Labor Supply Models: Unobserved Heterogeneity, Nonparticipation and Dynamics

Richard Blundell*, Thomas MaCurdy† and Costas Meghir‡

December 2006

Abstract

This chapter is concerned with the identification and estimation of models of labor supply. The focus is on the key issues that arise from unobserved heterogeneity, nonparticipation and dynamics. We examine the simple 'static' labor supply model with proportional taxes and highlight the problems surrounding nonparticipation and missing wages. The difference in differences approach to estimation and identification is developed within the context of the labour supply model. We also consider the impact of incorporating nonlinear taxation and welfare programme participation. Family labor supply is looked at from botht e unitary and collective persepctives. Finally we consider intertemporal models focusing on the difficulties that arise with participation and heterogeneity.

Keywords: Labour supply, consumption, taxation, microeconometrics.

JEL: D1, D9, J2.

*University College London and Institute for Fiscal Studies

†Department of Economics, Stanford University and Hoover Institute.

‡University College London and Institute for Fiscal Studies

1. Introduction

This chapter is concerned with the identification and estimation of labor supply models. The specification and estimation of such models has already been the subject of numerous studies and surveys. So why this one? The overall objective of this chapter is to consider models that allow policy evaluation and simulation allowing for individual heterogeneity. Evaluation concerns the assessment of reforms that have taken place. Policy simulation concerns the assessment of proposed reforms. For the most part it is the latter that has been the central concern of empirical researchers. That is to construct a model that can reliably be used for the assessment of proposed reforms. Since many policy proposals involve the reform of highly nonlinear budget constraints and impact decisions that are discrete and cover the whole life-cycle, we argue that a fully specified dynamic structural model is the ideal. In particular, it is of central importance to consider how labour supply and savings decisions interact and how policy affects both labour supply decisions within a period as well as intertemporally. However, this ideal has a number of practical and theoretical difficulties. In certain situations, the evaluation of existing reforms can be analyzed using much simpler and potentially more robust techniques.

To best convey the set of issues surrounding estimation of labour supply models we start with the simplest static framework and build up to the more complete dynamic models, adding important elements such as nonlinear budget sets on the way. Thus, the layout of the chapter is as follows. Section 2 presents an assessment of the estimation issues underlying the simple 'static' labor supply model with proportional taxes and highlights the problems surrounding nonparticipation and missing wages. In section 3 we consider the natural experiment and difference-in-differences approaches to estimation and evaluation of reforms, laying out the identifying assumptions underlying interpretation of the results. We consider estimation of a simple discrete policy response parameter as well as the estimation of income and substitution effects. In section 4 we examine the impact of incorporating nonlinear taxation and welfare programme participation. Section 5 considers some of the specific issues that relate to family labor supply, including the development of the collective approach and welfare programme participation as previously articulated. Section 6 discusses intertemporal labor supply models.

This section reviews the various approaches taken to dynamic modeling and examines the difficulties that arise with participation and heterogeneity. Section 7 concludes the chapter.

2. Estimation and Identification with Participation with Proportional Taxes

We begin by considering the simple static model of hours and consumption choices. We leave the discussion of nonlinear budget sets to section 4.

2.1. Static Specifications

2.1.1. The Allocation of Hours and Consumption

Utility is defined over hours of work h and consumption c, both of which are restricted to be non-negative and h is restricted to be below a maximal amount of an available time endowment. Formally, this discussion is easily extended to the case of family labor supply decisions where h is a vector of household labor supplies. However, there are many specific issues relating to joint participation decisions and to the allocation of resources within the family that are central to any study of family labor supply; we leave our discussion of family labor supply models to section 5. Equally, consumption decisions can be disaggregated. This disaggregation is central to the analysis of nonseparability of goods and leisure. We turn to this below.

If we let y represent the total unearned income available for consumption, and w the real wage rate, then the optimal choices for c and h are given by the solution to

$$\max_{c,h} \{U(c,h)|c{-}wh = y; c{\geq}0; h \geq 0\} \tag{2.1}$$

where $U(c,h)$ is a quasiconcave utility index defined increasing in c and $-h$. The resulting labor supply has the form

$$h = h(w, y). \tag{2.2}$$

In the static model y is taken to be income from other sources. However it turns out that the precise definition of y is crucial: If y is measured as the difference between total consumption expenditure and earnings, $c - wh = y$, it is consistent both with intertemporal two-stage budgeting both in the absence of liquidity constraints and with the presence of liquidity constraints that are unrelated to labour supply. This is discussed in a subsection below.

The indirect utility representation of preferences is given by

$$V(w, y) \equiv U(wh(w, y) + y, h(w, y)) \tag{2.3}$$

which is linear homogeneous, quasi-concave in $\mathbf{p}$, w and y, decreasing in $\mathbf{p}$ and w and increasing in y. The various representations of preferences (direct or indirect utility) detailed below are going to be particularly useful in specifying empirical models and defining the likelihood function.

2.1.2. Two-Stage Budgeting Specifications and Within Period Allocations

Labor supply and consumption models are frequently analyzed in a two good framework. Such modelling is less restrictive than it sounds because under Gorman's (1959, 1968) two stage budgeting, this labor supply model can be seen as the top stage where "full income" is shared between consumption and leisure and then the consumption budget is split among goods. However, for such an interpretation with all goods being represented by one or two price indices, we require some conditions on preferences.

Suppose utility is defined over hours of work h and a vector of goods $\mathbf{q}$. Assume the individual has a within period utility function of the form

$$v_t = v(c_t, h_t, \mathbf{p}_t) = max_{\mathbf{q},h} \left\{ u\left(\mathbf{q}_t, h_t\right) | \mathbf{p}_t' \mathbf{q}_t = c_t \right\} \tag{2.4}$$

where $\mathbf{p}_t$ is a vector of prices corresponding to the disaggregated commodity vector $\mathbf{q}_t$. The function v_t is a conditional indirect utility function which is increasing in total consumption expenditure c_t, decreasing and concave in prices and decreasing in hours of work h_t.

We say that $\mathbf{q}_t$ is weakly separable from h_t if the marginal rate of substitution between goods $\mathbf{q}_t$ does not depend on h_t. In this case the utility function can be written as $u\left(u_1(\mathbf{q}_t), h_t\right)$ where u_1 is a sub-utility function. If in addition the marginal utilities of $\mathbf{q}_t$ and h_t do not depend on each other then we say that the utility function is additively separable, in which case the utility function can be written as $u_1(\mathbf{q}_t) + u_2(h_t)$. Blackorby, Primont and Russell (1978) have a detailed analysis of the concepts of separability and Deaton (1978) analyses the empirical implications of the additive separability assumption.

Gorman has shown that if a set of goods x_1 is separable from goods x_2 then it is possible to express the demands for goods x_1 simply as a function of the total expenditure allocated to this group (x_1) and the prices of these goods alone (say p_1). In addition, if preferences can be expressed in the generalized Gorman polar form, then it is possible to express the overall expenditure allocations to each group as a function of the price indices for each group. This theorem can justify considering the allocation of total expenditure to overall consumption and leisure separately from the problem of how expenditure is allocated to goods. However, it has to be borne in mind that the justification which allows us to write labor supply as a function of the real wage alone (rather than of all relative prices) does imply restrictions on preferences.

These results offer a justification of the static model within an intertemporal context since the concept of separability can extend both over goods and over time.[1] Typically we impose additive separability over time in which case the marginal utility of consumption or hours of work in one period is unaffected by consumption and hours in any other time period. Additive intertemporal separability has the implication that we can use two stage budgeting to characterize consumption choices: given the level of consumption and separability, the within period demands for goods $\mathbf{q}_t$ only depend on the prices of those goods and on wages (if the goods are not separable from hours). The indirect utility function defined by (2.4) then becomes the criterion function for allocating consumption (and hours) over the life-cycle.[2]

It is well known that taking a monotonic transformation of the utility function does not change the observed within period allocations. In an intertemporal context this issue acquires a special importance: taking a monotonic transformation does not alter the way that consumption and hours are allocated within period,

[1] See Gorman (1959), MaCurdy (1983), Blundell and Walker (1986) and Arellano and Meghir (1992)

[2] Utility (2.4) implicitly assumes separability over time thus ruling out habits and / or adjustments costs (see Hotz, Kydland and Sedlacek (1988) and Meghir and Weber (1996)).

under intertemporal separability. However, it does potentially change the marginal rate of substitution between periods. Hence, as we will discuss further below, estimating intertemporal preferences generally requires intertemporal data.

Noting that modeling the monotonic transformation is modeling intertemporal preferences, we use the slightly more elaborate notation

$$v_t = \psi\left[U(c_t, h_t | z_{1t}), z_{2t}\right] \tag{2.5}$$

where $\psi[\cdot]$ is a monotonic function of its first argument u and where z_1 and z_2 are variables (observed or otherwise) that affect preferences over consumption and hours of work. In particular, z_{2t} affects intertemporal allocations but not within period ones (unless it contains common elements with z_{1t}). Our focus in this section is on within period allocations. The discussion here should make it clear that one can work with the utility function (2.1) to represent within period allocations of consumption and hours of work consistent with life-cycle choices.

2.1.3. Empirical Labor Supply Specifications

Preferences can be represented by direct utility functions, indirect utility functions or by the labor supply equation itself. In each case the function has satisfy some basic properties to be consistent with theory. Here we briefly review of some standard specifications of the static labor supply model (2.2) and relate them to their indirect utility function. Such specifications are usually chosen for ease of estimation and here we simply consider the specifications and their underlying model of preferences. With unobserved heterogeneity and nonparticipation it is useful, if not essential, to have some relatively simple parametric specification in mind.

The *linear labor supply* model

$$h = \alpha + \beta w + \gamma y \tag{2.6}$$

has indirect utility

$$V(w, y) = e^{\gamma w}(y + \frac{\beta}{\gamma} w - \frac{\beta}{\gamma^2} + \frac{\alpha}{\gamma}) \text{ with } \gamma \leq 0 \text{ and } \beta \geq 0. \tag{2.7}$$

Although popular (see Hausman (1981,1985), for example), it is arguable that this linear specification allows too little curvature with wages.

Alternative *semilog specifications* and their generalizations are also popular in empirical work. For example, the semilog specification

$$h = \alpha + \beta \ln w + \gamma y \tag{2.8}$$

with indirect utility

$$V(w,y) = \frac{e^{\gamma w}}{\gamma}(\gamma y + \alpha\frac{\beta}{\gamma} + \beta \ln w) - \frac{\beta}{\gamma}\int_{\gamma y} \frac{e^{\gamma y}}{\gamma y} d(\gamma y) \text{ with } \gamma \leq 0 \text{ and } \beta \geq 0. \tag{2.9}$$

Moreover, the linearity of (2.8) in α and $\ln w$ makes it particularly amenable to an empirical analysis with unobserved heterogeneity, endogenous wages and nonparticipation as discussed below. Consequently, this specification is used extensively in our discussion of estimation that follows.

Neither (2.6) nor (2.8) allow backward bending labor supply behavior although it is easy to generalize (2.8) by including a quadratic term in $\ln w$. Note that imposing integrability conditions at zero hours for either (2.6) or (2.8) implies positive wage and negative income parameters. A simple specification that does allow backward bending behavior, while retaining a three parameter linear in variables form, is that used in Blundell, Duncan and Meghir (1994)

$$h = \alpha + \beta \ln w + \gamma \frac{y}{w} \tag{2.10}$$

with indirect utility

$$V(w,y) = \frac{w^{\beta+1}}{\beta+1}(\frac{y}{w}(1+\gamma)^2 + \beta \ln w + \alpha - \frac{\beta}{1+\gamma}) \text{ with } \gamma \leq 0 \text{ and } \beta \geq 0. \tag{2.11}$$

This form has similar properties to the MRS specification of Heckman (1974).

Generalizations of the *Linear Expenditure System or Stone-Geary preferences* are also attractive from certain points of view. For example suppose the indirect utility function for individual i in period t takes the form

$$V_{it} = \left[\frac{wH + y - a(w)}{b(w)}\right] \tag{2.12}$$

where H is the maximum amount of hours available to be allocated between hours and leisure. This is the quasi-homothetic "Gorman polar form". The linear expenditure system belongs to this class. However, there is no need to impose additive separability between consumption and hours of work as would be the case under Stone-Geary/LES preferences. Indeed, such separability assumptions severely constrain the time path of consumption and hours of work and can lead to the impression that the life-cycle model is unable to explain a number of observed phenomena, see Heckman (1974). In particular we may specify

$$a(w) = a_0 + a_1 w + 2a_2 w^{\frac{1}{2}} \tag{2.13}$$

and

$$b(w) = w^{\beta} \tag{2.14}$$

which is a *Generalized Leontief* model. Preferences are additive and reduce to LES if $a_2 = 0$.

The implied labor supply function using (2.12), (2.13) and (2.14) can be derived using Roy's identity and takes the form

$$h_{it} = (H - a_1) - a_2 w^{-\frac{1}{2}} - \frac{\beta}{w}(M - a_0 + a_1 w + 2a_2 w^{\frac{1}{2}}) \tag{2.15}$$

where $M = wH + y$. Unobserved heterogeneity can also easily be allowed for, as well as measurement error in hours of work (but not in hourly wages) and or consumption. For example, we can allow a_1 to be heterogeneous across individuals and time, i.e. $a_1 = \bar{a}_1 + \varepsilon$. Under the simplifying assumption that a_1 is the only source of heterogeneity the error term in the earnings equation now becomes $\nu = -\varepsilon(1 + \beta)$.

2.2. Estimation of the Static Labor Supply model

The main estimation issue, ignoring problems related to participation and nonlinear taxation (discussed below) is the endogeneity of wages w and unearned income y. Wages may well be endogenous because unobservables affecting preferences for work may well be correlated with unobservables affecting productivity and hence wages. Unearned income may be endogenous for a number of reasons: If y represents asset income, then individuals who work harder (because of unobserved preferences for work) are also likely to have accumulated more assets.[3]

[3] If μ also represents income from spouses, positive assortative mating will imply that hard working individuals will tend to marry. Hence unobserved preferences for work will correlate with spousal income reflected in μ.

Take as a simple example the semilog model of labor supply as above, i.e.

$$h_i = \alpha' x_i + \beta \ln w_i + \gamma y_i + u_i \tag{2.16}$$

The subscript i denotes an individual. The variables x denote observables which determine preferences. We avoid using the log of y because it is conceivable that it is zero and, in some cases, even negative. We add to this system a wage equation

$$\ln w_i = \delta_1' x_i + \delta_2' z_i + v_i$$

and a reduce form equation for unearned income

$$y_i = \zeta_1' x_i + \zeta_2' z_i + \varepsilon_i \ .$$

Identification requires that the dimension of the variables excluded from the labor supply equation, z_i, is at least two. It also requires that the matrix $[\delta_2' \ \zeta_2']$ has rank 2. In this linear framework estimation is straightforward - Two stage least squares is the obvious choice. However, we will see below that it is convenient to estimate the three reduce forms first and then impose the parametric restrictions to recover the structural coefficients using minimum distance. The reduced form labor supply model is

$$h_i = (\alpha + \beta\delta_1 + \gamma\zeta_1)' x_i + (\beta\delta_2 + \gamma\zeta_2)' z_i + u_i$$

Given estimates of all the reduced form coefficients the restrictions can then be imposed using minimum distance. Thus let

$$\alpha_1 = (\alpha + \beta\delta_1 + \gamma\zeta_1)$$

$$\alpha_2 = (\beta\delta_2 + \gamma\zeta_2)$$

$$\alpha_3 = [\delta_1' \ \delta_2' \ \zeta_1' \ \zeta_2' \ ]'$$

and let Ω represent the covariance matrix of the OLS estimator of the three equation reduced form system. Finally let $\alpha(\theta) = [\alpha_1 \ \alpha_2 \ \alpha_3]'$ where θ.represents the set of parameters in the labor supply model, the wage equation and the unearned income equation. Then the optimal minimum distance estimator is

$$\hat{\theta} = \arg\min_{\theta} \{(\hat{\alpha} - \alpha(\theta))' \Omega^{-1} (\hat{\alpha} - \alpha(\theta))$$

The resulting estimator is efficient, to the extent that the first step estimator is efficient.

When the labor supply model is non linear this straightforward procedure is no longer available. In this case an alternative approach is Maximum likelihood or semi-parametric instrumental variables. Maximum likelihood will be discussed below in the context of the labor supply model with corner solutions and nonlinear taxation. Hence we avoid duplication by deferring discussion until then.

In the absence of censoring one can use non-parametric Instrumental variables as in Newey and Powell (2003) and Daroles, Florens and Renault (2000). Consider the case where the labor supply is an unknown function of w and y

$$h_i = h(w_i, y_i) + u_i$$

The object is to estimate the function h. Suppose we have a set of instruments z (at least two if we are to treat both the wage and other income as endogenous). If we assume that the error in the labor supply function satisfies the rank condition $E(u_i|z_i) = 0$. In addition one needs a strong identification assumption ensuring that *any* function of w, y can be explained by the instruments z. Under these conditions solving the moment condition

$$E(h_i - h(w_i, y_i)|z_i) = 0$$

for the function $h(w_i, y_i)$ provides a non-parametric estimator.

In the context of censoring due to non-participation a control function approach turns out to be more useful. However, it is important to note that the assumptions underlying the control function are different than those underlying the IV approach above, unless the instruments are independent of the unobservables.[4] A form of the control function approach relies on the assumption that

$$E(u_i|z, x, w, y) = g(v_i, \varepsilon_i)$$

where v_i and ε_i are the error terms from the wage and unearned income equations respectively.[5] With unknown h identification also requires measurable separability

[4] Florens, Heckman, Meghir and Vytlacil (2003).

[5] For a more general case with unknown h see Newey, Powell and Vella (1999) or Heckman, Florens, Meghir and Vytlacil (2003) who derive conditions for identification..

which ensures that the functions g and h vary independently and is the equivalent of the rank condition. In a parametric framework the requirements are less stringent since we are restricting attention to specific functional forms. One approach to estimation would be to take a series expansion of g. Alternatively we could use some kernel estimator. The procedure works under a generalized rank condition; however the important point to note is that even under non-linearity we do not require explicit distributional assumptions, other than the restriction on the conditional expectation of u.[6] Nevertheless it should be noted that in practice it may be difficult to motivate the control function assumption, which contrasts with the orthogonality conditions above that are often derived from economic theory.

2.3. The Censored Regression Model

Labor market participation raises two key questions for modelling labor supply. First, what market wage distribution, should be used for nonparticipants? Second, are there features of the labor market that make labor supply behavior on the extensive margin (participation) fundamentally different from behavior on the intensive margin (hours of work)? These questions are not wholly unrelated since, without further restrictions on the distribution of offered wages among nonparticipants, it is difficult to separately identify a process for nonparticipation and for hours of work.

Among the most compelling reasons for separating these two margins is fixed costs of work - either monetary or time. We take up the issue of fixed costs in section 2.5, and begin by working through a model without fixed costs. We consider first with semiparametric estimation in a model with missing wages.

Suppose individual heterogeneity in tastes for work is represented by the random variable v. Observed hours of work (2.2) in the censored regression case can be represented by

$$h = \max\{f(w, y, x, v), 0\}. \tag{2.17}$$

where $f(w, y, x, v)$ represents desired hours of work

$$f(w, y, x, v) \equiv h^*. \tag{2.18}$$

[6]Two functions $g(e)$ and $h(v)$ are measurably separable iff whenever $g(e) - h(v) = 0$ a.s. implies $g(e)$ and $h(v)$ are constant functions.

and where y represents some measure of current period unearned income.

The censored labor supply model implies the reservation wage condition

$$h > 0 \Leftrightarrow w > w^*(y, x, v) \tag{2.19}$$

where w^* is defined implicitly by

$$0 = f(w^*, y, x, v). \tag{2.20}$$

The existence and uniqueness of the reservation wage in this simple world is guaranteed by revealed preference arguments. Given the market wage w, (2.17) also defines a threshold condition on the unobservable heterogeneity term v given by

$$h > 0 \Leftrightarrow v \geq v^*(w, y, x)$$
$$\Rightarrow \Pr(h > 0) = \int_{v \geq v^*} g(v) dv.$$

where $g(v)$ is the density function for v.

To implement this censored regression specification we define the index I_i as an indicator variable that is unity if individual i participates[7] and zero otherwise. Observable hours of work then follow the rule

$$h_i = \left\{ \begin{array}{c} h_i^* \text{ if } I_i = 1 \\ 0 \text{ otherwise} \end{array} \right\}. \tag{2.21}$$

That is

$$I_i = 1 \Longleftrightarrow h_i^* > 0 \tag{2.22}$$
$$= 1\{h_i^* > 0\}. \tag{2.23}$$

This implies that participation in work follows a simple *corner-solution* framework and is equivalent to assuming there are no fixed costs.[8]

[7] By participation we mean participation in paid work.

[8] By contrast, the fixed costs framework retains

$$I_i = 1 \Longrightarrow h_i^* > 0$$

but not the reverse. As Cogan (1980) shows, fixed costs are equivalent to a positive reservation hours of work. We elaborate on this in section 3.2 below.

The log likelihood for an independently distributed random sample of n individuals in the censored model is given by

$$\ln L(\theta) = \sum_{i=1}^{n} \left(I_i \ln g(\nu;\theta) + (1 - I_i) \ln \int_{v \geq v^*} g(v;\theta) dv \right) \tag{2.24}$$

where θ are the unknown parameters of preferences and g is the distribution of v. In a linear specification with a normal iid assumption on v, this is equivalent to the Tobit censored regression specification.

The likelihood specification (2.24) makes two implicit assumptions on the wage distribution. First, that wages are observed for all individuals irrespective of their labor market status. Second, that wages are exogenous for labor supply. Neither of these are apriori reasonable.

2.4. Missing Wages

Wages are not observed if $h = 0$. Suppose the model for wages can be written as

$$\ln w = \gamma_1 x + \gamma_2 q + \eta \tag{2.25}$$

where q are a set of variables that are exclusive to the determination of real wages and where η is an iid error term with distribution $g_w(\eta)$. The likelihood contribution for $h = 0$ becomes

$$\begin{array}{ll} h > 0 & \ell_0 = g(v) g_w(\eta) \\ h = 0 & \ell_0 = \int^{\infty} \int^{v^*} g(v) g_w(\eta) dv d\eta \end{array} \tag{2.26}$$

By writing the joint distribution of ν and η as a product of the two marginals we have implicitly maintained that wages are exogenous for labor supply. This implies that the density of wages can be estimated separately; In a labor supply model linear in log wages this further implies that we can simply impute wages for all non workers and estimate the model as if wages are observed (correcting the standard errors of course for generated regressor bias) However, if we wish relax this assumption and permit w to be endogenous in the hours equation, the sample likelihood becomes

$$\ln L(\phi) = \sum_{i=1}^{n} \left(I_i \ln g_{hw}(v, \eta) + (1 - I_i) \ln \int^{\infty} \int^{v^*} g_{hw}(v, \eta) dv d\eta \right) \tag{2.27}$$

where $g_{hw}(v, \eta; \phi)$is the joint distribution of ν and η.

The resulting estimator simplifies enormously if we assume a parametric specification that permits an explicit reduced form for desired hours of work. A popular example of such a specification is the semi-log labor supply model to which we now turn

2.4.1. A Semi-log Specification

Suppose we write the optimal labor supply choice for individual i as

$$h_i^* = \beta_1 \ln w_i + \beta_2 y_i + \beta_3 x_i + v_i \tag{2.28}$$

where β_1, β_2 and β_3 are unknown parameters of labor supply. labor supply and wages are now completely described by the triangular system consisting of (2.25) and the following reduced form for desired hours of work

$$\begin{aligned} h_i^* &= (\beta_1 \gamma_1 + \beta_3) x_i + \beta_1 \gamma_2 q_i + \beta_2 y_i + \beta_1 \eta_i + v_i & (2.29) \\ &= \pi_1 x_i + \pi_2 q_i + \pi_3 y_i + \omega_i & (2.30) \\ &= \pi z_i + \omega_i & (2.31) \end{aligned}$$

2.4.2. Semi-parametric Estimation

If it can be assumed that v_i and η_i are distributed independently of the explanatory variables x, q and y then semiparametric identification and estimation can take the following simple step wise procedure.

The π coefficients in (2.31) can be estimated from a standard censored regression estimation procedure. If $g_\omega(\omega)$ describes the density of ω then the sample likelihood for a random sample of $i = 1, ..., n$ individuals is given by

$$L(\pi) = \prod_{i=1}^{n} \{g_\omega(\omega|\pi))\}^{I_i} \left\{ 1 - \int_{-\pi' z_i} g_\omega(\omega|\pi) d\omega \right\}^{1-I_i} \tag{2.32}$$

which is equivalent to the sample likelihood for the Tobit model when ω is homoscedastic normal. Root-n consistent and asymptotically normal estimators of π can be derived under much weaker assumptions on g_ω, see Powell(1984, 1986b), Horowitz (1986).

Given π, the conditional mean of (2.25) for participants can be used to estimate the wage equation parameters. This is the Heckman (1976, 1979) selectivity framework. Suppose we assume

$$E(\eta_i|I_i > 0) = \lambda_\eta(\pi z_i) \tag{2.33}$$

then the conditional mean of (2.25) given $I_i > 0$ is simply written

$$E(\ln w_i|z, I_i > 0) = \gamma_1 x_i + \gamma_2 q_i + \lambda_\eta(\pi z_i) \tag{2.34}$$

If a joint normal distribution is assumed for v_i and η_i then estimation can follow the two-step selectivity estimation approach developed by Heckman (1979). Alternatively, a $\sqrt{N}$ consistent and asymptotically normal semiparametric estimator can be constructed.

To consider the semiparametric estimator notice that the conditional expectation of (2.34) for participants given πz_i is

$$E(\ln w_i|\pi z_i, I > 0) = \gamma_1 E(x_i|\pi z_i) + \gamma_2 E(q_i|\pi z_i) + \lambda_\eta(\pi z_i). \tag{2.35}$$

Subtracting this from (2.34) eliminates the $\lambda_\eta(\pi z_i)$ term yielding

$$\begin{aligned} & E(\ln w_i|z, I_i > 0) - E(\ln w_i|\pi' z_i, I > 0) \\ = \; & \delta'(x_i - E(x_i|\pi' z_i)) + \phi'(q_i - E(q_i|\pi' z_i)). \end{aligned} \tag{2.36}$$

The conditional expectation terms $E(\ln w_i|\pi z_i)$, $E(x_i|\pi z_i)$ and $E(q_i|\pi z_i)$ in (2.36) can then be replaced by their unrestricted Nadaraya-Watson kernel regression estimators.[9]

The parameters of (2.34) can then be recovered by an instrumental variable regression. Robinson (1988), suggests regressing $\ln w - \widehat{\ln w}^h(\pi z)$ on $x - \widehat{x}^h(\pi z)$

[9] e.g.

$$E\widehat{(q_i|\pi' z_i)} = \widehat{q}^h(\pi z) = \frac{\widehat{r}(\pi z)}{\widehat{f}(\pi z)}$$

and $q - \widehat{q}^h(\pi z)$ using $I[\widehat{f}(\pi z) > b_N]x$ and $I[\widehat{f}(\pi z) > b_N]q$ as the respective instrumental variables, where $I[\widehat{f}(\ln x) > b_N]$ is an indicator function that trims out observations for which $\widehat{f}(\ln x) < b_N$, for some sequence of trimming constants b_N which tend to zero with the sample size at some appropriate rate. An alternative estimator, due to Powell (1987), is to use $\widehat{f}(\pi z).x$ and $\widehat{f}(\pi z).q$ as instruments. This effectively removes the random denominators from the kernel regression estimators.

Finally, given the $\gamma_1, \gamma_2, \pi_1, \pi_2$ and π_3 parameters, the structural labor supply parameters β_1, β_2 and β_3 can be recovered by minimum distance. In general, these steps can be combined to improve efficiency. Provided a suitable instrumental variable is available, this procedure can also be extended to control for the endogeneity of other income y_i. We consider this in more detail below.

2.5. Fixed Costs

2.5.1. A Structural Model of Fixed Costs

Fixed costs imply that participation does not simply follow the corner solution condition (2.22). Instead participation will depend on the determinants of fixed costs as well as the determinants of h_i^*. For example, suppose there is a fixed monetary cost of working S, this implies that non-labor income in the budget constraint becomes

$$\begin{array}{ll} y - S & \text{if} \quad h > 0 \\ y & \text{if} \quad h = 0 \end{array}$$

and the distribution of S is only partially observable. If we denote utility in work at the optimal hours point by the indirect utility level: $\nu(w, y, v)$ and utility at

---

in which

$$\widehat{r}(\pi z) = \frac{1}{n} \sum_i K_h (\pi z - \pi z_i) \, q_i,$$

and

$$\widehat{f}(\pi z) = \frac{1}{n} \sum_i K_h (\pi z - \pi z_i) \,,$$

where $K_h(\cdot) = h^{-1}k(\cdot/h)$ for some symmetric kernel weight function $k(.)$ which integrates to one. The bandwidth h is assumed to satisfy $h \to 0$ and $nh \to \infty$ as $n \to \infty$. Under standard conditions the estimator is consistent and asymptotically normal, see Härdle (1990) and Härdle and Linton (1994).

$h = 0$ by the direct utility at $h = 0$: $U(Y, 0, v)$. The decision to work follows from

$$\nu(w, y, v) \geq U(Y, 0, v).$$

Note that if $S > \underline{S} >> 0$ then there will be a discontinuity in the hours distribution at low wages which should reflect itself as a "hole" at the low end of the hours distribution.[10] This model is further developed in Section 4, here we analyse index models empirical models that are motivated by the presence of fixed costs of work.

Cogan (1981) defines reservation hours h_0 such that

$$\begin{aligned} U^1(T - h_0, y - S + wh_0, x, v) &= U^0(T, y) \\ h_0 &= h_0(y - S, w, x, v) \geq 0 \end{aligned} \tag{2.37}$$

and the participation decision becomes

$$\Pr\ (work) = \Pr(h > h^0). \tag{2.38}$$

For any v and η, nonparticipation will occur if fixed costs are sufficiently high $S > S^*(v, \eta)$.

Suppose we continue to assume wage equation (2.25) and also assume the specification of fixed costs to be

$$S = \theta_1 x + \theta_2 m + s \tag{2.39}$$

where m are a set of variables exclusive to the determination of fixed costs and s represents unobserved heterogeneity in the distribution of fixed costs. In terms of the likelihood contributions we have 'no work' regime:

$$\ell_0 = \int_{-\infty}^{\infty} \int_{-\infty}^{v^*} \int_{S^*}^{\infty} g(v, \eta, s) ds dv d\eta \tag{2.40}$$

work regime:

$$\ell_1 = \int_{v^*}^{\infty} \int_{0}^{S^*} g(\varepsilon, v, \eta, s) ds dv. \tag{2.41}$$

[10] This may not be visible since heterogeneity in fixed costs and in unobserved tastes may imply a different possition for the discontinuity for different individuals, smoothinng out the unconditional distribution. Hence looking for such "odd" features in the hours distributionn may not be a very good empirical strategy for detecting fixed costs. However such features can be seen in the distribution of relatively homogenous groups, e.g. Single women with no children or single men.

Given some parametric specification of direct (and indirect) utility, all the structural parameters of fixed costs, preferences and wage determination are identified from a likelihood based on the contributions (2.40) and (2.41).

Finally note that if we specify a model on the basis of the indirect utility or cost function we may not have an analytical expression for the direct utility function. Consequently this has to be obtained numerically. One way of doing this is to find the standard reservation wage when hours are zero and the fixed costs have not been incurred. Evaluating the indirect utility function at that reservation wage and non-labor income then provides us with the utility value of not working. Another important difficulty is then to derive the probability of participation given that the direct utility function at zero hours of work will depend on unobserved heterogeneity both directly and via the reservation wage - hence it is likely to be a highly non-linear function of the underlying error term. In practice, as we argue later, it may be easier to work with a direct utility specification when we have to deal with such non-convexities.

2.5.2. Semi-parametric Estimation in the Fixed Costs Model

Although the (semiparametric) censored regression approach to estimation of the hours equation described above is no longer valid in this fixed costs case, a semiparametric procedure applied to hours of work among the participants can be used as an approximation to the fixed costs model. The optimal choice of hours of work among those individuals who decide to join the labor market will have the form

$$\begin{aligned} h_i^* &= \beta_1 \ln w_i + \beta_2 (y_i - S_i) + \beta_3 x_i + v_i && (2.42) \\ &= (\beta_1 \gamma_1 + \beta_2 \theta_1 + \beta_3) x_i + \beta_1 \gamma_2 q_i + \beta_2 y_i && (2.43) \\ &\quad + \beta_2 \theta_2 m_i + \beta_1 \eta_i + \beta_2 s_i + v_i && (2.44) \\ &= \widetilde{\pi}_1 x_i + \widetilde{\pi}_2 q_i + \widetilde{\pi}_3 y_i + \widetilde{\pi}_4 m_i + u_i && (2.45) \\ &= \widetilde{\pi} \widetilde{z}_i + u_i && (2.46) \end{aligned}$$

where again the β_1, β_2 and β_3 are unknown parameters of labor supply. labor supply and wages are now completely described by the triangular system consisting of (2.25) and (2.46).

Assume that the participation condition (2.38) can be well approximated by the discrete index model

$$I_i = 1 \Longleftrightarrow \phi\widetilde{z}_i + e_i > 0 \tag{2.47}$$

where $\widetilde{z}_i$ contain all the exogenous variables determining reservation hours, log wages and desired hours of work. The term e_i is a random unobservable whose distribution F_e is normalized up to scale and assumed to be independent of $\widetilde{z}_i$. Parameters ϕ will be convolution of parameters of fixed costs, the wage equation and preferences. They can be identified through the condition

$$E(I_i = 1|\widetilde{z}_i) = \int_{-\phi\widetilde{z}_i} dF_e(e). \tag{2.48}$$

The ϕ coefficients in (2.47) can be estimated up to scale from a standard binary choice estimation procedure which replaces the censored regression rule (2.32) in this fixed costs model. The sample likelihood for a random sample of $i = 1, ..., n$ individuals is given by

$$\Im(\phi) = \prod_{i=1}^{n} \left\{ \int_{-\phi\widetilde{z}_i} dF(e) \right\}^{I_i} \left\{ 1 - \int_{-\phi\widetilde{z}_i} dF(e) \right\}^{1-I_i} \tag{2.49}$$

which is equivalent to the Probit likelihood when e is homoscedastic normal. $\sqrt{N}$ consistent and asymptotically normal estimators of ϕ up to scale can be derived under much weaker index assumptions on f, see Klein and Spady(1993) for example.

Given ϕ, the conditional mean of (2.25) for participants can be used to estimate the wage and hours equation parameters. This is the Heckman (1976, 1979) selectivity framework. Suppose we assume the single index framework

$$E(\eta_i|I_i > 0, \widetilde{z}_i) = \lambda_\eta(\phi\widetilde{z}_i) \tag{2.50}$$

and

$$E(u_i|I_i > 0, \widetilde{z}_i) = \lambda_u(\phi\widetilde{z}_i). \tag{2.51}$$

then the conditional mean of (2.25) and (2.31) given $I_i > 0$ is simply written

$$E(\ln w_i|I_i > 0) = \gamma_1 x_i + \gamma_2 q_i + \lambda_\eta(\phi\widetilde{z}_i) \tag{2.52}$$

and

$$E(h_i|I_i > 0, \widetilde{z}_i) = (\beta_1\gamma_1 + \beta_2\theta_1 + \beta_3)x_i + \beta_1\gamma_2 q_i + \beta_2 y_i + \beta_2\theta_2 m_i + \lambda_u(\phi\widetilde{z}_i). \quad (2.53)$$

where $\widetilde{z}_i$ is taken to include all exogenous variables. If a joint normal distribution is assumed for v_i, η_i and s_i then estimation can follow the two-step selectivity estimation approach developed by Heckman (1979).

Notice that (2.52) and (2.53) together only identify γ_1, γ_2, β_1 and β_2, the parameters of fixed costs and β_3 are not identified without more information on fixed costs. A $\sqrt{N}$ consistent and asymptotically normal semiparametric estimator of these parameters can be constructed from a natural extension of the procedures described above for the censored labor supply model.

For participants we have

$$E(h_i|\phi\widetilde{z}_i, I > 0) = \widetilde{\pi}_1 E(x_i|\phi\widetilde{z}_i,\ I > 0) + \widetilde{\pi}_2 E(q_i|\phi\widetilde{z}_i,\ I > 0) \quad (2.54)$$
$$+\widetilde{\pi}_3 E(y_i|\phi\widetilde{z}_i,\ I > 0) + \widetilde{\pi}_4 E(m_i|\phi\widetilde{z}_i,\ I > 0) + \lambda_u(\phi\widetilde{z}_i). \quad (2.55)$$

The nonparametric term describing the selection of participants can be eliminated as in (2.36) and root-n estimation of the unknown index parameters can also follow the same semiparametric techniques.

Finally, we should note that endogeneity of y_i can be handled in a similar fashion. Suppose a reduced form for y is given by

$$y_i = \vartheta' d_i + \zeta_i \quad (2.56)$$

since y_i is continuously observed for all individuals, ϑ can be estimated by least squares. Now suppose we also assume that

$$E(u_i|I_i > 0, y_i, \widetilde{z}_i) = \delta_y\zeta_i + \lambda_u(\phi\widetilde{z}_i). \quad (2.57)$$

Then adding the estimated residual from the regression (2.56) into the selection model (2.53) appropriately corrects for the endogeneity of y_i. This is an important consideration given the consumption based definition of y_i in the life-cycle consistent specification.

3. Difference-in-Differences, Natural Experiments and Grouping Methods

One of the central issues in labour supply is the endogeneity of marginal (post-tax) wages and other incomes. The work incentives facing individuals are usually endogenous. Consider as an example a world with a progressive tax system, as will be examined in detail in the next section. In this case individuals earning more face a higher rate of tax and hence a lower marginal incentive to work. Now take two individuals both of which have the same pre-tax wage but different tastes for work. The person working longer hours will earn more and will face a higher tax rate, which translates to a lower post-tax marginal wage. In a simple regression framework we would estimate a negative effect of the wage on hours of work since the person with higher hours (because of tastes) will be facing a lower wage. This kind of endogeneity has prompted researchers to seek exogenous sources of variation in policy that resemble experimental situations with a "treatment" group affected by the policy and a "control" or "comparison" group which is unaffected. The impact of incentives is then estimated by comparing the change n hours between the two groups before and after the policy is implemented.

Using this basic idea one can attempt to estimate a "causal" impact of the policy on labour supply, ignoring any structural considerations. Alternatively one can think of the policy changes as an attempt to obtain qausiexperimental conditions for estimating the structural parameters themselves. The former approach attempts to ignore the underlying theory and wishes to go straight to the effects of the particular policy. The latter is after structural parameters that can be used for extrapolation to other policy situations, assuming the theory is a good approximation of reality.

In the following sections we describe this approach to estimating the impact of incentives on labour supply. We also discuss the validity of the approach under different circumstances. As one may expect, even the "atheoretical" approach which seeks to estimate the impacts of policy without reference to a model does implicitly make strong assumptions about behaviour and/or the environment, and we discuss this. We also discuss conditions under which the quasi-experimental approach, which is a form of Instrumental Variables, can provide estimates of structural parameters. We go through the difference in differences estimator and a more general grouping estimator considering also the effects of selection due to nonparticipation.

3.1. Difference-in-Differences and Fixed Effects Models

Suppose one is interested in estimating the influence of a policy instrument on an outcome for a group, say outcome h_{it} measuring hours of work or participation. The group consists of individuals $i = 1, ..., N$, with these individuals observed over a sample horizon $t = 1, 2, ...$ Suppose further that a policy instrument of interest changes in a particular period t for only a segment of the individuals. Let δ_{it} be a zero-one indicator that equals unity if the policy change was operative for individual i in period t. Members of the group who experience the policy change react according to a parameter γ. A framework for estimating expressed in terms of a conventional fixed-effect model takes the form

$$h_{it} = \gamma\delta_{it} + \eta_i + m_t + \varepsilon_{it} \tag{3.1}$$

where i is a time-invariant effect unique to individual i, m_t is a time effect common to all individuals in period t, and ε_{it} is an individual time-varying error distributed independently across individuals and independently of all η_i and m_t.

The least squares estimator of γ in (3.1), which regresses h_{it} on δ_{it} and a set of individual and time dummy variables, is precisely the difference in differences estimator for the impact of the reform. It can be given the interpretation of a causal impact of the reform if $E(\varepsilon_{it}|\eta_i, m_t, \delta_{it}) = 0$. In a heterogeneous response model

$$h_{it} = \gamma_i\delta_{it} + \eta_i + m_t + \varepsilon_{it} \tag{3.2}$$

the least squares dummy variable estimator recovers the average of the response parameters γ_i for those affected by the policy. Since the error term ε_{it} may be correlated both over time and across individuals, this should be taken into account when constructing standard errors.

Now suppose that the policy does not affect everyone in a treatment group, but that the chance of being affected is higher among them $(g = T)$, than it is among a control group $(g = C)$. The error structure can be more general than above. Consider a specification in which

$$h_{it} = \gamma\delta_{it} + u_{it}. \tag{3.3}$$

where u_{it} represents an individual level heterogeneity term which may be fixed for that individual over time or may vary over time. Moreover it may be correlated across individuals and may have a cross section mean that is non zero. The implicit macro component and the average group characteristics to which the individual belongs may be correlated with δ_{it}. Suppose that limited time series data is available across individuals, either in terms of repeated cross-sections or as a panel data source. Under the following assumption, and the presumption that the policy is introduced only for one group, the impact of the policy may be identified by using two time periods of data, one before the reform and one after. The assumption we require is that

$$\mathbf{A1} : E[u_{it}|g,t] = a^g + m_t \tag{3.4}$$

which can be interpreted as saying that in the absence of the reform the changes in group means are the same across the two groups. Then with two groups and two time periods the slope coefficient γ can be written as

$$\gamma = \frac{\Delta E(h_{it} \mid T,t) - \Delta E(h_{it} \mid C,t)}{\Delta E(\delta_{it} \mid T,t) - \Delta E(\delta_{it} \mid C,t)}$$

the difference-in-differences estimator is the sample analog given by

$$\widehat{\gamma} = \frac{\Delta \overline{h}_t^T - \Delta \overline{h}_t^C}{\Delta \Pr(\delta_{it} = 1 \mid T,t) - \Delta E(\delta_{it} = 1 \mid C,t)} \tag{3.5}$$

where the 'bar' denotes sample average, Δ the first difference and the superscript the group for which first differences are taken. $\widehat{\gamma}$ is consistent for γ. This estimator is an instrumental variables estimator with excluded instruments the group time interactions. If the effect of the treatment is heterogeneous and if the policy does not decrease the chance of obtaining the treatment δ for anyone (monotonicity) then the difference in differences estimator above identifies the impact of the policy for those obtaining treatment as a result of the policy (Imbens and Angrist, 1994)

Assumption **A1** is very strong indeed. Failure will result if there is a change in group composition of unobservable individual effects over time or if there is a differential impact of macro shocks across groups. Again it will depend critically on the choice of groups which is a key issue in this framework. **A1** implies:

(i) time invariant composition for each group
and
(ii) common time effects across groups.

3.2. Estimating a Structural Parameter

Here we consider the use of this method in the estimation of a simple labor supply model (ignoring income effects for notational simplicity, we return to this below)

$$h_{it} = \alpha + \beta \ln w_{it} + u_{it} \tag{3.6}$$

Again u_{it} represents an individual level heterogeneity term which may be fixed for that individual over time or may vary over time. Moreover it may be correlated across individuals and may have a cross section mean that is non zero. This represents the impact of macro shocks to preferences on individual i's labor supply. Both the implicit macro component and the idiosyncratic heterogeneity may be correlated with the log wage ($\ln w_{it}$).

Make the following assumptions

$$\mathbf{A1} : E[\varepsilon_{it}|g,t] = a^g + m_t \tag{3.7}$$

$$\mathbf{A2} : [E[\ln w_{it}|g=T,t] - E[\ln w_{it}|g=C,t]] - [E[\ln w_{it}|g=T,t-1] - E[\ln w_{it}|g=C,t-1]] \neq 0 \tag{3.8}$$

Then with two groups and two time periods the slope coefficient β can be written as

$$\beta = \frac{\Delta E(h_{it} \mid T,t) - \Delta E(h_{it} \mid C,t)}{\Delta E(\ln w_{it} \mid T,t) - \Delta E(\ln w_{it} \mid C,t)}$$

The difference-in-differences estimator is the sample analog given by

$$\widehat{\beta} = \frac{\Delta \overline{h}_t^T - \Delta \overline{h}_t^C}{\Delta \overline{\ln w}_{it}^T - \Delta \overline{\ln w}_{it}^C} \tag{3.9}$$

and is consistent for β.

Assumption **A2** is simply analogous to a rank condition and should hold if the groups are chosen to reflect some systematic reason for a differential growth in $\ln w_{it}$ across groups. The choice of groups in this difference in differences approach usually reflects some policy change which affects the real wage. A tax change, for example, that can be argued to be incident on individuals in one group $i \in [\ g = T]$ but not on individuals in another $i \in [\ g = C]$. It is clear, however, that the assumption **A1** may be strong in some circumstances. However note the big difference with the previous section. In the previous section the policy was assumed to have no effect on wages of the treatment group relative to the control group; this is the assumption implicit in the fact that we only need to condition on time and group effects. Here we are conditioning on wages and we are adding the assumption from economic theory, that log wages and taxes share the same coefficient. Hence if the policy implicitly affecting incentives, changes pre-tax wages as well, this is allowed for; this in itself makes the assumptions underlying the difference in differences approach more credible. (see more on this below).

This method has some attractive features. It allows for correlated heterogeneity and for general common time effects. Although for many choices of grouping, often precisely those associated with some policy reform, assumption **A1** is likely to be invalid, there are possible grouping procedures for estimating labor supply models that are more convincing. This approach is also closely related to the natural experiment or quasi-experimental estimators that typically employ before and after comparisons relating directly to a policy reform.

Before moving on to consider these developments, we first simply outline how this approach can be extended to allow for many groups, for many time periods (or many reforms), for participation and for the inclusion of income terms and other regressors.

3.3. Grouping Estimators

Suppose individuals can be categorized in one of a finite number of groups g each sampled for at least two time periods. For any variable x_{it}, define D_x^{gt} as the residual from the following regression

$$E(x_{it}|P_{it}, g, t) = \sum_{g=1}^{G} \zeta_g d_g + \sum_{t=1}^{T} \xi_t d_t + D_x^{gt} \ , \tag{3.10}$$

where P_{it} indicates that the individual is observed working, that is $P_{it} \equiv \{I_{it} = 1\}$ and where d_g and d_t are time and group dummies respectively. Analogously with

A1 and **A2** we make assumptions

$$\mathbf{A1.1} \;\; E(u_{it}|P_{it}, g, t) \;\; = \;\; a_g + km_t \tag{3.11}$$

$$\mathbf{A2.1} \;\; E[D_w^{gt}]^2 \;\; \neq \;\; 0. \tag{3.12}$$

Assumption **A1.1** summarizes the exclusion restrictions for identification; it states that the unobserved differences in average labor supply across groups can be summarized by a permanent group effect a_g and an additive time effect m_t . In other words differences in average labor supply across groups, given the observables, remain unchanged over time. It also says that any self selection into employment (the conditioning on P_{it}) can be controlled for by group effects and time effects additively. Assumption A2.1 is again equivalent to the rank condition for identification; it states that wages grow differentially across groups; this is because the assumption requires that after we have taken away time and group effects there is still some variance of wages left. For example, if there is a tax reform between two periods, affecting the post tax wages of the two groups in different ways, and assuming that tax incidence does not fully counteract the effects of the reforms, identification of the wage elasticity will be guaranteed.

With these assumptions we can implement a generalized Wald estimator (see, Heckman and Robb, 1985). Defining the sample counterpart of D_x^{gt} as $\tilde{x}_{gt}$, i.e. the residual from regressing the time-group cell mean on time and group dummies, we can write the estimator as

$$\widehat{\beta} = \frac{\sum_g \sum_t \left[\tilde{h}_{gt}\right] \left[\widetilde{\ln w_{gt}}\right] n_{gt}}{\sum_g \sum_t \left(\widetilde{\ln w_{gt}}\right)^2 n_{gt}} \tag{3.13}$$

where n_{gt} is the number of observations in cell (g, t). The implementation of this estimator is simple; group the data for workers by g and by time and regress by weighted least squares the group average of hours of work on the group average of the log wage, including a set of time dummies and group dummies. An alternative that gives numerically identical results is as follows: regress using OLS the log after-tax wage rate on time dummies interacted with the group dummies, over the sample of workers only and compute the residual from this regression. Then use the original data to regress hours of work on the individual wage, a set of time dummies and group dummies and the wage residual. The t-value on the coefficient of the latter is a test of exogeneity, once the standard errors have been corrected for generated regressor bias and intra group dependence. It is also important to allow for the possibility of serial correlation and correlation of idiosyncratic shocks across individuals when computing the standard errors.

3.3.1. Controlling for Participation

A potential problem with the approach above is that it assumes that the composition effects from changes in participation can be fully accounted for by the additive time and group effects, $a_g + m_t$. First, changes in m_t will cause individuals to enter and leave the labor market. Second, with non-convexities, a tax policy reform may lead to changes in participation. This will be particularly true if fixed costs are large relative to the non-taxable allowance. The presence of composition effects is equivalent to saying that $E(u_{it}|P_{it}, g, t)$ is some general function of time and group and does not have the additive structure assumed in **A1.1**.

To control for the possibility that $E(u_{it}|P_{it}, g, t)$ may vary over time requires structural restrictions. A parsimonious specification is to make the assumption of linear conditional expectation. For example, we may extend **A1.1** and **A2.1** by assuming that

$$\begin{aligned} \textbf{A1.2}\ E(u_{it}|P_{it}, g, t) &= a_g + m_t + \delta\lambda_{gt} && (3.14) \\ \textbf{A2.2}\ E[D_w^{gt\lambda}]^2 &\neq 0. && (3.15) \end{aligned}$$

where λ_{gt} is the inverse Mills' ratio evaluated at $\Phi^{-1}(L_{gt})$, Φ^{-1} being the inverse function of the normal distribution and L_{gt} being the proportion of group g working in period t.[11] Finally $D_w^{gt\lambda}$ is defined by the population regression

$$E(w_{it}|P_{it}, g, t) = \sum_{g=1}^{G} \zeta_g d_g + \sum_{t=1}^{T} \xi_t d_t + \delta_w \lambda_{gt} + D_x^{gt\lambda} \;, \tag{3.16}$$

Assumption **A1.2** models the way that composition changes affect differences in the observed labor supplies across groups. It implies that

$$E(h_{it}|P_{it}, g, t) = \beta E(\ln w_{it}|P_{it}, g, t) + a_g + m_t + \delta\lambda_{gt} \tag{3.17}$$

[11] See Gronau (1974) and Heckman (1974, 1979).

where all expectations are over workers only. Assumption **A2.2** states that wages must vary differentially across groups over time over and above any observed variation induced by changes in sample composition. We have also implicitly assumed that $E[D_\lambda^{gt}]^2 \neq 0$. If this is not the case, there is no selection bias on the coefficients of interest (here the wage effect) because composition effects can be accounted for by the linear time and group effects. In this case we can use (3.13).

We can now estimate the wage effect using a generalization of (3.13), i.e.

$$\widehat{\beta} = \frac{\sum_g \sum_t \left[\tilde{h}_{gt\lambda}\right]\left[\widetilde{\ln w_{gt\lambda}}\right] n_{gt}}{\sum_g \sum_t \left(\widetilde{\ln w_{gt\lambda}}\right)^2 n_{gt}} \tag{3.18}$$

As before this estimator can be implemented using a residual addition technique. We can add an estimate of λ_{gt} as well as the residual of the wage equation estimated on the sample of workers (with no correction for sample selection bias as implied by (3.17) to an OLS regression of individual hours on individual wages, time dummies and group dummies.

To determine whether (3.18) or (3.13) should best be used we can test the null hypothesis that $E[D_\lambda^{gt}]^2 = 0$ which implies that the group effects a_g and the time effects m_t adequately control for any composition changes (given our choice of groups). If we do not reject this we can use (3.13).

The assumption in **A1.2** is worth some discussion. First note that where all regressors are discrete and a full set of interactions are included in the selection equation, use of the normal distribution to compute $\hat{\lambda}_{gt}$ imposes no restrictions. However, the linear conditional expectation assumption implies that a term linear in $\hat{\lambda}_{gt}$ is sufficient to control for selection effects and is potentially restrictive. Using the results in Lee (1984) in general we have that

$$E(u_{it}|P_{it}, g, t) = a_g + m_t + \sum_{k=1}^{K} \delta_k \lambda_{gt}^{(k)} \tag{3.19}$$

where $\lambda_{gt}^{(k)}$ are generalized residuals of order k. The linearity reduces the number of parameters to be estimated and hence the number of periods over which we require exogenous variability in wages. If it is found that $E[D_\lambda^{gt}]^2 \neq 0$ then one can experiment by including higher order generalized residuals after checking that they display sufficient independent variability.

3.3.2. Income Effects

Income effects are important for labor supply and we need to take them into account for at least two reasons. First, the wage elasticity cannot in general be interpreted as an uncompensated wage elasticity, unless we control for other income. Second, income effects are important if we wish to compute compensated wage elasticities for the purpose of evaluating the welfare effects of tax reforms. It is straightforward to extend the estimator in (3.18) to allow for extra regressors, such as other income. This involves regressing $\widetilde{h}_{gt\lambda}$ on $\widetilde{\ln w_{gt\lambda}}$ and $\widetilde{y}_{gt\lambda}$ where y is household other income. The rank condition for identification is now more stringent: It requires that the covariance matrix $V = Ez_{gt\lambda}z'_{gt\lambda}$ is full rank, where $z_{gt} = [D_w^{gt\lambda},\, D_y^{gt\lambda}]'$.

This is equivalent to requiring that the matrix of coefficients on the excluded exogenous variables in the reduced forms of log wage and other income, after taking into account of composition effects, is rank 2. A necessary but not sufficient condition for this to be true is that these coefficients are non-zero in each of the reduced forms - i.e. that $E(D_w^{gt\lambda})^2$ and $E(D_y^{gt\lambda})^2$ are non-zero. As before if we accept the hypothesis that $E(D_\lambda^{gt})^2 = 0$ we need to consider whether the rank of $V^* = Ez^*_{gt}z^{*\prime}_{gt}$ is two, where $z^*_{gt} = [D_w^{gt},\, D_y^{gt}]'$. In this case we estimate the model using the sample counterparts of z^*_{gt} as regressors.

3.4. The Difference-in-Difference Estimator and Behavioral Responses

As we have seen the simplest implementation of the difference-in-differences approach simply includes a policy reform dummy. This avoids directly specifying a structural model in the sense that the effect of the policy is not tied to a wage or income effect. The idea is that the policy should be evaluated directly without the intermediation of an economic model.

Suppose again there are simply two periods and two groups. Suppose the policy reform is a tax change in which τ is the change in the marginal tax rate for the treatment group. The *natural experiments* approaches simply includes a policy dummy $\delta_t^g = 1\{g = T, t = A\}$ in the hours regression

$$h_i = \alpha + \beta\delta_t^g + \zeta_{it}. \tag{3.20}$$

The quasi-experimental estimator in this case is just the difference-in-differences estimator applied to (3.20).

To interpret this estimator suppose the hours equation has the simple form (3.6). Suppose that pre and post reform wages are defined by:

	Before Reform	*After Reform*
$i \subset$ Treated	$\ln w_{iB}$	$\ln((1-\tau)w_{iA})$
$i \subset$ Control	$\ln w_{iB}$	$\ln w_{iA}$

Assuming **A1** and **A2,** taking group means we find

$$\overline{h}_t^g = \alpha + \beta \ln(1-\tau)\delta_t^g + \beta \overline{\ln w_t^g} + a^g + m_t \tag{3.21}$$

If $\delta_t^g = 1\{g = T, t = A\}$ is all that is included in the regression then the difference in differences estimator will only recover β if log wages have the group and common time effect form

$$\overline{\ln w_t^g} = \widetilde{a}^g + \widetilde{m}_t. \tag{3.22}$$

This seems a particularly strong assumption given empirical knowledge about the differential trends in wage group across different groups in the economy. Clearly, the cost of including simply the policy reform dummy $\delta_t^g = 1\{g = T, t = A\}$ alone is that the common time effects and time invariant composition effects assumptions become even more difficult to satisfy.

4. Estimation with Nonlinear Budget Constraints

A problem encountered in many analyses of consumption and labor supply involves the presence of intricate nonlinearities in budget sets arising from wages and prices that vary as a function of quantities. Tax and welfare programs constitute a prominent source of such functional relationships in analyses of labor supply, for these programs induce net wages to vary with the number of hours worked even when the gross wages remain constant. Hedonic environments and price schedules dependent upon quantities give rise to comparable sources of distortions in budget sets in many consumption settings.

To address the issues encountered with nonlinear budget sets, there has been steady expansion in the use of sophisticated statistical models characterizing distributions of discrete-continuous variables that jointly describe both interior choices and corner solutions in demand systems. These models offer a natural framework for capturing irregularities in budget constraints, including those induced by the institutional features of tax and welfare programs.

This section briefly describes approaches for estimating models incorporating such features, keeping the context general enough to illustrate how these models can readily accommodate a wide variety of nonlinearities in price and wage structures. The discussion begins with a brief overview of the methods implemented to model budget constraints involving nonlinearities, and then goes on to survey instrumental-variable procedures applied in the literature to estimate behavioral relationships in the presence of such constraints. We summarize the general approach for using maximum likelihood procedures to estimate the more sophisticated variants of these models with either convex or nonconvex budget sets. We provide simple illustrations of maximum likelihood methods to estimate familiar specifications of labor supply with convex constraints. We outline why the implementation of maximum likelihood procedures imposes interesting and important restrictions on behavioral parameters in the presence of nonlinear budget constraints. We then integrate the analysis of nonparticipation into our analysis of non-linear budget constraints and discuss estimation when the availability of welfare programs affect the shapes of budget sets, which induces not only nonconvexities but also opportunities for participating in multiple programs. Finally, we consider computational simplifications adopted in the literature to render maximum likelihood estimation feasible.

4.1. Modeling Nonlinear Features of Budget Constraints

A general formulation for the economic problem considered in the subsequent discussion specifies an agent as solving the following optimization problem:

$$\text{Max} \quad U\ (c, h, z, \nu) \quad \text{subject to} \quad b\ (c, h, W, Y) = 0 \tag{4.1}$$

where $U()$ delineates the utility function, c and h measure consumption and hours of work, the quantities z and ν represent respectively the observed and unobserved factors influencing choices beyond those incorporated in budget sets, and the function $b()$ specifies the budget constraint with W and Y designating the real gross wage per hour and nonlabor income (note that we use upper case to distinguish from marginal wage and virtual nonlabor income). For the moment, we restrict the economic framework to be static and the quantities c and h to be single goods rather than multidimensional vectors. In many applications the budget function, b, is not differentiable, and in some it is not even continuous.

For the familiar linear specification of the budget constraint, b takes the form:

$$b\ (c, h, W, Y) = Wh + Y - c. \tag{4.2}$$

Solving (4.1) for this form of b yields the following labor supply and consumption functions:

$$\begin{aligned} h &= \ell\ (W, Y, Z, \nu) \\ c &= c\ (W, Y, Z, \nu), \end{aligned} \tag{4.3}$$

which correspond to the standard demand functions for nonmarket time (i.e., leisure) and consumption. (The subsequent analysis often suppresses the Z argument in the functions $U()$, $\ell()$ and $c()$ to simplify notation.)

Another popular specification of $b()$ incorporates income or sales taxes in characterizing choices, with the budget constraint written as some variant of

$$b\ (c, h, W, Y) = Wh + Y - c - \tau(Wh, Y)\ \ , \tag{4.4}$$

where the function $\tau()$ gives the amount paid in taxes. This formulation for b admits different tax rates on earnings (Wh) and nonlabor income (Y). If these income sources are instead taxed the same, then (4.4) further simplifies to

$$b\ (c, h, W, Y) = Wh + Y - c - \tau(I)\ \ , \tag{4.5}$$

where tax payments $\tau(I) = \tau(I(h))$ where $I(h)$ = taxable income $= Wh + Y - D$ with D designating allowable deductions. Different marginal tax rates in the various income brackets combined with the existence of nonlabor income creates inherent nonlinearities in budget sets

The literature relies on two approaches for modeling nonlinearities in budget sets: piecewise-linear characterizations and smooth differentiable functions. To illustrate these approaches, the subsequent discussion principally focuses on the income-tax formulation of b given by (4.4) and (4.5) to illustrate central concepts.

4.1.1. Piecewise Linear Constraints

As a simple characterization of piecewise budget sets, Figure 4.1 shows a hypothetical budget constraint for an individual faced with a typical progressive income tax schedule defined by a series of income brackets. In this diagram, h denotes hours of work, and C measures total after-tax income or the consumption of market goods. The budget constraint is composed of several segments corresponding to the different marginal tax rates that an individual faces. In particular, he faces a tax rate of t_1 between $\overline{h}_o$ hours and $\overline{h}_2$ hours (segment 1 of his constraint) and tax rates of t_3 and t_5 respectively in the intervals $(\overline{h}_2, \overline{h}_4)$ and $(\overline{h}_4, \overline{h}_6)$ (segments 3 and 5 in the figure). Thus, with the variable W denoting the individual's gross wage rate, the net wages associated with each segment are: $w_1 = (1-t_1)W$ for segment 1; $w_3 = (1-t_3)W$ for segment 3; and $w_5 = (1-t_5)W$ for segment 5. Also, each segment has associated with it a virtual income (i.e., income associated with a linear extrapolation of the budget constraint) calculated as:

$$\begin{aligned} y_1 &= Y - \tau(0, Y) \; ; \\ y_j &= y_{j-2} + W(t_{j-2} - t_j)\overline{h}_{j-1} \;\; \text{for} \;\; j = 3, 5, \ldots \; . \end{aligned} \tag{4.6}$$

So, $y_3 = y_1 + (w_1 - w_3)\overline{h}_2$;and $y_5 = y_3 + (w_3 - w_5)\overline{h}_4$. Changes in tax brackets create the kink points which are designated 0, 2, 4, and 6.

Figure 4.2 illustrates stylized features of a budget constraint modified to incorporate an earned income tax credit (EITC) in conjunction with an income tax[12], and Figure 4.3 shows a prototype budget set induced by a conventional welfare program (or social security tax).[13] In Figure 4.2, the EITC increases benefits

[12] An earned income tax credit (EITC) constitutes a negative income tax scheme, which induces two kinks in a person's constraint in the simplest case: one where the proportional credit reached its maximum (h_2 in Figure 4.2), and one at the break even point where the credit was fully taxed away (h_4 in the figure). The tax rates associated with the first two segments are t_A, which is negative, and t_B, which is positive. Thereafter, the EITC imposed no further tax.

[13] A welfare program pays a family some level of benefits at zero hours of work, and then "taxes" this nonlabor income at some benefit reduction rate until all benefits are gone. Figure 4.3 assumes a proportional benefit reduction rate applies on earnings until benefits decline to zero, after which the family pays normal income taxes whihc here too is assumed to be a proportional rate. Thus, Figure 4.3 shows a constraint with a single interior kink (given by h_1 in the figure) corresponding to the level of earning when welfare benefits first become zero. The

until an individual reaches h_1 hours of work, and then benefits decline until hours attain h_2 when the regular income tax schedule applies. In Figure 4.3, welfare benefits start at $y_1 - y_2$ when a family does not work, they steadily decline as the family increases its hours of work until its earnings reach the value implied at h_1 hours when the family becomes disqualified for welfare. Each of these low-income support programs introduces regressive features in the overall tax schedule faced by a family, which in turn induces nonconvex portions in the budget sets.

In real world applications of piecewise budget constraints, the combination of various tax and public assistance programs faced by families implies budget sets have two noteworthy features. First, the constraint faced by a typical individual includes a large number of different rates. Translated into the hours-consumption space, this implies a large number of kink points in the budget constraint. Second, for most individuals the tax schedule contains nonconvex portions, arising from four potential sources. The first arises from the EITC program, as illustrated in Figure 4.2. A second source arises if a worker's family participates in any welfare program, within nonconvexities arise as benefits are withdrawn as earnings increase as illustrated in Figure 4.3. Third, social security taxes phase out after a fixed level of earnings, so they too induce a budget set similar in structure to that given by Figure 4.3. Finally, the standard deduction present in most income tax programs, wherein no taxes are paid on sufficiently low levels of income, creates yet another source of regressivity in the tax schedule and corresponding nonconvexities in the budget constraint.

4.1.2. Constructing Differentiable Constraints

Several approaches exist for approximating the piecewise-linear tax schedules by a differentiable function. A convenient method for constructing this function is to fit the marginal tax rate schedule–a step function–by a differentiable function. This approximation must itself be easily integrable to obtain a simple closed form for the tax function.

An elementary candidate for constructing a differentiable approximation that can be made as close as one desires to the piecewise-linear tax schedule has been

tax rate on the segment leading up to that kink is t_A, switching to t_B on the second segment. The Social Security systems induces a similar effect on the budget constraint.

applied in MaCurdy et al.(1990). To understand the nature of the approximation, return to Figure 4.1. One can represent the underlying schedule as follows:

$$\begin{aligned} \tau_e(Wh, Y) &= t_1 \quad \text{from } I(\overline{h}_0) \text{ to } I(\overline{h}_2) \\ &= t_3 \quad \text{from } I(\overline{h}_2) \text{ to } I(\overline{h}_4) \\ &= t_5 \quad \text{above } I(\overline{h}_4)\ , \end{aligned} \tag{4.7}$$

$$\begin{aligned} \text{where}\ \ \tau_e(Wh, Y) &= \text{the marginal tax rate on earnings} \\ I(h) &= \text{taxable income at h hours of work, and} \\ t_i &= \text{marginal tax rate, i = 1,3,5.} \end{aligned}$$

For expositional simplicity, suppose that $t_1 = 0$. Consider the following approximation of this schedule:

$$\widehat{\tau}_e(Wh, Y) = t_3\ ((\Phi_1(I(h)) - \Phi_3(I(h)) + t_5\Phi_3(I(h))\ \ . \tag{4.8}$$

This formulation for the marginal tax rate switches among three flat lines at the heights t_1 $(= 0)$, t_3 and t_5. The weight functions $\Phi_i(I(h))$ determine the rate at which the shift occurs from one line to another, along with the points at which the switches take place. Candidate weight functions are given by $\Phi_{_i}(I(h)) =$ the cumulative distribution function with mean y_i and variance σ_i^2, $i = 1, 3$. The middle segment of the tax schedule has height t_3 and runs from taxable income $I(\overline{h}_2)$ to $I(\overline{h}_4)$. To capture this feature, parameterize $\Phi_1(\cdot)$ and $\Phi_3(\cdot)$ with means $y_1 = I(\overline{h}_2)$ and $y_3 = I(\overline{h}_4)$, respectively, with both variances set small. The first distribution function, $\Phi_1(\cdot)$ takes a value close to zero for taxable income levels below $I(\overline{h}_2)$ and then switches quickly to take a value of one for higher values. Similarly, $\Phi_3(\cdot)$ takes a value of zero until near $I(\overline{h}_4)$ and one thereafter. The difference between the two equals zero until $I(h_2)$, one from $I(\overline{h}_{_2})$ to $I(\overline{h}_4)$ and zero thereafter. Thus, the difference takes a value of one just over the range where t_3 is relevant. Notice that we can control when that value of one begins and ends by adjusting the values y_1 and y_3. Also, one can control how quickly this branch of the estimated schedule turns on and off by adjusting the variances of the cumulative distribution functions, trading off a more gradual, smoother transition against more precision. In general, adjusting the mean and variance parameters allows one to fit each segment of a schedule virtually exactly, switch quickly between segments, and still maintain differentiability at the switch points.

A generalization of this approximation takes the form

$$\widehat{\tau}_e(Wh, Y) = \sum_{i=1,3,\ldots} \left[\Phi_{i-2}(I(h)) - \Phi_i(I(h))\right] \ t_i(I(h)) \tag{4.9}$$

where the functions $t_i(I(h))$ permit tax schedules to be nonconstant functions of taxable income within brackets. With the Φ_i denoting many cdfs associated with conventional continuously distributed distributions, function (4.9) yields closed form solutions when it is either integrated or differentiated.[14] Integrating (4.9) yields a formulation for the budget constraint $b(c, h, W, Y)$. The resulting approximation can be made to look arbitrarily close to the budget set boundary drawn in Figures 4.1, 4.2 or 4.3, except that the kink points are rounded.

Formula (4.9) can be extended to approximate virtually any specification of $b\ (c, h, W, Y)$. One can readily allow for distinct relationships describing the derivatives for each of the arguments of this function, and nonconvexities in budget sets cause no particular problems.

4.2. Simple Characterizations of Labor Supply and Consumption with Differentiable Constraints

A useful solution exists for the hours-of-work and consumption choices associated with utility maximization when budget constraints form a set with a

[14]Total taxes are given by: $\tau(I) = \int \tau'(I)dI$. The following relations enable one to calculate an explicit form for $\tau(X)$:

$$\begin{aligned}
\int \Phi dI &= I\Phi + \varphi \\
\int I\Phi dI &= \frac{1}{2}I^2\Phi - \frac{1}{2}\Phi + \frac{1}{2}I\varphi \\
\int I^2\Phi dI &= \frac{1}{3}I^3\Phi + \frac{2}{3}\varphi + \frac{1}{3}I^2\varphi \\
\int I^3\Phi dI &= \frac{1}{4}I^4\Phi - \frac{3}{4}\Phi + \frac{3}{4}I\Phi + \frac{1}{4}I^3\varphi
\end{aligned}$$

In this expression, Φ refers to any Φ_i's, and φ designates the density function associated with Φ_i.

twice-differentiable boundary. Specify the marginal wage rate as:

$$\omega = \omega(h) = b_h(c, h, W, Y) = b_h \tag{4.10}$$

and "virtual" income as:

$$y = y(h) \quad \text{which solves the equation} \quad b(hb_h + y, h, W, Y) = 0. \tag{4.11}$$

This solution for y satisfies:

$$y = y(h) = c - \omega h \ .$$

For the familiar specification found in analyses taxes and labor supply given by $b(c, h, W, Y) = Wh + Y - c - \tau(Wh + Y)$ with the function τ constituting the amount paid in taxes at before-tax income $Wh + Y$, the expressions for marginal wage and virtual income y simplify to:

$$\begin{aligned} \omega &= \omega(h) = (1 - \tau')W \\ y &= y(h) = Wh + Y - \omega h - \tau = Y + \tau' Wh - \tau \end{aligned} \tag{4.12}$$

where τ and τ' (the derivative of the tax function with respect to income) are evaluated at income level $I = I(h) = Y + Wh$ which directly depends on the value of h.

Utility maximization implies solutions for hours of work and consumption that obey the implicit equations:

$$\begin{aligned} h &= \ell\ (\omega, y, z, v) = \ell\ (\omega(h), y(h), z, v) \\ c &= c\ (\omega, y, z, v) = c\ (\omega(h), y(h), z, v) \end{aligned} \tag{4.13}$$

where ℓ and c represent the same conventional forms for labor supply and consumption demand functions given by (4.3). Figures 4.1 and 4.3 illustrate this representation of the solution for optimal hours of work and consumption. The characterization portrays an individual as facing a linear budget constraint in the presence of nonlinear tax programs. This linear constraint is constructed in a way to make it tangent to the actual nonlinear opportunity set at the optimal solution for hours of work. The implied slope of this linearized constraint is $\omega(h)$ and the corresponding value of virtual income is $y(h)$.

Relationships (4.11) constitute structural equations that determine hours of work and consumption. By applying the Implicit Function Theorem to specification (4.11), we can solve this implicit equation for h in terms of W, Y, and other variables and parameters entering the functions b and U. This operation produces the labor supply and consumption functions applicable with general forms of nonlinear budget sets.

4.3. Instrumental-Variable Estimation

The inclusion of taxes provide an additional reason for allowing for the endogeneity of (after tax) wages and other income. Writing the labor supply function as

$$h = \ell\left(\omega(h), y(h), z, \upsilon\right) = \ell^{*}\left(\omega(h), y(h), z\right) + \upsilon \quad . \tag{4.14}$$

makes the point explicitly. The Instrumental variable approach described earlier can be applied as well as the grouping methods (which of course is just an application of IV). The implementation of IV procedures imposes no parametric restrictions and it allows one to consider a wide variety of exogeneity assumptions. The fact that the error term does not interact with the wage and other income is critical for the interpretation of IV as identifying the structural parameters of the model.

4.3.1. Including Measurement Error

In many data sets there are serious suspicions that hours of work and wages are reported with error. This issue acquires added importance when we are dealing with non-linear tax schedules since this creates a problem of observability of the correct tax rate, which is the reason we introduce the issue here.

Suppose H denotes measured hours of work and that the function

$$H = H(h, \varepsilon) \tag{4.15}$$

relates H to actual hours h and to a randomly distributed error ε Typically, analyses presume that the state $h = 0$ is observed without error.

Measurement errors in hours often induce errors in observed wage rates since researchers construct wages by dividing total labor earnings, E, by hours worked in the period. Whereas $W = E/h$ defines the true hourly wage rate, $\tilde{W} = E/H$ designates the data available on wages. Measured wages $\tilde{W}$ are contaminated by reporting errors even when E provides accurate quantities for each individual's total labor earnings and wages are indeed constant for hours worked over the period. This formulation presumes a reciprocal relation in the measurement error linking data on hours and wages. More generally, suppose $\tilde{W}$ links to the true wage rate according to the relationship

$$\tilde{W} = \tilde{W}(W, h, \varepsilon) \quad . \tag{4.16}$$

In the reciprocal measurement error example, $\tilde{W} = Wh \ / \ H(h, \varepsilon)$ where $H(h, \varepsilon)$ comes from (4.15).

The presence of measurement errors in hours typically invalidates use of non-linear IV procedures to estimate the structural labor supply equation given by (4.14). Expressing this equation in terms of H rather than h involves merely substituting (4.15) into (4.14); and if measurement error also carries over to wages, then substitutions must be made for wages as well. These replacements typically result in a variant of structural equation (4.14) that cannot be transformed into a form that is linear in disturbances. Measurement errors in hours invariably render the marginal tax rate unobservable, which in turn makes both the marginal wage ($\omega(h)$) and virtual income ($y(h)$) also unobservable. Sophisticated adjustments must be included to account for such factors. These complications motivate many researchers to turn to maximum likelihood procedures to estimate hours-of-work specifications as we do below. However with some additional assumptions IV procedures are still possible, at least when the issue of censoring does not arise.

Suppose measurement error is of the multiplicative kind

$$H = H(h, \varepsilon) = he^{\varepsilon} \quad \text{with} \quad \tilde{W} = E/H, \tag{4.17}$$

In the presence of such error, specifications can also be found that allow for use of IV procedures to estimate substitution and income parameters. Incorporating the multiplicative measurement error model (4.17) into the semilog functional form of labor supply given in relation yields the empirical specification:

$$H = \overline{u} + Z\gamma + \alpha \ln \omega^m + \beta y + u \tag{4.18}$$

where

$$\begin{aligned}
\ln \omega^m &= \ln(E/H) + \ln(1-\tau') \\
\widetilde{y} &= y - \alpha\sigma_\varepsilon^2/2 \\
u &= \nu + \alpha(\varepsilon - E(\varepsilon)) + (H-h) = \nu + \alpha(\varepsilon - E(\varepsilon)) + h(e^\varepsilon - 1).
\end{aligned}$$

The disturbance u possesses a zero mean since $E(e^\varepsilon) = 1$. Virtual income $y(h)$ and the marginal tax rate $\tau\prime$ are not contaminated by measurement error because they are only functions of Y and $hW = H\ \tilde{W}$, quantities which are both perfectly observed (by assumption). The variable $\ln \omega^m$ represents the natural logarithm of the after-tax wage rate evaluated at observed hours, which differs from the actual marginal wage due to the presence of reporting error in hours. Assuming the error ε is distributed independently of all endogenous components determining h, including the heterogeneity disturbance v, the instrumental variables X applicable for estimation of the original specification can also serve as the instrumental variables in estimating the coefficients of (4.18) by familiar IV methods.

4.3.2. Sources of Unobservables in Budget Sets

An important class of models not widely recognized in the literature involves budget constraints that vary across individuals in ways that depend on unobserved factors. The modification required in the above analysis to account for such factors replaces budget function $b()$ appearing in (4.1) by:

$$b\ (c, h, W, Y, z, \xi) = 0\ . \tag{4.19}$$

The quantity z captures the influence of measured characteristics on budget sets. Classic examples include family characteristics that alter the form of the tax function relevant for families. The error component ξ represents unobserved factors shifting budget sets. Classic examples here include unmeasured components of fixed costs, prices, and elements determining tax obligations.

The presence of ξ in $b()$ typically renders IV methods inappropriate for estimating parameters of the labor supply function ℓ. The usual problem comes about since structural variants of ℓ cannot be found that are linear in disturbances, and this is especially true when nonlinearities exist in tax schedules. When ξ appears as a component of $b()$, researchers typically rely on the maximum likelihood methods summarized in the subsequent discussion to conduct estimation of behavioral models of hour of work and consumption.

4.3.3. Complications of IV Estimation with Piecewise-Linear Constraints

Naive application of instrumental-variable methods with piecewise-linear budget constraints generally produces inconsistent estimates of behavioral parameters, even ignoring the potential presence of measurement error. Section 4.6 below presents the structural specification–see (4.69)–implied for hours of work when Figure 4.1 designates the budget set and everyone works. As noted in Section 4.1, this budget set is convex and consists of three segments. Inspection of structural equation (4.69) reveals that the structural error is $\sum_{j=1,2,3} d_j v$ where d_j represents an indicator variable signifying whether an individual selects segment $j = 1, 2,$or 3. If the individual occupies any kink, then $\sum_{j=1,2,3} d_j v = 0$. Suppose X includes the set of instrumental variables presumed to satisfy $E(\nu|X) = 0$. The corresponding conditional expectation of the structural error implied by equation (4.69) is $\sum_{j=1,2,3} \Pr(d_j \mid X) E(v \mid d_j = 1, X)$. This expectation is typically not zero, a condition required to implement IV techniques. To use IV procedures in the estimation of equation (4.69) necessitates the inclusion of sample selection terms adjusting for the nonzero expectation of $\sum_{j=1,2,3} d_j v$.

4.3.4. Non Participation and Missing Wages

In the earlier sections we discussed how the estimation approach needs to be generalized so as to allow for non-participation and for missing wages, which present further complications for estimation. We argued that standard Instrumental Variables are not appropriate in this context. We now turn to Maximum likelihood estimation which we set up to deal with the problems introduced above, namely non-linear taxes, measurement error, missing and/or endogenous wages and other income and of course non-participation.

4.4. Maximum Likelihood Estimation: A General Representation

The instrumental variable estimator, developed in the last section, required exclusion restrictions to consistently estimate the parameters of the labor-supply and consumption models involving nonlinear budget sets. In contrast, Maximum Likelihood estimation exploits the precise structure of the budget constraint and need not rely on exclusion restrictions to identify parameters. Even though marginal wages and virtual incomes are endogenous, nonlinearities introduced through distributional assumptions provide a valuable source of identification. However, exclusion restrictions are only avoided in this approach if gross wages and incomes are assumed to be exogenous and in many applications of maximum likelihood researchers also impose stringent distributional and independence assumptions on sources of errors capturing heterogeneity and measurement error. Nonetheless, one can entertain a wide array of nonlinearities in budget sets and decision processes, along with rich specifications for heterogeneity and mismeasurement of variables.

The following discussion begins with a general presentation describing the application of maximum likelihood methods in hours of work and consumption analyses allowing for flexible distributional assumptions and intricate forms of nonlinearities in both preferences and budget constraints. This analysis draws heavily upon Appendix A. Later subsections cover simple illustration of techniques, many of which have been implemented in the empirical literature.

4.4.1. Dividing the Budget Constraint into Sets

Irrespective of whether one considers differentiable or piecewise-linear formulations for budget constraints, the essential idea underlying development of likelihood functions in the presence of nonlinear constraints involves defining a set of "states of the world". Each state designates a particular segment of the budget set boundary, with states being mutually exclusive and states jointly covering all parts of budget constraints. One interprets individuals as being endowed with a set of attributes determining of their tastes, resources and constraints, with these attributes viewed as random variables continuously distributed across the population. Based on the realizations of these variables, an individual selects consumption and hours of work to optimize utility.

Regarding the distribution of these variables in the previous discussion, suppose unobserved heterogeneity influencing preferences, ν, the unmeasured factors determining wages, η, and the unobservables incorporating budget sets, ξ, possess the following joint density:

$$\varphi\ (v, \eta, \xi) \equiv \varphi\ (v, \eta, \xi \mid X) \qquad \text{for} \quad (v, \eta, \xi) \in \Omega. \tag{4.20}$$

When errors, ε, contaminate the measurement of hours, the relevant joint distribution becomes:

$$\varphi\ (v, \eta, \xi, \varepsilon) \equiv \varphi\ (v, \eta, \xi, \varepsilon \mid X) \qquad \text{for} \quad (v, \eta, \xi, \varepsilon) \in \Omega \tag{4.21}$$

Both these expressions admit conditioning on a set of exogenous variables X, but the subsequent analysis suppresses X to simplify the notation. The set Ω designates the domain of these random variables.

In this setting, n states of the world can occur. The discrete random variable δ_i signifies whether state i happens, with $\delta_i = 1$ indicating realization of state i and $\delta_i = 0$ implying that some state other than i occurred. A state refers to locations on boundaries of budget sets, to be explained further below. Consequently, the value of δ_i depends on where (v, η, ξ) falls in its domain determined by the rule:

$$\delta_i = \begin{cases} 1 & \text{if } (v, \eta, \xi) \in \Omega_i \\ 0 & \text{otherwise} \end{cases} \tag{4.22}$$

where the set Ω_i constitutes that subset of the sample space Ω for which utility maximization yields a solution for consumption and hours that lies within the $\delta_i = 1$ portion of the budget. The mutually exclusive and exhaustive feature of the sets Ω_i for $i = 1, ..., n$ implies $\bigcup_{i=1}^{n} \Omega_i = \Omega$ and $\Omega_i \cap \Omega_j = \emptyset$ for $i \neq j$.

A central requirement invoked in dividing a budget constraint into its various sections involves ensuring that unique solutions exist for c and h for any $(v, \eta, \xi) \in \Omega_i$. Consumption and hours of work may take on discrete values when $(v, \eta, \xi) \in \Omega_i$. Alternatively, there may be a continuous mapping relating C and h to (v, η, ξ) within the set Ω_i, but inverses must exist for the consumption and labor supply functions

$$\begin{aligned} h &= \ell\ (\omega, y, z, v) = \ell\ (\omega(h), y(h), z, v) \\ c &= c\ (\omega, y, z, v) = c\ (\omega(h), y(h), z, v) \end{aligned}$$

expressed in terms of components of v. (These functions correspond directly to those in (4.13) except that marginal wage ω and virtual income y now are functions of ξ, the unobservable components b.) Considering the labor supply function ℓ, this requirement implies existence of the inverse function

$$v = v^h\left(h, \omega\left(h\right), y\left(h\right), z\right) \equiv \ell^{-1}\left(h, \omega\left(h\right), y\left(h\right), z\right) \tag{4.23}$$

for values of v within the set Ω_i. If v is in fact multi-dimensional (i.e., v' is a vector), then an inverse must exist of the form

$$v_1 = v^h\left(h, \omega\left(h\right), y\left(h\right), z, v_2'\right) \equiv \ell^{-1}\left(h, \omega\left(h\right), y\left(h\right), z, v_2\right) \tag{4.24}$$

for some decomposition $v' = (v_1, v_2')$.

Division of the budget constraint into the events $\delta_i = 1$ for $i = 1, ..., n$ generally creates two varieties of sets. First, differentiable segments of the budget constraint over which consumption and hours vary continuously in response to variation in preferences and constraint variables. Second, kink points at which consumption and hours of work take fixed discrete values implied by the location of the kink.

4.4.2. Maximum Utility Determines Placement on Budget Constraint

The portion of a budget constraint selected by an individual depends on the level of utility assigned to this state. The following discussion first characterizes maximum utility attainable on differentiable segments, and then considers evaluations at kink points.

For the differentiable segments of the constraint, utility is determined by function

$$\begin{aligned}
V &= U\ \left(c\ (\omega, y, z, \nu), \ell\ (\omega, y, z, \nu), z, \nu\right) \qquad (4.25)\\
&= U\ \left(c\ (\omega(h,\xi), y(h,\xi), z, \nu, \xi), \ell\ (\omega(h,\xi), y(h,\xi), z, \nu, \xi), z, \nu\right)\\
&\equiv V\ \left(\omega\left(h,\xi\right), y\left(h,\xi\right), z, v\right)\\
&= V\ \left(\omega, y, z, v\right)
\end{aligned}$$

evaluated at optimal points in the specified set. The function $V\ (W, Y, z, v)$ represents the conventional indirect utility function associated with maximizing $U\ (C, h, Z, \nu)$

in (4.1) subject to the linear form of the budget constraint given by (4.2). Roy's Identity specifies that the labor supply function ℓ can be written as

$$\ell\ (\omega, y, z, \nu) \equiv \frac{V_\omega\ (\omega, y, z, \nu)}{V_y\ (\omega, y, z, \nu)} \tag{4.26}$$

with V_ω and V_y denoting the partial derivatives of V. Suppose the interval $(\bar{h}_{i-1}, \bar{h}_{i+1})$ identifies the differential segment under consideration. The subsequent discussion refers to this segment as state i. Then the utility assigned to state i corresponds to the maximum value of V achievable for hours falling in the interval $(\bar{h}_{i-1}, \bar{h}_{i+1})$.

Difficulty in determining the achievable value of V depends on characteristics of the budget function $b(c, h, W, Y)$. For the most general specifications of b, inspection of relations (**??**) and (4.11) defining ω and y reveals that each depends on both c and h through the derivative b_h. If utility maximization occurs at an interior point of $(\bar{h}_{i-1}, \bar{h}_{i+1})$ given the realization of (υ, η, ξ), then the implied values of c and h solve the system

$$\begin{aligned} h =&\ \ \ell\ (\omega, y, z, \upsilon) \in (\bar{h}_{i-1}, \bar{h}_{i+1}) \\ b\ (c, h, W, Y, \xi) =&\ \ 0\ . \end{aligned} \tag{4.27}$$

Consequently, the maximum utility attainable on the interval $(\bar{h}_{i-1}, \bar{h}_{i+1})$ is V (or U) evaluated at these solutions for c and h. Define this maximum utility as $V_{(i)}$, where the (i) subscript on V signifies utility assigned to state i. If one extends state i to include either of the exterior points $\bar{h}_{i-1}$or $\bar{h}_{i+1}$, and uniqueness and differentiability continue to hold at these points, then the above procedure still applies in assigning a value for $V_{(i)}$. The subsequent discussion ignores such easily-handled extensions to simplify the exposition.

Use of indicator functions provides an expression for $V_{(i)}$. One can characterize the set of values of C and h satisfying equations (4.27) as

$$\{\ (c, h)\ \mid\ I\ [\ \ h = \ell\ (\omega, y, z, \upsilon) \in (\bar{h}_{i-1}, \bar{h}_{i+1})\ \ ;\ \ b\ (c, h, W, Y) =\ 0\ \ ] = 1\ \}\ , \tag{4.28}$$

where I denotes the indicator function defined by

$$I\ [\ conditions\ ] = \begin{cases} 1 & \text{if } [\ \textit{all conditions}\ ] \text{ are true} \\ 0 & \text{if } [\ \textit{any condition}\ ] \text{ is false}\ . \end{cases}$$

The indicator function I in (4.28) depends on satisfaction of 2 conditions. Using I, a simple expression for the maximum utility attainable in state i is given by

$$V_{(i)} = \; V \;\; (\omega, y, z, v) * I \; [\; h = \ell \; (\omega, y, z, v) \in (h_{i-1}, h_{i+1}) \; ; \; b \; (c, h, W, Y) = \; 0 \;] \quad . \tag{4.29}$$

For values of v, η and ξ not yielding a solution in state i, $V_{(i)} = 0$. It is possible in this analysis for $V_{(i)} = 0$ for all values of admissible values of (v, η, ξ) (i.e., $\Omega_i = \emptyset$). Throughout this discussion, we assume a utility function normalized so that $U \; (c, h, z, \nu) > 0$ for all admissible values of variables. So the event $V_{(i)} = 0$ always means that some state other than i has a higher assigned utility.

The most popular specifications of the budget function $b(c, h, W, Y)$ have derivatives b_h that depend on h but not on c. Examples include those specifications incorporating income or sales taxes given by (4.4). Under these circumstances, the first equation in (4.27) alone can be solved for h. Thus, $V_{(i)}$ simplifies to

$$V_{(i)} = \; V \;\; (\omega, y, z, v) * I \; [\; h = \ell \; (\omega, y, z, v) \in (h_{i-1}, h_{i+1}) \;\;] \tag{4.30}$$

This expression serves as the principal formulation used in the subsequent discussion.

The portion of a budget constraint selected by an individual depend on the level of utility assigned to this state. At kink points, utility takes the value

$$V_{(i)} = U \; (\bar{C}_i, \bar{h}_i, z, \nu) \tag{4.31}$$

where $\bar{C}_i$ and $\bar{h}_i$ designate the values of consumption and hours at the kink point associated with state i.

An individual occupies that portion of the budget constraint corresponding to state i if the assigned utility is highest for this state. According to (4.22), the subspace of (v, η, ξ) yielding this realization is the set Ω_i. Correspondingly, one can represent Ω_i as

$$\Omega_i = \{ \; (v, \eta, \xi) \;\; | \;\; V_{(i)} > V_{(j)} \quad \text{for all} \quad j \neq i \; \} \quad . \tag{4.32}$$

Relationships (4.29) (or (4.30)) and (4.31) define $V_{(i)}$ depending on characteristics of the state. For expositional simplicity without loss of generality, the subsequent discussion ignores equalities $V_{(i)} = V_{(j)}$ in defining the sets Ω_i since these events are zero probability events.

4.4.3. Density Functions for Hours and Wages

The distribution of consumption and hours of work depends on where individuals locate on the budget constraints. The probability that an individual makes selections falling within the state i portion of the budget equals:

$$\begin{aligned} P\ (\delta_i = 1) &= P\ ((v, \eta, \xi) \in \Omega_i\ ) \\ &= \int \cdots \int_{\Omega_i} \varphi\ (v, \eta, \xi)\ dv\ d\eta\ d\xi \\ &\equiv \int_{\Omega_i} \varphi\ (v, \eta, \xi)\ dv\ d\eta\ d\xi\ \ . \end{aligned} \tag{4.33}$$

The notation $\int \cdots \int_{\Omega_i}$ denotes integration over the set Ω_i, which the third line of this equation expresses in the shorthand notation $\int_{\Omega_i}$. The joint distribution of the δ_i's takes the form:

$$P\,(\delta_1, ..., \delta_m) = \prod_{i \in M} \left[P(\delta_i = 1)\right]^{\delta_i}$$

where the set M refers to the set of all possible states i that comprise the entire budget constraint. As noted previously, the events $\delta_i = 1$ may refer to either kinks or differentiable constraints.

When an optimum occurs at a kink point, the distribution of hours conditional on this event is

$$P\left(h = \bar{h}_i \mid \delta_i = 1\right) = 1\ \ . \tag{4.34}$$

This distribution is, of course, discrete.

On differentiable segments of the constraint, the distribution for hours is continuous. Performing a conventional change of variables yields the density

$$\begin{aligned} f\,(h, \eta, \xi) &= \frac{dv^h}{dh} \varphi\ \left(v^h, \eta, \xi\right) \\ &= \frac{dv^h}{dh} \varphi\ \left(v^h\,(h, \omega, y, z)\,, \eta, \xi\right) \end{aligned} \tag{4.35}$$

where

$$v^h = v^h\,(h, \omega\,(h)\,, y\,(h)\,, Z) = \ell^{-1}\,(h, \omega\,(h)\,, y\,(h)\,, z) \tag{4.36}$$

refers to the inverse of labor supply function (4.3), and the quantity

$$\frac{dv^h}{dh} = (\frac{\partial \ell}{\partial \omega}\frac{\partial \omega}{\partial h} - \frac{\partial \ell}{\partial y}\frac{\partial y}{\partial h})\ (\frac{\partial \ell}{\partial v})^{-1} \tag{4.37}$$

represents the Jacobian associated with this inverse.[15] The terms $\frac{\partial \ell}{\partial \omega}$ and $\frac{\partial \ell}{\partial y}$ correspond to the economic concepts of substitution and income effects, and the quantity $\frac{\partial \ell}{\partial v}$ determines how unobserved components of preferences influence labor supply. (In applying the change-of-variables formula, Jacobians must be constructed to be uniquely signed for densities to be properly defined. This result follows here because the selection of budget-set partitions ensures a unique solution exists for c and h for each partition combined with the innocuous assumption that unobserved components enter preferences such that $\frac{\partial \ell}{\partial v} > 0$). For the remaining terms in (4.37), differentiation of the budget constraint implies:

$$\frac{\partial w}{\partial h} = b_{hh} \tag{4.38}$$

and

$$\frac{\partial y}{\partial h} = -\frac{b_h}{b_c} - b_h - b_{hh}h \tag{4.39}$$

where the subscripts on the budget function b signify partial derivatives.[16] Assuming the popular tax form for the budget function given by (4.5), expressions (4.38) and (4.39) simplify to

$$\frac{\partial \omega}{\partial h} = (1 - \tau')\ \ W \tag{4.40}$$

and

$$\frac{\partial y}{\partial h} = \tau''\ W\ ^2\ h \tag{4.41}$$

[15] Derivation of this Jacobian follows from total differentiation of relation (4.4.x) with respect to h–treating ν as a function of h–and solving for $\frac{\partial \nu}{\partial h}$.

[16] Derivation of the expression for $\frac{\partial y}{\partial h}$ follows from total differentiation of the relation (4.10) defining y which yields

$$b_c \left(b_h + b_{hh}h + \frac{\partial y}{\partial h} \right) + b_h = 0$$

and solving this equation.

where τ' and τ'' denote the marginal tax rate and its derivative. Division of the budget set into states ensures that inverse (4.36) and its Jacobian (4.37) exist in the space defined by each state.

The implied density of h conditional on δ_i is

$$f(h \mid \delta_i = 1) = \frac{\int_{\Phi_{i|h}} f\ (h, \eta, \xi)\ d\eta\ d\xi}{P(\delta_i = 1)} \qquad \text{for} \qquad h \in \Theta_{i\cdot\ h} \tag{4.42}$$

where the set $\Theta_{i\cdot\ h} = (h_{i-1}, h_{i+1})$ designates the domain of h given occurrence of $\delta_i = 1$ and the notation $\int_{\Phi_{i|h}}$ denotes integration of (η, ξ) over the set

$$\Phi_{i|h} = \{\ (\eta, \xi)\ :|\ \ I\ [\ h = \ell\ (\omega, y, z, v)\ ;\ (v, \eta, \xi) \in \Omega_i\ ] = 1\ \}\ . \tag{4.43}$$

The set $\Phi_{i|h}$ treats h as fixed and, therefore, is a function of h.

Performing a further change of variables for wages yields the following joint density for hours and wages

$$\begin{aligned} f\ (h, W, \xi) &= \frac{d\eta^w}{dW}\ f\ (h, \eta^w, \xi) \\ &= \frac{d\eta^w}{dW}\ f\ (h, \eta^w(W, Q), \xi) \end{aligned} \tag{4.44}$$

where

$$\eta^w = \eta^w(W, Q) = W^{-1}(W, Q) \tag{4.45}$$

denotes the inverse of wage function, and the quantity

$$\frac{d\eta^w}{dW} = [\ \frac{\partial W}{\partial \eta}\ ]^{-1} \tag{4.46}$$

represents the Jacobian associated with this inverse. (For expositional convenience, and without loss of generality, this analysis assumes that a monotonically increasing relationship links W to η; so, (4.46) positive.)

The density of h and W conditional on δ_i is

$$f(h, W \mid \delta_i = 1) = \frac{\int_{\Phi_{i|h,W}} f\ (h, W, \xi)\ d\xi}{P(\delta_i = 1)} \qquad \text{for} \qquad (h, W) \in \Theta_{i\cdot\ h,W} \tag{4.47}$$

where the notation $\int_{\Phi_{i|h,W}}$ denotes integration of ξ over the set

$$\Phi_{i|h,W} = \{ \ (\xi) \ \mid \ I \ [\ h = \ell \ (\omega, y, z, \upsilon) \ ; \ W = W(Q, \eta) \ ; \ (\upsilon, \eta, \xi) \in \Omega_i \ ] = 1 \ \} \quad . \tag{4.48}$$

The set $\Phi_{i|h,W}$ is a function of h and W. One can express the set $\Theta_{i\cdot \ h,W}$ appearing in (4.47) as

$$\Theta_{i\cdot \ h,W} = \{ \ (h, W) \quad : \quad \xi \in \Phi_{i|h,W} \ \} \quad ,$$

which specifies the domain of h and W assuming occupancy of the state i part of the budget constraint.

If υ is multi-dimensional as specified in labor supply function (4.13), then (4.47) becomes

$$f\,(h, W \mid \delta_i = 1) = \frac{\int_{\Phi_{i|h,W}} \ f \ \ (h, \upsilon_2, W, \xi) \ \ d\upsilon_2 \ d\xi}{P\,(\delta_i = 1)} \qquad \text{for} \qquad (h, W) \in \Theta_{i\cdot \ h,W} \tag{4.49}$$

where $f \ (h, \upsilon_2, W, \xi)$ has a form analogous to (4.44), and the notation $\int_{\Phi_{i|h,W}}$ now denotes integration of (υ_2, ξ) over the set

$$\Phi_{i|h} = \{ \ (\upsilon_2, \xi) \quad : \quad I \ [\ h = \ell \ (\omega, y, Z, \upsilon) \ ; \ W = W(Q, \eta) \ ; \ (\upsilon, \eta, \xi) \in \Omega_i \ ] = 1 \ \} \quad .$$

The set $\Phi_{i|h,W}$ still remains a function of h and W .

Finally, when an individual selects an optimum at a kink point and $h = \bar{h}$ is discrete, then the distribution of wages takes the form

$$f \ \left(\bar{h}, W \mid \delta_i = 1\right) = \frac{\int_{\Phi_{i|\bar{h},W}} \ f \ \ (\nu, W, \xi) \ \ d\nu \ d\xi}{P\,(\delta_i = 1)} \qquad \text{for} \qquad W \in \Theta_{i\cdot \ W} \tag{4.50}$$

where the density $f \ (\nu, W, \xi)$ is specified analogous to (4.44), and the set $\Phi_{i|\bar{h},W}$ is a function of $\bar{h}$ and W defined by

$$\Phi_{i|\bar{h},W} = \{ \ (\nu, \xi) \quad : h = \bar{h}_i \ ; \ W = W(Q, \eta) \quad (\upsilon, \eta, \xi) \in \Omega_i \ \} \quad . \tag{4.51}$$

The domain $\Theta_{i\cdot \ W}$ of W in (4.50) corresponds to that part of the overall range of W consistent with being at kink $\bar{h}_i$.

4.4.4. Likelihood Functions for Hours and Wages

Appendix A presents the results required to develop a complete specification of the joint likelihood function for hours (h) and wages (W). Suppose the state $\delta_0 = 1$ refers to an individual choosing not to work; the states $\delta_i = 1$ for $i \in M_c$ designate those circumstances when the person works and selects optimums on differentiable segments of budget constraints; and the states $\delta_i = 1$ for $i \in M_d$ denote those events when an individual chooses hours located at a kink point. Hours (h) are continuously distributed for states in the set $i \in M_c$, and h is discretely distributed in the no-work state and for states in the set $i \in M_d$. Hours possess a combined continuous/discrete distribution. Knowledge of the value of h entirely determines the values of $\delta_0, ..., \delta_n$ where $n+1$ designates the total number of states.

Formula (A.26) of Appendix A implies that the following specification delimits the joint likelihood function of (h, W):

$$\mathcal{L}(h, W) = \mathcal{L}(h, W, \delta_0, ..., \delta_n) = \begin{array}{l} [P\ ((v, \eta, \xi) \in \Omega_0\ )]^{\delta_0} * \\ \prod\limits_{i\ \in\ M_c} \left[\int_{\Phi_{i|h,W}} f\ (h, v_2, W, \xi)\ dv_2\ d\xi \right]^{\delta_i} * \\ \prod\limits_{i\ \in\ M_d} \left[\int_{\Phi_{i|\bar{h},W}} f\ (\nu, W, \xi)\ d\nu\ d\xi \right]^{\delta_i} . \end{array} \tag{4.52}$$

The first line of this expression delineates the probability of not working; the second line–comprised of the numerators of (4.49)–designates the densities of (h, W) unconditional on $\delta_i = 1$; and the third line–encompassing the numerators of (4.50)–demarcates the probability that $h = \bar{h}_i$ combined with the density of W unconditional on $\delta_i = 1$.

4.4.5. Density Functions Accounting for Measurement Error

With measurement error contaminating hours of work, h is no longer observed and one instead has data on measured hours H specified by relation (4.15). Without loss of generality, suppose (4.15) constitutes a monotically increasing relationship that links H to the measurement error component ε. The joint density function (4.21) relates the distribution of ε to the distributions of the structural errors ν, η, and ξ.

On differentiable segments of the budget constraint, the density function for true hours and wages is $f(h, W, \xi, \varepsilon)$ which has a form entirely analogous to (4.44). Performing a conventional change of variables using relationship (**??**) yields the density

$$\begin{aligned} f(h, W, \xi, H) &= \frac{\partial \varepsilon^H}{\partial H} f\ \left(h, W, \xi, \varepsilon^H\right) \\ &= \frac{\partial \varepsilon^H}{\partial H} f\ \left(h, W, \xi, \varepsilon^H(H, h)\right) \end{aligned} \tag{4.53}$$

where

$$\varepsilon^H = \varepsilon^H(H, h) = H^{-1}(H, h) \tag{4.54}$$

refers to the inverse of measurement error function (4.15), and the quantity

$$\frac{\partial \varepsilon^H}{\partial H} = [\ \frac{\partial H}{\partial h}\ ]^{-1} \tag{4.55}$$

designates the Jacobian associated with this inverse. The corresponding density of H and W conditional on $\delta_i = 1$ is

$$f(H, W \mid \delta_i = 1) = \frac{\int_{\Theta_{i\cdot\ h}} \int_{\Phi_{i|h,\tilde{W}}} f\ (h, W, \xi, H)\ d\xi\ dh}{P(\delta_i = 1)} \qquad \text{for} \qquad (H, W) \in \Theta_{i\cdot\ H,W} \tag{4.56}$$

where $\int_{\Theta_{i\cdot\ h}}$ denotes integration over the set $\Theta_{i\cdot\ h}$ which corresponds to the domain of h conditional on $\delta_i = 1$.

When wages are also measured with error through mismeasurement of hours as characterized by relation (4.16), then (4.53) is replaced by

$$f\left(h, \tilde{W}, \xi, H\right) = f\ \left(h, \tilde{W}^{-1}(\tilde{W}, h, H), \xi, H\right) \tag{4.57}$$

where

$$W = \tilde{W}^{-1}\left(\tilde{W}, h, \varepsilon^H(H, h)\right) \equiv \tilde{W}^{-1}\left(\tilde{W}, h, H\right)$$

refers to the inverse of measurement error function (4.16). The corresponding density of H and $\tilde{W}$ conditional on $\delta_i = 1$ becomes

$$f\left(H, \tilde{W} \mid \delta_i = 1\right) = \frac{\int_{\Theta_{i \cdot\ h}} \int_{\Phi_{i|h,W}} f\ \left(h, \tilde{W}, \xi, H\right)\ d\xi\ dh}{P\left(\delta_i = 1\right)} \quad \text{for} \quad (H, \tilde{W}) \in \Theta_{i \cdot\ H, \tilde{W}} \tag{4.58}$$

No change of variables occurs in deriving this expression since $\tilde{W}$ is fully known given values for h, W and H.

A similar situation applies to incorporating measurement error when an individual selects an optimum at a kink point of the budget set. Conditional on realization of the state $\delta_i = 1$, the value of ε is known since one sees H and $h = \bar{h}_i$ with probability one. Defining the $f\ (\nu, W, \xi, \varepsilon)$ as the generalization of the joint density function appearing in (4.50) incorporating measurement error, then substitution of the inverse functions $\tilde{W}^{-1}(\tilde{W}, \bar{h}_i, H)$ and $\varepsilon^H\left(H, \bar{h}_i\right)$ introduced above into this joint density yields

$$f\left(\nu, \tilde{W}, \xi, H\right) = f\ \left(\nu, \tilde{W}^{-1}(\tilde{W}, \bar{h}_i, H), \xi, \varepsilon^H\left(H, \bar{h}_i\right)\right)\ . \tag{4.59}$$

Following the steps above, one can readily verify that the density of $(H, \tilde{W})$ conditional on $\delta_i = 1$ takes the form

$$f\ \left(H, \tilde{W} \mid \delta_i = 1\right) = \frac{\int_{\Phi_{i|\bar{h},\tilde{W}}} f\ \left(\nu, \tilde{W}, \xi, H\right)\ d\nu\ d\xi}{P\left(\delta_i = 1\right)} \quad \text{for} \quad \tilde{W} \in \Theta_{i \cdot\ \tilde{W}}. \tag{4.60}$$

Clearly, both specifications (4.59) and (4.60) depend directly on $\bar{h}_i$, but as in representation of other specifications the only arguments included in the function are those variables which are random in the state; $\bar{h}_i$ is fixed and known given $\delta_i = 1$.

4.4.6. Likelihood Functions for Measured Hours and Wages

Formulating the likelihood function for $(H, \tilde{W})$ is complicated by the fact that a researcher does not observe precisely which portion of the budget constraint an individual selects since this decision reveals h and this quantity is unknown. Thus, when a person works, one cannot distinguish which individual state i occurs. On the other hand, a researcher does observe when a person does not work. Expressed in terms of the endogenous dummy variables δ_i, these circumstances imply that the data reveal the event $\delta_o = 1$ but not the individual events $\delta_i = 1$ for $i \in M = M_c \cup M_d$. Instead, one merely observes whether $\overline{\delta}_1 \equiv \sum_{i \in M} \delta_i = 1$ or $\overline{\delta}_1 = 0$.

Appealing to formula (**??**) of Appendix A, the following specification represents the joint likelihood function of (h, W):

$$\mathcal{L}\left(H, \tilde{W}\right) = \left[P\ \left((v, \eta, \xi) \in \Omega_0\ \right)\right]^{\overline{\delta}_0} *$$
$$\left[\sum_{i \in M_t} \int_{\Theta_i} \left[\int_{\Phi_{i|h,W}} f\ \left(h, v_2, \tilde{W}, \xi, H\right)\ dv_2\ d\xi\right]\ dh + \sum_{i \in M_t} \int_{\Phi_{i|\bar{h},W}} f\ \left(\nu, \tilde{W}, \xi, H\right)\right.$$

The first line of this expression delineates the probability of not working; and the second line designates the density of $(H, \tilde{W})$ unconditional on $\overline{\delta}_1$. Accordingly, both H and $\tilde{W}$ are continuously distributed throughout the range on $H > 0$.

4.5. Maximum Likelihood: Convex Differentiable Constraints with Full Participation

Developing specifications for likelihood functions when budget sets are convex and have differentiable boundaries is straightforward, especially assuming labor force participation is not a factor for the population under investigation. The following discussion presents two examples of such specifications to illustrate elementary versions of the general formulas presented above.

4.5.1. Specifications for Linear Parameterizations of Labor Supply

Derivation of likelihood functions assuming a linear specification for hours of work when (4.5) describes the budget constraint–wherein tax payments depend only on a single taxable income quantity–follows directly from the previous results. Assuming no measurement error (i.e., $H = h$), a change in variables from the heterogeneity error ν to actual hours h using relation (**??**) yields the likelihood function for h:

$$f_h(h) = \frac{d\nu}{dh} \varphi_\nu \left(h - y_\nu - z\gamma - \alpha\omega - \beta y\right) \tag{4.62}$$

where $\varphi_v(v)$ denotes the density of the heterogeneity component ν, and the Jacobian term is

$$\frac{d\nu}{dh} = 1 + (\alpha - \beta h) W^2 \frac{\partial \tau'}{\partial I}. \tag{4.63}$$

This Jacobian term is restricted to be non-negative over the admissible range. Maximizing (4.62) yields maximum-likelihood estimates for the parameters of the labor supply function, ℓ, which provide the information needed to infer the work disincentive effects of taxation.

If hours are indeed contaminated by additive measurement error, then the likelihood function for observed hours $H = h + \varepsilon$ is given by:

$$f_H(H) = \int_0^{\text{maxhours}} \varphi_\varepsilon(H - h)\varphi_h(h)dh \tag{4.64}$$

where $\varphi_\varepsilon(\varepsilon)$ denotes the density of the heterogeneity component ε. This expression resembles relation (4.62) except that integration occurs over hours to account for the existence of reporting error, and H replaces actual hours h in the Jacobian term in (4.63).

4.5.2. Specifications for Multiplicative Measurement Error

Now consider maximum-likelihood estimation of the semilog specification of labor supply. Suppose the heterogeneity-error-component ν in structural labor-supply equation and the disturbance ε in the measurement-error equation for hours of work possess the joint distribution $\varphi_{\nu\varepsilon}(\nu, \varepsilon)$, where $\varphi_{\nu\varepsilon}$ designates a density function. For the moment, suppose (ν, ε) are independently distributed of the gross wage and other income. Using relations (4.35) and (4.44) to perform a standard change in variables from the errors ν and ε to the variables h and H produces the likelihood function needed to compute maximum-likelihood estimates. The transformation from (ν, ε) to (h, H) is monotonic for a wide range of functional forms for ℓ as long as the underlying preferences satisfy quasiconcavity and budget sets are convex.

Without measurement error, the likelihood function for hours of work, h, takes the form

$$f_h(h) = \frac{d\nu}{dh}\varphi_\nu (h - y - z\gamma - \alpha \ln W - \alpha \ln(1 - \tau') - \beta y) \tag{4.65}$$

where φ_ν is the marginal density for ν, and the Jacobian term is

$$\frac{d\nu}{dh} = 1 + \left(\left(\frac{\alpha}{W(1-\tau')} - \beta h \right) W^2 \frac{\partial \tau'}{\partial I} \right) \tag{4.66}$$

which is required to be non-negative. In these expressions, the derivative τ' is evaluated at $I = Wh + Y - \tau(Wh + Y)$.

With multiplicative measurement error, the likelihood function for observed hours H becomes

$$\mathcal{L} = \int_0^{\text{maxwage}} \int_0^{\text{maxhours}} \frac{d\nu}{dh} f_{\nu\varepsilon w} \left(h - y - z\gamma - \alpha \ln \omega - \beta y, \ln H - \ln h, W \right) dh \; dW \tag{4.67}$$

where integration occurs over the hourly wage, which is unobserved, using the joint density $f_{\nu\varepsilon w}\ (\nu, \varepsilon, W)$. The non-negativity of the Jacobian term clearly places restrictions on the behavioral parameters and we discuss these restrictions further below.

4.6. Maximum Likelihood: Convex Piecewise-Linear Constraints with Full Participation

The majority of empirical labor-supply studies incorporating taxes treat the tax schedule as a series of brackets implying a piecewise-linear budget set. With such a tax function, the familiar change-in-variables techniques implemented in conventional maximum likelihood do not apply due to the nonexistence of the Jacobian over measurable segments of the sample space arising from nondifferentiability of functional relationships characterizing hours-of-work choices. Moreover, a piecewise-linear budget set creates endogenous variables (hours and after-tax wages) that are both discrete and continuous in character. Section 4.4 covers specifications for likelihood functions for such endogenous variables.

4.6.1. Characterization of Labor Supply with Piecewise-Linear Constraints

To illustrate the derivation of an estimable labor supply model using the piecewise-linear approach assuming the linear structural specification for hours of work, consider the simple case of a budget set with only three segments as presented in Figure 4.1. (To simplify the exposition here, we number the kink points as 0,1,2, and 3 rather than as 0, 2, 4, and 6 .) The preceding discussion defines the variables y_j, ω_j and $\overline{h}_j$ appearing in this figure. To locate the kinks and slopes of the budget constraint for an individual, a researcher must know the individual's level of non-labor income, gross wage rate, hours of work, and the structure of the tax system. The hours of work at which kinks occur are given by $\overline{h}_j = (I_j - Y + D)/W$, where Y and D, respectively, represent taxable non-labor income and deductions, and I_j is the maximum taxable income for segment j. The slope of each segment is given by the marginal wage rate for that segment: $\omega_j = W(1 - t_j)$, where j denotes the segment, t_j signifies the marginal tax rate for that segment, and W is the gross wage rate per hour. Finally, the non-labor income at zero hours of work - the intercept of the budget line - is $y_1 = Y - \tau(Y - D)$, where $\tau(\cdot)$ is the tax function evaluated at the individual's taxable income at zero earnings. Given this intercept value, virtual incomes or the intercepts associated with successive budget segments are computed by repeated application of the formula: $y_j = y_{j-1} + (\omega_{j-1} - \omega_j)\overline{h}_{j-1}$.

Given a convex budget constraint, an individual's optimization problem amounts to maximizing $U(C, h)$ subject to

$$\begin{aligned} C &= y_1 && \text{if } h = 0 \\ &= \omega_1 h + y_1 && \text{if } \overline{h}_0 < h \leq \overline{h}_1 \\ &= \omega_2 h + y_2 && \text{if } \overline{h}_1 < h \leq \overline{h}_2 \\ &= \omega_3 h + y_3 && \text{if } \overline{h}_2 < h \leq \overline{h}_3 \\ &= \omega_3 \overline{h}_3 + y_3 && \text{if } h = \overline{h}_3 \end{aligned} \tag{4.68}$$

The solution of this maximization problem decomposes into two steps. First, determine the choice of h conditional on locating on a particular segment or a kink. This step yields the solution:

$$\begin{aligned} h &= 0 && \text{if } h = 0 && \text{(lower limit)} \\ &= \ell(\omega_1, y_1, \nu) && \text{if } 0 < h < \overline{h}_1 && \text{(segment 1)} \\ &= \overline{h}_1 && \text{if } h = \overline{h}_1 && \text{(kink 1)} \\ &= \ell(\omega_2, y_2, \nu) && \text{if } \overline{h}_1 < h < \overline{h}_2 && \text{(segment 2)} \\ &= \overline{h}_2 && \text{if } h = \overline{h}_2 && \text{(kink 2)} \\ &= \ell(\omega_3, y_3, \nu) && \text{if } \overline{h}_2 < h < \overline{h}_3 && \text{(segment 3)} \\ &= \overline{h}_3 && \text{if } h = \overline{h}_3 && \text{(kink 3 = upper limit)} \end{aligned} \tag{4.69}$$

Second, determine the segment or the kink on which the person locates. The following relations characterize this solution: choose:

$$
\begin{array}{ll}
0 & \text{if } \ell(\omega_1, y_1, \nu) \leq 0 \\
\text{(Segment 1)} & \text{if } \overline{h}_0 < \ell(\omega_1, y_1, \nu) < \overline{h}_1 \\
\text{(Kink 1)} & \text{if } \ell(\omega_2, y_2, \nu) \leq \overline{h}_1 < \ell(\omega_1, y_1, \nu) \\
\text{(Segment 2)} & \text{if } \overline{h}_1 < \ell(\omega_2, y_2, \nu) < \overline{h}_2 \\
\text{(Kink 2)} & \text{if } \ell(\omega_3, y_3, \nu) \leq \overline{h}_2 < \ell(\omega_2, y_2, \nu) \\
\text{(Segment 3)} & \text{if } \overline{h}_2 < \ell(\omega_3, y_3, \nu) < \overline{h}_3 \\
\text{(Kink 3)} & \text{if } \ell(\omega_3, y_3, \nu) \geq \overline{h}_3
\end{array}
\tag{4.70}
$$

Combined, these two steps imply the values of h and C that represent the utility-maximizing solution for labor supply and consumption.

All studies implementing the piecewise-linear approach assume the existence of measurement error in hours of work. With the linear measurement error model given by (4.16), observed hours $H = h + \varepsilon$. As long as the measurement error component ε is continuously distributed, so is H. In contrast to information on h, knowledge of H suffices neither to allocate individuals to the correct branch of the budget constraint nor to identify the marginal tax rate faced by individuals, other than at zero hours of work. The state of the world an individual occupies can no longer be directly observed, and one confronts a discrete-data version of an errors-in-variables problem. The interpretation of measurement error maintained in this analysis is that ε represents reporting error that contaminates the observation on h for persons who work.[17]

With measurement error, the linear specification of ℓ given by (**??**) with $\widehat{h}_j \equiv$

[17] Note that expected hours of work, in this convex piece-wise linear case, is additive in each hours choice weighted by the probability of each segment or kink. Each term in this sum being at most a function of two marginal wages and two virtual incomes. Blomquist and Newey (1997) exploit this observation to develop a semi-parametric estimator for hours of work imposing the additivity through a series estimator.

$\mu+\alpha\omega_j+\blacksquare y_j+Z\gamma$ implies the following stochastic specification for labor supply:

$$H=\begin{cases}\widehat{h}_1+\nu+\varepsilon & \text{if } 0<\widehat{h}_1+\nu\leq\overline{h}_1 \quad \text{(segment 1)}\\ \overline{h}_1+\varepsilon & \text{if } \widehat{h}_2+\nu<\overline{h}_1<\widehat{h}_1+\nu \quad \text{(kink 1)}\\ \widehat{h}_2+\nu+\varepsilon & \text{if } \overline{h}_1<\widehat{h}_2+\nu\leq\overline{h}_2 \quad \text{(segment 2)}\\ \overline{h}_2+\varepsilon & \text{if } \widehat{h}_3+\nu<\overline{h}_2<\widehat{h}_2+\nu \quad \text{(kink 2)}\\ \widehat{h}_3+\nu+\varepsilon & \text{if } \overline{h}_2<\widehat{h}_3+\nu\leq\overline{h}_3 \quad \text{(segment 3)}\\ \overline{h}_3+\varepsilon & \text{if } \widehat{h}_3+\nu\geq\overline{h}_3 \quad \text{(upper limit)}\end{cases} \tag{4.71}$$

This represents a sophisticated variant of an econometric model that combines discrete and continuous choice elements.

4.6.2. Likelihood Function with Measurement Error When All Work

The log-likelihood function for this model is given by $\sum_i \ln f_H(H)$, where i indexes observations. Defining $\underline{\nu}_j=\overline{h}_{j-1}-\widehat{h}_j$ and $\overline{v}_j=\overline{h}_j-\widehat{h}_j$, the components $f_H\ (H)$ are given by

$$\begin{aligned} f_H(H) \;=\; & \sum_{j=1}^{3}\int_{\underline{\nu}_j}^{\overline{v}_j}\varphi_2\ (H-\widehat{h}_j,\nu)\ d\nu \quad \text{(segments 1, 2, 3)} \\ & +\sum_{j=1}^{2}\int_{\overline{v}_j}^{\underline{\nu}_{j+1}}\varphi_1\ (H-\overline{h}_j,\nu)\ d\nu \quad \text{(kinks 1, 2)} \\ & +\int_{\overline{v}_3}^{\infty}\varphi_1\ (H-\overline{h}_3,\nu)\ d\nu \quad \text{(upper limit)} \end{aligned} \tag{4.72}$$

where $\varphi_1(\cdot,\cdot)$ and $\varphi_2(\cdot,\cdot)$ are the bivariate density functions of (ε,ν) and $(\varepsilon+\nu,\nu)$, respectively. Maximizing the log-likelihood function produces estimates of the coefficients of the labor supply function ℓ. These estimates provide the information used to infer both substitution and income responses, which in turn provide the basis for calculating the work disincentive effects of income taxation.

4.6.3. Shortcomings of Conventional Piecewise-Linear Analyses

The piecewise-linear approach for estimating the work disincentive effects of taxes offers both advantages and disadvantages relative to other methods. Concerning the attractive features of this approach, piecewise-linear analyses recognize that institutional features of tax systems induce budget sets with linear segments and kinks. This is important if one believes that a smooth tax function does not provide a reasonably accurate description of the tax schedule. The piecewise-linear approach admits randomness in hours of work arising from both measurement error and variation in individual preferences and it explicitly accounts for endogeneity of the marginal tax rate in estimation, but so do the instrumental-variable and differentiable likelihood methods discussed above. As we will see below, the piecewise-linear approach more readily incorporates fixed costs of holding a job, regressive features of the tax code, and multiple program participation than other procedures due to the discrete-continuous character of hours-of-work choices induced in these environments. These features of the piecewise-linear method make it a vital approach in empirical analyses of labor supply.

On the other hand, the following shortcomings of the piecewise-linear procedure raise serious doubts about the reliability of its estimates of work disincentive effects. First, the piecewise-linear methodology assumes that both the econometrician and each individual in the sample have perfect knowledge of the *entire* budget constraint that is relevant for the worker in question. Errors are permitted neither in perceptions nor in measuring budget constraints. Taken literally, this means that: all income and wage variables used to compute each sample member's taxes are observed perfectly by the econometrician; individuals making labor supply choices know these variables exactly prior to deciding on hours of work; each individual and the econometrician know when the taxpayer will itemize deductions and the amount of these itemizations; and each taxpayer's understanding of the tax system is equivalent to that of the econometrician (e.g., the operation of such features as earned-income credits). Clearly, given virtual certainty that most of these assumptions are violated in empirical analyses of labor supply, the estimates produced by methods relying on these assumptions must be interpreted very cautiously. The differentiable-likelihood methods rely on the same assumptions. The instrumental-variable methods do not, so they are likely to be more robust.

Second, measurement error plays an artificial role in econometric models based on the piecewise-linear approach. Its presence is needed to avoid implausible predictions of the model. The statistical framework induced by the piecewise-linear approach implies that bunching in hours of work should occur at kink points if hours precisely measure h. However, for the vast majority of data sources currently used in the literature, only a trivial number of individuals, if indeed any at all, report hours of work at interior kink points. Unless one presumes that the data on hours do not directly represent h, such evidence provides the basis for immediately rejecting the distributional implications of the above specifications. Considering, for example, the labor-supply characterization proposed in equation (4.69), almost any test of the distributional assumptions implied by this specification would be readily rejected because observed hours would take the values $\overline{h}_0, \overline{h}_1, \overline{h}_2$ and $\overline{h}_3$ with only a trivial or zero probability. Instead, observed hours essentially look as if they are distributed according to a continuous distribution. When a continuously-distributed measurement error ε is added to the model, observed hours H are continuously distributed. This provides an essential reason for introducing measurement error in the data, for without it, the piecewise-linear structure provides a framework that is grossly inconsistent with the data. Of course, several sound reasons exist for admitting measurement error in a labor supply model, including the widespread suspicion that reporting error contaminates data on hours of work. However, measurement error in hours of work implies measurement error in wages, since they are typically computed as average hourly earnings. Current applications of the piecewise-linear analysis mistakenly ignore this by assuming perfectly measured budget constraints.[18] The unnatural role played by measurement error raises questions about the credibility of findings derived from the piecewise-linear approach. In contrast to the piecewise-linear approach, it is not essential to introduce measurement error in either the differentiable-likelihood or the instrumental-variable approach because hours in the distribution of h are continuous without measurement error.

Third, existing research implementing the piecewise-linear methodology relies on very strong exogeneity assumptions. Other than hours of work, all variables involved in the calculation of taxes are presumed to be exogenous determinants of labor supply behavior, both from a statistical and from an economic perspective. These variables include gross wages, the various components of non-labor income, and deductions. In light of the evidence supporting the view that wages and income are endogenous variables in labor supply analyses, particularly in the case

[18] It is possible to argue that this error does not result in measurement error in the hourly wage, if the measurement error is interpreted as an "optimization" error.

of wages, suspicions arise regarding the dependability of estimated substitution and income effects based on procedures that ignore such possibilities. Most of the exogeneity assumptions are also maintained in the differentiable-likelihood approach, but are easily relaxed when applying instrumental-variable procedures (given the availability of a sufficient number of other instrumental variables).

Fourth, some concerns about the reliability of estimates produced by the piecewise-linear approach ensue due to the static behavioral framework maintained in the formulation of empirical relations. Piecewise-linear studies invariably rely on the textbook one-period model of labor supply as a description of hours-of-work choices, and impose it to estimate parameters. Existing implementations of the differentiable-likelihood approach suffer from the same problem. Everyone acknowledges that individuals are not simply myopic optimizers; they transfer income across periods to achieve consumption plans that are infeasible without savings. A serious question arises concerning the relevance of such considerations in estimating substitution and income effects used to predict responses to tax policy.

4.7. Maximum Likelihood Estimation Imposes Restrictions on Behavioral Responses

The implementation of maximum likelihood procedures imposes interesting and important restrictions on behavioral parameters in the presence of non-linear budget constraints. These restrictions come about in defining the statistical model to be coherent, requiring probabilities to fall in the [0,1] interval and densities to be nonnegative.

4.7.1. Restrictions Imposed Under Piecewise-Linear Constraints

The econometric model produced by the piece-wise linear formulation given by (4.72) implicitly imposes parametric restrictions that constraint the signs of estimated substitution and income effects. As developed in MaCurdy *et al.* (1990), particular inequality restrictions must hold in the application of estimation procedures with piecewise-linear budget constraints for likelihood functions to be defined (i.e., to ensure that the components of these functions are non-negative). More specifically, in applications of such procedures, the Slutsky condition must be locally satisfied at all interior kink points of budget sets that represent feasible options for any individual in the sample such that the compensated substitution effect must be positive. For the linear specification of the labor supply function considered in the preceding discussion, the specific inequality constraints imposed are

$$\alpha - \beta \overline{h}_{jk} \geq 0, \quad \forall\ j,k \tag{4.73}$$

where the quantities $\overline{h}_{jk}$ represent the hours-of-work values that correspond to interior kink points j on a sample member k's budget set. Because many values of $\overline{h}_{jk}$ exist in most analyses of piecewise-linear constraints, fulfillment of relations (4.73) essentially require global satisfaction of the Slutsky condition by the labor supply function. Such a requirement, in essence, globally dictates that the uncompensated substitution effect of a wage change on hours of work must be positive for the labor supply specification considered in the preceding discussion, and the income effect for hours of work must be negative. The imposition of these restrictions, especially for men, is highly suspect given the available evidence from other studies. These restrictions carry over to more general labor supply functions.

4.7.2. Restrictions Imposed Under Differentiable Constraints

Maximum likelihood estimation with differentiable constraints induces comparable restrictions. Consider, for example, likelihood function (4.62). For this specification to be a properly-defined likelihood function, the Jacobian $\frac{d\nu}{dh}$ must be non-negative. Violation of this condition implies that the density function for h is negative, which obviously cannot occur. Non-negativity of $\frac{d\nu}{dh}$ translates into the property

$$\frac{\partial \ell}{\partial w} - \frac{\partial \ell}{\partial y} h \geq - \left(\frac{\partial \tau}{\partial I}\ W^{\ 2} \right)^{-1} \leq 0, \tag{4.74}$$

where ℓ refers to the labor supply function. The left-hand side of this inequality is the Slutsky term. This inequality result does not require compensated substitution effects to be positive as quasi-concave preferences mandate, only that these effects cannot become too negative.

Maximum likelihood procedures yield nonsensical results unless equation (4.74) holds. Without measurement error, estimated parameter values cannot imply a violation of equation (4.74) at any of the data combinations $(h, w(h), y(h))$

actually observed in the sample. If a violation occurs, then the evaluation of (4.62) for the observation associated with this combination would result in a non-positive value which causes the overall log likelihood function to approach minus infinity which clearly cannot represent a maximum.

With measurement error, maximum likelihood estimation applied to function (4.64) ensures that a weighted average of density functions appearing in (4.64) holds, with weighting occurring over all combinations of hours, marginal wages, and virtual income lying in the feasible range of the budget constraint of any individual included in the sample. Since maximum likelihood procedures assume the validity of such restrictions when calculating estimates of the coefficients of ℓ, the resulting estimated labor supply function can be expected to exhibit compensated substitution effects that obey inequality (4.74.) over a very wide range of hours, wages, and incomes.[19]

4.8. Maximum Likelihood: Accounting for Participation and Missing Wages

As mentioned in previous sections, some applications of the piecewise-linear approach incorporate fixed costs to working - costs such as transportation that must be paid for any amount of work but which may vary across individuals. This significantly complicates the analysis because the optimized level of work under the budget constraint while working may not represent the optimal choice overall; one must explicitly consider the option of not working and thus avoiding the fixed cost. For any level of fixed costs, a minimum number of hours worked is implied creating an attainable range in the observable hours of work distribution; individuals will not work unless the gain is large enough to overcome the fixed costs. In essence, these complications arise because the budget constraint is not convex, invalidating simple maximization procedures.

[19] It is, of course, computationally feasible to use (4.5.3) in estimation and not require f_h to be defined over the entire range of its support. Computationally one merely requires f_h to be non-negative over a sufficiently large region to ensure (4.5.3) > 0. Of course, not requiring $f_h \geq 0$ over its relevant range produces a nonsensical statistical model.

4.8.1. Fixed Costs of Working

If an individual must pay fixed monetary costs, F, to work, then non-labor income, Y, in the above budget constraints is replaced by

$$\begin{aligned} Y-F, \quad &\text{if } h > 0, \\ Y, \quad &\text{if } h = 0. \end{aligned} \tag{4.75}$$

F is partially unobservable and, thus, modelled as a stochastic element, varying across individuals. Hence, we see that the budget constraint discontinuously jumps down by F when the individual chooses to work.

To solve for the optimum when faced with this budget constraint, two regimes must explicitly be considered: working and not working. Estimation proceeds by finding the maximum utility under each regime and then comparing these to determine which option is chosen. In neither regime, the utility function $U(C,h,\nu)$ - where we explicitly note the unobserved component, ν - is maximized subject to optimization problem (4.1) with (4.4) modified by (4.75).

In the no-work regime, the solution is simple. We know h is 0, so utility is given by $U(Y-\tau(Y-D),0,\nu)$.

The solution in the work regime closely follows the solution presented in Section 4.6. Again utilizing the labor supply function, $\ell(\omega,y,\nu)$ yields the solution for h given in (4.69), where the virtual income y now subtracts fixed costs F. However, to compute maximum utility in this regime requires associating a utility level with each possible hours choice. Utility along any segment, j, is given by the indirect utility function, $V(\omega_j,y_j,\nu)$. At kinks, the direct utility function must be used, so the utility at kink j is given by $U(\omega_j\overline{h}_j+y_j,\overline{h}_j,\nu)$. Hence, utilizing exactly the same solution procedure exploited in Section 4.6, we can define maximized utility when working, V^*:

$$V^*(w,y,v) = \begin{array}{ll} -\infty & \ell_1 \le 0 \\ V(\omega_1,y_1,\nu) & 0 < \ell_1 < \overline{h}_1 \\ U(\omega_1\overline{h}_1+y_1,\overline{h}_1,\nu) & \ell_2 < \overline{h}_1 \le \ell_1 \\ V(\omega_2,y_2,\nu) & \overline{h}_1 < \ell_2 < \overline{h}_2 \\ U(\omega_2\overline{h}_2+y_2,\overline{h}_2,\nu) & \ell_3 < \overline{h}_2 \le \ell_2 \\ V(\omega_3,y_3,\nu) & \overline{h}_2 < \ell_3 < \overline{h}_3 \\ U(\omega_3\overline{h_m}+y_3,\overline{h_m},\nu) & \ell_3 \ge \overline{h_m} \end{array} \tag{4.76}$$

where

$$\ell_j \equiv \ell(\omega_j, y_j, \nu) \equiv \frac{V_\omega(\omega_j, y_j, \nu)}{V_y(\omega_j, y_j, \nu)} \tag{4.77}$$

with V_ω and V_y denoting the partial derivatives of V; relation (4.77) is, of course, Roy's identity defining the labor supply function, ℓ, evaluated at wage and income levels ω_j and y_j. The use of $-\infty$ for $h = 0$ simply indicates that $h = 0$ is not included in this regime and, thus, selecting it indicates that the no-work regime is preferred. Given functional forms for V and U, finding V^* is straightforward.

Given maximized utility under each regime, the final step in the solution is to compare the two regimes. An individual chooses to work at the hours specified by the solution in (4.69) if

$$V^*(\omega, y, \nu) \geq U(Y - \tau(Y - D), 0, \nu) \tag{4.78}$$

and chooses not to work otherwise. For any level of ν, treating equation (4.78) as an equality implies a critical level of fixed costs, $F^*(\nu)$ above which the individual will choose not to work; F enters this relation through the virtual income variable y. Because desired hours of work increase with ν, this critical value will generally be increasing in ν - greater propensity to work implies that higher fixed costs are required to prefer the no-work option. If restrictions are placed on the support of F, such as $F > \underline{F}$, there will be values of ν low enough to rule out the work regime, thus implying a hole at the low end of the h distribution.

As a final step before deriving the likelihood function, note that in the no-work regime, gross wage, W, is not observed and, thus, the budget constraint cannot be derived. Hence, W must be endogenized. Such a step amounts to modeling the offered gross wage rate as being generated by a variant of equation (**??**) which presumes that W is randomly distributed across the population depending on measured characteristics Q and unobservable components η. To simplify the discussion below, we assume that the linear variant of specification (**??**) (i.e., $W = W^*(Q) + \eta$) generates W.

4.8.2. Likelihood Function Incorporating Fixed Costs

To derive the likelihood function, first consider the likelihood contribution of an individual who does not work. We assume this no-work decision can be observed, so there is no measurement error. In the no-work case, one of two situations applies: (i) fixed costs are sufficiently high with $F > F^* \equiv F^*(\nu, \eta)$ for any given ν and η, or (ii) if this fixed-cost threshold falls below the lowest admissible value for F (i.e. $F^* \leq \underline{F}$), then desired hours are sufficiently low with $\nu < \nu^* \equiv \nu^*(\eta)$ for any η.[20] The probability of this event is

$$\mathcal{L}_0 = \int_{-\infty}^{\infty} \int_{-\infty}^{\nu^*} \int_{F^*}^{\infty} \varphi_{\nu\eta F}\ (\nu, \eta, F)\ dF\ d\eta\ d\nu \tag{4.79}$$

where $\varphi_{\nu\eta F}$ is joint density of (ν, η, F).

For the work regime, the likelihood contribution looks very much like that derived in specification (4.72), as we continue to assume the linear hours of work function and the form of measurement error assumed there. The only changes are the addition of terms for δ and F (accounting for the fact that $F < F^*(\nu)$) and the removal of the term for the lower limit which is no longer part of that regime and is now perfectly observable. Using φ_1 and φ_2 to denote the distribution of $(\varepsilon, \nu, \eta, F)$ and $(\varepsilon + \nu, \nu, \eta, F)$ yields:

$$\begin{aligned} \mathcal{L}_1 \quad = \quad & \sum_{j=1}^{3} \int_{\underline{\nu}_j}^{\overline{\nu}_j} \int_{0}^{F^*} \varphi_2\ (H - \widehat{h}_j, \nu, W - W^*(Q), F)\ dF\ d\nu \\ & + \sum_{j=1}^{2} \int_{\overline{\nu}_j}^{\underline{\nu}_{j+1}} \int_{0}^{F^*} \varphi_1(H - \overline{h}_j, \nu, W - W^*(Q), F)\ dF\ d\nu \\ & + \int_{\overline{\nu}_3}^{\infty} \int_{0}^{F^*} \varphi_1\ (H - \overline{h}_3, \nu, W - W^*(Q), F)\ dF\ d\nu \end{aligned} \tag{4.80}$$

where

$$\begin{aligned} \underline{\nu}_j \text{ solves the equation } \ell(\omega_j, y_j, \underline{\nu}_j) \quad &= \quad \overline{h}_{j-1} \\ \overline{\nu}_j \text{ solves the equation } \ell(\omega_j, y_j, \overline{\nu}_j) \quad &= \quad \overline{h}_j. \end{aligned} \tag{4.81}$$

[20] The critical value v^* solves relation (4.8.4) treated as an equality with virtual income y evaluated at $\underline{F}$.

All variables appearing in these expressions are defined as in Section 4.6.

The likelihood function for an individual is given by

$$\mathcal{L} = (\mathcal{L}_1)^{\delta_E} (\mathcal{L}_0)^{1-\delta_E} \tag{4.82}$$

where $\delta_E = 1$ if the individual works and $= 0$ otherwise. Estimation proceeds by maximizing the sum of log likelihoods across individuals, as always. This is quite complex in this case, requiring knowledge of both the direct utility U and the indirect utility, V, and also requiring comparisons across regimes for all individuals and all parameter values.

4.9. Welfare Participation: Maximum Likelihood with NonConvex Constraints

A common source of non-linearity in budget constraints involves participation in welfare programs. To illustrate this situation, consider the simplest case in which the only taxes faced by an individual result from benefit reduction on a single welfare program. Figure 4.3 presents this scenario. Under most welfare programs, individuals face very high effective tax rates when they initially work due to large reductions in their benefits occurring when earnings increase. Once benefits reach 0, the tax rate drops to a lower level, creating a non-convex kink in the budget constraint. This non-convexity invalidates the simple procedures exploited in Section 4.6 implemented to divide sample spaces into locations on budget sets.

4.9.1. Simple Nonconvex Constraints with No Welfare Stigma

Following the picture portrayed in Figure 4.3, an individual maximizes $U(c, h, \nu)$ subject to the budget constraint

$$c = Wh + Y + b(I(h)), \tag{4.83}$$

where benefits are given by the simple benefit schedule:

$$B(I(h)) = \begin{cases} G - \rho Wh, & \text{if } G - \rho Wh > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4.84}$$

G gives the guarantee amount which is reduced at the benefit reduction rate ρ as the earnings, Wh, increase. This implies a kink point at $\overline{h}_1 = G/\rho W$ where benefits reach 0 and, thus, the marginal wage rises to W. So, the individual faces two segments: segment 1 has $h < \overline{h}_1$ with net wage $\omega_1 = (1-\rho)W$ and virtual income $y_1 = Y + G$; and segment 2 has $h > \overline{h}_1$ with net wage $\omega_2 = W$ and virtual income $y_2 = Y$.[21]

Because the budget constraint is non-convex, the solution cannot be characterized simply by finding a tangency with the budget constraint as it was in Section 4.6. Multiple tangencies are possible and these must be directly compared to determine the optimum. Hence, one requires the regime shift approach summarized in Section 4.4.

Consider first the regime in which positive benefits are received; that is, $h < \overline{h}_1$. Maximization, given the effective wage and income, on this linear segment follows the approach of Section 4.4. We can characterize the optimal choice according to the function $\ell(\omega_1, y_1, \nu)$. Denotes the value of ν which implies $\ell(\omega_1, y_1, \nu) = 0$ as ν_0. Then the optimal hours choice along that segment is given by

$$h = \ell_1 = \ell(\omega_1, y_1, \nu), \quad \nu > \nu_0, \quad h = 0, \quad \nu \leq \nu_0. \tag{4.85}$$

The optimized value on this segment (including the zero work option), accounting for the fact that $h > \overline{h}_1$ is not allowed, is given by

$$V_1^*(\omega_1, y_1, \nu) = \begin{cases} V(\omega_1, y_1, \nu), & 0 < \ell_1 \leq \overline{h}_1 \\ U(y_1, 0, \nu), & \ell_1 \leq 0 \\ -\infty, & \ell_1 > \overline{h}_1, \end{cases} \tag{4.86}$$

where equation (4.85) defines ℓ_1.

Next, consider the regime without benefits, that is with $h \geq \overline{h}_1$. Again the optimal choice, given the wage and income, on this segment is given by the labor supply function $\ell_2 = \ell(\omega_2, y_2, \nu)$. The optimized value, accounting for the fact that $h < \overline{h}_1$ is not admissible, is given by[22]

$$V_2^*(\omega_2, y_2, \nu) = \begin{cases} V(\omega_2, y_2, \nu), & \ell_2 \geq \overline{h}_1, \\ -\infty, & \ell_2 < \overline{h}_1. \end{cases} \tag{4.87}$$

[21] We ignore any upper bound on hours worked for simplicity.

[22] In the following formulation, we implicitly assume that the event $\ell_2 \geq \overline{H}$ occurs with zero probability.

Hence, the individual selects regime 1, with welfare receipt, if $V_1^* > V_2^*$, and regime 2 otherwise. Since work propensity increases with ν, this can be characterized by a cutoff value, ν^*, defined by

$$V_1^*(\omega_1, y_1, \nu^*) = V_2^*(\omega_2, y_2, \nu^*). \tag{4.88}$$

For values of ν above ν^*, regime 2 is chosen; and for values below ν^*, regime 1 is realized.

We can define three sets, Ω_0, Ω_1, and Ω_2, such that for $\nu \in \Omega_0$ the individual chooses not to work, for $\nu \in \Omega_1$ the individual locates on segment 1 receiving benefits with positive hours of work, and for $\nu \in \Omega_2$ the individual locates on segment 2. We must consider two cases to define these sets exactly. First, suppose $\nu^* > \nu_0$. Then we have

$$\begin{aligned} \Omega_0 &= \{\nu \mid \nu \leq \nu_0\} \\ \Omega_1 &= \{\nu \mid \nu_0 < \nu \leq \nu^*\} \\ \Omega_2 &= \{\nu \mid \nu > \nu^*\} \end{aligned} \tag{4.89}$$

Alternatively, if $\nu^* \leq \nu_0$, then the switch to regime 2 occurs before positive hours are worked in regime 1, that is

$$\begin{aligned} \Omega_0 &= \{\nu \mid \nu \leq \nu^*\} \\ \Omega_1 &= \emptyset \\ \Omega_2 &= \{\nu \mid \nu > \nu^*\} \quad . \end{aligned} \tag{4.90}$$

Hence, for certain individuals and parameter values, no value of ν exists such that they will locate on segment 1 with positive hours of work.

To characterize the likelihood function we again need a functional form for the gross wage of the form $W = W(Z) + \eta$. We ignore measurement error here for simplicity, and because there is no problem with individuals failing to locate at the kink in this non-convex case. Define $\delta_B = 1$ if the individual receives benefits, and $\delta_E = 1$ is the individual works, both 0 otherwise. The likelihood function is given as follows, incorporating $\varphi_{\eta\nu}\ (\eta, \upsilon)$ and the general inverse function $\nu = \nu(h)$:

$$\begin{aligned}
\delta_B &= 1, \quad \delta_E = 1, \quad \mathcal{L}_{11} = \frac{\partial \nu}{\partial h}\varphi_{\nu\eta}\left(\nu(h), W - W(Z)\right)\ I(\nu \in \Omega_1), \quad (4.91) \\
\delta_B &= 0, \quad \delta_E = 1, \quad \mathcal{L}_{01} = \frac{\partial \nu}{\partial h}\varphi_{\nu\eta}\left(\nu(h), W - W(Z)\right)\ I(\nu \in \Omega_2), \\
\delta_B &= 1, \quad \delta_E = 0, \quad \mathcal{L}_{10} = \int_{\Omega_0} \varphi_{\nu\eta}(\nu, \eta) d\nu d\eta.
\end{aligned}$$

where $I(\cdot)$ represents an indicator function equal to 1 if the condition in the parentheses is true. Because the value of ν implied by the hours choice may be inconsistent with the value implied by the regime choice, it is possible to have "holes" in the hours distribution around the kink point. For example, an individual on segment 1 must have $\nu \leq \nu^*$. If his hours choice is too close to the kink, this may imply a value of $\nu > \nu^*$ and thus an observation with zero likelihood.

The overall likelihood function is given by

$$l = (\mathcal{L}_{11})^{(\delta_B)(\delta_E)}\ (\mathcal{L}_{01})^{(1-\delta_B)(\delta_E)}\ (\mathcal{L}_{10})^{(\delta_B)(1-\delta_E)}. \quad (4.92)$$

Estimation proceeds by maximizing the sum of log likelihoods across individuals, as always. This is quite complex in this case, requiring knowledge of both the direct utility U and the indirect utility, V, and also requiring comparisons across regimes for all individuals and all parameter values.

4.9.2. Welfare Stigma Implies Selection of Budget Constraint

The above analysis assumes that all individuals eligible for welfare are on welfare. Individuals working less than h_0 but failing to receive welfare are operating below the implied budget constraint, a possibility not permitted in the analysis. Yet, many individuals are in exactly this situation. This is generally explained by assuming the existence of some utility loss or stigma associated with welfare.

To capture welfare stigma the utility function is modified to take the form

$$U = U(c, h, \nu) - \delta_B\ \zeta, \quad (4.93)$$

where ζ is the level of welfare stigma which is greater than 0 and varies across individuals.[23] Two unobserved components now enter preferences, ν and ζ. Such cases were considered in the general analysis of Section 4.4. In terms of the notation of this preceding section, in relationship (4.49) ν corresponds to ν_1 and ζ corresponds to ν_2.

With this modification we again consider the welfare and non-welfare regimes. Since the welfare stigma term does not affect the marginal decisions, given that the individual is on welfare, the discussion of hours of work presented above for regime 1 is still valid. The optimal utility is now given by

$$V^*(\omega_1, y_1, \nu) = \begin{cases} V_1(\omega_1, y_1, \nu) - \zeta, & 0 < \ell_1 \leq \overline{h}_1, \\ U(y_1, 0, \nu) - \zeta, & \ell_1 \leq 0, \\ -\infty, & \ell_1 > \overline{h}_1. \end{cases} \tag{4.94}$$

The analysis for regime 2 is altered in this case, because an individual can be observed not receiving welfare for any value of h - that is, given welfare stigma, it is possible to observe an individual with $h < \overline{h}_1$, but $\delta_B = 0$. So regime 2 is now defined solely by $\delta_B = 0$. Optimal hours of work, given ω_2 and y_2, are given by $\ell(\omega_2, y_2, \nu)$. Defining the value of ν for which $\ell(\omega_2, y_2, \nu) = 0$ as ν^+, hours of work under this regime are now given by

$$\begin{aligned} h &= \ell_2 = \ell(\omega_2, y_2, \nu), \quad \nu > \nu^+, \\ h &= 0, \quad \nu \leq \nu^+. \end{aligned} \tag{4.95}$$

Optimized utility is now

$$V_2^*(\omega_2, y_2, \nu) = \begin{cases} V(\omega_2, y_2, \nu), & \ell_2 > 0 \\ U(y_2, 0, \nu), & \ell_2 \leq 0. \end{cases} \tag{4.96}$$

Choice of regime still proceeds by comparing V_1^* and V_2^*, as done in relationship (4.88). For any ν in the sets Ω_0 or Ω_1 defined by expressions (4.89) or (4.90), there is now some critical level of $\zeta^* = \zeta^*(\nu)$, which depends on ν, such that regime 2 is chosen when $\zeta > \zeta^*$; regime 1 is chosen otherwise.

[23] This additive form is used for simplicity. More general forms can be used, but change none of the substantive points presented here.

Given this characterization, we can derive the likelihood function for each combination of δ_B and δ_E, using the joint densities $\varphi_{\nu\zeta\eta}\ (\nu, \zeta, \eta)$ and $\varphi_{\nu\eta}\ (\nu, \eta)$:

$$
\begin{array}{llll}
\delta_B = 1, & \delta_E = 1, & \mathcal{L}_{11} = \frac{\partial \nu}{\partial h} \int_0^{\zeta^*} \varphi_{\nu\zeta\eta}(\nu(h), \zeta, W - W(z)) \ \ I(\nu \in \Omega_1)\ d\zeta \\
\delta_B = 0, & \delta_E = 1, & \mathcal{L}_{01} = \frac{\partial \nu}{\partial h} \varphi_{\nu\eta}(\nu(h), W - W^*(Z))\ I(\nu \in \Omega_1) \\
& & \qquad + \frac{\partial \nu}{\partial h} \int_{\zeta^*}^{\infty} \varphi_{\nu\zeta\eta}(\nu(h), \zeta W - W^*(Z))\ I(\nu \in \Omega_1)\ d\zeta \\
\delta_B = 1, & \delta_E = 0, & \mathcal{L}_{10} = \int_{-\infty}^{\infty} \int_{\Omega_0} \int_0^{\zeta^*} \varphi_{\nu\zeta\eta}(\nu, \zeta, \eta)\ d\zeta\ d\nu\ d\eta \\
\delta_B = 0, & \delta_E = 0, & \mathcal{L}_{00} = \int_{-\infty}^{\infty} \int_{-\infty}^{\nu^*} \int_0^{\zeta^*} \varphi_{\nu\zeta\eta}(\nu, \zeta\eta)\ d\zeta\ d\nu\ d\eta \quad .
\end{array}
\tag{4.97}
$$

Estimation proceeds as in the non-stigma case by selecting the appropriate likelihood branch for each individual and then maximizing the sum of the log likelihoods.

As with the fixed cost case, the likelihood function is complex even in this extremely simplified welfare case. For each possible set of parameter values, the maximum must be computed for each regime and then compared to compute ζ^*. Adding the tax codes, with their implied kinks, increases computational complexity. As a result, the literature has adopted a simplifying methodology which we present in Section 4.10 below.

4.9.3. Multiple Program Participation

In principle, the extension to the case of multiple program participation is straightforward. For simplicity, we consider a case in which the individual can choose between participating in no welfare programs, participating in welfare program 1, participating only in program 2, or participating in both welfare programs 1 and 2. We extend the utility function as follows:

$$U = U(c, h, \nu) - \delta_1 \zeta - \delta_2 \chi \tag{4.98}$$

where $\delta_1 = 1$ if the individual participates in program 1, and $\delta_2 = 1$ if the individual participates in program 2.[24] Benefits from program j, $B_j(I(h))$, are given:

$$B_j(I(h)) = \begin{cases} G_j - \rho_j W h, & \text{if } G_j - \rho_j W h > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{4.99}$$

[24] The use of two additive errors is a simplifying assumption which ensures that the stigma from both programs is higher than stigma from program 1 alone.

Benefits from both together are given as

$$B_1(I(h))+B_2(I(h)) = \begin{cases} G_1 + G_2 - \rho_1 Wh - \rho_2 Wh = G - \rho Wh, & \text{if } G - \rho Wh > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{4.100}$$

where $G = G_1+G_2$ and $\rho = \rho_1+\rho_2$. In general, the benefit functions for programs 1 and 2 will have different breakeven points, implying the values of hours defining kinks ($\overline{h}_1$ in Figure 4.3) will not be the same.

This formulation expands the model considered in Sections 4.6 and 4.9.2. To adapt this earlier model, one must designate three distinct regimes in place of regime 1 specified above: regime $1a$ indicating an individual participates only in program 1, regime 1b signifying this person collects benefits only from welfare program 2, and regime $1c$ designating participation in both programs. Optimal hours and utility for participation in a regime are given by (4.85), (4.86), (4.94), (4.95), and (4.96), with net wages and virtual income in these formulations specified as $\omega_j = W(1-\rho_j)$ and $y_j = Y + G_j$, with $j = 1a$, $1b$, or $1c$. In particular, relations analogous to (4.85) and (4.86) define the labor supply and utility functions for each of the new regimes for the "on-welfare" segments associated with relevant combination of welfare programs. Relations (4.95) and (4.96) still define the labor supply and utility functions for the non-welfare regime. The set of relations define thresholds for ν demarcating the regions of unobserved tastes determining when a person works (ν_0 in (4.85) and ν^+ in (4.95)). Maximization again requires selection of a regime. Relations analogous to (4.94) and (4.96) characterize utilities corresponding to the various regimes. Conditional on values ν, these relations in turn imply thresholds for the stigma errors ζ, χ, and $\zeta + \chi$ that determine individuals' welfare participation. The likelihood function for this model takes a form similar to equation (4.97), with more branches appearing in the function reflecting the additional regimes analyzed in this formulation.

Again, note the complexity of these extremely simplified welfare cases, even these involve significantly financial burden. For each possible set of parameter values, one must compute the maximum for each regime, account for the benefit structure, and then compare these to compute the error ranges for the likelihood function. When the individual is unemployed, one must perform these calculations for all possible wage values and all values of ν consistent with the no-work decision. Adding the tax code, with its implied kinks, increases computational difficulties. Introducing additional sources of unobserved heterogeneity enlarges the number of dimensions over which one must calculate integrals, requiring sophisticated numerical procedures and considerable computer resources. As a result, the literature has adopted simplifying methodologies, a topic to which we now turn.

4.10. Computational Simplification by Making Hours Choices Discrete

To make estimation problems manageable, a popular method is to presume that consumers face only a limited set of hours choices. For example, a worker may choose only full-time work, part-time work, or no work, with each of these options implying a prescribed number of hours. Formally, this is done by assuming that unobservable tastes components, ν, possess a discrete distribution, usually characterized as a multinomial distribution conditional on covariates. Combined with a 0/1 welfare decision, this finite set of hours choices yields a relatively small set of discrete states, say a set of S states, over which the utility function must be maximized.

Given a specific form for the preference function, utility can be readily evaluated at each of the hours choices and the maximum can be determined. Given an assumed joint distribution for unobservable taste components, ν, for the error component determining wages, η, and for welfare stigma, ζ, one can compute a probability that a family selects alternative j. This in turn defines a sample log likelihood of the form

$$\mathcal{L} = \sum_{j \,\in\, S} d_j \; \ln P(j \mid X, \theta) \tag{4.101}$$

where d_j is an indicator for whether individual i chooses alternative j, X is a vector of observable characteristics, and $P(j|\, X, \theta)$ is the probability of choosing alternative j with θ the set of unknown parameters. Such formulations are substantially less complicated than the specifications considered above because one avoids the intricate process of calculating thresholds and dealing with combined continuous-discrete endogenous variables; only discrete choices are allowed for here.

This formulation requires each individual to be placed into a limited set of preassigned work states, even though observed hours worked take many more values, making hours look as if they were continuously distributed. To overcome this issue, analyses applying this approach necessarily introduce measurement error in hours of work to admit hours to deviate from the discrete values assumed for the choice set. Hence, conditional on ν, each alternative j contributes some positive probability $P(j|\, X, \theta, \nu)$ which now depends on the value of the unobservable measurement error variables.

We illustrate this approach by considering the linear measurement error model given by (4.16) where the reporting error $\varepsilon \sim \varphi_\varepsilon$, with ε and ν independent. Further, as typically assumed, we specify that hours are not subject to measurement error in no-work states. The likelihood function for hours now takes the form

$$\mathcal{L} = \left(\sum_{j \,\in\, S_0} d_j \ln P(j|\, X, \theta) \right)^{1-\delta_E} \left(\sum_{j \,\in\, S_1} d_j \;\ln \left(\varphi_\varepsilon(H - h_j)\; P(j|\, X, \theta)\right) \right)^{\delta_E} \tag{4.102}$$

where δ_E denotes a 0/1 variable with 1 indicating that the individual works, S_0 designates the set of all states associated with the individual not working, the set S_1 includes all states in which the individual works, and h_j denotes the admissible values of true hours. Earnings depend on the values of h_j and wages. In (4.102), observed hours (H) are continuously distributed among workers.

5. Family Labor Supply

The study of family labor supply is motivated by a need to understand how a couple responds to tax and welfare benefit incentives when the benefit rules create links in the incentive structure as well as the need to understand how welfare is distributed within the household, so as to design the targeting of benefits appropriately. Indeed the structure of family labour supply has changed quite substantially and this may be partly due to changes in the benefit structure as well as a re sult of changes in relative wages. For example, in the UK there has been a large increase in the participation rate of married women and a decrease in the participation of men. These changes have been accompanied by an increase in the number of families where no one works. This is perhaps predictable given the structure of the benefit system. However the design of income maintenance programs that target the right households and offer the right incentive system is of course important and crucially relies our knowing the way that family labor supply is determined.

The basic family labor supply model for a married couple is the unitary model where the household is seen as maximizing one (household) utility function whose arguments are male and female labor supply and consumption. Applying demand analysis one can derive the implications of changes in wages and unearned income on behavior. Since taxes can be viewed as changes in wages and unearned income, such models can be used to simulate the labor market effects of changes in the tax system or welfare benefits. However in this model intra-household distribution has little meaning and of course the model has nothing to say about this. In addition it is unclear how the household utility function can come about from the interaction of two individuals with incentives that are not necessarily perfectly aligned. This has led to the recognition that even when dealing with households we need to account for individuals within households and we need to model the way they share resources. This leads to potentially richer models of behavior that are capable of explaining much more than the standard household model.

In the sections that follow we outline the two models and some of their implications in greater detail.

5.1. The Standard 'Unitary' Family Labor Supply Model

Consider the family labor supply and consumption problem

$$\max U(c, h_1, h_2, x)$$

$$\text{st} \qquad c = y + w_1 h_1 + w_2 h_2$$

where U is a strictly quasiconcave function of consumption C and the two labor supplies h_i. The budget constraint equates household consumption to total income, consisting of unearned income (y) and the two earnings $(w_i h_i)$, T being total time available for market work and w_i the two wages. In addition to the budget constraint, leisure cannot exceed T and hence labor supply must be positive or zero (h_i). This is a standard demand analysis problem with the complication that there may be corner solutions and wages being individual specific are not observed when the individual is working.

The first order conditions for an interior solution simply state that the marginal rate of substitution between the two leisures will equal the ratio of wages

$$\frac{U_{h_1}}{U_{h_2}} = \frac{w_1}{w_2} \tag{5.1}$$

An implication of this model is that behavior is neutral to within household lump sum redistributions of income. Thus paying a benefit to the male or the female will have exactly the same effect, so long as it does not distort wages. This is often termed the income pooling hypothesis and we revisit the issue when we discuss the collective model. Here it suffices to note that the symmetry condition and the income pooling hypothesis are properties of the unitary model and may not be satisfied in the collective one.

5.1.1. Nonparticipation

In this subsection we show how to deal with non-participation and missing wages in the family labour context. The issues are very similar to those already discussed in the single person labour supply model.

The first issue to be addressed is allowing for unobserved heterogeneity in the parameters of the utility function. Typically this can be addressed in a number of ways. One way would be to assume that the marginal rate of substitution for each of the two leisures with consumption included a multiplicative error term (see Heckman, 1974 for example). In this case we could write the first order conditions as

$$\begin{aligned} \log\left(-\frac{U_{h_1}}{U_C}\right) &= \log\left(w_1\right) + \varepsilon_1 \\ \log\left(-\frac{U_{h_2}}{U_C}\right) &= \log\left(w_2\right) + \varepsilon_2 \end{aligned} \tag{5.2}$$

We can also assume a (bivariate) density for the wage rates, say $f(w_1, w_2|z)$ where z are the observable characteristics that drive wages and ε_1 and ε_2 will be taken to be independent of them. Typically one would assume a distribution function for $\varepsilon = [\varepsilon_{1,}\ \varepsilon_2]'$, for example $N(0, \Omega)$.

The functions 5.2 together with the distributional assumption for the unobserved heterogeneity define the distribution of hours of work. Hence the likelihood contribution for a couple where both are participating is simply the joint density of hours of work and wages for the two of them.

$$\ell(h_1, h_2, w_1, w_2) =$$

$$\mid J \mid g\left(\log\left(-\frac{U_{h_1}}{U_C}\right) - \log(w_1), \log\left(\frac{U_{h_2}}{U_C}\right) - \log(w_2) \,|w_1, w_2, x\right) \times f(w_1, w_2|z)$$

$$J = \frac{\partial \varepsilon}{\partial h'} \qquad [Jacobian]$$

where x are observables affecting individual preferences and $h = [h_1 \ h_2]'$. When one or both partners are not working, hours of work are censored and the respective wage is unobserved. Take as an example the case where one of the two is not working (say 1). In this case note that $\varepsilon_1 < \log\left(-\frac{U_{h_1}}{U_C}\right) - \log(w_1)$, where $-\frac{U_{h_1}}{U_C}$ is the marginal rate of substitution evaluated at hours $h_1 = 0$. The likelihood contribution must be written taking this censoring into account. We will write this in terms of the joint density of hours and wages given above. Thus the likelihood contribution for this case is

$$\ell(h_1 = 0, h_2, w_1, w_2) = \int_{w_1} \int_{h_1 \leq 0} \ell(h_1, h_2, w_1, w_2) dh_1 dw_1$$

The integration with respect to the wage takes place over the entire range of wages. The contributions to the likelihood for the case of the other partner not working or both not working can similarly be derived. The sample likelihood is then the product of all contributions. In a similar fashion one can construct the likelihood contribution for the case where neither member of the household is working. The sample likelihood is then the product of the contributions for each observation. This is th basic likelihood structure. We next discuss issues relating to introducing taxation in this framework.

5.2. Discrete Hours of Work and Program Participation

It is straightforward to allow for proportionate taxes, or even piece-wise linear taxes, so long as these lead to a budget constraint that is convex and so long as the endogeneity of the tax rate is taken into account. However, most welfare programs are designed is such a way that they define a non-convex budget set: Implicit marginal tax rates are higher at low hours of work, where increases in earnings lead to a rapid withdrawal of benefits and lower at higher hours where the individual pays the usual taxes. As we showed earlier, this is a complex problem itself and in the family labor supply context even more so because the benefits may be interdependent.

To simplify the problem it has now become almost standard to discretize hours of work. Then the problem of utility maximization becomes one of choosing packages of consumption and earnings - consumption is defined by the earnings of the individual, the tax system and the benefit system. Within this context we can also account for fixed costs of work (another non-convexity) and for the decision to participate (or not) in a welfare program. (Hoynes (1996) and Keane and Moffitt (1998))

We start with a utility function defined over hours of work H_1 and H_2 and we discretize the distribution of hours. For example hours can take the discrete values H =$\{0,20,40\}$. Suppose we write family utility at hours $H_1 = h_i$, $H_2 = h_j$ where h_i and h_j are the ith and jth point of the discrete hours distribution respectively.

$$U_{h_i h_j} = U(H_1 = h_i, H_2 = h_j, c, \varepsilon) - \eta P_B + u_{h_i h_j}$$

where P_B is a 0-1 program participation dummy. The term ηP_B reflects the utility costs of program participation such as "stigma". This may be randomly distributed over the population. The term ε reflects unobserved heterogeneity in preferences and the term $u_{h_i h_j}$ hours specific unobserved heterogeneity. Given the associated wage the discrete hours imply a corresponding set of earnings for each individual.

The budget constraint incorporates all relevant aspects of the tax and benefit system to define the resulting level of consumption.

$$c = w_1 H_1 + w_2 H_2 + y - T(y, w_1 H_1, w_2 H_2) + B(y, w_1 H_1, w_2 H_2) P_B$$

where

T	-	tax function
B	-	program benefit function

The likelihood is derived taking into account program participation. First note that the observation on whether an individual is participating in welfare programs or not is informative about the range in which the participation cost η lies. Note also that for any given η the utility function and the budget constraint define whether the person will be a participant or not. At each observation we can derive the probability that the chosen point $pr(U_{h_i h_j} > U_{h_k h_s}, \forall\ k \neq i\ and\ s \neq j|w_1, w_2, \eta, \varepsilon)$ is optimal, conditional on η and the heterogeneity terms ε. If the person is eligible for a welfare program at the observed point and he does actually participate (i.e. receives benefits) then the range in which η lies is defined by the fact that the utility gain from participating is higher than the cost η. For the non-participants η lies in the complement of this set. This allows us to integrate out η over the relevant range. When the person is ineligible at the observed point no information is available for η and we integrate over its entire range. In this case as we move over different values of η the probabilities change not only because of the direct effect of η through the utility function but also because it induces different potential participation decisions at each discrete hours points, thus changing both optimal hours and consumption. Thus consider the likelihood contribution for a couple where both work and participate in a welfare program - (in work benefits). This will take the form

$$\left\{ \int_{\eta \epsilon Q} \int_{\varepsilon} pr(U_{h_i h_j} > U_{h_k h_s}, \forall k \neq i\ and\ s \neq j|w_1, w_2, \eta, \varepsilon) d\varepsilon d\eta \right\} f(w_1, w_2|z)$$

where Q is the set of η such that program participation is optimal at the point of observation. The form of the probability is defined by $u_{h_i h_j}$. Imposing a logistic is not restrictive if we allow for unobserved heterogeneity through the ε (Manski and McFadden, 1981). The contribution to the likelihood for a non-worker must also take into account the fact that the wage will not be observed in that case. This is done as before by integrating over all possible wages. Of course the practical difficulty is that the probability of participation is a complicated function of the wage rate through the formulae of the tax and welfare benefit system.

The models estimated in this way have the great attraction that they allow us to simulate policies allowing for possible changes in the take-up of means tested benefits. To the extent that there is sufficient genuine exogenous variation in the data allowing us to identify the factors that determine take up these can be very useful for the *ex ante* evaluation of welfare policies.

5.3. Collective Models of Family Labor Supply

The family labour supply model presented above treats the household as a single optimising decision unit, and has nothing to say about within household allocations. It also imposes stronger restrictions than necessary, such as symmetry. An alternative approach, the Collective model, looks upon the household as a set of individuals with their own preferences, whoa have to decide how to share the overall set of resources available to them. Within this framework we can have private goods (enjoyed by the members separately), public goods and household production. The main empirical issue is that of identification: What can we learn about individual preferences and the sharing rule when we observe aggregate household consumption. This has led to a number of important theoretical results by Chiappori (1988, 1992) recently extended by Blundell, Chiappori, Magnac and Meghir (2006) to allow for corner solutions and to discuss identification in the presence of unobserved heterogeneity.

The framework we describe here is the collective model with two household members and no public goods or household production.[25] Each member supplies labor h^i $(i = m, f)$ and consumes a private good (C_i). A critical assumption in the collective approach as introduced by Chiappori is that the household only takes Pareto-efficient decisions. That is, for any set of male and female wages and unearned income (w_f, w_m, y), there exist some level of male utility $\bar{u}^m(w_f, w_m, y)$ such that labor supply and consumption for each household member (h^i, C^i) is a solution to the program :

$$\max_{h^f, h^m, C^f, C^m} U^f\left[1 - h^f, C^f\right] \tag{5.3}$$

$$U^m\left[1 - h^m, C^m\right] \geq \bar{u}^m(w_f, w_m, y)$$

$$C = w_f.h^f + w_m.h^m + y$$

$$0 \leq h^i \leq 1, \quad i = m, f$$

where the labor supply has been normalized to lie between 0 and 1. The function $\bar{u}^m(w_f, w_m, y)$ defines the level of utility that member m can command when the

[25] Blundell, Chiappori and Meghir (2005) further extend the model to discuss identification conditions with public goods.

relevant exogenous variables take the values w_f, w_m, y. Underlying the determination of $\bar{u}^m$ is some allocation mechanism (such as a bargaining model) that leads to Pareto efficient allocations. The nice thing about the collective approach is that there is no need to be explicit about such a mechanism; identification does not rely on specific assumptions about the precise way that couples share resources.

Suppose first that preferences are such that there are never any corner solutions. It is assumed that we observe aggregate household consumption $C = C^m + C^f$ and that we know the locus of labor supplies as a function of (w_f, w_m, y).Then Chiappori (1988) proves the following:

Proposition 5.1. *(Chiappori 1988) Assume that h^m and h^f are twice differentiable functions of wages and non labor income. Generically, the observation of h^m and h^f allows to recover individual preferences and individual consumptions of the private good up to an additive constant.*

There are two critical issues to be resolved following this proposition: One is what happens with corner solutions and with discrete labor supply. The other is what happens with unobserved heterogeneity in preferences, i.e. when we do not know the exact loci h^m and h^f.

Blundell, Chiappori, Magnac and Meghir (2006) set up a framework where the male decision is discrete (work or not) and the female is continuous - however she can choose not to work. The framework underlying the proposition above exploits the fact that the marginal rates of substitution between consumption and labor supply for each agent will be equalized within the household, under efficiency. This result cannot help when one of the labor supplies is discrete. Define the participation frontier to be the set of male and female wages and unearned income y so that member m is indifferent between working and on working. Blundell et al (2006) then exploit the following implication of efficiency:

Definition and Lemma DI (double indifference): *The participation frontier L is such that member m is indifferent between participating or not. Pareto efficiency then implies that f is indifferent as well about whether m participates or not.*

Technically, this amounts to assuming that in the program above, $\bar{u}^m$ is a *continuous function* of both wages and non labor income. This will imply that the behavior of the female will depend on the male market wage even when he is not working. This continuity assumption restricts the set of possible behavior and plays a key role for identification. We will not go through the technical details, all of which are available in the paper referenced above. However, identification of preferences and the consumption sharing rule (up to an additive constant) follows from the assumption that all goods are private (no public goods and no household production) as well as from the assumption above. Blundell, Chiappori and Meghir (2005) discuss results in the presence of public goods. The essence of the results there is that full identification of preferences over private and public goods and the sharing rule follows when preferences over private consumption and labor supply are weakly separable from the public good. In any case it is shown that some aspects of the public good must be observable.

The next important obstacle for identification here is unobserved heterogeneity. The results outlined above relate to the case where we know the locus of the observable endogenous variables (labor supplies, the public good etc.) as functions of wages and unearned income. However for empirical purposes we need to establish identification in the presence of unobserved heterogeneity in preferences. This is generally complicated by the fact that any unobserved components affecting individual preferences are likely to affect the sharing rule. Since this can take any form (more or less) we may well end up with error terms that are non-separable, which of course may lead to lack of identification in general. Identification problems are compounded by the specific context of labor supply where wages are only observed for workers. Blundell et al. (2006) have established identification in the special case where the labor supplies and the sharing rule are linear in log wages and all have additive unobservables. Even in this case the proof is not trivial because they do not rely on distributional assumptions. One conclusion of this study is that identification in more complex preference structures will have to be established on a case by case basis. Nevertheless, the dividends of such an exercise are probably very high. Blundell et al (2006) reject the unitary model, while the collective model is not rejected and gives interesting insights into the way that resources are split up within the household. Further empirical work needs to include public goods and household production. This will allow an extension of this analysis to households with children. Finally, this framework needs to be extended to deal explicitly with the issues of taxation and means tested benefits, which the previous analysis of the collective model has not developed.

6. Intertemporal Models of labor Supply

The models discussed up to now focused on the work decision within a period. The lifecycle and dynamic issues have not been addressed. However, studying dynamics is of critical importance because of the numerous intertemporal dependencies in labor supply their implications and for the design of policy.

The most obvious intertemporal dependence comes through borrowing and saving. In this framework the credit market is used to shift labor income across periods of the life-cycle so that labor supply can be concentrated in periods when the relative benefit of supplying labor is highest or costs are lowest. This allows a reduction in labor supply during college, during childbirth and during retirement while consumption can be maintained at a level consistent with expectations and overall uncertainty. An additional reason for changes in labor supply over the life-cycle is the precautionary motive, which implies more labor supply when one is young and less when one is older and some of the uncertainty has been resolved (Low 1999)

However, intertemporal dependence may be more direct. Labor supply preferences may depend on past actions (habit formation); current work may improve future wages through learning by doing; current work may increase a future pension entitlement. Since a rational individual will take into account the impact of current actions on future budgets or preferences, the standard static labor supply model does not tell the complete story and may in fact be misleading. With intertemporal dependencies the individual may find it rational to work in circumstances where the static model would exclude such a possibility. For example, it may still be worth working when welfare benefits are reduced one for one with earnings, because work offers future returns in the form of higher wages.

The recent intertemporal labor supply literature has developed along two lines. This is reflected in these two intertemporal aspects on labor supply - through credit markets and saving, and through intertemporal non-separabilities. In the former case applications exploit the continuity of consumption and saving to derive Euler equation conditions for intertemporal labor supply. In the later case the focus is more on participation and intertemporal nonseprabilities, largely ignoring savings decisions.

This classification of approaches is necessarily too restrictive. There are intertemporal substitution applications that allow nonseparability over time, but these are few and typically do not account for fixed costs and nonparticipation. Also there are examples of dynamic programing models that account for savings decisions but to date these have been quite rare and based on very specific assumptions concerning preferences and markets.

This section presents dynamic models of labor supply and consumption and discusses their estimation. We start by presenting the standard dynamic framework, followed by the empirical models of MaCurdy (1981) and Heckman and MaCurdy (1980). We then discuss issues to do with intertemporal non-separability, unobserved heterogeneity in the context of incomplete insurance markets. We conclude with the presentation of a framework in which all these aspects are taken into account in a theoretically coherent fashion.

6.1. Intertemporal Labor Supply with Saving

As we have mentioned in section 2, the 'static' labor supply model can be made consistent with an additively separable life cycle model under uncertainty using the two-stage budgeting framework. However, this does not recover all of the parameters necessary for intertemporal analysis and for that we need to look directly at the first order conditions for intertemporal optimization. Before moving to consider the problems of unobserved heterogeneity in the context of uncertainty and with the possibility of corner solutions we consider a simpler model.

Using the framework of Heckman and MaCurdy (1980) and MaCurdy (1981) we discuss estimation of lifecycle labor supply models in a complete markets setting, i.e. with no uninsurable uncertainty and no aggregate shocks. We start by expositing the case of no corner solutions, where all individuals work. We then allow for non-participation. Next we introduce uncertainty, first by considering the no corners case and later allowing for corners as well. Finally we discuss the issue of unobserved heterogeneity in models with uncertainty and corner solutions and present an estimation framework based on the complete dynamic programming characterisation of the problem..

6.1.1. The Life-Cycle Model

Before discussing the identification and estimation issues in the dynamic models of labor supply and consumption we present the standard life-cycle model.

The individual maximizes expected life-time utility subject to an intertemporal budget constraint. We assume that future wage rates, prices and interest rates are uncertain and that labor market risk is uninsurable. Define A_t to be the assets, denominated in the same units as consumption. Letting i_t denote the nominal interest rate and p_t the price level, we define the real rate of return on assets to be $1 + r_t = \frac{p_t}{p_{t+1}}(1 + i_t)$. Thus r_t is to be taken as uncertain in period t. The real wage rate is denoted by w_t.

Denote by E_t the expectations operator with respect to the distribution of uncertain future variables conditional on information in period t. These include interest rates, wages, the price level, possible preference shocks and other variables which affect choices either through their impact on expectations or directly. Denote the collection of such *state* variables by S_t. The state variables contain all the information that is needed to summarize the individual's position at any point in time. Thus, conditional on the state variables the past otherwise is irrelevant. We can also think of the taste shifter variables z_{1t} and z_{2t} as being uncertain in future periods, in which case expectations are taken with respect to their distribution as well. We abstract from issues relating to uncertain date of death and the presence or absence of perfect annuity markets. Hence we take the personal discount factor β to be constant over time as a simplifying assumption.

We can write the intertemporal optimization problem as

$$V_0 = \max_{h_t, c_t} \left\{ E_0 \sum_{t=0}^{T} \beta^t \psi \left[U(c_t, h_t | z_{1t}), z_{2t} \right] \qquad | \qquad \sum_{t=0}^{T} \frac{1}{\prod_{s=0}^{t}(1 + r_s)} (c_t - w_t h_t) \geq 0 \right\}$$

where the second part in the expression is the intertemporal budget constraint. The way it is written implies that the individual can borrow and lend freely at a market rate of interest r_t.

The additive structure of this problem is viewed from the perspective of period 0. However, since there exists uninsurable uncertainty the individual will replan in each period as news arrives. In this context and since the problem is recursive (trivially since it is additive over time) it is more convenient to use the Bellman equation formulation

$$V_t(A_t|S_t) = max_{h_t,c_t}\{\psi\left[U(c_t, h_t|z_{1t}), z_{2t}\right] + E_t\beta V_{t+1}(A_{t+1}|S_{t+1})\} \qquad (6.1)$$

where $V_t(A_t|S_t)$ is the optimum value function given information up to period t and S_t are relevant state variables which help predict future uncertain income, interest rates and characteristics.

In the absence of credit market restrictions the intertemporal budget constraint implies that

$$A_{t+1} = (1 + r_t)(A_t + w_t h_t - c_t)$$

with the terminal value of assets fixed at some value (say zero).[26] This implies that the revenues and expenditures need to balance over the entire lifecycle but not necessarily at any point in time.

The first order conditions for labor supply and consumption can be written as

$$-u_h' \geq \lambda_t w_t \quad h_t \geq 0$$

$$u_c' \geq \lambda_t \quad c_t \geq 0$$

Usually an Inada condition is imposed which ensures that optimal consumption will always be strictly positive. However, optimal labor supply may be zero which leads to a corner solution.

For individuals with an interior solution the optimal allocation between consumption and hours of work within a period, equates the marginal rate of substitution to the real wage rate. The important point is that even in this dynamic context the marginal rate if substitution is the ratio of *within period* marginal utilities. Thus consumption and labor supply satisfy

$$\cap\frac{u_h'}{u_c'} = w_t \qquad (6.2)$$

where u_x' is the marginal utility of x. The important point to note is that the within period marginal rate of substitution between consumption and hours of work does

[26] We abstract from issues relating to portfolio choices and r_t is the return to the market portfolio.

not depend directly on any expectations about the future, nor does it depend on interest rates. Crucially, it does not depend on the monotonic transformation of the utility function ψ. This is important because it implies that in general we can not estimate the parameters governing intertemporal allocations just by using within period ones. Condition 6.2 is the basis of the life-cycle consistent 'static' labor supply model of the earlier sections.

We can apply the envelope condition for assets on (6.1) to characterise the link between decisions over time . This gives

$$V_t' = E_t \left\{ \beta(1+r_t)V_{t+1}' \right\}$$

Since the first order conditions also imply that

$$\psi_t' U_{ct}' = E_t \left\{ \beta(1+r_t)V_{t+1}' \right\}$$

and

$$\psi_t' U_{ht}' = -E_t \left\{ \beta(1+r_t)w_t V_{t+1}' \right\}$$

we can characterize the intertemporal rates of substitution for consumption and hours of work for interior solutions as

$$\psi_t' U_{ct}' = E_t \left\{ \beta(1+r_t)\psi_{t+1}' U_{ct+1}' \right\} \tag{6.3}$$

$$\psi_t' U_{ht}' = E_t \left\{ \beta(1+r_t)\frac{w_t}{w_{t+1}}\psi_{t+1}' U_{ht+1}' \right\} \tag{6.4}$$

The object of the exercise is to estimate the parameters of $\psi\left[U(c_t, h_t|z_1), z_2\right]$ from observations of consumption and labor supply over time. It turns out that we need to use two of the three conditions (6.2), (6.3) and (6.4). At this point note that the variables in z_1 affect both the within period marginal rate of substitution and intertemporal allocations. The z_2 variables only affect directly intertemporal allocations because they cancel out of the monotonic transformation. Of course they do affect within period allocations indirectly and in a full solution the consumption and labor supply functions will depend on all variables affecting tastes, expectations and the budget.

6.1.2. A simplification: A model with full participation

Before complicating the matters with non-participation we consider the estimation problem in a simpler model presented by MaCurdy (1981) where everybody works. The utility specification he used does not allow for corner solutions and takes the form

$$U_t = B_t C_t^{\gamma} - A_t H^{\alpha} \qquad 0 < \gamma < 1,\ \alpha > 1 \tag{6.5}$$

where H_t corresponds to hours of work (rather than leisure) and C_t to consumption. The range of parameters ensures positive marginal utility of consumption, negative marginal utility of hours of work and concavity in both arguments. Applying exactly the same analysis as above the implied intertemporal Frisch labor supply becomes

$$\log H_t = A_t^* + \log \lambda + \frac{1}{\alpha - 1} \ln w_t + \frac{\rho - r}{\alpha - 1} t \tag{6.6}$$

where the use of log hours of work presumes that all individuals work and hence $H > 0$. In (6.6) λ is the shadow value of the lifetime budget constraint and t is the age of the individual. Finally A_t^* reflects preferences and is defined by $A_t^* = -\frac{1}{\alpha-1} \log A_t$.

This equation is the Frisch labor supply equation. The important insight is that under certainty (complete markets - no aggregate shocks) all relevant future variables, such as wages are summarized by the fixed effect λ. So this equation has a simple message: Hours of work are higher at the points of the lifecycle when wages are high ($\frac{1}{\alpha-1} > 0$). Moreover if the personal discount rate is lower than the interest rate, hours of work decline over the lifecycle. Finally, hours of work will vary over the life-cycle with A_t^*, which could be a function of demographic composition or other taste shifter variables.

Specifying $A_t^* = \gamma' x_t + \eta_1 + u_t$ we obtain an econometric equation of the form

$$\log H_t = \gamma' x_t + \frac{1}{\alpha - 1} \ln w_t + \frac{\rho - r}{\alpha - 1} t + \left[\frac{1}{\alpha - 1} \log \lambda + \eta_1 \right] + u_t \tag{6.7}$$

where $\left[\frac{1}{\alpha-1} \log \lambda + \eta_1\right]$ is a fixed unobservable individual effect consisting of the marginal utility of wealth and of a permanent unobserved preference component.

u_t is an idiosyncratic shock to individual preference. For simplicity we take this as serially uncorrelated.

As it is, this equation presents a problem for estimation to the extent that the fixed unobservable effect (or the idiosyncratic shock u_t) is correlated with the hourly wage rate w_t. Because λ is a function of all wages over the life-cycle and because wages are highly persistent it is not tenable to assume that the fixed unobservable is not correlated with wages. The simplest case here is to assume that all right hand side variables, including wages are strictly exogenous, namely that $E(u_t|x_s, \log w_s \forall s = 1, ..., T) = 0$ in which case the model can be estimated using within groups: variables are transformed into deviations from their individual specific time-mean and OLS is applied on

$$\widetilde{\log H}_t = \gamma' \tilde{x}_t + \frac{1}{\alpha - 1} \widetilde{\ln w}_t + \frac{\rho - r}{\alpha - 1} \tilde{t} + \tilde{u}_t \tag{6.8}$$

where $\tilde{z}_t = z_t - \bar{z}$ represents the deviation of an individual specific variable from the time mean for this individual. the first difference operator (i.e. $\Delta z_t = z_t - z_{t-1}$). This model is estimable using panel data with a relatively small number of repeated observations for each of many individuals.[27] Here Ordinary Least Squares on the transformed model is consistent and fully efficient.

This empirical strategy is sensitive to measurement error for the right hand side variables. Suppose that log wages are measured with additive and serially uncorrelated (classical) measurement error. In this case the strict exogeneity assumption is violated and 6.7 cannot be estimated by within groups. An alternative approach in this case would be to take first differences, thus eliminating the fixed effect and then using instrumental variables to estimate the parameters based on the transformed equation. The instruments would have to be dated t-2 or earlier because the error in the first difference equation will have an MA(1) structure. Thus, under the assumptions made, valid instruments would be hours and wages lagged at least two periods. However, these instruments will only be valid if they are able to explain future growth in wages ($\Delta \log w_t$); hence this rank condition needs to be tested (see Steiger and Stock (1998), for example).

[27] FIxed T and large N asymptotics

6.1.3. The Heckman and MaCurdy Study

The MaCurdy (1981) paper set out the first clear analysis of issues to do with estimating intertemporal labour supply relationships. However the approach did not deal with corner solutions, which is particularly relevant for women. The first attempt to do so in the context of a life-cycle model of labor supply and consumption is the paper by Heckman and MaCurdy (1980). In this model women are endowed with an explicitly additive utility function for leisure L and consumption C in period t, of the form:

$$U_t = A_t \frac{L_t^\alpha - 1}{\alpha} + B_t \frac{C_t^\gamma - 1}{\gamma} \qquad \alpha, \gamma < 1 \tag{6.9}$$

Consumers are assumed to maximise life-cycle utility

$$V_t = \sum_{t=1}^{T} \beta^t U_t$$

subject to the lifetime budget constraint

$$\sum_{t=1}^{T} \frac{1}{(1+r)^t} \left[w_t h_t - c_t \right] \geq 0$$

where $h_t = \overline{L} - L_t$, $\overline{L}$ being maximal time available for work and where w_t is the hourly wage rate. Note that now utility depends on leisure and is well defined at the point where hours are zero since there one obtains maximum leisure.

Optimization is assumed to take place under perfect foresight. Solving for the first order conditions we obtain the following equation for leisure

$$lnL_t \begin{array}{ll} = & A_t^* + \frac{1}{\alpha-1} lnw_t + \frac{\rho-r}{a-1} t + \lambda^* \qquad \textit{when the woman works} \\ = & ln\overline{L} \qquad \textit{otherwise} \end{array} \tag{6.10}$$

where

$$\lambda^* = \frac{1}{\alpha-1} ln\lambda \text{ and } A_t^* = -\frac{1}{\alpha-1} lnA_t \tag{6.11}$$

and where we have approximated $ln\frac{1+\rho}{1+r} \approx \rho - r$. As before in (6.11) λ is the shadow value of the lifetime budget constraint and t is the age of the individual which again is a fixed effect because of the complete markets assumption.

Estimation with non-participation To estimate the model, Heckman and MaCurdy specify $B_t = \gamma' x_t + \eta_1 + u_{1t}$ where u_{1t} is normally distributed and where η_1 is a fixed effect reflecting permanent unobserved differences in tastes across individuals.

Given λ^*, η_1 and wages w_t this gives rise to a Tobit model, with censoring whenever the interior solution requires more hours of leisure than are available $(L_t > \bar{L})$. There are two main difficulties with this however. First, hourly wage rates are not observed for non-workers. Second, λ^* and η_1 are unobserved and cannot be differenced out in a conventional manner since the Tobit model is essentially nonlinear. Finally, a problem addressed only indirectly before (through the treatment of measurement error) is that of the endogeneity of wages. To solve these problems and to take into account that wages may be endogenous we may specify a wage equation of the form

$$lnw_t = z_t' \beta_2 + \eta_2 + u_{2t}$$

with η_2 being an unobserved fixed effect reflecting permanent productivity characteristics of the individual and u_{2t} being normally distributed. Endogeneity may arise if either the fixed effects in the wage and labor supply equations are correlated or if the idiosyncratic components are correlated (or both). In the former case (correlated fixed effects) treating the problem of fixed effects will also solve the endogeneity problem. In this sense we can think of wages as being endogenous in the case where we dealt with no corner solutions.

To proceed we can use the approach described earlier in the context of the static labor supply models. The wage equation is substituted into the structural labor supply equation and the conditions for an interior solution or otherwise is given in terms of the reduced form, i.e. not conditional on the wage rate. Hence we get

$$\begin{array}{lll} & = & \gamma' x_t + \frac{1}{\alpha-1} z_t' \beta_2 + \frac{\rho-r}{a-1} t + f + \upsilon_t \qquad when \\ lnL_t & & \upsilon_t < ln\overline{L} \ - (\gamma' x_t + \frac{1}{\alpha-1} z_t' \beta_2 + \frac{\rho-r}{a-1} t + f) \\ & = & ln\overline{L} \qquad otherwise \end{array}$$

where $f = \lambda^* + \eta_1 + \eta_2$, $\upsilon = u_{1t} + u_{2t}$. This gives rise to a Tobit model for the reduced form parameters. However, two important difficulties need to be addressed. The first relates to estimating this reduced form. The second to recovering the structural parameters characterising labour supply.

The reduced form labour supply includes a fixed effect f. In a linear model and with strict exogeneity the within groups estimator is consistent and efficient. The model here is nonlinear because of censoring. Heckman and MaCurdy (1980) treated them as parameters to be estimated. Formally speaking, when the model is nonlinear, this estimator is not consistent as the number of individuals N grows, while the number of time periods per individual T remains fixed. This is because the number of (incidental) parameters grows with the sample size. In practice the estimator is likely to work well with strictly exogenous regressors for moderate to large T. Heckman and MaCurdy provide Monte Carlo evidence showing that in their context the bias involved when using this approach is likely to be minimal for moderate T. However, this is not a general result and it depends very much on the model, the data and the number of time periods available. For example with lagged endogenous variables the biases could be substantial. Such lagged endogenous variables could appear in time non-separable models and in models with incomplete insurance markets as we will see subsequently. Thus the complete markets assumption turns out to be particularly powerful as far as identification is concerned.

An alternative approach is to use a semi-parametric LAD estimator introduced by Honore (1992). This estimator, relies on symmetry of the difference of the errors $(u_{it} - u_{it-1})$ conditional on the sum of the errors $(u_{it} + u_{it-1})$ and on the regressors, which is weaker than the assumption of normality combined with iid errors.

We have described how the reduced form labour supply equation can be estimated. This does not provide the parameters of the structural model because they are a function of the parameters of the wage equation. The next step is to recover the structural parameters. The difficulty here is that we first need to identify the parameters of the wage equation. This is not a simple problem because wages are observed for workers only, who are endogenously selected. In addition both the selection mechanism and probably the wage equation depend on fixed effects. Before we discuss estimation first we need to ensure that the parameters are identified. A necessary condition is that the wage equation includes variables that are excluded from the structural labour supply equation. Under normality no further restrictions are required. However, if one applies a semi-parametric estimation framework, that relaxes the normality assumption one also requires variables included in the labour supply equation that are excluded from the wage equation. One approach to completing estimation is to apply the Kyriazidou (1997) estimator to the wage equation. This controls for selection allowing for fixed effects in both the wage and the participation equations. Once the parameters of the wage equation have been recovered, one can use minimum distance to back out the parameters of the labour supply equation, which were estimated as above.

An alternative approach, and one followed by Heckman and MaCurdy is to use maximum likelihood treating the fixed effects as parameters to be estimated jointly (as discussed above). We turn to this approach now.

Maximum Likelihood Estimation The first step is to specify the joint distribution of hours of work and wages, conditional on the observables and the fixed unobserved effects. This is denoted by

$$g_{hw}(h, w|z, f, \eta) = g_h(h|x, f, w)g_w(w|z, \eta_1) \tag{6.12}$$

where z are the observed variables in the wage equation, which include all those in the labor supply equation (x) and more for identification purposes. In the above $g_h(h|x, f, w)$ is the conditional density of hours of work given wages, x, and f and $g_w(w|z, \eta_1)$ is the conditional distribution of wages given z and η_1. Thus the model likelihood is bivariate with the wage equation estimated at the same time.

The likelihood has the general form

$$\begin{aligned} L \;=\; & \prod_{workers} g_h(h|x, f, w)g_w(w|z, \eta_1). \\ & \prod_{non-workers} \int_{h<0} \int_w g_h(h|x, f, w)g_w(w|z, \eta_1)dwdh \end{aligned} \tag{6.13}$$

The first part of the likelihood relates to workers, where both wages and hours are jointly observed. The second part of the likelihood refers to non-workers where all we know is that desired hours are negative. Hence we integrate over $h < 0$ and over the entire support of the wage distribution, since for any wage rate there is a configuration of unobservables that would make the person a non-participant - being a non-worker conveys no information about wages. This likelihood can recover the parameters in the reduced form labor supply equation and in the wage equation.

As mentioned above, to identify the structural parameters of labor supply and the wage equation it is necessary to impose exclusion restrictions or some other form of parametric restrictions. Moreover, note that any variables that are fixed cannot be used for identification since they will be absorbed by the fixed effect. Heckman and MaCurdy exclude education/age interactions and aggregate unemployment from the labor supply equations and husband's labor market behavior from the wage equation. The former restriction effectively implies that differences in tastes across education groups *vis a vis* labor supply do not change with age. Consequently any change in observed behavior across education groups at different ages is attributed to education specific changes is individual productivity and hence to wages. The business cycle indicator (the unemployment rate) serves to identify wages for the non-workers through the aggregate price of human capital. Note however, that given the functional form assumptions the model is then overidentified.

The Heckman and MaCurdy model presented above offers a way of handling unobserved heterogeneity and corner solutions and even allows for persistent heterogeneity and endogenous wages. These properties have been delivered at a cost. Preferences between consumption and female hours are explicitly additively separable and no uncertainty is allowed for. The explicit additivity implies that, given consumption data, all parameters could be identified in principle using just within period allocations. This is worrying since it implies that intertemporal allocations are tied to the way that resources are allocated within period - an implication that does not come from economic theory. However, this assumption is testable since we can compare the estimates obtained from data on within period and data on intertemporal allocations. Finally, the perfect foresight assumption which is equivalent to complete markets with no aggregate shocks is also strong given the available evidence.

However, easy as it may be to criticize such an approach, it turns out that it is very hard to generalize. In what follows we discuss how the existing literature has attempted to build on this and what are the successes and shortcomings of these attempts. We start by describing an estimation strategy for a model of consumption and labor supply with corner solutions but with no explicit treatment of unobserved heterogeneity. As we argue below, an explicit treatment of unobserved heterogeneity places extensive requirements on data and an approach based on the complete solution of the life-cycle model, rather than on Euler equations.

6.1.4. Estimating the intertemporal substitution elasticity and other preference parameters under uncertainty

We now consider explicitly estimation in the presence of uninsurable uncertainty. Estimation will be based on two marginal conditions: One defines the within period allocations and the other the intertemporal allocation. Combining these two conditions in a suitable way can allow us to identify all parameters while accounting for corner solutions.

We start by characterizing within period preferences using the indirect utility function and appealing to two-stage budgeting. The within period indirect utility function is defined by

$$\psi\left[v_t(w,y)|z_t\right] = \max_{h,c}\left\{\psi\left[U_t(h,c)|z_t\right]|c_t = w_t h + y_t\right\} \tag{6.14}$$

where the variables z_t are shown explicitly to emphasize that intertemporal allocations will typically depend on taste shifter variables. As explained earlier in the chapter, the variable y_t reflects net saving or dissaving. Because c_t is realized consumption and $w_t h_t$ are actual earnings this amount (y_t) will only equal unearned income (e.g. from transfers or income from investments) if there is neither borrowing nor saving by the individual. Based on Roy's identity it is possible to derive the implied within period (or Marshallian) labor supply function, i.e.

$$h(w,y) = \frac{\partial v/\partial w}{\partial v/\partial y} \tag{6.15}$$

This labor supply function is conditional on y_t which reflects intertemporal decisions.

The labor supply function originating from (6.15) can be estimated using the methods described in earlier sections. The estimation of the within period labor supply function allows us to estimate all the parameters characterizing within period preferences, i.e. the function $v_t(w,y)$ in (6.14) but not the parameters of the function ψ. The latter affects intertemporal allocations only.

Thus we now need data on intertemporal allocations to estimate the parameters implicit in the monotonic transformation ψ, which characterizes savings behavior and intertemporal substitution in labor supply.

Consider again the Euler equation in an environment with uninsurable risk. This equates the marginal utility of consumption today with the expected marginal utility of consumption tomorrow

$$\psi'_t v'_{yt} = E_t \left\{ \beta(1 + r_t)\psi'_{t+1} v'_{yt+1} \right\}.$$

The term $v'_{yt} = \frac{\partial v}{\partial y}$ is the marginal utility of money, and $\psi'_t = \frac{\partial \psi[v_t(w,y)|z_t]}{\partial v_t}$ reflects the monotonic transformation of the utility function, which determines the intertemporal substitution. The marginal utility of money v'_{yt} can be estimated at a first step based on observations relating to within period allocations. We denote the estimated quantity by $\hat{v}'_{yt}$. The next step is to parameterize the function ψ which can then be estimated using the Euler condition. To write the Euler condition based on the indirect utility function we can use the envelope theorem to see that $U'_{ct} = v'_{yt}$ where U'_{ct} is the marginal utility of consumption which appears in the Euler condition 6.3.Based on this we can estimate the parameters characterizing ψ'_t using the following equation

$$\psi'_t \widehat{v}'_{yt} = \beta(1 + r_t)\psi'_{t+1} \widehat{v}'_{yt+1} + u_{it+1} + \varepsilon_{it} \tag{6.16}$$

where ε_{it} represents the estimation error due to the fact we are replacing v'_{yt} with its estimated value. Under the hypothesis of rational expectations any variable dated t or earlier will be orthogonal to u_{it+1}. This observation can serve as a basis for estimation using GMM (see Hansen, 1982 and Hansen and Singleton, 1983). Asymptotically ε_{it} will become irrelevant if the first step estimator is consistent, but can have serious implications in small samples.

With uninsurable uncertainty and in the presence of aggregate shocks it is imperative to estimate 6.16 using long enough time series. The innovation to the marginal utility of wealth u_{it+1} reflects uninsurable idiosyncratic risk and aggregate uncertainty. As Altug and Miller (1990 and 1998) have shown, the moment conditions do not hold in the cross section. In fact, the conditional expectation $E(u_{it+1}|t, z_{it}) = m(z_{it})$ where z_{it} represents the vector of instruments. Consequently with idiosyncratic uninsurable risk and aggregate uncertainty the model is not identifiable using methods that rely on fixed T; we require methods that rely on large T asymptotics and in practice we need long enough time series of data that allows the aggregate shocks to average out. The suitable time series dimension depends on the variance of such shocks, with longer series required the higher the variance. However, we do not require to observe the same individual for a large number of time periods; just that the data covers long T.[28] Moreover, aggregate shocks cannot be accounted for using time dummies as emphasized by Altug and Miller (1990) unless there is no idiosyncratic uncertainty.

Linearising the Euler equation A simpler way to go about estimation is to loglinearise 6.16.

$$-\Delta \ln \hat{v}'_{iyt+1} + \log(1+r_t) = d_{it} + \Delta \ln \psi'_{t+1} + \varepsilon_{it} \tag{6.17}$$

where $d_{it} = \log\left[E_t\left\{\beta(1+r_t)\psi'_{t+1}\widehat{v}'_{yt+1}\right\}\right] - E_t\log\left[\beta(1+r_t)\psi'_{t+1}\widehat{v}'_{yt+1}\right]$ .In the simplest case where the discounted marginal utility of consumption $mu_{it+1} = \beta(1+r_t)\psi'_{t+1}\widehat{v}'_{yt+1}$ is a log-normal random variable d_{it} will be proportional to its variance conditional on information in period t, i.e. $d_{it} = Var_t\left\{mu_{it+1}\right\}^2$. It is precisely this point that gives rise to the identification issue since the conditional variance will depend on variables relevant for predicting future income or wage realizations. However, if we are willing to restrict what the conditional variance depends on (and hence the stochastic process governing wages), this linearization offers a great simplification and often makes it easier to deal with measurement error in the underlying variables forming the marginal utility. Under non-normality d_{it} will also depend on higher order moments of the marginal utility of consumption mu_{it+1}.

Log linearization has been widely used in the empirical analysis of consumption. However, identification in this case, requires more restrictions than those implied by the theory. Its usage has been controversial (see Carroll, 1997 and Ludvigson and Paxson, 2001) precisely because the basic exclusion restrictions used for identification in 6.16 may no longer be valid in 6.17. Implicitly linearization imposes restrictions on expectation formation and on the underlying process of uncertainty. Attanasio and Low (2002) examine these issues using Monte Carlo

[28]Meghir and Weber (1996) discuss this point in relation to estimating Euler equations.

analysis in a wide variety of settings and conclude that in practice linearization is unlikely to bias the results in a serious way.

Accounting for corner solutions with no fixed costs When hours of work are at a corner solution the Euler condition 6.16 does not hold when evaluated at market prices. However, we can use the results of Heckman (1974) and Neary and Roberts (1980) to keep the Euler equation representation evaluated at shadow prices. Here we assume that there are no fixed costs of work and no search frictions and consequently that participation decision is fully characterized by the standard reservation wage condition (Heckman, 1974). In particular non-workers have a negative desired labor supply at the market wage corresponding to their skills, while workers have a positive desired labor supply, which is observed. It is easy to show that the intertemporal first order conditions still hold, so long as we evaluate the indirect utility function at the shadow (reservation) wage w_{it}^R defined by

$$h(w_{it}^R, y_{it}) = 0 \tag{6.18}$$

Estimating the 'static' within period labor supply function as described in earlier sections allows us to obtain a labor supply model that can then be solved for the reservation wage as in 6.18. In the next step the consumption Euler equation can be estimated using observed market wages for workers and shadow wages for non-workers.

An Example Consider the labor supply model

$$h_{it} = \alpha(z_{it}) + \beta \ln w_{it} + \gamma \frac{y_{it}}{w_{it}} \tag{6.19}$$

where z_{it} are preference shifters such as household characteristics. This corresponds to a particular form of the indirect utility function presented in an earlier section. The term y is defined by $y = c - wh$, where w is the after tax wage and c is total household (non-durable) consumption, and hence is endogenous. The utility index can be computed by using the formula in 2.7. This gives the value of $\hat{\nu}_t$,from which $\widehat{v}'_{yt}$ can be calculated. For workers the relevant wage will be the observed wage. For non-workers the relevant wage at which to evaluate within period utility is the reservation wage which is given by the positive solution for w

in equation 6.19 when $h = 0$, for given y. This has to be solved for numerically in this example. Using the reservation wage, is equivalent to computing the direct utility function when hours are zero. This calculation is only valid if there are no fixed costs of work.

In the next step we can specify the part of the utility function that is not revealed by within period choices. This is the monotonic transformation. One simple possibility would be to use a linear transformation; for example $\psi\left[v_t(w_{it}, y_{it})|z_{it}\right] = a(z_{it})v(c_{it}, h_{it})$, which would be interpretable as saying that characteristics z_{it} affect the discount rate. A more general alternative would be to allow characteristics to also affect the intertemporal substitution elasticity; for example $\psi\left[v_t(w_{it}, y)|z_{it}\right] = \frac{a(z_{it})}{1+\rho(z_{it})} v_{it}(w_{it}, y_{it})^{1+\rho(z_{it})}$, for some negative valued function $\rho(z_{it})$. The fact that all or some of the characteristics z affect within period allocations does not imply that they will not also affect risk aversion or the way the future is discounted.

To obtain an example specification let $a(z_{it}) = 1$ and $\rho(z_{it})= \rho_0 + \rho_1 fs_{it}$ where fs_{it} is family size for household i in period t. Using the utility function 2.7 term $\hat{v}'_{yt} = (1+\hat{\gamma})^2 \frac{w^{\hat{\beta}}}{\hat{\beta}+1}$ which can be evaluated at the estimated parameters. In this case the Euler equation for consumption over time will take the form

$$\hat{v}'_{iyt} \hat{v}_{it}^{\rho_0 + \rho_1 fs_{it}} = E_t \left\{ \beta(1+r_t) \hat{v}'_{iyt+1} \hat{v}_{it+1}^{\rho_0+\rho_1 fs_{it}} \right\}. \tag{6.20}$$

This can be estimated using non-linear GMM treating the estimated marginal utility of money $\hat{v}'_{iyt}$ and the within period utility index $\hat{v}_t$ as known (see Hansen (1982) and Hansen and Singleton (1983)). The fact that the expression depends on estimated parameters does not affect consistency because as the sample size goes to infinity the parameters estimated on the first stage converge to the true values. Inference however requires us to correct the standard errors for the fact that we are relying on pre-estimated parameters.

The linearized version of the Euler equation here takes the form

$$-\Delta \log \hat{v}'_{iyt+1} - \log(1+r_t) = d_{it} + \log\beta + \rho_o \Delta \log \hat{v}_{it+1} + \rho'_1 \Delta \log fs_{it} \hat{v}_{it+1} + \varepsilon_{it}$$

which, given the assumptions implied by the log-linearisation can be estimated by linear GMM.

Testing for liquidity Constraints One key issue for the interpretation of intertemporal behaviour is the extent to which individuals are liquidity constrained which is defined as being able to borrow and save freely at a constant interest rate. It has been observed from very early on that consumption seems to track income, which is a fact often cited as evidence for liquidity constraints. However, this phenomenon can be explained within the model we have presented.

First, Heckman (1974) has argued that such income tracking can be induced by non-separabilty of consumption and labor supply: If consumption and leisure are strong enough substitutes, higher amounts of consumption will be related to higher levels of labor supply and hence higher income.

Second, family size and demographics, which affect consumption and labor supply allocations, evolve very much alongside income over the life-cycle, with family size growing when income grows most and declining when income declines (probably endogenously see Blundell, Browning and Meghir, 1994). By allowing for this in our model we have effectively accounted for another reason for tracking.

Finally, the evolution of the conditional variance of the marginal utility d_{it} also leads to consumption growth. This variance is likely to decline over the lifecycle as uncertainty is revealed. This is particularly true if shocks to wages are permanent or highly persistent. Thus a high d_{it} when young and a lower d_{it} when old will imply rapid consumption growth early on declining later, much like the evolution of income over the lifecycle (Carroll and Samwick (1998), Attanasio, Banks, Meghir and Weber, 1999)

The empirical challenge is to find sources of predictable income growth not already included in the model to account for preferences (e.g. non-separability) and to test the hypothesis that they do not affect consumption growth. Browning and Collado (2001) use the powerful idea of predictable changes in income due to pre-announced and regular seasonal bonuses in Spain and establish that consumption growth is not sensitive to these totally predictable changes in income. However we are not always as fortunate as that and we need to use other perhaps less compelling sources of predictable growth. One possibility is to include labor income growth. This is a useful source of variation for two reasons: Conditional on the wage rate labor income would have variability because hours of work may change in a predictable way for other exogenous reasons. Second hours should not enter the Euler equation once we also include wages. Nevertheless it is still an issue of what the exogenous source of hours would be that has not to do with preferences or changes in wages. Another possibility is to use predictable changes in other income. The problem is that income from investments etc. are likely to be positive only for the wealthier individuals who are unlikely to want to borrow anyway.

Tests of liquidity constraints find no evidence of their importance once non-separabilities and demographics are allowed for. This should not be interpreted as saying that anyone can borrow any amount they wish at a fixed rate; after all the lack of complete markets is now generally accepted with moral hazard as its most probable source. However it may well mean that the lack of perfect credit markets is not important because individuals do not wish to borrow much against future income growth anyway when they would most need it (i.e. when young) because of uncertainty.

6.2. Further Issues in the Specification and Estimation of Dynamic Models of labor Supply and Consumption

The model we have presented up to now in the context of intertemporal optimisation lacks a number of potentially important features. These include unobserved preference heterogeneity, fixed costs of work and non-separability over time. We now discuss these issues in turn and we complete our chapter by presenting the estimation of a model containing potentially all these features.

6.2.1. Unobserved heterogeneity

Allowing for unobserved preference heterogeneity seems like a natural step in constructing realistic models. Thus for example, both MaCurdy (1981) and Heckman and MaCurdy (1980) recognise this and include fixed effects in their models. They recognise that preference heterogeneity could be persistent and may well be correlated with wages. The question is how to account for unobserved heterogeneity in a model without complete markets. The key difficulty stems from the fact that it is not possible to specify a model where both the Euler equation and the within period condition have additive errors without restricting the structure of intertemporal preferences. Inevitably a model with unrestricted intertemporal preferences and unobserved heterogeneity will be non-separable in unobservables.

Standard orthogonality conditions do not suffice for identification in this case. In the HM study the errors are effectively non-separable because of the corner solutions. However, the complete markets assumption meant that a fixed effects Tobit estimator worked well even with moderate T.

There is a developing literature on the identification and estimation of models with nonseparable errors and endogenous regressors (e.g. Florens, Heckman, Meghir and Vytlacil, 2006, Imbens and Newey 2003, Blundell and Powell, 2003), which provide alternative identifying conditions in this case. Even if one is to impose these stronger assumptions there remains the problem of finding suitable instruments, which are an ingredient of all such methods. The problem is particularly acute if unobserved heterogeneity is serially correlated , since the instruments are likely to be predetermined decisions. These difficulties will lead us to an estimation method based on a complete solution of the dynamic programming model.

6.2.2. Estimating the intertemporal substitution model with fixed costs of work

Fixed costs of work or other non-convexities in the budget constraint pose a very serious challenge to the empirical analysis, even within a static framework. In this context the labor supply function is discontinuous at low hourly wage rates. Moreover as Cogan (1981) pointed out the standard reservation wage which sets labor supply to zero does not generate a participation condition. Generally the participation and hours margins are explained by different models, which could be the result of the existence of fixed costs of work or of search frictions. The separation between the intensive and extensive margins (hours of work) requires extra identifying assumptions.

Within an intertemporal context fixed costs pose additional difficulties for modelling the participation decision: This involves a comparison between the lifecycle utility of work and non-work, which requires solving the life-cycle model conditional on the person working and conditional on the person not working. Such a solution allows one to evaluate the current and future welfare consequences of the two decisions.

In the presence of fixed costs we can follow two empirical strategies. The first is a partial one and seeks to estimate the subset of parameters that are identifiable if one keeps labor supply behaviour fixed. As such it cannot be informative for policy questions whose answer relies on the quantification of the complete labor supply and consumption response. However, it offers a way of testing some aspects of the life-cycle model in a relatively general setting and may be a first step in a stepwise approach for identifying the complete set of preferences.

The second approach specifies a complete structural model of labor supply and participation and uses methods from dynamic discrete choice to estimate labor supply responses. Before moving to a discussion of the full solution approach we briefly outline the conditional approach.

6.2.3. The Conditional Euler Equation for Consumption

Consider the definition of the indirect utility function within period, based on a vector of goods $\mathbf{q}_t$ and prices $\mathbf{p}_t$ conditional on labor supply behavior h_t

$$\upsilon_t = \psi\left[v(c_t|\mathbf{p}_t, h_t), h_t\right] = max_q \left\{\psi\left[u\left(\mathbf{q}_t|h_t\right), h_t\right] | \mathbf{p}_t'\mathbf{q}_t = \mathbf{c}_t\right\} \tag{6.21}$$

We can then base the analysis of the intertemporal allocations on the utility index $\upsilon_t = \psi\left[v(c_t|\mathbf{p}_t, h_t), h_t\right]$. As in the case of the joint labor supply and consumption model presented earlier, all parameters implicit in $v(c_t|\mathbf{p}_t, h_t)$ can be estimated using a conditional (on h_t) within period demand system (see Browning and Meghir, 1991). This will on h_t if and only if the goods $\mathbf{q}$ are nonseparable from h_t. Under weak separability h_t will not affect demands directly. However, the intertemporal allocations can still depend on h_t without this having any implications for the structure of the within period marginal rate of substitution functions between goods. This point has been noted now in several papers, all of which have demonstrated its empirical importance.[29]

The estimation approach is broadly similar to the one described above so we do not go over it again in detail. Once the within period demand system characterizing the conditional choice of $\mathbf{q}_t$ has been estimated, we can construct the utility index $v(c_t|\mathbf{p}_t, h_t)$. The Euler equation for c_t can then be used to estimate the parameters of the function ψ up to an explicitly additive function of h_t. In

[29] Attanasio and Weber (1993, 1996) and Blundell, Browning and Meghir (1994), Meghir and Weber (1996) all strongly reject the hypothesis that intertemporal allocations do not depend directly on observed labour supply.

general, the Euler equation as well as the demand system will be a function of h. This can include both hours of work as a continuous variable and indicators of whether the person is working or not, or other functions of h that are considered relevant. The crucial point to recognize however, is that labor supply is endogenous both for within period and for intertemporal allocations. Thus estimation requires suitable instruments. One possibility is to use lags in labor supply for this purpose. In the absence of unobserved heterogeneity the approach is valid. However, if persistent preference shocks have been ignored this approach could lead to inconsistent parameter estimates.

The conditional Euler equation for consumption provides a very powerful vehicle for testing the lifecycle model in relation to consumption behavior and for estimating some of the parameters in a way that is robust to the specific model of labour supply. In principle, hours of work can be determined in a number of ways, which we do not have to specify, subject to the proviso that we can specify instruments that can "predict" labour supply. However, from a policy perspective, the conditional Euler equation for consumption is of limited interest because it does not provide the full set of parameters required to answer even a simple partial equilibrium question. Thus a complete analysis of intertemporal labor supply and consumption needs to address directly estimation of a model for the determination of hours of work.

6.2.4. Intertemporal non-separability

A final issue is whether preferences should be taken as separable over time. It is well documented that labor supply behaviour is very persistent which may be interpreted as being due to non-separability, although the source of persistence could well be unobserved heterogeneity. Another source of non-separability can be the structure of the intertemporal budget constraint since current behaviour may affect eligibility for welfare programs. Finally, if wages depend on past work experience, current work affects future earning prospects, which also leads to intertemporal non-separability. These issues are considered in the next section.

6.3. Dynamic Discrete Choice Models and Intertemporal Non-Separability

To address many of the issues presented above in a coherent and unified way we need to consider a complete model of lifecycle labor supply and consumption. This can be very complex and demanding on data. Thus in our presentation we start with a simplified model along these lines which ignores the savings decision but offers a way forward on the issue of fixed costs and non-separability. We subsequently build on this to present a more complete model that includes savings.

One of the first attempts to model the dynamics of participation decisions when choices are discrete is given by Eckstein and Wolpin (1989). Their model concerns the labor supply of women. Husband's income is taken as exogenous. The within period utility function, which is non-separable in consumption c_t and participation p_t takes the form

$$U_t = c_t + a_1 p_t + a_2 c_t p_t + a_3 p_t K_{t-1} + \Sigma_{j=1}^{J} a_{4j} N_{tj} p_t + a_5 p_t S \tag{6.22}$$

where K_{t-1} is the number of periods worked in the past; depending on the sign of a_3 this may turn out to reinforce work habits or not. The law of motion of K_t is simply $K_t = K_{t-1} + p_t$. Finally, S represents years of schooling and N_{tj} represents the number of children in age group j. This utility function in itself gives rise to intertemporal dependencies since current participation affects future preferences and a forward looking individual will take this into account when making participation decisions. Further dynamics are induced by the budget constraint. This takes the form

$$y_t^w p_t + y_t^h = c_t + \Sigma_{j=1}^{J} \kappa_j N_{tj} p_t + b p_t \tag{6.23}$$

where κ_j are costs relating to children in the jth age group and b is a fixed cost of work and y_t^h is husbands income, which is taken to be an exogenous stochastic process, affecting female utility only through total resources. The female wage y_t^w, depends on past work decisions

$$ln y_t^w = \beta_1 + \beta_2 K_{t-1} + \beta_3 K_{t-1}^2 + \beta_4 S + \varepsilon_t \tag{6.24}$$

where ε_t is an independently and identically distributed normal shock to wages. Hence the implied dynamics in this model are quite intricate: Past work decisions produce human capital and enhance earnings potential. This should lead to increases in participation. On the other hand, past work decisions change preferences, either dampening down or reinforcing the effects due to enhanced human capital.

At this stage the only source of stochastic variation is the iid shock to wages ε_t.This formulation has the undesirable feature that the minimum observed wage is a consistent estimator for the reservation wage; this is because preferences are homogeneous in the population. To overcome this problem Eckstein and Wolpin allow observed wages to be measured with error, which turns out to be particularly important empirically. Thus observed wages satisfy

$$lny_t^{w*} = lny_t^w + u_t \tag{6.25}$$

Eckstein and Wolpin assume that u_t is normally distributed.

In such dynamic discrete choice models estimation is complicated by the fact that participation in this period confers benefit/costs in future period. Thus the future impact of current choices needs to be computed explicitly in order to compute the probability of participation. Eckstein and Wolpin follow a maximum likelihood approach where the parameters of the participation decision, of wages and of the measurement error process are estimated simultaneously.

Their estimation approach can be described as follows: An individual participates if the utility from doing so is higher than the utility from not working. To illustrate the approach we simplify further their model by assuming additive separability between consumption and participation. In this case the husband's income will not affect female labor supply. For notational simplicity we also drop the schooling (S) and household composition terms (N_{tj}). In this simplified framework, utility when participating can be written as

$$\begin{aligned} V_t^{(1)} &= y_t^w + y_t^h - b + a_1 + a_3 K_{t-1} + \delta E V_{t+1}(K_{t-1}+1) \\ &= exp\left(\beta_1 + \beta_2 K_{t-1} + \beta_3 K_{t-1}^2 + \beta_4 S + \varepsilon_t\right) \\ &\quad + y_t^h - b + a_1 + a_3 K_{t-1} + \delta E V_{t+1}(K_{t-1}+1) \end{aligned} \tag{6.26}$$

while the utility from nonparticipation is given by

$$V_t^{(0)} = y_t^h + \delta E V_{t+1}(K_{t-1}) \tag{6.27}$$

where δ is the personal discount factor. Note that when the woman participates in this period, human capital increases by one and does not increase otherwise. This is what gives rise to the difference in the future values associated with the current actions. In the expressions above the expectation is taken over the uncertain realizations of ε_t (and of the husband's income). This expectation is conditional on information known in period t. However, since the shock is iid conditional and unconditional expectations coincide.

A participation rule can be derived now from these two expressions written in terms of thresholds for the unobserved shock ε_t. Workers are individuals with wage shocks such that[30]

$$\begin{aligned}
\varepsilon_t &\geqslant ln\left[b - a_1 - a_3 K_{t-1} + \delta\left(EV_{t+1}(K_{t-1}) - EV_{t+1}(K_{t-1}+1)\right)\right] \\
&\quad -(\beta_1 + \beta_2 K_{t-1} + \beta_3 K_{t-1}^2 + \beta_4 S) \qquad (6.28) \\
&or \\
\varepsilon_t &\geqslant \varepsilon_t^*(K_{t-1})
\end{aligned}$$

Given a distributional assumption on ε_t this leads to a probability of participation. Note however, that the expression in 6.28 depends on the future expected gain from working. Hence to estimate the model this gain needs to be computed. This is achieved by backwards induction.

For a given set of parameters of the utility function and the distribution of the unobservable ε_t the value of participation and non participation is constructed in a terminal period, given all possible values of the state variables (in this case K). For each K we then compute $EV_T(K) = E\left[max(V_T^{(1)}, V_T^{(0)})\right]$ where the expectation is over the realizations ε_T. Computing the value in period T is very simple since the problem is essentially static then.

The only way by which past decisions affect the future is through the state variable K. Hence the future gain from working this period when the current experience stock is K is simply $EV_T(K+1) - EV_T(K)$. Whether this is positive or negative will depend on the effect of an extra unit of human capital on wages and on preferences. Given the terminal value function we can now compute the

[30]In our simplified model the husbands income plays no role in the wife's decision. This is not a feature of the Eckstein and Wolpin model but a result of our simplified exposition in which we have assumed additive separability.

values in period $T-1$ for all possible K accumulated by period $T-1$ and so on until we reach period t. This computation is a simple recursion. The procedure requires one to specify a terminal period (age) T. It also require us to be specific about what happens beyond that period. In models that require backwards induction it is often necessary to parameterize separately a terminal value function. In Eckstein and Wolpin the value beyond the last decision period T is assumed to be zero.

Given a way to compute $EV_{t+1}(K_{t-1})-EV_{t+1}(K_{t-1}+1)$ we can now easily construct the likelihood function. For non-workers this is simply $Pr(\varepsilon_t < \varepsilon_t^*(K_{t-1})) = \Phi(\varepsilon_t^*(K_{t-1}))$ where Φ is the standard normal distribution function. For workers the contribution to the likelihood function is the joint density of wages (driven by the sum of the shock ε_t and the measurement error u_t) and the probability that $\varepsilon_t > \varepsilon_t^*(K_{t-1})$. Hence estimation proceeds as follows: For an initial set of parameters the future gains from work are computed. Then the observed event is computed and the likelihood function is constructed for each observation. A Newton type algorithm can then be used to update the parameters. The value functions need to be recomputed at each iteration when updated parameters are available - this is what makes dynamic discrete choice computationally burdensome.

Estimation of this model requires observations on K_{t-1} and the choice p_t as well as wages. In general retrospective information on periods worked can be used, although entire work histories constructed over time as event unfold would reduce the chance of measurement error. Administrative data has now become available which improves the data situation substantially (see Adda, Dustmann, Meghir and Robin, 2006).

The dynamic discrete choice model described above is a coherent and powerful way of modelling the dynamics of participation and the evolution in wages. However, it does not allow for unobserved heterogeneity and thus all dependence on the past is in effect assumed to be pure state dependence.

The model by Eckstein and Wolpin is a prototype on which other researchers have built, drawing also from the experience gained in the analysis of discrete choice in other fields or in labor supply (Rust, 1987 Pakes 1986, Hotz and Miller 1988, Berkovec and Stern 1991). One of the most important subsequent contributions in the field of labor supply is the paper by Rust and Phellan (1997). The crucial aspect of this paper is that it models explicitly the relationship between work and future social security entitlements, thus building a model that can be used to evaluate the impact of policy reforms. An important feature, which complicates the model and makes it much harder to implement is that the individual's choice depends on a large number of state variables that evolve stochastically. In the Eckstein and Wolpin prototypical model there was basically only one state variable: the number of periods worked in the past. Here the state space includes health status, own earnings, spouse's earnings and social security income. Some of these variables are affected by past decisions. Hence the intertemporal non-separabilities in this model are primarily induced by the structure of the budget set: Current work decisions affect both future earnings and future social security receipts.

The principle of estimating such a model does not differ fundamentally from that of estimating the Eckstein and Wolpin model: The stochastic process for the exogenous state variables is estimated from the data. Then, following the specification of a distribution for the unobservables, the probability of observed choices is constructed, which depends on the future and current utility gains from this choice. As before, for each set of parameter values and at each value of the state variables the model has to be solved and the optimal choice determined. The probabilities at each data point are combined in the usual way to form the sample likelihood function. However, the problem is more complicated because of the many sources of uncertainty, originating from the large number of stochastically evolving state variables. These components are critical additions because they recognize explicitly that there are events such as the possibility of death or taste shifter variables such as health that affect behavior but are fundamentally uncertain. Such uncertainty is very likely to affect labor supply and retirement behavior of individuals.

6.4. Estimation with Savings, Participation and Unobserved Heterogeneity

We conclude our chapter by a brief discussion of estimation of dynamic models with savings in the absence of complete markets, which bring together the entire set of issues we have identified as challenges in estimating labor supply models and takes us right against the research frontier in this field.

6.4.1. Estimation with Complete Markets

Altug and Miller (1998) specify a model of consumption and labour supply, where preferences are non-separable over time and where wages depend on past labour supply (experience). In a departure from the earlier literature, savings are explicitly taken into account as are aggregate shocks. Moreover, the estimation methods proposed are relatively simple since they exploit a modified version of the conditional choice probability estimator developed in Hotz and Miller (1993). The key assumption that allows them to estimate such a complex model is that markets are complete. They also assume that preferences for leisure and consumption are additive. Finally the problem is simplified further by assuming that preference shocks are independently and identically distributed over time (and individuals) and there is no source of persistent heterogeneity in preferences.

The complete markets assumption allows them to express consumption allocations as a function of a fixed effect and an aggregate time effect. This solves at one go the problem of dealing with aggregate shocks when the time period is short (Chamberlain, 1984) and the problem of having to simulate alternative consumption paths explicitly when solving the dynamic programming problem.

In Altug and Miller the complete markets assumption can be viewed as an approximation that allows them to estimate a more general economic model than the ones considered earlier in the literature. Indeed their model is particularly rich, because it allows for endogenous human capital accumulation, for non-separable preferences as well as savings. However, the complete markets assumption is resoundingly rejected whenever it is tested (Cochrane, 1991 and Attanasio and Davis, 1996). It is not known how much bias the assumption would introduce in the parameter estimates. Nevertheless, the real empirical challenge is to relax both the complete markets assumption and relax the structure of unobserved heterogeneity. In the next section we review the issues surrounding this challenge.

6.4.2. Estimation with Uninsurable idiosyncratic risk

We consider an economy where some idiosyncratic risk remains uninsurable. However we assume that perfect credit markets are available.[31] Consider a utility function depending on hours of work h_{it}, on participation p_{it} (to reflect fixed costs) and on consumption c_{it}.

$$U_{it} = U_1(c_{it}, h_{it}, p_{it}, f_i|z_{it}) + U_2(h_{it}, f_i|z_{it}) + \gamma(z_{it})p_{it} + p_{it}\nu_{it}^{(1)} + (1-p_{it})\nu_{it}^{(2)} \quad (6.29)$$

where $z_{it} are$taste shifter variables and where f_i, $\nu_{it}^{(1)}$ and $\nu_{it}^{(2)}$ are heterogeneity terms, the first being time invariant. Assets accumulate according to the difference equation

$$A_{it+1} = (1 + r_t)(A_{it} + w_{it}h_{it} - c_{it})$$

The terminal condition for assets is

$$A_{iT} = 0$$

where T is the last period of the planning horizon. We do not discuss retirement explicitly. However, early retirement can be induced by the availability of pensions later in life, the accumulation of private assets, by aspects of the welfare system such as easily available disability insurance and/or by a decline in wages at an older age.

We assume wages take the form

$$\ln w_{it} = d_t^e + \kappa_i + \zeta^{ed'} x_{it} + e_{it}$$

where d_t^e is the log price of human capital for education group e, x_{it} denotes observable characteristics, some of which may be common with x_{it}, κ_i is a fixed effect and e_{it} is an iid shock with a known distribution, say normal.[32]

Suppose the function U_1 in 6.29 is non-additive in participation p,hours h and consumption c with no components that are additive in p or h. In this

[31] Some may view this as a contradiction. However, given uncertainty, most individuals will typically not want much uncolateralised borrowing, making the modelling of LCs probably redundant for all practical purposes. This may be why many tests for Liquidty Constraints fail to reject the null of no constraints.

[32] Richer stochastic structures are in principle possible, but they do increase the state space substantially.

case it is possible to estimate U_1 and U_2 based on the conditional Euler equation for consumption and on the within period labour supply decision as discussed earlier, subject to being able to deal with unobserved heterogeneity. However, the parameter γ cannot be identified in this way. This missing component will be key to simulating counterfactual employment, hours and consumption paths for individuals. Despite the relative simplicity of preferences and the wage function, both of which exclude intertemporal dependencies, the estimation of all relevant parameters requires the full solution of the dynamic optimisation problem: the probability of working is a function of the utility gain from doing so. To compute this utility gain one must know the consumption in the counterfactual state. With incomplete markets and idiosyncratic shocks this is not as straightforward as in the Altug and Miller case. We outline a possible approach.

We start by simplifying the model and assume a constant interest rate $r_t = r$. Next specify the conditional distribution governing the evolution of all other state variables, i.e. $g_s(S_{it}|S_{it-1}, ..., S_{t-p})$, where S includes all stochastically time varying in x and $z$taken to be exogenous. In general g_s can be estimated separately and we can condition on it during estimation of the rest of the model.

In general heterogeneity in the wage rate κ_i will be correlated with the heterogeneity in preferences f_i. This implies that wages are endogenous for both labor supply and consumption and this reflects the idea that unobserved productivity and the tastes for work are related. A simplifying assumption could be made reducing the dimension of heterogeneity, i.e. $f_i \propto \kappa_i$.

In this model assets are the only endogenous state variable, which in principle should include all sources of household wealth, including housing and pension wealth. This causes a very serious measurement problem. Leaving this aside, given suitable data the model is solved numerically to obtain the value of consumption conditional on the person's labour market state. Denote the optimal solutions as follows; workers : $c_{it}^{(1)} = c_t^{(1)}(w_{it}, A_{it}|S_{it}, f_i, p_{it} = 1)$, non-workers $c_{it}^{(0)} = c_t^{(0)}(A_{it}|S_{it}, f_i, p_{it} = 0)$ and $h_{it}^{(1)} = h_t(w_{it}, A_{it}|S_{it}, f_i, p_{it} = 1)$. In general there will be no closed form solutions to these functions and they will need to be computed numerically during estimation. To compute these policy functions we need to solve for the future optimal policies. One approach, for this finite horizon problem is to use backwards induction. Starting from some terminal period, the optimal policies are evaluated for all possible values of the state variables backwards up until the current period. At his point we have all the ingredients to evaluate th probability of work, including $c^{(1)}$ and $c^{(0)}$ and the future values conditional on current actions1 $(EV^{(1)}_{it+1})$ and not working $(EV^{(0)}_{it+1})$.The current value of working and not working is then given by

$$\begin{aligned} V^{(1)}_{it} &= U(c^{(1)}_{it}, h^{(1)}_{it}, p_{it} = 1, f_i) + \nu^{(1)}_{it} + \beta E_t V^{(1)}_{it+1} \\ V^{(0)}_{it} &= U(c^{(0)}_{it}, h_{it} = 0, p_{it} = 0, f_i) + \nu^{(0)}_{it} + \beta E_t V^{(0)}_{it+1} \end{aligned}$$

which now allows us to specify the probability of working as

$$Pr(p_{it} = 1|A_{it}, S_{it}, f_i) = Pr(\nu^{(1)}_{it} - \nu^{(0)}_{it} > U^{(0)}_{it} - U^{(1)}_{it} + \beta \left[E_t V^{(0)}_{it+1} - E_t V^{(1)}_{it+1}\right]$$

The consumption and labour supply as derived above are deterministic given the fixed effect f_i. The reason for this is that the time varying heterogeneity terms $v^{(1)}$ and $v^{(0)}$ do not affect the marginal utility of hours (given participation) or consumption. One simple way to enrich the stochastic specification is to allow for measurement error in consumption and hours. This will induce a density of observed hours m_h among workers and observed consumption $m^{(1)}_c$ for workers and $m^{(0)}_c$ for nonworkers.Thus the likelihood conditional on the heterogeneity term is

$$\begin{aligned} L &= \prod_{i=1}^{N} \prod_{t=1}^{T_i} \{ \left[m^{(1)}_h m^{(1)}_c g\left(w_{it}|\kappa_i, S_{it}\right) Pr(p_{it} = 1|w_{it}, A_{it}, S_{it}, f_i)\right]^{p_{it}} \\ & \left[\int \left[m^{(0)}_c g\left(w_{it}|f_i, S_{it}\right)\left(1 - Pr(p_{it} = 1|w_{it}, A_{it}, S_{it}, f_i)\right)\right] dw_{it}\right]^{1-p_{it}} \} \\ & \prod_{i=1}^{N} \prod_{t=1}^{T_i} L_{it}(f_i) \end{aligned}$$

where $g(\cdot)$ is the density of wages, N is the number of individuals and T_i is the number of time periods over which individual i is observed and $L_{it}(f_i)$ is the likelihood contribution for individual i. The stochastic dependence between the various elements in the likelihood is driven by the unobserved component f_i, which needs to be integrated out. .

Allowing for persistent unobserved heterogeneity is complicated by the fact that at any point in time f_i will be correlated with: these are the outcome of past decisions, themselves a function of f_i.Thus in a panel of individual data the initial value of assets cannot be taken as exogenous in general. To solve this problem we need to specify a model for the initial value (A_{i0}), conditional on a set of variables assumed themselves to be exogenous. Denote the distribution of initial assets by $g_A(A_{i0}|\zeta_i, z_{it})$ where z_{it} are a set of instruments explaining initial assets, which are excludable from the participation probability. Finding such instruments is not straightforward. One possibility could be to use random shocks that affected wealth at some point, but did not change preferences, such as, for example, parental health. The unobserved variables ζ_i and f_i may be correlated, which is the source of endogeneity of initial assets. If these are exogenous, f_i and ζ_i would be independent of each other.[33]

Given a model for initial assets and using a discrete mixture as an approximation to the distribution of the pair (f_i, ζ_i) (see Heckman and Singer, 1983) the likelihood function now becomes

$$L = \prod_{i=1}^{N} \sum_{k=1}^{K} \sum_{s=1}^{S} \left\{ pr_{ks} g_A(A_{i0}|\zeta_s, z_{it}) \prod_{t=1}^{T_i} L_{it}(f_k) \right\}$$

where K and S are the number of points of support for the distribution of f_i and ζ_i respectively and pr_{ks} is the probability mass at a point of the (f_i, ζ_i) distribution.

The computational burden in these models arises from having to solve the model at each iteration and each individual type (defined by the observable and unobservable characteristics) for all values of the state variables. If these are continuous (such as assets) they need to be discretised.

Macroeconomic shocks

The model allows for macroeconomic shocks through wages. In its simplest form there is just one type of human capital and the time effect on the wage reflects its value relative to the consumption good. In a richer setting there are different types of human capital with relative prices that vary. To allow for macro-shocks in the model we require a model that predicts forward prices as a function of current observables. In principle, this process will have to be estimated simultaneously with the model, because of the changes in labour force composition over time, which the model accounts for.

[33] See Ham and LaLonde (1996) and Meghir and Whitehouse (1997), for applications in dynamic transition models.

6.4.3. Why allow for savings

Allowing for savings is complicated both computationally and empirically. Allowing for a linear utility in consumption would eliminate the complications. So why should we get into all this trouble? The answer lies in the fact that individuals are risk averse and risk in not fully insurable. Modelling savings in this context is important for understanding a number of issues, including self-insurance for events such unemployment (Low 1999) and more importantly pensions and retirement. For example, to understand the policy impact of changes in pension arrangements we need to understand how such policies interact with savings. The extent to which public policies crowd out private savings can only be studied in a model that accounts for both. Similar issues will arise when studying the impact of policies such as taxes and tax credits. The complete labour supply effect cannot be understood if we do not know how savings behaviour will be affected. On the other hand, there are many questions relating to whether our fully rational forward looking model is a good enough representation of reality. Ignoring the issue is, however, not the way forward.

7. Summary and Conclusions

The study of labour supply is valuable from a number of perspectives. The analysis of the impact of taxes and benefits is perhaps the best established motivation. Within this field we are concerned with the impact of taxes on effort as well as the role of taxes and benefits in affecting education decisions; in this latter case labour supply is seen as an alternative to school or training for younger individuals. From a more dynamic perspective, focus recently has also shifted to labour supply as a way of responding to uncertainty and mitigating the amount of saving as well a for understanding the evolution of consumption over the life-cycle: without allowing for changes in labour supply, it is very difficult to rationalize the observed behaviour of consumption. Finally, the relationship of consumption and labour supply is critical for understanding issues to do with optimal taxes and the design of benefits - in work benefits in particular. For all the above reason, it is clearly important to understand the way labour supply is determined and how this relates to intertemporal considerations, such as savings

This chapter outlines a number of approaches to the study of labour supply beginning with the original static models and ending with dynamic ones that allow for savings and possibly intertemporal non-separabilities. Along the way we have discussed incorporating taxes and allowing for non-convex budget sets and the importance of unobserved heterogeneity. Allowing for the latter has proved particularly important empirically for estimating reliable models that are capable of fitting the data and accounting for the large persistence in labour supply patterns. Empirically labour supply analysis poses significant challenges not only because of the non-convexities but also because of the endogeneity of the main variables whose effect we are attempting to measure. High effort people, are likely to have invested more in human capital and thus have higher wages. They also accumulate more wealth making asset income potentially endogenous as well. Adding dynamics and allowing for non-convexities in the budget sets compounds the difficulties. We have attempted to provide a flavour of these difficulties and point to solutions. However, it is clear that there is more to be done. One relatively new and important area of research which we did not touch upon is modelling the entire career, starting with education choice and continuing with labour supply over the life-cycle. This is likely to be of key importance for understanding the longer term impact of public policy: programmes, such as tax credits, that encourage labour supply may well discourage education. Trading off these two margins of adjustment is important and requires reliable models for both. Thus considering the dynamics of labour supply and developing reliable modeling methods will continue to be of key importance for policy purposes.

Acknowledgements: We would like to thank Martin Browning, the editors and participants at the London Handbook meeting for helpful comments. This study is part of the research program of the ESRC Centre for the Microeconomic Analysis of Fiscal Policy at the IFS.

References

[1] Abbott, M. and Ashenfelter, O. (1976), "Labor Supply, Commodity Demand and the Allocation of Time", Review of Economic Studies, 42, 389-411.

[2] Ackum-Agell, S. and Meghir, C. (1995), "Male labor Supply in Sweden:

Are Incentive Effects Important?", Tax Reform Evaluation Report No.12, National Institute of Economic Research.

[3] Adda, J. C. Dustmann, C. Meghir and J.M. Robin (2006) Career progression and formal versus on-the-job training IFS Working paper W06/16.

[4] Altonji, J.G. (1982), "The intertemporal substitution model of labour market fluctuations: an empirical analysis", Review of Economic Studies, 49, 783-824.

[5] Altonji, J.G. (1986), "Intertemporal substitution in Labor supply: Evidence from Micro data", Journal of Political Economy, 94, June, Part II, S176-S215.

[6] Altug, S. and R. Miller, (1990), "Household choices in Equilibrium", Econometrica, 58, May 1990, 543-70.

[7] Altug, S. and R. Miller, (1998), "The Effect of Work Experience on Female Wages and Labor Supply Review of Economic Studies, 65, January, 45-85.

[8] Amemiya, T. (1973), "Regression Analysis When the Dependent Variable is Truncated Normal", Econometrica, 41, 997-1017.

[9] Angrist, J. (1991) "Grouped Data Estimation and Testing in Simple Labor Supply Models", Journal of Econometrics, Vol 47, pp. 243-265.

[10] Apps, Patricia, F. and Rees, Ray (1988), "Taxation and the Household", Journal of Public Economics, 35, 355-369.

[11] Apps, Patricia, F. and Rees, Ray (1993), "Labor Supply, Household Production and Intra-Family Welfare distribution", Australian National University discussion paper, January 1993.

[12] Apps, Patricia, F. and Rees, Ray (1997), "Collective Labor Supply and Household Production", Journal of Political Economy, 105, 178-190.

[13] Arellano, M. and Meghir, C. (1992), "Female Labour Supply and On-the-Job Search: An Empirical Model Estimated using Complementary Data Sets", Review of Economic Studies, 59(3),No. 200, 537-59.

[14] Arrufat, J.L. and Zabalza, A.(1986), "Female Labour Supply with Taxation, Random Preferences, and Optimization Errors", Econometrica, 54, 47-63.

[15] Ashenfelter, O. and J. Ham (1979), "Education, Unemployment and Earnings", Journal of Political Economy, 87, S99-S166.

[16] Ashenfelter, O. (1983), "Determining participation in Income-Tested Social programs", Journal of the American Statistical Society, 78, 517-525.

[17] Ashenfelter, O. and J. Ham (1979), "Education, Unemployment and Earnings", Journal of Political Economy, 87, S99-S166.

[18] Ashenfelter, O. and Heckman, J.J. (1974),"The Estimation of Income and Substitution Effects in a Model of Family Labor Supply", Econometrica, 42, 73-85.

[19] Attanasio, O, J. Banks, C. Meghir and G. Weber (1999) "Humps and Bumps in Lifetime Consumption," Journal of Business and Economic Statistics 17:1 (1999), 22–35.

[20] Attanasio, O. and S. Davies (1996) Relative Wage Movements and the Distribution of Consumption *The Journal of Political Economy*, Vol. 104, No. 6. (Dec., 1996), pp. 1227-1262.

[21] Attanasio, O. and H. Low (2002) "Estimating Euler Equations", IFS working paper WP02/06

[22] Attanasio, O.P. and MaCurdy, T. (1997), "Interactions in Family Labor Supply and their Implications for the Impact of EITC", Mimeo, Stanford University, February.

[23] Attanasio, Orazio P., and Guglielmo Weber (1993) , "Consumption Growth, the Interest Rate and Aggregation," Review of Economic Studies 60

[24] Attanasio, Orazio P., and Guglielmo Weber (1995) Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey Orazio P. Attanasio; Guglielmo Weber *The Journal of Political Economy*, Vol. 103, No. 6. (Dec., 1995), pp. 1121-1157.

[25] Becker, Gary S. (1965), "A theory of the Allocation of Time", Economic Journal, 75, 493-517..

[26] Blau, D. and P. Robins, (1988), "Child-Care Costs and Family Labor Supply", Review of Economics and Statistics, 70, 374-381.

[27] Blomquist, N.S. (1983), "The Effect of Income Taxation on the Labour Supply of Married Men in Sweden", Journal of Public Economics, 22, 169-97.

[28] Blomquist, S. and U. Hansson-Brusewitz (1990), "The Effect of Taxes on Male and Female Labour Supply in Sweden", Journal of Human Resources, 25, 317-357.

[29] Blomquist, S. (1996) "Estimation Methods for Male Labour Supply Functions: How to take account of nonlinear taxes", Journal of Econometrics, 70, 383-405.

[30] Blomquist, S. and W. Newey (2000), "Nonparametric Estimation of Labor Supply Functions Generated by Piece Wise Linear Budget Constraints", Working Paper 2000, Department of Economics, Uppsala University, Sweden.

[31] Blundell, R.W. (1986), "Econometric Approaches to the Specification of Life-Cycle Labour Supply and Commodity Demand Behaviour", Econometric Reviews, 6(1), 147-51.

[32] Blundell, R.W., Browning, M. and Meghir, C. (1994), "Consumer Demand and the Life-Cycle Allocation of Household Expenditure", Review of Economic Studies, 161. 57-80.

[33] Blundell, R., P. A. Chiappori, T. Magnac and C. Meghir (2006),"Collective Labor Supply and Participation", forthcoming *Review of Economic Studies.*

[34] Blundell, R., P. A. Chiappori, C. Meghir (2005),"Collective Labour Supply with Children ", *Journal of Political Economy*, Volume 113, Number 6, 1277-1306, December 2005.

[35] Blundell, R.W., Duncan, A. and Meghir, C. (1992), "Taxation and Empirical Labour Supply Models: Lone Parents in the UK", Economic Journal, 102, 265-78.

[36] Blundell, R.W., Duncan, A. and Meghir, C. (1998), "Estimating Labour Supply Responses using Tax Policy Reforms", Econometrica, 66, 827-861.

[37] Blundell, R.W. and MaCurdy, T. (1999), "Labor Supply: A Review of Alternative Approaches", in Ashenfelter, O. and Card, D. (eds), *Handbook of Labor Economics*, North Holland.)

[38] Blundell, R.W. and Meghir, C. (1986), "Selection Criteria for a Microeconometric Model of Labour Supply", Journal of Applied Econometrics, 1, 55-81

[39] Blundell, R.W. and Meghir, C. (1987), "Bivariate Alternatives to the Tobit Model", Journal of Econometrics, 34, 179-200.

[40] Blundell, R.W., Meghir, C., Symons, E. and Walker, I. (1988), "Labour Supply Specification and the evaluation of tax reforms", Journal of Public Economics, 36, 23-52.

[41] Blundell, R., Meghir, C. and Neves, P. (1993), "Labour Supply: An Intertemporal Substitution", Journal of Econometrics, 59, 137-60.

[42] Blundell, R.W. and Walker, I. (1982), "Modelling the Joint Determination of Household Labour Supplies and Commodity Demands", Economic Journal, 92, 58- 74.

[43] Blundell, R.W. and Walker, I. (1986), "A Life Cycle Consistent Empirical Model of Labour Supply using Cross Section Data", Review of Economic Studies, 53, 539-58.

[44] Bound, J. Jaeger, D. A. and Baker, R.M. (1995), "Problems with Instrumental Variables Estimation when the Correlation between the Instrument and the Endogenous Explanatory Variable is Weak, *Journal of the American Statistical Association*, 90, 443-450.

[45] Browning, M. (1998),"M-Labour Supplies and Commodity Demands", mimeo, University of Copenhagen.

[46] Browning, M., Deaton, A. and M. Irish (1985), "A Profitable Approach to Labor Supply and Commodity Demand over the Life Cycle", Econometrica 53, May 1985, 503-43.

[47] Browning, M. and C. Meghir (1991), "Testing for Separability between Goods and Leisure using conditional Demand Systems", Econometrica, 59, July, 925-52.

[48] Card, D. (1994), "Intertemporal Labour Supply: An Assessment", in C.Sims (ed), Advances in Econometrics, Cambridge University Press.

[49] Card, D. and Robins, P.K. (1998), "Do Financial Incentives Encourage Welfare Participants to Work", Research in Labor Economics.

[50] Carroll, C (1997), "Death to the Log-Linearized Euler Equation! (And Very Poor Health to the Second Order Approximation," NBER working paper 6298 (1997b).

[51] Carroll, Christopher D., and Andrew A. Sawmwick, "How Important Is Precautionary Saving?" Review of Economics and Statistics 80:3 (1998), 410–419.

[52] Chamberlain G. (1984)

[53] Chiappori, Pierre-Andre, (1988), "Rational Household Labor supply", Econometrica, 56, 63-90.

[54] Chiappori, Pierre-Andre, (1992), "Collective Labor Supply", Journal of Public Economics, 437-467.

[55] Chiappori, Pierre-Andre, (1997),"Introducing Household Production in Collective Models of Labor Supply" Journal of Political Economy, 105, 191-209.

[56] Cochrane, J. A Simple Test of Consumption Insurance *The Journal of Political Economy*, Vol. 99, No. 5. (Oct., 1991), pp. 957-976.

[57] Cogan, J. (1980), "Labor Supply with Costs of Labor Market Entry", in Female Labor Supply: Theory and Estimation, ed. By James Smith. Princeton:Princeton University Press, 327-364.

[58] Cogan, J.F. (1981), "Fixed costs and labor supply", Econometrica, 49, 945-64.

[59] Eckstein, Z. and Wolpin, K. (1989), "Dynamic Labour Force Participation of Married Women and Endogenous Work Experience", Review of Economic Studies, 56, 375- 90.

[60] Eissa, N. and Liebman, J.(1995), "Labor Supply Responses to The Earned Income Tax Credit", NBER working paper no. 5158.

[61] Fortin, Bernard and Lacroix, Guy (1997), "A Test of Neoclassical and Collective Models of Household Labor Supply", Economic Journal, 107, 933-955.

[62] Fraker, T. and R. Moffitt, (1988), "The Effect of Food Stamps on Labor Supply: A Bivariate Selection Model", Journal of Public Economics, 35, 25-56.

[63] Gorman, W.M. (1959), "Separable Utility and Aggregation", *Econometrica,* 21, 63-80.

[64] Gorman, W.M. (1968), “The Structure of Utility Functions”, The Review of Economic Studies, 35, 369-390.

[65] Gosling, A. S. Machin and C. Meghir (2000) “The Changing Distribution of Male wages in the UK”, Review of Economic Studies,

[66] Gronau, R. (1974),“Wage Comparisons, A Selectivity Bias”, *Journal of Political Economy*, 82(6), 1119-1144.

[67] Ham, J. (1986a), “Testing whether unemployment represents life-cycle labor supply behaviour”, Review of Economic Studies, LIII, 559-78.

[68] Ham, John and Robert LaLonde (1996) *Econometrica*

[69] Hausman, J. (1980), “The effect of Wages, Taxes and Fixed costs on Women”s Labor Force participation”, Journal of Pubic Economics, 14, 161-94.

[70] Hausman , J. (1981),“Labor Supply. How Taxes Affect Economic Behavior”, in (eds) H. Aaron and J. Pechman, Washington, D.C. Brookings Institution.

[71] Hausman, J. (1985a), “The Econometrics of Noninear Budget Sets”, Econometrica, 53, 1255-82.

[72] Hausman, J. (1985b), “Taxes and Labor Supply”,In Handbook of Public Economics, vol 1., (eds) A. Auerbach and M. Feldstein, Amsterdam: North Holland.

[73] Hausman, J. and Ruud, P. (1986), “Family Labor Supply with Taxes”, American Economic Review (Papers and Proceedings), 74, 242-248.

[74] Heckman, J.J. (1974a), “Shadow prices, market wages and labor supply”, Econometrica, 42, 679-94.

[75] Heckman, J.J., (1974b), “Life-Cycle Consumption and Labor Supply: An Explanation of the Relationship between Income and Consumption over the Life Cycle”, American Economic Review, 64, March, 188-94.

[76] Heckman, J.J., (1974c), "Effects of Child-Care Programs on Women"s Work Effort", Journal of Political Economy, 82(2), S136-S163.

[77] Heckman, J.J., (1976), "Life-Cycle Model of Earnings, Learning and Consumption" Journal of Political Economy, 84, S11-S44.

[78] Heckman, J.J., (1978), "Dummy Endogenous Variables in a Simultaneous Equation System", Econometrica, 46, 931-59.

[79] Heckman, J. J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System", *Econometrica*, 46, 931-60.

[80] Heckman, J.J. (1979), "Sample Selection Bias as a Specification Error', *Econometrica*, 47, 153-161.

[81] Heckman, J.J. and MaCurdy, T.E. (1980), "A Life-cycle Model of Female Labour Supply", Review of Economic Studies, 47, 47-74.

[82] Heckman, J.J. (1993), "What Has Been Learned about Labor Supply in the Past Twenty Years", A.E.R. Papers and Proceedings, 83, 116-21.

[83] Heckman, J.J. (1995) "Instrumental variables: A cautionary Tale", mimeo

[84] Heckman, J.J. and Robb, R. (1985), "Alternative Methods for Evaluating the Impact of Interventions', in Heckman and Singer (eds), Longitudinal Analysis of Labor Market Data, Econometric Society Monograph 10, Cambridge University Press.

[85] Honore, Bo E. (1992) "Trimmed Lad and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects" *Econometrica* Vol. 60, No. 3 (May, 1992), pp. 533-565

[86] Hotz, V.J., Kydland, F.E. and G.L. Sedlacek, (1988), "Intertemporal Substitution and Labor Supply,", Econometrica , 56, March 1988, 335-60

[87] Imbens, G. and J. Angrist (1994) Identification and Estimation of Local Average Treatment Effects (in Notes and Comments) *Econometrica*, Vol. 62, No. 2. (Mar., 1994), pp. 467-475.

[88] Johnson, T.R. and J.H. Pencavel (1984), "Dynamic Hours of Work Functions for Husbands, Wives and Single Females", Econometrica , 52, 363-389.

[89] Keane, M.P. and Moffitt, R. (1998), "A Structural Model of Multiple Welfare Program Participation and Labor Supply", International Economic Review.

[90] Keane and Wolpin (1997) The Career Decisions of Young Men, *The Journal of Political Economy,* Vol. 105, No. 3. (Jun., 1997), pp. 473-522

[91] Killingsworth, M. (1983), Labor Supply. Cambridge: Cambridge University Press.

[92] Killingsworth, M. and J. Heckman (1986), "Female Labor Supply: A Survey" in Handbook of Labor Economics, Vol.1, ed. By O. Ashenfelter and R. Layard. Amsterdam: North-Holland, 103-204.

[93] Kooreman, P. and Kapteyn, A. (1990), "On the Empirical Implementation of Some Game Theoretic Models of Household Labour Supply", Journal of Human Resources, 25, 584-98.

[94] Lee, L. F. (1984), "Tests for the Bivariate Normal Distribution in Econometric Models with Selectivity', *Econometrica*, 52, 4, 843-864.

[95] Low, Hamish (1999) Self-insurance and unemployment benefit in a life-cycle model of labour supply and savings IFS Working Paper 99/24.

[96] Lundberg, S. (1988), "Labor Supply of Husbands and Wives", Review of Economics and Statistics, 70, 224-35.

[97] Ludvigson, Sydney and C. Paxson. "Approximation Bias in Linearized Euler Equations" , Review of Economics and Statistics, 83(2), pp. 242-256, May, 2001.

[98] MaCurdy, T.E. (1981), “An Empirical Model of Labour Supply in a Life-Cycle Setting”, Journal of Political Economy, 89, 1059-85.

[99] MaCurdy, T.E. (1983), “A simple scheme for estimating an intertemporal model of labor supply and consumption in the presence of taxes and uncertainty”, International Economic Review, 24, 265-89.

[100] MaCurdy, T.E. (1985), “Interpreting Empirical Models of Labour Supply in an Intertemporal Framework with Uncertainty”, in J.J. Heckman and B. Singer (eds.), Longitudinal Analysis of Labour Market Data, Cambridge University Press, ch. 3.

[101] MaCurdy, T.E. (1992), “Work Disincentive Effects of Taxes: A Re-examination of Some Evidence”, American Economic Review, 82, 243-49.

[102] MaCurdy, T., Green, D. and Paarsch, H. (1990), “Assessing Empirical Approaches for Analyzing Taxes and Labour Supply”, Journal of Human Resources, 25, 415-90.

[103] Manski, C and D. McFadden (1981) Structural Analysis of Discrete Datat with Econometric Applications, MIT Press: Cambridge, MA, 1981.

[104] McElroy, M.B. (1981), “Empirical Results from Estimates of Joint Labor Supply functions of Husbands and Wives”, in Research in Labor Economics, Vol. 4. Ed. By R.G. Ehrenberg, Greenwich, Connecticut: JAI Press, 53-64.

[105] McElroy, M. (1990), “The Empirical Content of Nash-Bargained Household Behavior”, Journal of Human Resources, 25, 559-583. Moffit, R. (1983), ”An Economic model of Welfare Stigma”, American Economic Review, 73, 1023-1035. Amsterdam, 1978

[106] Meghir, C. and G. Weber (1996) “Intertemporal non-separability or liquidity constraints: An empirical analysis using as US consumption panel”, *Econometrica*

[107] Meghir, C. and E. Whitehouse (1997) *Journal of Econometrics*

[108] Moffit, R. (1986), "The Econometrics of Piecewise-Linear Budget Constraints: Survey and Exposition of the Maximum Likelihood Method", Journal of Business and Economic Statistics, 4, 317-27.

[109] Moffitt, R. (1992), "Estimating the Value of an In-Kind Transfer: The Case of Food Stamps", Econometrica, 57 (March), 385-409.

[110] Moffitt, R. (1992b), "Incentive Effects of the US Welfare System: A Review", Journal of Economic Literature, 15, 1-161.

[111] Moffit, R. (1993), "Identification and Estimation of dynamic Models with a Time series of Repeated Cross Sections", in Journal of Econometrics, 59, 99-124.

[112] Mroz, T.A. (1987), "The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions", Econometrica, 55, 765-800.

[113] Neary, P. and K. Roberts (1980) European Economic Review

[114] Newey, W.K., Powell, J.L. and Walker, J.R. (1990), "Semiparametric Estimation of Selection Models: Some Empirical Results", American Economic Review, 80(2), 324- 8.

[115] Pencavel, J. (1986), "Labor Supply of Men: A Survey", in Handbook of Labor Economics, Vol. 1, ed. By O. Ashenfelter and R. Layard. Amsterdam: North-Holland, 3-102.

[116] Ransom, M.R. (1987), "An Empirical Model of Discrete and Continuous Choice in Family Labor Supply", The Review of Economics and Statistics, 59, 465-472.

[117] Rust, J and C. Phelan (1997) "How Social security and Medicare affect retirement in a World of Incomplete Markets *Econometrica* 65(4) 781-831

[118] Shaw, K. (1989), "Life-cycle Labour Supply with Human Capital Accumulation", International Economic Review, 30(2), 431-57.

[119] Smith, R.J. and Blundell, R.W. (1986), "An Exogeneity Test for the Simultaneous Equation Tobit Model", Econometrica, 54, 679-85.

[120] Stern, N. (1986), "On the specification of labor supply functions" in R.W. Blundell and I. Walker (eds.), Unemployment, Search and Labour Supply, Cambridge University Press, 121-42.

[121] Van Soest, A(1995), "Structural Models of Family Labor Supply: A Discrete Choice Approach", Journal of Human Resources, 30, 63-88

Wagner, Burkhuser and Behringer (1993),

[122] Wales, T.J. and A. Wodland, A. (1976), "Estimation of Household Utility Functions and Labor Supply Response", International Economic Review, 17, 69-80.

[123] Zabalza, A. (1983), "The CES Utility Function, Nonlinear Budget Constraints and Labor Supply", Economic Journal, 93, 312-330.

8. Appendix

This appendix reviews general formulations for likelihood functions applicable to econometric models involving any combination of five types of endogenous variables: (1) discrete, (2) continuous, (3) censored, (4) truncated and (5) continuous/discrete. The subsequent discussion opens with an overview of the statistical framework considered here. It next considers increasingly complex variants of this framework, starting with models incorporating just discrete variables, adding in continuous variables, and then including endogenous variables of a combined continuous-discrete character. The analysis proceeds to cover specifications appropriate when one does not observe all states of the world but instead only knows whether various combination of states has occurred. The concluding subsection presents alternative representations of likelihood functions commonly found in the literature comparable to the specifications presented here, as well as presenting simple extensions of specifications that allow for dependence on exogenous variables.

8.1. Overview of Statistical Framework

The basic idea at the foundation of econometric models characterizing distributions of discrete-continuous variables relies on the notion that all endogenous quantities depend on the values of an underlying set of continuously-distributed random variables. Specify these underlying variables by the vector U, assumed to include r linearly-independent components. This $rx1$ vector possesses the joint density function:

$$\varphi(U) \quad \text{for} \quad U \in \Omega \tag{A.1}$$

where the set Ω designates the sample space or domain of the random variables U.

In this model, m states of the world can occur. The discrete random variable δ_i signifies whether state i happens, with $\delta_i = 1$ indicating realization of state i and $\delta_i = 0$ implying that some state other than i occurred. The value of δ_i depends on where U falls in its sample space; specifically,

$$\delta_i = \begin{cases} 1 & \text{if } U \in \Omega_i \\ 0 & \text{otherwise} \end{cases} \tag{A.2}$$

where the set Ω_i represents a nontrivial subset of the entire sample space Ω. Without loss of generality, assume that the sets Ω_i for $i = 1, ..., m$ are mutually exclusive and exhaustive, meaning $\bigcup_{i=1}^{m} \Omega_i = \Omega$ and the sets $\Omega_i \cap \Omega_j = \emptyset$ for $i \neq j$ (i.e., the sets Ω_i and Ω_j are disjoint).

In association with state i, there exists n_i continuously distributed random variables designated Y_{ji}, $j = 1, ..., n_i$. The following equations determine the values of these continuous variables:

$$Y_{ji} = g_{ji}(U) \ . \tag{A.3}$$

Stacking these individual random variables into a vector yields

$$Y_i = \begin{pmatrix} Y_{1i} \\ \vdots \\ Y_{n_i i} \end{pmatrix} = \begin{pmatrix} g_{1i} \\ \vdots \\ g_{n_i i} \end{pmatrix} = g_i \ . \tag{A.4}$$

To avoid introduction of redundant or ill-defined Y_{ji}'s, assume there exists an inverse of g_i such that

$$U_{(i)} = g_i^{-1}(Y_i, U_i) \tag{A.5}$$

for some subvector $U_{(i)}$ comprised of any n_i components of U.[34] The subvector U_i includes those elements of U not included in $U_{(i)}$. Designate Φ_i as the domain of (Y_i, U_i) and Θ_i as the domain of Y_i .

Another interesting class of random variables consists of quantities that take a fixed single value in some states and a continuous set of values in others. Denote these discrete/continuous variables as Z_{ji}, with the index i signalling the state realized and $j = 1, ..., k_i$ signifying the particular Z realized in this state. The value of Z_{ji} follows a rule of the form:

$$Z_{ji} = \begin{cases} Y_{ji} & \text{for } \ j \in K_{ci} \\ Z_{ji}^{*} & \text{for } \ j \in K_{di} \end{cases} \tag{A.6}$$

where the set K_{ci} indexes those Z_{ji} taking the form of a continuous variable in state i, and the set K_{di} identifies those Z_{ji} equaling a constant value Z_{ji}^{*} in state i. Define Z_i as the vector containing the Z_{ji} $j = 1, ..., k_i$ as elements analogous to Y_i specified in (A.4).

Finally, form all the unique variables appearing in any of the Y_i's into the vector Y, assumed to be of dimension $nx1$, and all the variables making up the Z_i's into the vector Z, assumed to be of dimension $kx1$. For any event $\delta_i = 1, Y$ consists of two sets of components: the vector Y_i incorporating all the continuous random variables registering in state i, and $Y_{(i)}$ made up of all other continuous variables unobserved in this state but seen in some other state $j \neq i$. Similarly, Z consists of the vector Z_i and $Z_{(i)}$ defined analogously. In some states i, all of the elements of Y and Z may be observed, and in others none may be.

The subsequent discussion characterizes formulations of conditional and unconditional likelihood functions associated with Y, Z and combinations of the δ_i's.

[34] "Assuming existence of the inverse of g_i in () is not as restrictive as one might first surmise. If an inverse does not exist on set Φ_i , then one can replace Φ_i with a further segment of this set inverses defined on each of these smaller sets. The subsequent analysis can then be carried out for this expanded decomposition of Φ."

As briefly noted at the end of this appendix, one can readily introduce the presence of exogenous variables both in specifying the distribution of U and in defining the regions of definition of δ_i. An exogenous variable in this analysis must be observed in all states; otherwise, this variable must be include as a component of Y or Z.

8.2. Discrete Variables: All and Combinations of States

Initially consider empirical frameworks in which one observes only discrete variables whose outcomes register the realization of m distinct regimes determined by the relative values of U.

A common formulation specifies that a researcher sees exactly which state m occurs, implying that one observes all individual δ_i, $i = 1, ..., m$. From (A.2) we see that the probability that $\delta_i = 1$ equals:

$$\begin{aligned} P\ (\delta_i = 1) &= P\left(U \in \Omega_i\right) \\ &= \int \cdots \int_{\Omega_i} \varphi\left(U\right) dU \\ &\equiv \int_{\Omega_i} \varphi\left(U\right) dU\ . \end{aligned} \tag{A.7}$$

The notation $\int \cdots \int_{\Omega_i}$ denotes integration over the set Ω_i, which the third line of this equation expresses in the shorthand notation $\int_{\Omega_i}$. The joint distribution of the δ_i's takes the form:

$$\begin{aligned} P\left(\delta_1, ..., \delta_m\right) &= \prod_{i\ =1}^{m} \left[P(\delta_i = 1)\right]^{\delta_i} \\ &= \prod_{i \in M} \left[P(\delta_i = 1)\right]^{\delta_i} \end{aligned} \tag{A.8}$$

In the second line of this equation, the notation $M = \{i : i = 1, ..., m\}$ refers to the set of all possible states i.

In other formulations, a researcher does not observe or chooses to ignore each state individually. Instead, one accounts for only whether some combination of states has been realized. More specifically, suppose one knows that at least one $\delta_i = 1$ when $i \in M_t \subset M$, but one does not account for which particular δ_i in this group actually occurred. So,

$$\text{if } i \in M_t \text{ , then } \quad \overline{\delta}_t \equiv \sum_{i \in M_t} \delta_i = 1 \quad \text{; otherwise, } \ \overline{\delta}_t = 0 \ . \tag{A.9}$$

The sets M_t, $t = 1, ..., \tau$, are mutually exclusive and exhaustive (i.e., $\bigcup_{t=1}^{\tau} M_i = M$ and $M_t \cap M_j = \emptyset$ for $t \neq j$). The probability of the occurrence of group state t equals

$$P\left(\overline{\delta}_t = 1\right) = \sum_{i \in M_t} P\left(\delta_i = 1\right) \ . \tag{A.10}$$

The joint distribution of the $\overline{\delta}_t$'s takes the form:

$$P\left(\overline{\delta}_1, ..., \overline{\delta}_\tau\right) = \prod_{t \in T} \ \left[P(\overline{\delta}_t = 1)\right]^{\overline{\delta}_t} \tag{A.11}$$

where the notation $T = \{t : t = 1, ..., \tau\}$ refers to the set of all possible "group" states t.

8.3. Continuous Variables: All States Observed

Consider those models in which one observes each individual δ_i along with vectors Y_i of continuously distributed random variables for states $i \in M_y \subseteq M$. Conditional on occurrence of a state, the components of Y_i may either be truncated or censored. The truncated elements of Y_i refer to those that lie in a strict subset of their overall domain given realization of the selection mechanism $U \in \Omega_i$ (or, equivalently, $(Y_i, U_i) \in \Phi_i$). The censored elements consist of those that instead range over their entire domain. The set $\Theta = \bigcup_{i=1}^{m} \Theta_i$ defines the sample space of Y. So, if Y_i includes truncated components, then $\Theta_i \subset \Theta$.

The first step in formulating specifications for the distributions of the Y_i's involves recognizing that the density of underlying random variables U conditional on the event $\delta_i = 1$ takes the form:

$$\varphi\left(U \mid \delta_i = 1\right) = \frac{\varphi\left(U\right)}{P\left(\delta_i = 1\right)} \tag{A.12}$$

where relationship (A.7) gives the formula for $P\left(\delta_i = 1\right)$. An alternative expression (A.7) is given by:

$$\begin{aligned} P\left(\delta_i = 1\right) &= P\left(U \in \Omega_i\right) \\ &= P\left(\left(Y_i, U_i\right) \in \Phi_i\right) \\ &= \int_{\Phi_i} h_i\left(Y_i, U_i\right)\ dY_i dU_i \end{aligned} \tag{A.13}$$

where the set $\Phi = \bigcup\limits_{i=1}^{n} \Phi_i$ defines the domain of $(Y, U_1, ...U_n)$.

Application of a conventional change-in-variables formula exploiting relations (A.3) and (A.5) yields the following specification for the density of Y_i conditional on δ_i:

$$f\left(Y_i \mid \delta_i = 1\right) = \frac{\int_{\Phi_{i|Y_i}} h_i\left(Y_i, U_i\right) dU_i}{P\left(\delta_i = 1\right)} \qquad \text{for} \qquad Y_i \in \Theta_i \tag{A.14}$$

where

$$h_i\left(Y_i, U_i\right) = J_i\ \varphi\left(g_i^{-1}\left(Y_i, U_i\right), U_i\right) \qquad \text{with} \qquad J_i = \left|\frac{\partial g_i^{-1}}{\partial Y_i{}'}\right|^{+}\ , \tag{A.15}$$

and the notation $\int_{\Phi_{i|Y_i}}$ denotes integration of U_i over the set

$$\Phi_{i|Y_i} = \left\{U_i : \left(Y_i, U_i\right) \in \Phi_i\right\}\ . \tag{A.16}$$

The term J_i in (A.*) represents the Jacobian of the transformation associated with (A.5) (i.e., J_i is the absolute value of the determinant of the matrix of partial derivatives $\frac{\partial g_i^{-1}}{\partial Y_i{}'}$). One can express the domain of Y_i as

$$\Theta_i \ = \ \Theta_{i\cdot\ Y_i} = \left\{Y_i : U_i \in \Phi_{i|Y_i}\right\}\ , \tag{A.17}$$

where the notation $\Theta_{i \cdot\ Y_i}$ simply signifies that this set is a suspace of Y_i.

A compact expression for the conditional density of Y is

$$f(Y \mid \delta_i\ ,\ i \in M_y) = \prod_{i \in M_y} \left[f(Y_i \mid \delta_i = 1)\right]^{\delta_i} \tag{A.18}$$

where as defined above M_y designates the set of states in which one observes at least one element of Y. An alternative representation for this conditional density takes the form:

$$f(Y \mid \delta_1, ..., \delta_m) = \prod_{i \in M_y} \left[f(Y_i \mid \delta_i = 1)\right]^{\delta_i} \prod_{i \in M_y^c} [1]^{\delta_i} \quad , \tag{A.19}$$

where the set M_y^c denotes the complement of M_y with respect to M. Realizations of $i \in M_y^c$ mean that all elements of Y are either undefined or unobserved.

The joint density of Y and $\delta_1, ..., \delta_m$ is the product of the conditional density of Y given by (A.19) and the joint probability of $\delta_1, ..., \delta_m$ given by (A.8) yielding:

$$\begin{aligned} f(Y, \delta_1, ..., \delta_m) &= \prod_{i \in M_y} \left[f(Y_i \mid \delta_i = 1)\ \ P(\delta_i = 1)\right]^{\delta_i} \prod_{i \in M_y^c} \left[P(\delta_i = 1)\right]^{\delta_i} \qquad \text{(A.20)} \\ &= \prod_{i \in M_y} \left[\int_{\Phi_{i|Y_i}} h_i(Y_i, U_i)\, dU_i\right]^{\delta_i} \prod_{i \in M_y^c} \left[\int_{\Omega_i} \varphi(U)\, dU\right]^{\delta_i} . \end{aligned}$$

The second line of this expression follows by substituting relationships from (A.7) and (A.14).

8.4. Discrete/Continuous Variables: All States Observed

Consider models in which one observes individual δ_i along with the vectors Z_i comprised of discrete/continuous random variables for states $i \in M_z \subseteq M$. The components included in Z_i are either distributed continuously or equal to constants according to the following rule:

$$Z_i = \begin{pmatrix} Z_{ci} \\ Z_{di} \end{pmatrix} = \begin{pmatrix} Y_i \\ Z_{di}^* \end{pmatrix} \quad \text{for } \ i \in M_z\ . \tag{A.21}$$

Inspection of (A.6) reveals that those individual Z_{ji} for $j \in K_{ci}$ make up the elements of the vector Z_{ci}; and those Z_{ji} for $j \in K_{di}$ form the vector Z_{di}. The set M_z comprises all states in which any component of Z is realized.

For states $i \in M_y$, one can express the distribution of Z_i conditional on $\delta_i = 1$ as:

$$\begin{aligned} f(Z_i \mid \delta_i = 1) &= f(Z_{ci}, Z_{di} \mid \delta_i = 1) \\ &= f(Z_{ci} \mid Z_{di}\ ,\ \delta_i = 1)\ P\,(Z_{di} \mid \delta_i = 1) \\ &= f(Y_i \mid Z^*_{di}\ ,\ \delta_i = 1) \end{aligned} \tag{A.22}$$

where the third line follows from

$$P\,(Z_{di} = Z^*_{di} \mid \delta_i = 1) = 1 \quad . \tag{A.23}$$

Formally, the argument Z^*_{di} in $f(Y_i \mid Z^*_{di}\ ,\ \delta_i = 1)$ is redundant since the event $\delta_i = 1$ already implies $Z_{di} = Z^*_{di}$; the argument is included merely to remind the reader that the density appearing the last row of (A.22) typically depends on Z^*_{di}. A compact expression for the conditional density of Z is

$$f\,(Z \mid \delta_i\ ,\ i \in M_z) = \prod_{i \in M_y} [\ f\ \ (Y_i \mid Z^*_{di}\ ,\ \delta_i = 1)]^{\delta_i} \prod_{i \in M_d} [1]^{\delta_i} \quad . \tag{A.24}$$

Realizations of $i \in M_y$ mean that some of the elements of Z_i are continuously distributed, whereas occurrence of $i \in M_d$ imply that all elements of Z_i are discrete. One can write an alternative representation for this conditional density as:

$$\begin{aligned} f\,(Z \mid \delta_1, ..., \delta_m) &= \prod_{i \in M_y} [f\,(Y_i \mid Z^*_{di}\ ,\ \delta_i = 1)]^{\delta_i} \prod_{i \in M_d} [1]^{\delta_i} \prod_{i \in M_u} [1]^{\delta_i} \\ &= \prod_{i \in M_y} [f\,(Y_i \mid Z^*_{di}\ ,\ \delta_i = 1)]^{\delta_i} \prod_{i \in M_d \cup M_{yu}} [1]^{\delta_i} . \end{aligned} \tag{A.25}$$

Realizations of $i \in M_u$ mean that all components of Z are either undefined or unknown.

The joint density of Z and $\delta_1, ..., \delta_m$ is the product of the conditional density of Z given by (A.25) and the joint probability of $\delta_1, ..., \delta_m$ given by (A.8) yielding:

$$\begin{aligned} f\left(Z, \delta_1, ..., \delta_m\right) &= \prod_{i \in M_y} \left[f\left(Y_i \mid Z^*_{di}\ ,\ \delta_i = 1\right)\ P(\delta_i = 1)\right]^{\delta_i} \prod_{i \in M_d \cup M_u} \left[P(\delta_i = 1)\right]^{\delta_i} \quad \text{(A.26)} \\ &= \prod_{i \in M_y} \left[\int_{\Phi_{i|Y_i}\ Z^*_{di}} h_i\left(Y_i, U_i\right) dU_i\right]^{\delta_i} \prod_{i \in M_d \cup M_u} \left[\int_{\Omega_i} \varphi\left(U\right) dU\right]^{\delta_i} . \end{aligned}$$

The second line of this expression follows by substituting relationships from (A.7) and (A.14), where the notation $\Phi_{i|Y_i\ Z^*_{di}}$ still refers to the set $\Phi_{i|Y_i}$ defined by (A.16) with emphasis added to indicate that this set also depends on Z^*_{di}.

8.5. Discrete/Continuous Variables: Combinations of States

An important category of models involves characterizing the distribution of continuous and discrete/continuous variables when one either observes or chooses to distinguish the occurrence of groups rather than individual states. Define the relevant groups of states by the $\overline{\delta}_t$'s specified in (A.9) for $t \in T$ as outlined in Section A.2.

Consider the distribution of the continuous random variable

$$Y_t = \sum_{i \in M_t} \delta_i Y_i \quad . \tag{A.27}$$

Relation (A.27) implicitly assumes that each Y_i is defined and of comparable dimension for $i \in M_t$. Application of the law of iterated expectations yields the following density for Y_t conditional on $\overline{\delta}_t = 1$:

$$\begin{aligned} f\ \left(Y_t \mid \overline{\delta}_t = 1\right) &= \sum_{i \in M_t} f\ \left(Y_t \mid \delta_i = 1, \overline{\delta}_t = 1\right)\ P\left(\delta_i = 1 \mid \overline{\delta}_t = 1\right) \quad \text{(A.28)} \\ &= \sum_{i \in M_t} f\ \left(Y_t \mid \delta_i = 1\right)\ P\left(\delta_i = 1 \mid \overline{\delta}_t = 1\right) \\ &= \sum_{i \in M_t} f\ \left(Y_t \mid \delta_i = 1\right) \frac{P\left(\delta_i = 1\right)}{P\left(\overline{\delta}_t = 1\right)} \end{aligned}$$

The latter two lines of this relationship follow from the assumptions that the individual states $\delta_i = 1$ for $i \in M_t$ making up the event $\overline{\delta}_t = 1$ are mutually exhaustive and exclusive.

Discrete/continuous variables are realized according to the following rule:

$$Z_t = \begin{pmatrix} Z_{ct} \\ Z_{dt} \end{pmatrix} = \begin{pmatrix} Y_t \\ Z_{dt}^* \end{pmatrix} \quad \text{for } \; t \in T_z \; . \tag{A.28}$$

The set $T_z = T_y \cup T_d$ comprises all group states in which any component of Z is realized. The set T_y includes those group states t in which Z_t incorporates the continuously-distributed vector Y_t specified by (A.27); and the set T_d includes those group states wherein all the components of Z_t equals constant values.[35]

For group states $t \in T_c$, the distribution of Z_t conditional on $\overline{\delta}_t = 1$ takes the form:

$$\begin{aligned} f(Z_t \mid \overline{\delta}_t = 1) &= f(Z_{ct}, Z_{dt} \mid \overline{\delta}_t = 1) \\ &= f(Z_{ct} \mid Z_{dt} \; , \; \overline{\delta}_t = 1) \; P\left(Z_{dt} \mid \overline{\delta}_t = 1\right) \\ &= f(Y_i \mid Z_{dt}^* \; , \; \overline{\delta}_t = 1) \quad , \end{aligned} \tag{A.29}$$

where this latter expression exploits the relationship:

$$P\left(Z_{dt} = Z_{dt}^* \mid \overline{\delta}_t = 1\right) = 1 \quad \text{for} \quad t \in T_d \; . \tag{A.30}$$

Analogous to (A.25), a compact expression for the conditional density of Z is

$$f\left(Z \mid \overline{\delta}_1, ..., \overline{\delta}_\tau\right) = \prod_{t \in T_y} \left[f\left(Y_t \mid Z_{dt}^* \; , \; \overline{\delta}_t = 1\right)\right]^{\overline{\delta}_t} \prod_{t \in T_d \cup T_u} [1]^{\overline{\delta}_t} \tag{A.31}$$

where the set T_u includes those state groups in which no Z_{jt} are either undefined or unknown.

[35] "For notational simplicity, the specification of the values of Z_t when $t \in T_d$ presumes that Z_t^* is common across the individual states $i \in M_t$ making up group state t. One can instead replace the common value Z_t^* by a set $\{Z_t^*\}$ consisting of several discrete values at the expense of introducing some complexity in specifying likelihood functions."

Multiplying the conditional density (A.25) by the joint probability of the events $\overline{\delta}_1$,...,$\overline{\delta}_\tau$ given by (A.8) generates the following joint density for Z and the $\overline{\delta}_t$'s:

$$\begin{aligned} f\left(Z,\overline{\delta}_1,...,\overline{\delta}_\tau\right) &= \prod_{t\in T_y}\left[f\left(Y_t \mid Z^*_{dt}\ ,\ \overline{\delta}_t=1\right)P(\overline{\delta}_t=1)\right]^{\overline{\delta}_t}\prod_{t\in T_d\cup T_u}\left[P(\overline{\delta}_t=1)\right]^{\overline{\delta}_t} \quad \text{(A.32)} \\ &= \prod_{t\in T_y}\left[\sum_{i\in M_t}\int_{\Phi_{i|Y_t}\ Z^*_{dt}} h_i\left(Y_t,U_i\right)\ dU_i\right]^{\overline{\delta}_t}\prod_{t\in T_d\cup T_u}\left[\sum_{i\in M_t}\int_{\Omega_i}\varphi\left(U\right)dU\right]^{\overline{\delta}_t} \end{aligned}$$

The second line of this expression follows from substitution of relationships from (A.7), (A.10), (A.14), and (A.28).

8.6. Accounting for Unobserved and Exogenous Variables

Specification (A.32) presents a general formulation for likelihood functions incorporating discrete, continuous and discrete/continuous variables. One often sees alternative representations of this specification in the literature that may at first not appear as a special case of (A.32).

One such representation defines a set of continuous or discrete/continuous variables Z that are then presumed to be unobserved and, therefore, must be eliminated as arguments of the f 's in (A.32). In particular, suppose Z consists of two components $Z_t' = (Z_{1t}'\ , Z_{2t}')$ where the variables Z_{1t}' are observed and those included in Z_{2t}' are not. Correspondingly, decompose $Y_t' = (Y_{1t}'\ , Y_{2t}')$ and $Z_{dt}' = (Z_{1dt}'\ , Z_{2dt}')$, with the random variables Y_{2t}' and Z_{2dt}' unobserved.

Integrating (or summing) the joint likelihood function (A.32) over Z_{2t}' produces the marginal for Z_{1t}'. This exercise yields:

$$\begin{aligned} f\left(Z_1,\overline{\delta}_1,...,\overline{\delta}_\tau\right) &= \prod_{t\in T_y}\int_{\Theta_{t\cdot}\ Y_{2t}} f(Y_{1t},Y_{2t} \mid Z^*_{dt},\overline{\delta}_t=1)dY_{2t}\ P(\overline{\delta}_t=1)^{\overline{\delta}_t}\prod_{t\in T_d\cup T_u}\left[P(\overline{\delta}_t=1)\right]^{\overline{\delta}_t} \quad \text{(A.33)} \\ &= \prod_{t\in T_y} f(Y_{1t} \mid Z^*_{d1t},\overline{\delta}_t=1)\ P(\overline{\delta}_t=1)^{\overline{\delta}_t}\prod_{t\in T_d\cup T_u}\left[P(\overline{\delta}_t=1)\right]^{\overline{\delta}_t} \quad . \end{aligned}$$

The second line of this expression exploits the relationship:

$$\begin{aligned} \int_{\Theta_{t\cdot}\ Y_{2t}} f\ (Y_{1t},Y_{2t} \mid Z^*_{dt},\overline{\delta}_t=1)dZ_{2t} &= \int_{\Theta_{t\cdot}\ Y_{2t}} f(Y_{2t} \mid Y_{1t}\ ,Z^*_{dt},\overline{\delta}_t=1)dY_{2t}\ f(Y_{1t} \mid Z^*_{dt},\overline{\delta}_t=1) \\ &= f(Y_{1t} \mid Z^*_{dt},\overline{\delta}_t=1) \quad , \end{aligned}$$

which follows since $\Theta_{t \cdot Y_{2t}}$ constitutes the domain of Y_{2t} given the event $\overline{\delta}_t = 1$. Clearly, the second expression in (A.33) is a special case of (A.32). This merely reflects the fact that an unobserved Y_{2t} has been reinterpreted as a component of the U_i's implicit in (A.32). The variables making up U_i in a state i (or t) may be observed as a Y_j in some other state.

Finally, throughout the above discussion one can readily interpret the distribution of U as being conditional on a set of exogenous variables X, as well as define the regions of definition of δ_i to depend on X (so, $\Omega_i = \Omega_i(X)$). To be deemed exogenous, each component of X must be observed in all states; otherwise, this variable must be treated as a component of Y or Z in the previous analysis. Modifying the above formula to admit exogenous X merely involves adding X as an argument of $f(\cdot)$ and interpreting the sample subspaces Ω_i, $\Phi_{i|Y_i}$, and Θ_i as functions of X.