1), 1–32.

**Castex, Gonzalo and Evgenia Dechter**, "A Model of Labor Supply, Fixed Costs and Work Schedules," *University of New South Wales, mimeo*, 2016.

**Chang, Yongsung and Sun-Bin Kim**, "From Individual to Aggregate Labor Supply : a Quantitative Analysis Based on a Heterogeneous Agent Macroeconomy," *International Economic Review*, 2006, *47* (1), 1–27.

**Chetty, Raj**, "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 2012, *80* (3), 969–1018.

**_ , Adam Guren, Day Manoli, and Andrea Weber**, "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins," *American Economic Review*, 2011, *101* (3), 471–475.

__ , **László Sandor, and Adam Szeidl**, "The Effect of Housing Portfolio Choice," *Journal of Finance*, 2017, *72* (3), 1171–1212.

**Choi, Hyun-Soo, Harrison Hong, and Jose Scheinkman**, "Speculating on Home Improvements," *Journal of Financial Economics*, 2014, *111*, 609–624.

**Christelis, Dmitris, Dmitris Georgarakos, and Tullio Japelli**, "Wealth Shocks, Unemployment Shocks and Consumption in the Wake of the Great Recession," *Journal of Monetary Economics*, 2015, *72*, 21–41.

**Christensen, Mette**, "Demand Patterns Around Retirement: Evidence from Spanish Panel Data," The School of Economics Discussion Paper Series 0809, Economics, The University of Manchester 2008.

**Clements, Kenneth W., Izan H. Y. Izan, and E. Antony Selvanathan**, "Stochastic Index Numbers: A Review," *International Statistical Review*, 2006, *74* (2), 235–270.

**Cloyne, James, Kilian Huber, Ethan Ilzetzki, and Henrik Kleven**, "The Effect of House Prices on Household Borrowing: A New Approach," *American Economic Review*, Forthcoming.

**Congressional Budget Office**, "Financing Long-Term Care for the Elderly | Congressional Budget Office," April 2004.

**Cooper, Daniel**, "Did Easy Credit Lead to Overspending? Home Equity Borrowing and Household Behavior in the Early 2000s," *Federal Reserve Bank of Boston Public Policy Discussion Papers 09-7*, 2010.

__ , "House Price Fluctuations: the Role of Housing Wealth as Borrowing Collateral," *The Review of Economics and Statistics*, 2013, *95* (4), 1183–1197.

**Crawford, Rowena and George Stoye**, "The Prevalence and Dynamics of Social Care Receipt," IFS Report R125, Institute for Fiscal Studies February 2017.

**Crossley, Thomas F. and Cormac O'Dea**, "The Wealth and Saving of UK families on the Eve of the Financial Crisis," IFS Report R71, Institute for Fiscal Studies July 2010.

_ **and Krishna Pendakur**, "The Common-Scaling Social Cost-of-Living Index," *Journal of Business & Economic Statistics*, 2010, *28* (4), 523–538.

**Dalén, J**, "Computing Elementary Aggregates in the Swedish Consumer Price Index," *Journal of Official Statistics*, 1992, *8*, 129–147.

_ , "A Note on the Variance of the Sample Geometric Mean," Research Report 1, Department of Statistics, Stockholm University, Stockholm 1999.

**D'Alessio, Giovanni and Ivan Faiella**, "Non-Response Behaviour in the Bank of Italy's Survey of Income and Wealth," Temi di discussione del Servizio Studi 462, Bank of Italy, Economic Research and International Relations Area 2002.

**Dalton, K, J. Greenlees, and K Stewart**, "Incorporating a Geometric Mean Formula into the CPI," *Monthly Labor Review*, 1998, *121*, 3–7.

**Danforth, John**, *Studies in the Economics of Search of Job Search*, New York: New-Holland,

**Davidson, Russell and James G. MacKinnon**, *Econometric Theory and Methods*, Oxford University Press, 2004.

**de Leeuw, Edith and Wim de Heer**, *Survey Nonresponse*, New York: Wiley,

**Deaton, Angus**, "Panel data from Time Series of Cross-sections," *Journal of Econometrics*, 1985, *30*, 109–126.

_ , "Measuring Poverty in a Growing World (Or Measuring Growth in a Poor World)," *The Review of Economics and Statistics*, 2005, *87* (1), 1–19.

_ **and John Muellbauer**, "An Almost Ideal Demand System," *The American Economic Review*, 1980, *70* (3), 312–326.

**DeNardi, Mariacristina, Eric French, and JohnB. Jones**, "Why Do the Elderly Save? The Role of Medical Expenses," *Journal of Political Economy*, 2010, *118* (1), 39–75.

**Department for Environment, Food and Rural Affairs, Office for National Statistics**, "Living Costs and Food Survey (Expenditure and Food Survey)," *SN: 2000028*, 2016.

**Diewert, W. E.**, "Afriat and Revealed Preference Theory," *The Review of Economic Studies*, 1973, *40* (3), 419–425.

_ , *Essays in Index Number Theory*, Amsterdam: North-Holland, 1992.

_ , "Fisher Ideal Output, Input, and Productivity Indexes Revisited," *Journal of Productivity Analysis*, September 1992, *3* (3), 211–248.

_ , "Axiomatic and Economic Approaches to Elementary Prices Indexes," Discussion Paper 1202, Department of Economics, University of British Columbia, Vancouver 1995.

_ , "On the Stochastic Approach to Index Numbers," Archive 9531, Department of Economics, University of British Columbia, Vancouver 1995.

_ , "Answers to Questions Arising from the RPI Consultation," Technical Report, Department of Economics, University of British Columbia, Vancouver 2012.

_ , "Consumer Price Statistics in the UK," Technical Report, Office for National Statistics, Newport 2012.

_ , "Irving Fisher and Index Number Theory," Discussion Paper 9501, Department of Economics, University of British Columbia, Vancouver 2012.

**Diewert, W. Erwin**, "The Consumer Price Index and index number purpose," *Journal of Economic and Social Measurement*, January 2001, *27* (3,4), 167–248.

**Disney, Richard and John Gathergood**, "House Price Growth, Collateral Constraints and the Accumulation of Homeowner Debt in the United States," *The B.E. Journal of Macroeconomics*, 2011, *11* (1), 1–30.

**_ , _ , and Andrew Henley**, "House Price Shocks, Negative Equity, and Household Consumption in the United Kingdom," *Journal of the European Economics Association*, 2010, *8* (6), 1179–1207.

**_ , Sarah Bridges, and John Gathergood**, "House Price Shocks and Household Indebtedness in the United Kingdom," *Economica*, 2010, *77*, 472–496.

**Dynan, Karen**, "Is a Household Debt Overhang Holding Back Consumption," *Brookings Papers on Economic Activity*, 2012.

**Edgar, Jennifer and Adam Safir**, *Consumer Expenditure Survey Anthology, 2011* number Statistics Report no. 1030, US Bureau of Labor Statistics, 2011.

**Edgeworth, F. Y.**, *Papers Relating to Political Economy*, Vol. 1, New York: Franklin,

**Eichhorn, W**, *Functional equations in economics*, Reading: Addison-Wesley, 1978.

**_ and J Voeller**, *Theory of the Price Index*, Berlin: Springer, 1976.

**Elliott, D, R ONeill, J Ralph, and R Sanderson**, "Stochastic and Sampling Approaches to the Choice of Elementary Aggregate Formula," Discussion Paper, Office for National Statistics, Newport 2012.

**Erosa, Andres, Luisa Fuster, and Gueorgui Kambourov**, "Towards a Microfounded Theory of Aggregate Labor Supply," *Review of Economic Studies*, 2016, *83* (3), 1001–1039.

**Fahle, Sean, Kathleen McGarry, and Jonathan Skinner**, "Out-of-Pocket Medical Expenditures in the United States: Evidence from the Health and Retirement Study," *Fiscal Studies*, 2016, *37* (3-4), 785–819.

**Feenberg, Daniel and Elisabeth Coutts**, "An Introduction to the TAXSIM Model," *Journal of Policy Analysis and Management*, 1993, *12* (1), 189–194.

**Fenwick, D**, "The Impact of Choice of Base Month and Other Factors on the Relative Performance of Different Formulae Used for Aggregation of Consumer Price Index Data at an Elementary Aggregate Level," Technical Report, 5th Meet. Ottawa Group, Reykjavik August 1999.

**Fernandez-Corugedo, Emilio and John Muellbauer**, "Consumer Credit Conditions in the United Kingdom," *Bank of England working papers 314*, 2006.

**Finkelstein, Amy, Erzo F. P. Luttmer, and Matthew J. Notowidigdo**, "Approaches to Estimating the Health State Dependence of the Utility Function," *American Economic Review*, May 2009, *99* (2), 116–121.

**_ , _ , and _** , "What Good Is Wealth Without Health? The Effect of Health on the Marginal Utility of Consumption," *Journal of the European Economic Association*, 2013, *11* (s1), 221–258.

**Fisher, Irving**, *The Purchasing Power of Money*, London: Macmillan, 1911.

_ , *The Making of Index Numbers*, 1 ed., Boston: Houghton Mifflin, 1922.

**Fuller, Wayne A.**, "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 1977, *45* (4), 939–953.

**Garner, Thesia I., George Janini, William Passero, Laura Paskkiewicz, and Mark Vendemia**, "The CE and the PCE: a Comparison," *Monthly Labor Review*, September 2006, pp. 20–46.

**_ , Robert McClelland, and William Passero**, "Strengths and Weaknesses of the Consumer Expenditure Survey," Draft presented at the National Bureau of Economic Research Summer Institute, Cambridge MA, July 13-14 2009.

**Ghent, Andra C. and Marianna Kudlyak**, "Recourse and Residential Mortgage Default: Evidence from US States," *Review of Financial Studies*, September 2011, *24* (9), 31393186.

**Goldenberg, Karen and Jay Ryan**, "Evolution and Change in the Consumer Expenditure Surveys: Adapting Methodologies to Meet Changing Needs," National Bureau of Economic Research Conference on Income and Wealth, Summer Institute Cambridge MA, July 13-14 2009.

**Goldin, Claudia and Joshua Mitchell**, "The New Life Cycle of Womens Employment: Disappearing Humps, Sagging Middles, Expanding Tops," *Journal of Economic Perspectives*, 2017, *31* (1), 161182.

_ **and Lawrence F. Katz**, "Long-Run Changes in the Wage Structure: Narrowing, Widening, Polarizing," *Brookings Papers on Economic Activity*, 2007, *2*, 135–167.

**Gooding, Philip**, "Consumer Prices Index and Retail Prices Index: the 2012 Basket of Goods and Services," "", Office for National Statistics, Newport 2012.

**Gorman, William M.**, "Separable Utility and Aggregation," *Econometrica*, 1959, *27* (3), 469481.

**Greenlees, J. S.**, "Random Errors and Superlative Indexes," Working Paper 343, Bureau of Labor Statistics, Washington DC 2001.

**Guner, Nezih, Remzi Kaygusuz, and Gustavo Ventura**, "Taxation and Household Labour Supply," *Review of Economic Studies*, 2012, *79*, 11131149.

**Gyourko, Joseph and Albert Saiz**, "Reinvestment in the Housing Stock: the Role of Construction Costs and the Supply Side," *Journal of Urban Economics*, 2004, *55*, 238–256.

**Haag, Berthold R., Stefan Hoderlein, and Krishna Pendakur**, "Testing and Imposing Slutsky Symmetry in Non-parametric Demand Systems," *Journal of Econometrics*, November 2009, *153* (1), 33–50.

**Hahn, Jinyong and Jerry Hausman**, "Weak Instruments: Diagnosis and Cures in Empirical Econometrics," *American Economic Review*, 2003, *93* (2), 118–125.

**_ , _ , and Guido Kuersteiner**, "Estimation with Weak Instruments: Accuracy of Higher-order Bias and MSE Approximations," *Econometrics Journal*, 2004, *7* (1), 272–306.

**Hall, Robert E.**, "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, December 1978, *86* (6), 971–987.

**Hansen, Lars Peter, John Heaton, and Amir Yaron**, "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business & Economic Statistics*, 1996, *14*, 262–280.

**Hardy, G. H., J. E. Littlewood, and G. Polya**, *Inequalities*, Cambridge: Cambridge University Press, 1934.

**Hausman, Jerry, Whitney K. Newey, Tiemen Woutersen, John C. Chao, and Norman R. Swanson**, "Instrumental Variable Estimation with Heteroskedasticity and Many Instruments," *Quantitative Economics*, 2012, *3* (2), 211–255.

**Hayashi, Fumio**, "Tests for Liquidity Constraints: A Critical Survey," in Truman Bewley, ed., *Advances in Econometrics, Fifth World Congress*, Vol. 2, Cambridge University Press, 1987, pp. 91–120.

**Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante**, "Unequal we stand: An Empirical Analysis of Economic Inequality in the United States, 19672006," *Review of Economic Dynamics*, January 2010, *13* (1), 15–51.

**Heckman, James J. and Thomas E. Macurdy**, "A Life Cycle Model of Female Labour Supply," *The Review of Economic Studies*, 1980, *47* (1), 47–74.

**Henderschott, Patric H., Gwilym Pryce, and Michael White**, "Household Leverage and the Deductibility of Home Mortgage Interest: Evidence from U.K House Purchasers," *Journal of Housing Research*, 2003, *14* (1), 49–82.

**Hirsch, Barry T.**, "Why Do Part-Time Workers Earn Less? The Role of Worker and Job Skills," *ILR Review*, July 2005, *58* (4), 525–551.

**Huggett, Mark, Gustavo Ventura, and Amir Yaron**, "Sources of Lifetime Inequality," *American Economic Review*, 2011, *101* (7), 2923–2954.

**Hurd, Michael D.**, "Mortality Risk and Bequests," *Econometrica*, 1989, *57* (4), 779–813.

**Hurst, Erik**, "The Retirement of a Consumption Puzzle," Working Paper 13789, National Bureau of Economic Research February 2008.

**Hyslop, Dean R.**, "Rising U.S. Earnings Inequality and Family Labor Supply: The Covariance Structure of Intrafamily Earnings," *American Economic Review*, 2001, *91* (4), 755–777.

**Imai, Susumu and Michael P. Keane**, "Intertemporal Labor Supply and Human Capital Accumulation," *International Economic Review*, 2004, *45* (2), 601–641.

**Inoue, Atsushi and Gary Solon**, "Two-sample Instrumental Variables Estimators," *The Review of Economics and Statistics*, 2010, *92* (3), 557–561.

**International Labor Organization**, *Consumer Price Index Manual: Theory and Practice*, Geneva: International Labor Office, 2004.

**Iskhakov, Fedor, Thomas H. Jørgensen, John Rust, and Bertel Schjerning**, "The Endogenous Grid Method for Discrete-continuous Dynamic Choice Models with (or without) Taste Shocks," *Quantitative Economics*, 2017, *8* (2), 317–365.

**Jaynes, E. T.**, "Information Theory and Statistical Mechanics," *Physical Review*, May 1957, *106* (4), 620–630.

—, "Information Theory and Statistical Mechanics. II," *Physical Review*, October 1957, *108* (2), 171–190.

—, *Papers on Probability, Statistics and Statistical Physics*, Dordrecht: Reidel, 1983.

**Jevons, W. S.**, *Investigations in Currency and Finance*, London: Macmillan,

**Jordá, Óscar, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan M. Taylor**, "The Rate of Return on Everything, 1870-2015," *NBER Working Paper No. 24112*, 2017.

**Keane, Michael P.**, "Labor Supply and Taxes: A Survey," *Journal of Economic Literature*, 2011, *49* (4), 961–1075.

_ **and Nada Wasi**, "Labor Supply: the Roles of Human Capital and the Extensive Margin," *The Economic Journal*, 2016, *126*, 578–617.

_ **and Richard Rogerson**, "Micro and Macro Labor Supply Elasticities: A Reassessment of Conventional Wisdom," *Journal of Economic Literature*, 2012, *50* (2), 464–476.

_ **and** _ , "Reconciling Micro and Macro Labor Supply Elasticities: A Structural Perspective," *Annual Review of Economics*, 2015, *7*, 89–117.

**Keynes, J. M.**, *Treatise on Money*, Vol. 1, London: Macmillan, 1930.

**Kimmel, Jean and Thomas J. Kniesner**, "New Evidence on Labor Supply: Employment versus Hours Elasticities by Sex and Marital Status," *Journal of Monetary Economics*, 1998, *42* (2), 289–301.

**Konüs, A. A.**, "The Problem of the True Index of the Cost of Living," *Econometrica*, 1939, *7* (1), 10–29.

**Krueger, Dirk, Fabrizio Perri, Luigi Pistaferri, and Giovanni L. Violante**, "Cross-sectional Facts for Macroeconomists," *Review of Economic Dynamics*, January 2010, *13* (1), 1–14.

**Laspeyres, E.**, "Die Berechnung einer mittleren Waarenpreissteigerung," *Jahrbcher fr Nationalkonomie und Statistik / Journal of Economics and Statistics*, 1981, *196* (3), 218–237.

**Lehnert, Andreas**, "Housing, Consumption, and Credit Constraints," *Board of Governers of the Federal Reserve System of Finance and Economis Discussion Paper*, 2004.

**Lillard, Lee and Yoram Weiss**, “Uncertain Health and Survival: Effects on End-of-Life Consumption,” *Journal of Business & Economic Statistics*, April 1997, *15* (2), 254–268.

**Ljungqvist, Lars and Thomas J. Sargent**, “A Labor Supply Elasticity Accord?,” *American Economic Review*, 2011, *101* (3), 487–491.

**Love, David A., Michael G. Palumbo, and Paul A. Smith**, “The Trajectory of Wealth in Retirement,” *Journal of Public Economics*, February 2009, *93* (1), 191–208.

**Low, Hamish**, “Self-insurance in a Life-cycle Model of Labour Supply and Savings,” *Review of Economic Dynamics*, 2005, *8* (4), 945–975.

**Luengo-Prado, Mara Jos and Almudena Sevilla**, “Time to Cook: Expenditure at Retirement in Spain,” *The Economic Journal*, 2013, *123* (569), 764–789.

**MaCurdy, Thomas**, “An Empirical Model of Labour Supply in a Life Cycle Setting,” *Journal of Political Economy*, 1981, *89*, 1059–1085.

_ , “A Simple Scheme for Estimating an Intertemporal Model of Labor Supply and Consumption in the Presence of Taxes and Uncertainty,” *International Economic Review*, 1983, *24* (2), 265–289.

**Mariano, Robert and Takamitsu Sawa**, “The Exact Finite-Sample Distribution of the Limited-Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables,” *Journal of the American Statistical Association*, 1972, *67* (337), 159–163.

**Meghir, Costas and David Phillips**, “Labour Supply and Taxes,” *IZA WP*, 2008, *3405.*

**Melzer, Brian**, “Mortgage Debt Overhang: Reduced Investment by Homeowners at Risk of Default,” *Journal of Finance*, 2017, *72* (2), 575–612.

**Merton, Robert**, "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case," *The Review of Economics and Statistics,*, 1969, *51* (3), 247–257.

**Meyer, Bruce D and James X Sullivan**, "Five Decades of Consumption and Income Poverty," Working Paper 14827, National Bureau of Economic Research March 2009.

**Mian, Atif and Amir Sufi**, "House Price, Home-Equity-Based Borrowing, and the US Household Leverage Crisis," *American Economic Review*, 2011, *101* (5), 2132–2156.

_ **, Kamalesh Rao, and Amir Sufi**, "Household Balance Sheets, Consumption, and the Economic Slump," *Quarterly Journal of Economics*, 2013, *128* (4), 1687–1726.

**Morgan, Darren and Philip Gooding**, "CPI and RPI: Increased Impact of the Formula Effect in 2010," Technical Report, Office for National Statistics, Newport 2010.

**National Audit Office**, "Adult social care in England: overview," Report by the Comptroller and Auditor General March 2014.

**Nelson, Charles R. and Richard Startz**, "Some Further Results on the Exact Small Sample Properties of the Instumental Variables Estimator," *Econometrica*, 1990, *58* (4), 967–976.

**OECD**, "Long-term Care for Older People," OECD Publishing, Paris, https://doi.org/10.1787/9789264015852-en. 2005.

**Office for National Statistics**, "Living Costs and Food Survey 2009 User Guide," Technical Report, Office for National Statistics 2010.

_ , "Implications of the Differences Between the Consumer Prices Index and the Retail Prices Index," Technical Report, Office for National Statistics, Newport 2011.

__ , *Consumer Price Indices Technical Manual*, Newport: Office for National Statistics, 2012.

__ , "National Statistician Announces Outcome of Consultation on RPI," Technical Report, Office for National Statistics, Newport 2012.

__ , "National Statistician's consultation on options for improving the Retail Prices Index," Technical Report, Office for National Statistics, Newport 2012.

**O'Shaughnessy, Carol V.**, "National Spending for Long-Term Services and Supports (LTSS), 2012," National Health Policy Forum [accessed on August 12, 2018]. Available at http://www.nhpf.org/library/details.cfm/2783 2014.

**Pacini, David and Frank Windmeijer**, "Robust inference for the Two-Sample 2SLS estimator," *Economics Letters*, 2016, *146*, 50–54.

**Palumbo, Michael G.**, "Uncertain Medical Expenses and Precautionary Saving Near the End of the Life Cycle.," *Review of Economic Studies*, 1999, *66* (2), 395–421.

**Passero, William, Thesia Garner, and Clinton McCully**, *Improving the Measurement of Consumption Expenditures*, Vol. 74 of *Studies in Income and Wealth*, University of Chicago Press,

**Peeters, Hans, Annelies Debels, and Rika Verpoorten**, "Excluding Institutionalized Elderly from Surveys: Consequences for Income and Poverty Statistics," *Social Indicators Research*, January 2013, *110* (2), 751–769.

**Pollak, Robert A.**, "Conditional Demand Functions and Consumption Theory," *The Quarterly Journal of Economics*, 1969, *83* (1), 60–78.

**Rendall, Michael, Encarnacion Aracil, Christos Bagavos, Christine Couet, Alessandra DeRose, Paola DiGiulio, Trude Lappegard, Isabelle Robert-Bobée, Marit Rønsen, Steve Smallwood, and Georgia Verropoulou**, "Increasingly Heterogeneous Ages at First Birth by Education in Southern European and

Anglo-American Family-Policy Regimes: A Seven-country Comparison by Birth Cohort," *Population Studies*, 2010, *64* (3), 209–227.

**Rogerson, Richard and Johanna Wallenius**, "Micro and Macro Elasticities in a Life Cycle Model with Taxes," *Journal of Economic Theory*, 2009, *144* (6), 2277–2292.

**Sabelhaus, John, David Johnson, Stephen Ash, David Swanson, Thesia I. Garner, John Greenlees, and Steve Henderson**, *Is the Consumer Expenditure Survey Representative by Income?*, Vol. 74 of *Studies in Income and Wealth*, University of Chicago Press,

**Sawa, Takamitsu**, "Finite-Sample Properties of the k-Class Estimators," *Econometrica*, 1972, *40* (4), 653–680.

**Selten, Reinhard**, "Properties of a Measure of Predictive Success," *Mathematical Social Sciences*, 1991, *21* (2), 153–167.

**Selvanathan, A. E. and D. S. Prasada Rao**, *Index Numbers: a Stochastic Approach*, Ann Arbor: University of Michigan Press, 1994.

**Shannon, C. E.**, "A Mathematical Theory of Communication," *Bell System Technical Journal*, 1948, *27* (3), 379–423.

**Slesnick, Daniel T.**, "Aggregate Consumption and Saving in the Postwar United States," *The Review of Economics and Statistics*, 1992, *74* (4), 585–597.

**Smith, Sarah**, "The Retirement-Consumption Puzzle and Involuntary Early Retirement: Evidence from the British Household Panel Survey*," *The Economic Journal*, 2006, *116* (510), C130–C148.

**Staiger, David and James Stock**, "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 1997, *65* (3), 557–586.

**Staiger, Douglas and James H. Stock**, "IV Regression with Weak Instruments," *Econometrica*, 1997, *65* (3), 557–586.

**Statistics Canada**, "Methodology of the Survey of Household Spending," Technical Report Catalogue no. 62F0026MIE-01003, Income Statistics Division, Statistics Canada 2001.

_ , "User Guide for the Public-Use Microdata File Survey of Household Spending, 2009," Technical Report Catalogue no. 62M0004XCB, Income Statistics Division, Statistics Canada 2011.

**Steinsson, Jòn and Emi Nakamura**, "Identification in Macroeconomics," *Journal of Economic Perspectives*, 2018, *32* (3), 59–86.

**Stock, James H. and Motohiro Yogo**, "Testing for Weak Instruments in Linear IV Regression," in Donald W. K. Andrews and James H. Stock, eds., *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge University Press, 2005, chapter 6, pp. 80–108.

**Tourangeau, Roger**, "Survey Research and Societal Change," *Annual Review of Psychology*, 2004, *55*, 775–801.

**Tremblay, J., J. Lynch, and G. Dubreuil**, "Pilot Survey Results from the Canadian Survey of Household Spending Redesign," Paper presented at the 2010 Joint Statistical Meetings of the American Statistical Association, Vancouver British Columbia 2010.

**University of Essex. Institute for Social and Economic Research**, "British Household Panel Survey: Waves 1-18, 1991-2009.," *7th Edition. UK Data Service. SN: 5151*, 2010.

**University of Essex. Institute for Social and Economic Research.**, "Understanding Society: Waves 1-6, 2009-2015.," *8th Edition. UK Data Service. SN: 6614*, 2016.

**U.S. Bureau of Economic Analysis (BEA)**, *Concepts and Methods of the U.S. National Income and Product Accounts*, Washington, D.C: BEA, 2016.

**Varian, Hal R.**, “The Nonparametric Approach to Demand Analysis,” *Econometrica*, July 1982, *50* (4), 945–973.

**Varian, Hal R**, “Revealed Preference with a Subset of Goods,” *Journal of Economic Theory*, October 1988, *46* (1), 179–185.

**von der Lippe, Peter**, “Covariances and Relationships between Price Indices: Notes on a Theorem of Ladislaus von Bortkiewicz in Linear Index Functions,” Paper 38566, Munich Personal RePec Archive, University of Munich, Munich 2012.

**Walsh, C. M.**, *The Measurement of General Exchange Value*, New York: Macmillan, 1901.

**Ziliak, James P. and Thomas J. Kniesner**, “Estimating Life-cycle Labor Supply Tax Effects,” *Journal of Political Economy*, 1999, *107* (2), 326–359.