Semiparametric Estimation of Random Coefficients in Structural Economic Models[*]

Stefan Hoderlein[†] Lars Nesheim [‡] Anna Simoni[§]

Boston College University College London CNRS - CREST

September 20, 2016

Abstract

This paper discusses nonparametric estimation of the distribution of random coefficients in a structural model that is nonlinear in the random coefficients. We establish that the problem of recovering the probability density function (*pdf*) of random parameters falls into the class of convexly-constrained inverse problems. The framework offers an estimation method that separates computational solution of the structural model from estimation. We first discuss nonparametric identification. Then, we propose two alternative estimation procedures to estimate the density and derive their asymptotic properties. Our general framework allows us to deal with unobservable nuisance variables, e.g., measurement error, but also covers the case when there are no such nuisance variables. Finally, Monte Carlo experiments for several structural models are provided which illustrate the performance of our estimation procedure.

Keywords: Nonlinear random coefficients, mixture models, structural models, nonparametric, semiparametric, heterogeneity, inverse problems.

[*] We thank the co-editor Yoon-Jae Whang and three anonymous referees whose comments greatly improved the paper. We have benefited from comments and discussions with Victor Aguirregabiria, Orazio Attanasio, Richard Blundell, Chris Carroll, Jeremy Fox, Emmanuel Guerre, Dirk Krueger, Arthur Lewbel, Enno Mammen, Rosa Matzkin, Peter Phillips, Jean-Marc Robin, Sami Stouli, Yuanyuan Wan, and Hal White. Lars Nesheim gratefully acknowledges financial support from the UK Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001. Anna Simoni gratefully acknowledges financial support from the University of Mannheim through SFB 884 and from ANR-13-BSH1-0004, and ANR-11-LABEX-0047.

[†] Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA, email: stefan_hoderlein@yahoo.com.

[‡] Department of Economics, University College London, Gower Street, London, WC1E 6BT, UK, email: l.nesheim@ucl.ac.uk

[§] CREST, 15 Boulevard Gabriel Péri, 92240 Malakoff (France), email: simoni.anna@gmail.com

1 Introduction

Many structural microeconomic models postulate that individual decision makers solve complicated optimization problems governed by a small number of structural parameters θ. While these parameters are fixed for every individual, economic theory does not postulate that they be the same for every individual. Yet, in most empirical applications, the extent to which individual decision makers are allowed to vary is severely constrained. These constraints on heterogeneity are typically not based on economic theory.

A natural way to relax the constraints and make structural model assumptions more appealing is to assume that the unobservable parameters θ in individuals' decision problems are random parameters drawn from a fully flexible nonparametric continuous distribution. In this paper, we propose and analyze a method to estimate a nonparametric distribution of random coefficients θ in general structural economic models in which the mapping from random coefficients to outcomes is nonlinear and may only be implicitly defined. We allow the random coefficients θ to be correlated with some of the explanatory variables, analyze identification and propose an estimation method that completely separates computational solution of the economic model from estimation.

To give a stylized example, consider the workhorse Euler equation model of the consumption literature, where for simplicity we have set the discount rate to the interest rate:

$$\partial_c u(C_t, \theta, \varepsilon_t) = \mathbb{E}\left[\partial_c u(C_{t+1}, \theta, \varepsilon_{t+1}) | W_t, Z_t, \theta, \varepsilon_t\right]. \tag{1.1}$$

Here, $\partial_c u$ denotes the derivative of instantaneous utility with respect to consumption, C_t is consumption in period t, and the random parameters θ may include preference parameters such as the coefficient of risk aversion or parameters defining beliefs about future states. Moreover, $(W_t, Z_t, \varepsilon_t)$ are endogeneous observable, exogenous observable, and unobservable state variables, respectively. W_t may also contain deterministic variables such as time or age.[1] Equation (1.1) implicitly defines the consumption function $C_t = \varphi(W_t, Z_t, \theta, \varepsilon_t)$. When we lack information about the probability distribution of heterogeneity in the population (for example the *pdf* $f_{\theta|W}$) but have knowledge about the structural equation (so that we can explicitly compute φ), we can use this knowledge to define a mapping from $f_{\theta|W}$ to the population *pdf* of observables $f_{C|WZ}$. This mapping can be written as an integral equation of the form

$$f_{C|WZ} = T f_{\theta|W}, \tag{1.2}$$

where T is a known integral operator derived from the economic model. Our estimators are based on recovering $f_{\theta|W}$ from (1.2).

This paper makes several contributions. First, it shows how to analyze nonlinear random coefficients in a structural economic model using tools from the theory of linear inverse problems. Second, it shows how to derive the estimating equation (1.2) from the economic model without requiring the structural function φ to be monotonic in ε. Third, based on (1.2), the paper proposes two estimators of $f_{\theta|W}$, derives rates of convergence and shows asymptotic normality. The estimators are based on a simple Tikhonov regularization method modified to impose the constraint that the estimate must be a density function. Our main contributions in this respect are: *(i)* to extend source conditions and provide the rate of convergence for the convexly-constrained Tikhonov estimator in a stochastic setting and *(ii)* to provide and study properties of a step-by-step procedure to compute the orthogonal projection of the unconstrained Tikhonov-regularized estimate onto the set of densities. Fourth, the paper studies nonparametric identification of $f_{\theta|W}$. We characterize the identified set when the model is not point-identified and provide a necessary and sufficient condition for point-identification that we call $\mathcal{T}$-completeness which is weaker than L^2- or bounded-completeness.

This research therefore extends the parametric structural models literature to allow for nonlinear and endogenous random coefficients. This literature is vast. For a recent survey of the consumption literature, which originally motivated this research, see Attanasio & Weber (2010). Our analysis provides insights into when identification is only partial and provides novel conditions for point-identification. In addition, our analysis makes clear that estimation in these contexts is fundamentally an ill-posed inverse problem.

Most closely related to our approach are nonparametric econometric models involving random parameters. In particular, there is a literature that considers linear single index nonparametric random coefficients models as in Beran & Millar (1994), Beran, Feuerverger, & Hall (1996), Ichimura & Thompson (1998), Hoderlein, Klemelä, & Mammen (2010), Gautier & Kitamura (2013), Dunker, Hoderlein, & Kaido (2013), Gautier & Hoderlein (2015), and Lewbel & Pendakur (2016). In these papers, the random coefficients are continuously distributed and fully independent of explanatory variables. Also related is Masten (2014) who considers a linear simultaneous equations model and focuses on estimation of the density of one random slope coefficient. We extend this literature by allowing for endogenous random coefficients that enter in a nonlinear way and by allowing for models in which the function mapping explanatory variables into outcomes is only implicitly defined.

Our work is also linked to the mixture models literature following Heckman & Singer (1984) (HS). In a duration model setting, HS analyze the equation

$$f_{C|WZ}(c,w,z)=\int_{\Theta} f_{C|WZ\theta}(c,w,z,\theta)f_{\theta|W}(\theta,w)d\theta. \tag{1.3}$$

They focus on estimating a finite dimensional parameter that impacts the *pdf* $f_{C|WZ\theta}(c,w,z,\theta)$ while treating $f_{\theta|W}$ as a nuisance parameter. Closely related to HS are Henry, Kitamura, & Salanié (2014) and Kasahara & Shimotsu (2009). In contrast to these references, our analysis focuses on $f_{\theta|W}$ and the kernel of the operator in (1.3) is derived from the economic model.

Our work is also related to the stochastic inverse problem literature. See Engl, Hanke, & Neubauer (2000) for an overview. In particular, recovering the probability density of θ from (1.2) is equivalent to solving a convexly constrained integral equation of the first kind. Integral equations of the first kind have been studied extensively in different areas of econometrics (see e.g. Carrasco, Florens, & Renault, 2007, for an overview). These areas include, among others: nonparametric instrumental regression estimation - see e.g. Florens (2003), Newey & Powell (2003), Hall & Horowitz (2005), Blundell, Chen, & Kristensen (2007), Darolles, Fan, Florens, & Renault (2011), Chen & Reiss (2011) and Canay, Santos, & Shaikh (2013) - and moment estimation and deconvolution - see e.g. Carrasco & Florens (2000), Ai & Chen (2003), Carrasco & Florens (2011) and Chen & Pouzo (2012). There are two important differences between our model and the models studied in these papers. First, the kernel of our integral operator is not estimated but is derived from a structural economic model. Second, we seek to estimate the density of random coefficients, not a function of observables.

Our estimating equation is also related to Hu & Schennach (2008). However, our model differs in many core aspects from their model, not least for the different object of interest (i.e., the distribution of random parameters), and for the structural nonseparability of the model considered. Moreover, our exclusion restrictions are different from theirs (e.g., we do not assume conditional independence of C and Z given θ) and are motivated by the structural economic application.

2 Example 1: a roadmap

To illustrate the structure of the model and the main results in this paper, consider a common specification in consumer demand. Let X measure the true log-expenditure on all nondurable goods, and let C^* measure the true log expenditure for one good. Assume that C^* is generated by a linear random coefficients model with

$$C^* = \theta_0 + \theta_1 Z_1 + \theta_2 X$$

where Z_1 is the log-price and $\theta = (\theta_0, \theta_1, \theta_2)$ is a vector of random parameters. If θ is independent from (Z_1, X) and if (Z_1, X) have support equal to $\mathbb{R}^2$, then the joint density of random coefficients is nonparametrically identified and can be estimated using, for example, Hoderlein et al. (2010). However, in consumer demand two important problems arise. First, since X is a choice variable, it is likely to be endogenous. The same preference parameters that determine C^* also determine X. Second, actual demand C is frequently measured with error, i.e., $C \neq C^*$.

To handle endogeneity, we follow the demand literature and use instruments in a control function fashion. Let Z_2 measure log-income and suppose there is a relation $X = g\left(Z_2, W\right)$, where g is a (identified) function that is strictly monotonic in W and where W is the percentile of log-expenditure conditional on Z_2. See Hoderlein (2011) for such a structure in a demand application. To handle measurement error, we assume that observed C is generated as $C = C^* + \eta$, where $\eta | \theta, W, Z \sim \exp(\lambda)$, with $\lambda > 0$. Finally, let $Z = (Z_1, Z_2)$ and assume that $Z \perp\!\!\!\perp (\theta, \eta) | \, W$. Substituting all elements into the outcome equation for observed C, we obtain

$$C = \theta_1 Z_1 + \theta_2 g(Z_2, W) + \varepsilon, \tag{2.1}$$

where $\varepsilon = \eta + \theta_0$ and $f_{\varepsilon|WZ\theta}(\varepsilon, \theta_0) = \lambda e^{-\lambda(\varepsilon - \theta_0)}$. This model is a special case of the general class of nonlinear models developed below. It serves to illustrate why we consider the specific structure put forward in this paper and to show how our assumptions lead to identification.

First, note that X is endogenous due to correlation between W and θ, ruling out use of standard approaches such as Hoderlein et al. (2010). The main idea we adapt from the control function literature is that, conditional on W, there is no endogeneity. Since $Z \perp\!\!\!\perp \theta | W$, equation (2.1) can be estimated conditional on W. Provided $g(Z_2, w)$ varies enough, the *pdf* $f_{\theta|W}(\cdot, w)$ is nonparametrically identified using standard arguments from the random coefficients literature.

Second, observe the dual sources of unobserved variation in the model, θ and ε. The former is the preference heterogeneity of interest and the latter contains measurement error. Since the latter part is a nuisance part of our model, we follow the measurement error and deconvolution literatures and assume a (partially) parametric model for ε.

Thus, our model introduces two general elements that are novel to this literature. First, θ may depend on some variables W while the instruments Z provide exogenous variation in the sense that $Z \perp\!\!\!\perp \theta | W$. As we show below in the Euler equation example, allowing some variables to be correlated with θ while having others be (conditionally) independent is a feature that arises more generally in heterogeneous structural models. Our strategy is thus to first recover $f_{\theta|W}$ by performing all steps conditional on W and then to obtain f_θ by integrating out W. The case without endogenous variables is obviously a special case in which we can directly obtain f_θ since $Z \perp\!\!\!\perp \theta$. The instruments Z, however, are generally necessary to identify f_θ, especially when θ is a vector. This is clear in example (2.1) above, as it is variation in Z that is used to trace out variation in θ.

The second general feature is the composite error ε, comprised of preference heterogeneity θ and a nuisance part η. In the "demand with measurement error" application as well as in the Euler equation application, this is a necessary feature of the structural model. In a more general version of the Euler equation model, this feature is even more important as it can capture time-varying unobserved states like transitory income shocks. See e.g., Attanasio & Weber (2010) or Hoderlein, Nesheim, & Simoni (2012). Our strategy in this part is motivated by the deconvolution literature and requires a parametric assumption on the nuisance error. While we believe this nuisance error to be an important feature of many applied models, our approach does not rely on the existence of ε. As we demonstrate in the online Supplement [2], all arguments go through with minor modifications if there is no nuisance unobservable ε.

The rest of the paper formalises these ideas for a general nonlinear model in which $C = \varphi(W, Z, \theta, \varepsilon)$ for a function $\varphi(\cdot)$ defined below. In Theorem 1, we show that there is an integral operator T which maps $f_{\theta|W}$ into $f_{C|WZ}$. After characterising some properties of this operator, in Theorem 2 we characterise the set of *pdf*'s $f_{\theta|W}$ that are compatible with a particular $f_{C|WZ}$ via T. Since φ and $f_{\varepsilon|WZ\theta}$ are known, this set can be computed. Next, for the general model, point-identification requires a completeness condition on the probability distribution characterizing the operator. The completeness condition we require is weaker than that required in the nonparametric IV literature since our object of interest is a probability density and not an unrestricted function. We call our condition $\mathcal{T}$-completeness. While it must be checked in every application, we provide a sufficient condition that is easier to check. For instance, in the demand example (2.1), the equation that identifies $f_{\theta|W}$ is

$$f_{C|WZ}(c,w,z) = \int_{\Theta} \lambda e^{-\lambda[c-\theta_0-\theta_1 z_1-\theta_2 g(z_2,w)]} 1_{\{c\geq\theta_0+\theta_1 z_1+\theta_2 g(z_2,w)\}}(c) f_{\theta|W} d\theta$$

where $1_A(c)$ is equal to 1 if c satisfies condition A and zero otherwise. The exponential function that characterizes the kernel of the operator can be rewritten as

$$\exp\left\{-\lambda\left[c-\theta_0-\theta_1 z_1-\theta_2 g\left(z_2,w\right)\right]\right\} = \exp\left(-\lambda c\right)\exp\left\{\lambda\left[1,z_1,g\left(z_2,w\right)\right]\theta\right\}$$

where $\theta = (\theta_0,\theta_1,\theta_2)'$. These expressions satisfy the sufficient conditions of Proposition 3 below with $h(\theta)=\lambda$, $m(\theta)=\theta$, $\tau(c,w,z)=\lambda\left[1,z_1,g\left(z_2,w\right)\right]'$ and $k(c,w,z)=\exp\left(-\lambda c\right)$, implying that $f_{\theta|W}$ is point-identified.

The plan of the rest of the paper is as follows. In Section 3, we present the model and assumptions. Then, we analyse identification in Section 4. Section 5 presents our two estimators and Section 6 concludes with results from two Monte Carlo simulations. Proofs of the main results are in Appendix A while minor and technical results are in the online Supplement.

3 The general structural model

Let $(\Omega,\mathcal{F},P)$ be a complete probability space and $(C,W,Z,\theta,\varepsilon)$ be a real-valued random vector defined on it, and partitioned into $C\in\mathbb{R}$, $W\in\mathcal{W}\subset\mathbb{R}^k$, $Z\in\mathcal{Z}\subset\mathbb{R}^l$, $\theta\in\Theta\subset\mathbb{R}^d$ and $\varepsilon\in\mathbb{R}$, with k, l and d finite integers. We denote by $\mathcal{B}_{\mathcal{C}}$, $\mathcal{B}_{\mathcal{W}}$, $\mathcal{B}_{\mathcal{Z}}$, $\mathcal{B}_{\Theta}$ and $\mathcal{B}_{\varepsilon}$ the corresponding Borel σ-fields in $\mathbb{R}$, $\mathbb{R}^k$, $\mathbb{R}^l$, $\mathbb{R}^d$ and $\mathbb{R}$, respectively, and use capital and lowercase Latin letters for observable random variables and their realizations. We use lowercase Greek letters for unobservable random variables as well as their realizations. For two random vectors A and B we write: $P_{A|B}$ for the conditional distribution of A given B and $f_{A|B}$ for the density function (*pdf*, hereafter) of $P_{A|B}$ with respect to Lebesgue measure. We use the convention that $f_{A|B}(a,b)=0$ if a is not in the support of $P_{A|B=b}$. We denote by $\mathcal{C}\subset\mathbb{R}$ (resp. $\mathcal{Z}\times\mathcal{W}$) the support of the marginal distribution of C (resp. (Z,W)).

To exploit desirable properties of Hilbert spaces, we develop our analysis in L^2 spaces. For this purpose, we introduce two non-negative weighting functions, π_θ and π_{cz}, with support on Θ and $\mathcal{C}\times\mathcal{Z}$ respectively.[3] Define the space $L^2_{\pi_\theta}$ (resp. $L^2_{\pi_{cz}}$) of real-valued functions defined on Θ (resp.

$\mathcal{C} \times \mathcal{Z}$) that are square integrable with respect to π_θ (resp. π_{cz}). That is,

$$\begin{aligned} L^2_{\pi_\theta} &= \left\{ h : \Theta \to \mathbb{R} \,\middle|\, \int_\Theta h^2(\theta)\pi_\theta(\theta)d\theta < \infty \right\}, \\ L^2_{\pi_{cz}} &= \left\{ \psi : \mathcal{C} \times \mathcal{Z} \to \mathbb{R} \,\middle|\, \int_{\mathcal{C}\times\mathcal{Z}} \psi^2(c,z)\pi_{cz}(c,z)dcdz < \infty \right\}. \end{aligned}$$

We denote the scalar product by $\langle \cdot, \cdot \rangle$ and the induced norm by $\| \cdot \|$ in both spaces without distinction. The sets of conditional *pdf*'s relevant for our analysis are defined as follows

$$\begin{aligned} \mathcal{F}_{\theta|W} = &\quad \left\{ f \text{ is a conditional } pdf \text{ on } (\mathbb{R}^d, \mathcal{B}_\Theta) \text{ given } W \text{ and } f \in L^2_{\pi_\theta} \text{ a.s. } \right\} \\ \mathcal{F}_{C|WZ} = &\quad \left\{ f \text{ is a conditional } pdf \text{ on } (\mathbb{R}, \mathcal{B}_\mathcal{C}) \text{ given } (Z,W) \text{ and } f \in L^2_{\pi_{cz}} \text{ a.s. } \right\}, \end{aligned}$$

and analogously for $\mathcal{F}_{C|WZ\theta}$.

The next assumption specifies the structural data generating process.

Assumption 1. *The random element $(C, W, Z, \theta, \varepsilon)$ satisfies a structural economic model*

$$\Psi(C, W, Z, \theta, \varepsilon) = 0 \quad a.s. \tag{3.1}$$

where Ψ is a ***known*** *Borel measurable real-valued function.*[4] *We assume that (3.1) has a unique global solution in terms of C:*

$$C = \varphi(W, Z, \theta, \varepsilon), \quad a.s. \tag{3.2}$$

where $\varphi : \mathbb{R}^{k+l+d+1} \to \mathbb{R}$ is a Borel-measurable function. In addition, we assume (3.2) has a finite number s of solutions in terms of ε almost surely.[5]

This assumption describes how our structural model links observables (C, W, Z) to unobservables (θ, ε). We distinguish between three different observables. C is the dependent variable, while W and Z denote variables that cause C. W is allowed to be correlated with θ while Z is assumed to be conditionally independent of θ given W. As discussed in Section 2, this distinction is motivated by applications in which some important explanatory variables W are endogenous. The distinction between the unobservable variables θ and ε is made to separate objects of interest from a nuisance error term ε. Consequently, the distribution of θ is allowed to be completely nonparametric while the distribution of ε is flexible but parametric.

Our approach does not require that the function φ be available in closed-form nor that it be globally monotone in ε. All that is required is that we can solve equation (3.1) numerically, and that the function φ be piecewise invertible. More precisely, for some set A, let $\mathrm{Im}\left(A|\, w, z, \theta\right)$ be the image of A through φ conditional on (w, z, θ). We can then define a finite partition of $\mathbb{R}$, $(\mathcal{E}_1, ..., \mathcal{E}_s)$, such that $\varphi_i^{-1}(w, z, \theta, \cdot) : \mathrm{Im}\left(\mathcal{E}_i|\, w, z, \theta\right) \to \mathcal{E}_i$ is one-to-one for each $i \in \{1, ..., s\}$. The elements of the partition and the inverse can be computed since they are implicitly defined by (3.1). In the following we denote, $\forall i \in \{1, \ldots, s\}$, by $\varphi_i^{-1}(w, z, \theta, \cdot)$ the function $\varphi^{-1}(w, z, \theta, \cdot)$ with domain $\varphi(w, z, \theta, \mathcal{E}_i)$ and image $\mathcal{E}_i$.

Allowing for this general form of the structural model is an important weakening of assumptions, as closed form expressions are frequently not available and monotonicity conditions are difficult to justify. In our Euler equation example, the consumption function is only implicitly defined and there is little reason to believe that there is a monotonic relationship between unobserved states and levels of consumption.

The only other assumption on Ψ is differentiability. Let $\partial_c \Psi(c, w, z, \theta, \varepsilon)$ and $\partial_\varepsilon \Psi(c, w, z, \theta, \varepsilon)$ denote the partial derivatives of Ψ with respect to C and ε respectively.

Assumption 2. *The structural function $\Psi : \mathbb{R}^{k+l+d+2} \to \mathbb{R}$ is almost everywhere differentiable in C and in ε with $\partial_c \Psi(c, w, z, \theta, \varepsilon) \neq 0$ and $\partial_\varepsilon \Psi(c, w, z, \theta, \varepsilon) \neq 0$ for every $(c, w, z, \theta, \varepsilon)$ in the support of $(C, W, Z, \theta, \varepsilon)$ except possibly on a set of $(c, w, z, \theta, \varepsilon)$ values whose Lebesgue measure is* 0.

Finally, the following three assumptions characterize the joint conditional distribution of (ε, Z, θ) given W.

Assumption 3. *The conditional probability distribution $P_{\varepsilon|WZ\theta}$ on $\mathcal{B}_\varepsilon$ given (W, Z, θ) admits a pdf $f_{\varepsilon|WZ\theta}$ with respect to the Lebesgue measure. This pdf $f_{\varepsilon|WZ\theta}$ is known up to a finite-dimensional parameter θ_ε, a subvector of the vector θ. Moreover, $f_{\varepsilon|WZ\theta}$ is strictly positive and bounded away from infinity a.s. on the support of $P_{\varepsilon|WZ\theta}$.*

Assumption 4. *The conditional probability distribution $P_{Z\theta|W}$ on $\mathcal{B}_Z \otimes \mathcal{B}_\Theta$ given W admits a pdf $f_{Z\theta|W}$ with respect to the Lebesgue measure. The pdf $f_{\theta|W}$ is strictly positive and bounded away from infinity a.s. on its support.*

Assumption 5. *The random element Z is conditionally independent of θ given W, i.e. $Z \perp\!\!\!\perp \theta | W$.*

Assumption 3 allows the conditional distribution of ε to depend on all variables in the model. Unlike in deconvolution, we can allow for ε and θ to be dependent. By allowing $f_{\varepsilon|WZ\theta}$ to be known up to a finite dimensional random parameter[6], we allow for cases where not everything is known about $f_{\varepsilon|WZ\theta}$. In theory, the specification for $f_{\varepsilon|WZ\theta}$ can be very close to a nonparametric specification. As in all semiparametric models there is a trade-off between flexibility and feasibility. Adding flexibility will generally increase the dimension of the estimation problem, reduce the convergence rate of the estimator and may lead to a failure of point identification.

Note that Z and θ are continuous random vectors while W may be discrete. If some elements of Z are discrete, then the analysis is unchanged as long as the *pdf* of Z is replaced with the probability mass function and integrals with respect to Z are replaced by sums. However, discrete Z are likely to have little identifying power. For example, in the linear random coefficients model $Y = \theta_1 + \theta_2 Z$, if Z is scalar and binary, then the two marginal distributions of θ_1 and $\theta_1 + \theta_2$ respectively are identified. All other features of the joint distribution of (θ_1, θ_2) are not identified. If some elements of θ are discrete and random with known support, then the analysis also is unchanged. In this case, all of the statements with respect to $f_{\theta|W}$ have their finite dimensional counterparts. If some elements of θ are deterministic (or discrete random variables with unknown, finite support), then the analysis is slightly different. We discuss this case in Section 5.3 and explain how to estimate the model when some components of θ are deterministic.

Finally, Assumption 5 is the key independence condition that is often required for point-identification of the *pdf* $f_{\theta|W}$. Strictly speaking Z is not required for point-identification. It is possible to specify a model in which $f_{\theta|W}$ is point-identified solely by nonlinearities in $f_{C|W\theta}$. When θ is a scalar, such a specification may be reasonable. However, especially when θ is a high dimensional vector, it is easy to specify models in which $f_{\theta|W}$ is not point-identified without exogenous variation in Z. The linear random coefficients model in Section 2 is a leading case. We now provide a second example that illustrates these points as well as how our model can be applied in a richer setting.

Example 2 (Intertemporal consumption model)**.** *Consider the constant absolute risk aversion (CARA) intertemporal utility maximization problem with finite horizon T, constant interest rate r and random parameter θ capturing heterogeneity in utility. Define $R = 1+r$. Let A_t be a consumer's assets and let Y_t be his/her income. Suppose income follows a random walk. Let $S_t = (A_t, Y_t)$ be the state vector and let $v_t(S_t, \theta)$ be the value function for a consumer of type θ at date t. Let the terminal value function be $v_{T+1}(S_{T+1}, \theta) = -\frac{e^{\gamma A_{T+1}}}{\gamma}$ and let $\theta = (\gamma, \beta)$ where γ is the coefficient of risk aversion and β is the discount factor. At each date $t \leq T$, a consumer's value function is defined by*

$$v_t(S_t,\theta) = \max_{\{C_t^*\}} \left\{ \begin{array}{c} -\frac{e^{-\gamma C_t^*}}{\gamma} + \beta \mathbb{E}_t\left[v_{t+1}(S_{t+1},\theta)\right] \\ \textit{subject to} \\ A_{t+1} = R\left(A_t + Y_t - C_t^*\right) \\ Y_{t+1} = Y_t + \eta_{t+1} \end{array} \right\}$$

where C_t^* is consumption, $\eta_t \sim N\left(0, \sigma_\eta^2\right)$, and $\mathbb{E}_t$ is the time t conditional expectation operator. We assume that σ_η^2 is either known or can be identified and hence estimated from other aspects of the data. Suppose observed consumption C_t satisfies $C_t = C_t^* + \varepsilon_t$ where ε_t is measurement error. Let $W_t = (A_t, Y_{t-1})$ and $Z_t = Y_t - Y_{t-1}$. Under the assumptions stated, the consumption function $\varphi(\cdot)$ in (3.2) *takes the form*

$$C_t = \varphi_{1t} W_t^1 + \varphi_{2t}\left(W_t^2 + Z_t\right) + m_t(\gamma,\beta) + \varepsilon_t \tag{3.3}$$

with

$$m_t(\gamma,\beta) = \varphi_{3t} + \varphi_{4t}\gamma + \varphi_{5t}\frac{\ln\beta}{\gamma}. \tag{3.4}$$

The parameters $(\varphi_{1t}, \varphi_{2t}, \varphi_{3t}, \varphi_{4t}, \varphi_{5t})$ in (3.3) *and* (3.4) *depend only on R, t and σ_η^2.*[7] *The vector $\theta = (\gamma, \beta)$ is assumed to be a time-invariant random coefficient vector heterogeneously distributed* *in the population. We assume that the income process* $(Y_t)_{t=1,..,T} \perp\!\!\!\perp \theta$ *and that* $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.

Because θ *determines both past and current consumption and savings, it is correlated with* W_t*. However, by assumption,* $Z_t = Y_t - Y_{t-1}$ *is independent of* θ*. In addition, with the choice* $W_t = (A_t, Y_{t-1})$ *and* $Z_t = Y_t - Y_{t-1}$*, we have* $\theta \perp\!\!\!\perp Z_t | \, W_t$ *which satisfies Assumption 5.*

We continue discussion of this example at the end of Section 4.1. We discuss simulation results based on this example in Section 6.2.

4 Identification of the distribution of parameters

In the following, we use $(f_{\varepsilon|WZ\theta} \circ \varphi_i^{-1})(c, w, z, \theta)$ to denote $f_{\varepsilon|WZ\theta}\left[\varphi_i^{-1}(w, z, \theta, c), w, z, \theta\right]$, we suppress the dependence of the operator T on W (where T is defined below), and, for a subset $\mathcal{A} \subset L^2_{\pi_\theta}$, use the notation $T|_{\mathcal{A}}$ to denote the operator T restricted to $\mathcal{A}$, that is, $T|_{\mathcal{A}} : \mathcal{A} \to L^2_{\pi_{cz}}$. We also use $\mathcal{R}(T)$ to denote the range of the operator T.

Theorem 1. *Let Assumptions 1 - 5 be satisfied. Then,*

$$f_{C|WZ} = T f_{\theta|W} \quad a.s. \tag{4.1}$$

where $\forall h \in L^2_{\pi_\theta}$,

$$Th = \int\limits_{\Theta} \sum_{i=1}^{s} (f_{\varepsilon|WZ\theta} \circ \varphi_i^{-1})(c, w, z, \theta) \left| \frac{\partial_c \Psi(c, w, z, \theta, \varphi_i^{-1}(w, z, \theta, c))}{\partial_\varepsilon \Psi(c, w, z, \theta, \varphi_i^{-1}(w, z, \theta, c))} \right| 1_{\mathcal{C}_i}(c) h(\theta) d\theta, \tag{4.2}$$

$\mathcal{C}_i = \{c \in \operatorname{Im}(\mathcal{E}_i \, | w, z, \theta)\}$ *and* $\operatorname{Im}(\mathcal{E}_i \, | w, z, \theta)$ *is the image of* $\mathcal{E}_i$ *through* φ *conditional on* (w, z, θ)*. This implies that* $f_{\theta|W}$ *is a solution of*

$$f_{C|WZ} = T f_{\theta|W} \quad \textit{subject to} \quad f_{\theta|W} \in \mathcal{F}_{\theta|W}, \quad a.s. \tag{4.3}$$

The operator T is a mixing operator and $f_{C|WZ}$ is an a.s. $f_{\theta|W}$-mixture of $f_{C|WZ\theta}$. Equation (4.2) provides an expression for the operator T that depends only on the elements of the structural model $\left(\Psi, \varphi, f_{\varepsilon|WZ\theta}\right)$. These elements are known by assumption and can be directly computed.

Equation (4.3) is the basis for our estimation strategy. This equation characterizes $f_{\theta|W}$ as the solution of a *convexly-constrained Fredholm integral equation of the first kind.* Under Assumptions 1-5 the existence of at least one solution to (4.3) is guaranteed since $f_{C|WZ} \in \mathcal{R}(T|_{\mathcal{F}_{\theta|W}})$. We note that recovering $f_{\theta|W}$ from (4.3) is an ill-posed inverse problem.

The properties of the solution (or solutions) to (4.3) depend on the properties of T and its adjoint. The next proposition characterizes the adjoint operator of T.

Proposition 1 (Adjoint of T)**.** *Let $T : L^2_{\pi_\theta} \to L^2_{\pi_{cz}}$ be the operator defined in (4.2). Assume that T is bounded. Then, the operator T^* defined for all $\psi \in L^2_{\pi_{cz}}$ as*

$$T^*\psi = \int_{\mathcal{C}}\int_{\mathcal{Z}} f_{C|WZ\theta}\,(c,w,z,\theta)\,\psi(c,z)\frac{\pi_{cz}(c,z)}{\pi_\theta(\theta)}dcdz,$$

with

$$f_{C|WZ\theta}(c,w,z,\theta) = \sum_{i=1}^{s}(f_{\varepsilon|WZ\theta}\circ\varphi_i^{-1})(c,w,z,\theta)\cdot\left|\frac{\partial_c\Psi(c,w,z,\theta,\varphi_i^{-1}(w,z,\theta,c))}{\partial_\varepsilon\Psi(c,w,z,\theta,\varphi_i^{-1}(w,z,\theta,c))}\right|1_{\mathcal{C}_i}(c)$$

exists and is the adjoint of T. The operator $T^* : L^2_{\pi_{cz}} \to L^2_{\pi_\theta}$ is bounded and linear.

Because the kernel of T is known, we can always choose two non-negative weight functions π_θ and π_{cz} so that T is bounded and compact with $\mathcal{R}(T) \subset L^2_{\pi_{cz}}$. For example, in the linear random coefficients model in consumer demand discussed in Section 2 (cf. (2.1)), let $g(z_2, w) = z_2 w$. Then the operator T takes the form: $\forall h \in L^2_{\pi_\theta}$, $Th = \int_\Theta \lambda e^{[-\lambda(c-\theta_0-\theta_1 z_1-\theta_2 z_2 w)]}1_{\{c\geq\theta_0+\theta_1 z_1+\theta_2 z_2 w\}}(c)h(\theta)d\theta$ where we have set $g(Z_2, W) = Z_2 W$, and $\Theta = [\underline{\theta}_0, \overline{\theta}_0] \times [\underline{\theta}_1, \overline{\theta}_1] \times [\underline{\theta}_2, \overline{\theta}_2]$ for some $\underline{\theta}_i < \infty$, $\overline{\theta}_i < \infty$, $i = 0, 1, 2$. Because C is nonnegative, we can choose the weight $\pi_{cz} = \pi_c \times \pi_z$ with $\pi_c(c) = \lambda_c e^{-\lambda_c c}$, $\lambda_c > 0$ and π_z a standard bivariate normal distribution.

One way to verify that T is bounded and compact is to verify that

$$\int_{\mathcal{Z}}\int_{\mathcal{C}}\int_{\Theta}\lambda^2 e^{[-2\lambda(c-\theta_0-\theta_1 z_1-\theta_2 z_2 w)]}1_{\{c\geq\theta_0+\theta_1 z_1+\theta_2 z_2 w\}}(c)\pi_{cz}(c,z)/\pi_\theta(\theta)d\theta dcdz$$

is finite. By taking π_θ equal to the Lebesgue measure on Θ it is easily seen that (by denoting $\theta_w = (\theta_1, \theta_2 w)'$ and $z = (z_1, z_2)'$):

$$\begin{aligned}
& \int_{\mathcal{Z}}\int_{\mathcal{C}}\int_{\Theta}\lambda^2 e^{[-2\lambda(c-\theta_0-\theta_w' z)]}1_{\{c\geq\theta_0+\theta_w' z\}}(c)\pi_{cz}(c,z)/\pi_\theta(\theta)d\theta dcdz \\
\leq\ & \int_{\mathcal{Z}}\int_{\mathcal{C}}\int_{\Theta}\lambda^2 1_{\{c\geq\theta_0+\theta_w' z\}}(c)\times\pi_{cz}(c,z)/\pi_\theta(\theta)d\theta dcdz \\
\leq\ & \int_{\mathcal{Z}}\int_{\Theta}\lambda^2 e^{-\lambda_c(\theta_0+\theta_w' z)}1_{\{\theta_0+\theta_w' z\geq 0\}}(c)\pi_z(z)/\pi_\theta(\theta)d\theta dz + \int_{\mathcal{Z}}\int_{\Theta}\lambda^2 1_{\{\theta_0+\theta_w' z<0\}}(c)\pi_z(z)/\pi_\theta(\theta)d\theta dz \\
\leq\ & \lambda^2\int_{\Theta}(1-\Phi(-\theta_0/\|\theta_w\|))\pi_\theta(\theta)d\theta + \lambda^2\int_{\Theta}\Phi(-\theta_0/\|\theta_w\|)\pi_\theta(\theta)d\theta \\
<\ & \infty
\end{aligned}$$

where $\Phi(u)$ denotes the cumulative distribution function of a $N(0,1)$. Therefore, this choice of π_θ and π_{cz} gives a compact T.

Assumption 6 gives sufficient conditions for compactness and boundedness of T in terms of Ψ, $f_{\varepsilon|WZ\theta}$, π_{cz} and π_θ.

Assumption 6. *The function* $2^{\frac{s-1}{2}} f_{\varepsilon|WZ\theta} |\partial_c \Psi / \partial_\varepsilon \Psi|^{1/2} \Big|_{c=\varphi(w,z,\theta,\varepsilon)}$ *is a.s. square integrable in* (ε, Z, θ) *with respect to* $\frac{\pi_{cz}}{\pi_\theta}\Big|_{c=\varphi(w,z,\theta,\varepsilon)}$, *where* $s < \infty$ *is the number of piecewise monotonic components of the inverse of* φ *as defined in Assumption 1.*

In the online Supplement, we show that this assumption ensures that T is compact and bounded (see Proposition 4 in the online Supplement.). This proposition is not necessary to define our estimator nor to derive its asymptotic properties. However, when it is true, one of our proposed estimators can be written simply in terms of the singular value decomposition of T. In practice, compactness can be checked by scrutinizing Assumption 6. To do this, simply compute the integral of the square of $2^{\frac{s-1}{2}} f_{\varepsilon|WZ\theta} |\partial_c \Psi / \partial_\varepsilon \Psi|^{1/2} \Big|_{c=\varphi(w,z,\theta,\varepsilon)}$ with respect to the weight functions.

Under Assumptions 1 - 6, T^*T is characterized by a countable number of eigenvalues which accumulate only at zero, admitting the following *singular value decomposition* (SVD):

$$T\varphi_j = \lambda_j \psi_j, \qquad T^*\psi_j = \lambda_j \varphi_j, \quad j \in \mathbb{N} \tag{4.4}$$

where $\{\lambda_j\}_{j\in\mathbb{N}}$ and $\{\varphi_j, \psi_j\}_{j\in\mathbb{N}}$ are the sequences of singular values and singular functions, respectively. The set of functions $\{\varphi_j\}_{j\in\mathbb{N}}$ (resp. $\{\psi_j\}_{j\in\mathbb{N}}$) is a complete orthonormal system of eigenfunctions of T^*T (resp. of TT^*) which spans $\overline{\mathcal{R}(T^*)} = \overline{\mathcal{R}(T^*T)}$ (resp. $\overline{\mathcal{R}(T)} = \overline{\mathcal{R}(TT^*)}$) where $\overline{\mathcal{R}(T^*)}$ is the closure of the range of the operator T^* in $L^2_{\pi_\theta}$. When $\mathcal{N}(T)$ is not a singleton, where $\mathcal{N}(\cdot)$ denotes the null space of an operator, we can complete this orthonormal system in order to form an orthonormal basis (o.n.b.) of $L^2_{\pi_\theta}$ denoted by $\left\{\{\varphi_j\}_{j\in\mathbb{N}}, \{\tilde{\varphi}_l\}_{l\in J_0}\right\}$ where $J_0 \subset \mathbb{N}$ and $\tilde{\varphi}_l$ are such that $\mathcal{N}(T) = span\{\tilde{\varphi}_l\}_{l\in J_0}$. In words, the null space is spanned by the elements in J_0.

In the following, we use the SVD to characterize the set of possible solutions of (4.3), i.e. the identified set. To gain intuition, consider the analogous case in which the support of θ is discrete with θ taking on only k distinct values. Suppose the support is known. In that case, if we evaluate T at a finite number of points (C, Z, W), the operator T is a finite dimensional matrix. The probability mass function of θ conditional on W is identified if the researcher can identify, for each value of W, k distinct values (C, Z) such that T is invertible. If the matrix is not invertible (because some of its eigenvalues equal zero), then the identified set can be computed using the singular value decomposition of T. In the limit, as k grows to infinity, the discrete case approaches the continuous case. Assumptions 4 and 6 are required to ensure that this discrete intuition holds true in the continuous case.

The *pdf* $f_{\theta|W} \in \mathcal{F}_{\theta|W}$ will be called *identified* (with respect to the class $\mathcal{F}_{\theta|W}$) if

$$T|_{\mathcal{F}_{\theta|W}}(f_{\theta|W}) = T|_{\mathcal{F}_{\theta|W}}(\tilde{f}_{\theta|W}) \quad \Rightarrow \quad f_{\theta|W} = \tilde{f}_{\theta|W}, \quad a.s. \tag{4.5}$$

for all $f_{\theta|W}, \tilde{f}_{\theta|W} \in \mathcal{F}_{\theta|W}$. In words, $f_{\theta|W}$ is point-identified if the operator $T|_{\mathfrak{D}}$ is injective, where $\mathfrak{D} = \{f_1 - f_2 \,|\forall f_1, f_2 \in \mathcal{F}_{\theta|W}\}$ is the set of functions that can be written as the difference between two densities. The injectivity of $T|_{\mathfrak{D}}$ depends on the injectivity of T but it is not equivalent. If T

is injective, that is, $\mathcal{N}(T) = \{0\}$, then $T|_{\mathfrak{D}}$ is injective as well. However, when T is non-injective the restricted operator $T|_{\mathfrak{D}}$ may be injective. This is possible when the domain of $T|_{\mathfrak{D}}$ is sufficiently restricted.

The following theorem characterizes the set of possible solutions of (4.3). We denote by $f^{\dagger}_{\theta|W}$ the minimum-norm solution of the *unconstrained* linear inverse problem $f_{C|WZ} = Tf_{\theta|W}$, that is, $f^{\dagger}_{\theta|W} = \arg\min\{\|h\| : h \in L^2_{\pi_\theta} \text{ and } f_{C|WZ} = Th\}$.

Theorem 2. *Under Assumptions 1-5, the set of all the solutions of (4.3) is:*

$$\Lambda = \left\{h \in \mathcal{F}_{\theta|W} \,\middle|\, f_{C|WZ} = Th \ a.s.\right\} = \left\{f^{\dagger}_{\theta|W} \oplus \mathcal{N}(T)\right\} \cap \mathcal{F}_{\theta|W}.$$

If in addition, Assumption 6 holds, then T is compact and there exist $\zeta_l \in \mathbb{R}$ for every $l \in J_0 \subset \mathbb{N}$ such that

$$\Lambda = \left\{ h(\theta, w) = f^{\dagger}_{\theta|W}(\theta, w) + \sum_{\{l \in J_0\}} \zeta_l \tilde{\varphi}_l(\theta, w); \sum_{\{l \in J_0\}} \zeta_l^2 < \infty \ and \ \sup_{\theta \in \Theta} h^-(\theta, w) = 0 \ a.s. \right\}.$$

where $h^-(\theta, w) = -\min(h(\theta, w), 0)$ denotes the negative part of h and span$\{\tilde{\varphi}_l\}_{l \in J_0} = \mathcal{N}(T)$.

The second part of this theorem characterizes the set Λ in terms of the SVD of T which is known and the density $f_{C|WZ}$ which can be easily estimated. When the null space of $T|_{\mathfrak{D}}$ is a singleton, Λ is a singleton as well and the model is point-identified. This occurs in two cases:

(i) the operator T is injective, i.e. $\mathcal{N}(T) = \{0\}$. Then, $f^{\dagger}_{\theta|W} \in \mathcal{F}_{\theta|W}$ and is the unique solution of (4.3);

(ii) the operator T is not injective, i.e. $\mathcal{N}(T) \neq \{0\}$, but $T|_{\mathfrak{D}}$ is injective, i.e. (4.5) holds. In this case we have $\Lambda = (f^{\dagger}_{\theta|W} + h_{\theta|W})$ where $h_{\theta|W} \in \mathcal{N}(T)$ is such that $\int_\Theta (f^{\dagger}_{\theta|W} + h_{\theta|W})(\theta, W) d\theta = 1$ and $(f^{\dagger}_{\theta|W} + h_{\theta|W})$ is non-negative *a.e.* on Θ, a.s. In this case we can also have $\Lambda = f^{\dagger}_{\theta|W}$ if $f^{\dagger}_{\theta|W}$ is a probability density function.

In our context, injectivity of $T|_{\mathfrak{D}}$ is determined by the structural economic model and depends on how C, Z and θ interact. When $T|_{\mathfrak{D}}$ is not injective, computation of Λ requires computation of the complete singular value expansion of the kernel of the operator T. In theory, because T is known and is not estimated, a researcher can compute the SVD of T, calculate the elements $\{\tilde{\varphi}_j\}_{j \in J_0}$ by a simple procedure of basis completion, like the Gram-Schmidt orthonormalization, and then characterize the null space of the operator, see Hansen (1988). In practice, a researcher must truncate the expansion at some point and impose that all singular values not computed equal zero. The error of this approximation can be bounded using methods in Hansen (1988).

It is well known that shape restrictions may provide identifying power. For example, see Matzkin (2007) or Blundell et al. (2007). Nonetheless, the econometric literature on inverse problems for the most part has not exploited the fact that point-identification can be obtained even without injectivity of T because $T|_{\mathfrak{D}}$ may be injective.[8] We discuss this formally in the next Section where we provide a necessary and sufficient condition for point-identification that we call $\mathcal{T}$-completeness. This condition is weaker than the conditions of completeness or bounded completeness that have been used in the previous econometric literature on inverse problems.

4.1 Identification and completeness

Define $\mathcal{F}_{\theta|CWZ} = \{f \,|f$ is a conditional *pdf* on $(\mathbb{R}^d, \mathcal{B}_\Theta)$ given $(C, W, Z)\}$ as the set of *pdf*'s of θ conditional on (C, W, Z). Provided that $f_{C|WZ}$ and $f_{\theta|W}$ are bounded away from zero and infinity, injectivity of the operator T is equivalent to the requirement that $\mathcal{F}_{\theta|CWZ}$ is $L^2_{\pi_\theta}$-complete (or bounded-complete) as noted in Florens, Mouchart, & Rolin (1990), Florens (2003), Newey & Powell (2003), Blundell et al. (2007) and Hu & Schennach (2008) in different setups.

It is well-known that, if the elements of $\mathcal{F}_{\theta|CWZ}$ belong to the exponential family, then $\mathcal{F}_{\theta|CWZ}$ is $L^2_{\pi_\theta}$-complete. However, in our framework, neither $L^2_{\pi_\theta}$-completeness nor bounded completeness are equivalent to identification of $f_{\theta|W}$. Because the solutions of (4.3) are constrained to be *pdf*'s, identification of $f_{\theta|W}$ is equivalent to completeness of $\mathcal{F}_{\theta|CWZ}$ with respect to a class of functions smaller than both $L^2_{\pi_\theta}$ and the class of bounded functions. This class, that we denote by $\mathcal{T}$, is the class of functions that equal the difference between two densities scaled by the true density of θ. That is, $\mathcal{T} = \left\{h \in L^2_{\pi_\theta} : h = \frac{f}{f_{\theta|W}} \quad \text{for some } f \in \mathfrak{D}\right\} \subset L^2_{\pi_\theta}$ where $f_{\theta|W}$ is the true conditional *pdf* of θ given W. Mandelbaum & Ruschendorf (1987) provide more background on *completeness* of a probability distribution with respect to a general family of functions $\mathcal{T}$.

Completeness with respect to $\mathcal{T}$, or $\mathcal{T}$-completeness, is necessary and sufficient for identification in our framework. We summarize this discussion with a definition and a proposition.

Definition 1 ($\mathcal{T}$-completeness)**.** *A family of distributions $\mathcal{F}_{\theta|CWZ}$ is complete with respect to $\mathcal{T}$ if and only if*

$$h \in \mathcal{T} \textit{ and } \int_\Theta h f_{\theta|CWZ} d\theta = 0, \quad \textit{a.s. implies } h = 0, \quad \textit{a.s.}$$

for all $f_{\theta|CWZ} \in \mathcal{F}_{\theta|CWZ}$.

Proposition 2 ($\mathcal{T}$-completeness)**.** *Under the assumptions of Theorem 1, (4.5) holds if and only if $\mathcal{F}_{\theta|CWZ}$ is complete with respect to $\mathcal{T} \subset L^2_{\pi_\theta}$.*

Since the set $\mathcal{T}$ is strictly smaller than $L^2_{\pi_\theta}$, identification can be achieved even when $L^2_{\pi_\theta}$-completeness fails. In general, whether Proposition 2 is satisfied is a computational issue that must be checked on a case by case basis. The next proposition, while stronger than required, provides a sufficient condition for identification that can be more easily checked in practice and that provides some intuition as to what variation is required to secure identification.

Proposition 3. *Let Assumptions 1-5 hold. Assume that* $\forall i = 1, \ldots, s$, $(f_{\varepsilon|\theta WZ} \circ \varphi_i^{-1})(c, w, z, \theta)$ *is of the form*

$$\exp\left[\tau_i\left(c, w, z\right)' m_i\left(\theta\right)\right] h_i(\theta) k_i(c, w, z), \qquad i = 1, \ldots, s$$

where h_i *is a positive function,* m_i *is a vector-valued invertible function whose image has dimension equal to* d *where* d *is the dimension of* θ*. The functions* τ_i *and* k_i *are real-valued and* k_i *is positive and bounded. Further, the rank of* $\mathbb{E}\left(\tau_i'\tau_i\right)$ *is equal to* d *and the vector* τ_i *varies over* $\mathbb{R}^d$*. Then,* $f_{\theta|W}$ *is identified with respect to the class* $\mathcal{F}_{\theta|W}$*.*

The conditions of this proposition are satisfied in the linear random coefficient model outlined in Section 2 as long as $g(Z_2, W)$ has support on the entire real line. Moreover, if Θ is bounded, then more limited variation of τ_i is sufficient to obtain the result of the lemma. The conditions of the proposition are also satisfied in the classical examples of the additively-closed and the location-scale one-parameter family of distributions. We detail these classes in the online Supplement. In contrast, the conditions are not satisfied in Example 2.

Example 2 (Continued)**.** *Suppressing the time subscript,* (3.3) *implies*

$$f_{C|WZ}\left(c, w, z\right) = \int_{\Theta} \frac{\exp\left\{-\frac{1}{2}\left[\frac{c - \varphi_1 w_1 - \varphi_2 (w_2 + z) - m(\gamma, \beta)}{\sigma_\varepsilon}\right]^2\right\}}{\sqrt{2\pi\sigma_\varepsilon^2}} f_{\gamma\beta|W}\left(\gamma, \beta, w\right) d\gamma d\beta. \qquad (4.6)$$

Define $\delta = m\left(\gamma, \beta\right)$*. Denote the supports of* (δ, γ) *by* D *and* Γ *respectively. Then* (4.6) *can be rewritten*

$$\begin{aligned} f_{C|WZ}\left(c, w, z\right) &= \int_D \frac{\exp\left\{-\frac{1}{2}\left[\frac{c - \varphi_1 w_1 - \varphi_2 (w_2 + z) - \delta}{\sigma_\varepsilon}\right]^2\right\}}{\sqrt{2\pi\sigma_\varepsilon^2}} \widetilde{f}_{\delta|W}\left(\delta, w\right) \left[\int_\Gamma \widetilde{f}_{\gamma|W\delta}\left(\gamma, \delta, w\right) d\gamma\right] d\delta \\ &= \int_D \frac{\exp\left\{-\frac{1}{2}\left[\frac{c - \varphi_1 w_1 - \varphi_2 (w_1 + z) - \delta}{\sigma_\varepsilon}\right]^2\right\}}{\sqrt{2\pi\sigma_\varepsilon^2}} \widetilde{f}_{\delta|W}\left(\delta, w\right) d\delta. \end{aligned}$$

The joint density of (γ, δ) *is not point-identified because any conditional density* $\widetilde{f}_{\gamma|W\delta}\left(\gamma, \delta, w\right)$ *is consistent with the data. The conditions of Proposition 3 are not satisfied. However, the marginal density* $\widetilde{f}_{\delta|W}\left(\delta, w\right)$ *is point-identified. The identified set* Λ *is the set containing all elements of the form*

$$f_{\gamma\beta|W}\left(\gamma, \beta, w\right) = \widetilde{f}_{\delta|W}\left[m\left(\gamma, \beta\right), w\right] \widetilde{f}_{\gamma|W\delta}\left[\gamma, m\left(\gamma, \beta\right), w\right] \left|\frac{\partial m}{\partial \beta}\right|$$

for some conditional density $\widetilde{f}_{\gamma|W\delta}$*. In Section 6, we show in simulations that despite this failure of point-identification of* $f_{\gamma\beta|W}$*, the model has identifying power because our estimate of* $\widetilde{f}_{\delta|W}\left(\delta, w\right)$ *places meaningful bounds on the identified set. For example, Figure 5 shows that the probability that* γ *is between 2 and 2.5 and* β *is between 0.95 and 0.96 is identified. Joint densities of* (β, γ) *that* *are inconsistent with the estimated probability of this event are ruled out.*

5 Estimation

Our estimation strategy is based on equation (4.3). While the solution of (4.3) need not be unique, there is a unique solution of minimal norm which we denote by $f_{\theta|W}^{\dagger c}$. This solution takes the form

$$f_{\theta|W}^{\dagger c} = T_{\mathcal{F}_{\theta|W}}^{\dagger} f_{C|WZ} \tag{5.1}$$

where $T_{\mathcal{F}_{\theta|W}}^{\dagger}$ denotes the constrained generalized inverse of the restricted operator $T|_{\mathcal{F}_{\theta|W}}$ (see e.g., Neubauer, 1988, Definition 2.1). The definition of $f_{\theta|W}^{\dagger c}$ differs from the definition of $f_{\theta|W}^{\dagger}$ since the latter is not constrained. However, in some cases (for instance in the point identified case): $f_{\theta|W}^{\dagger c} = f_{\theta|W}^{\dagger}$. Note that the operator $T_{\mathcal{F}_{\theta|W}}^{\dagger}$ is nonlinear and noncontinuous since, in general, $\mathcal{R}(T|_{\mathcal{F}_{\theta|W}})$ is not closed. As a result, the inverse problem of recovering $f_{\theta|W}$ from (4.3) suffers from ill-posedness. Consistent estimation of $f_{\theta|W}$ based on (5.1) requires regularization.

To implement our estimation procedure we assume that a nonparametric consistent estimator of $f_{C|WZ}$ is available.

Assumption 7. *Let (c_i, w_i, z_i), $i = 1, \ldots, n$ be an i.i.d. sample of (C, W, Z) that is used to construct an estimator $\hat{f}_{C|WZ}$ of $f_{C|WZ}$ such that $\hat{f}_{C|WZ} \in L^2_{\pi_{cz}}$ a.s. and $\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2 \to 0$ as $n \uparrow \infty$.*

Once an estimator $\hat{f}_{C|WZ}$ has been computed, we use a constrained Tikhonov-type estimator for $f_{\theta|W}$. This estimator is the infinite dimensional counterpart of Ridge regression. It is defined as the minimizer, with respect to h, of

$$||Th - \hat{f}_{C|WZ}||^2 + \alpha||h||^2, \qquad h \in \mathcal{F}_{\theta|W}, \tag{5.2}$$

where the *regularization* parameter $\alpha > 0$ decreases to 0 at a suitable rate as $n \uparrow \infty$.

Alternatively, estimation could be based on a semi-nonparametric sieve penalized maximum likelihood estimator (MLE). The Tikhonov-type estimator is computationally simple, is guaranteed to converge, and the eigenvalues and eigenfunctions are computed as part of the estimation procedure. In contrast, sieve penalised MLE may lack these features. One advantage of the sieve MLE approach is that it is relatively straightforward to impose the constraints of the model when estimating the density $f_{C|WZ}$. When the model is correctly specified, this may result in computational efficiency gains.

We develop the classical Tikhonov case where the penalty term $||h||^2$ is based on the $L^2_{\pi_\theta}$ norm. This penalty has the benefit of being easy to compute and well understood in the literature. From an economic point of view, since the minimum norm element is closest to the origin, heuristically, it may have the smallest impact on counterfactual predictions. Alternatively, if a researcher has a prior belief on $f_{\theta|W}$ based on previous research, then the penalty can be replaced by $\|h - f^o_{\theta|W}\|^2$ or by the entropy $\int \log(h/f^o_{\theta|W})hd\theta$ where $f^o_{\theta|W}$ is the researcher's prior belief about the density.

Since the norms in (5.2) depend on π_θ and π_{cz}, choice of the weighting functions can be important. As noted after Proposition 1, the weights can be chosen so that the operator T is compact. In addition, the weights π_{cz} and π_θ should be chosen to reflect the researcher's loss function. For example, a researcher may choose to place greater weight on some values of C or Z than others to reflect greater economic importance. Or, he/she may place greater weights on some values of θ to reflect prior beliefs about the distribution of θ. In our simulations in Section 6, we use constant weights that weight all values equally.

We propose two methods to compute the minimizer of (5.2). The first method is a two-step procedure that first computes the unconstrained Tikhonov regularized estimator and then projects it onto the closed and convex set $\mathcal{F}_{\theta|W}$. The second method uses numerical methods to directly solve the constrained minimization problem in (5.2).

The first estimator is simple. The first step has a closed-form and the second step consists of a simple iterative procedure. As a result, in many cases it will be very fast to compute. On the other hand, the two-step estimator is only consistent if $f^{\dagger c}_{\theta|W} = f^\dagger_{\theta|W}$. The second estimator we propose overcomes this problem. It does not have a closed form but works regardless of whether $f^{\dagger c}_{\theta|W} = f^\dagger_{\theta|W}$ holds or not.

When point-identification fails, our estimator converges to the minimum norm element in the identified set. This element is easy to compute, and, once computed, can be used to estimate the set Λ using the formula detailed in Theorem 2. The procedure is straightforward: estimate $f^\dagger_{\theta|W}$, compute the eigenfunctions of T, and then construct the identified set as described in Theorem 2. Because the operator T and its eigenfunctions are not estimated, consistency of the estimated set Λ is guaranteed by consistency of the estimator of $f^\dagger_{\theta|W}$ so that the estimated Λ will contain the true $f_{\theta|W}$ asymptotically.

The first step of our two-step estimator has been previously studied, for example by Darolles et al. (2011) and Carrasco & Florens (2011). In our setting, the expression for the estimator is somewhat different from that in Carrasco & Florens (2011).[9] We provide asymptotic properties of the two-step estimator and extend previous results by considering also the important case where the problem is severely ill-posed and the *pdf* $f_{\theta|W}$ is not analytic. Therefore, the rates given in Corollary 1 below, and the asymptotic normality results are new and not provided in the previous literature. These rates are given for the case where $\hat{f}_{C|WZ}$ is obtained using kernel smoothing.

We also provide the convergence rate for the constrained estimator and discuss how the regularity condition required for the unconstrained case must be modified to obtain the convergence rate in the constrained case. To the best of our knowledge, these results are available only for deterministic inverse problems and not for stochastic inverse problems which are relevant in econometrics.

5.1 Estimation of $f^{\dagger c}_{\theta|W}$: a two-step approach

The two-step estimator is computed as follows.

First step. Compute the solution, denoted by $\hat{f}^{\alpha}_{\theta|W}$, of the unconstrained problem:

$$\min_{h \in L^2_{\pi_\theta}} \left\{ ||Th - \hat{f}_{C|WZ}||^2 + \alpha ||h||^2 \right\}. \tag{5.3}$$

The solution is the classical Tikhonov regularized estimator:

$$\hat{f}^{\alpha}_{\theta|W}(\theta, w) = (\alpha I + T^* T)^{-1} T^* \hat{f}_{C|WZ} \tag{5.4}$$

where I denotes the identity operator in $L^2_{\pi_\theta}$. When T is compact, expression (5.4) simplifies to $\hat{f}^{\alpha}_{\theta|W}(\theta, w) = \sum_{j=1}^{\infty} \lambda_j (\alpha + \lambda_j^2)^{-1} \langle \hat{f}_{C|WZ}, \psi_j \rangle \varphi_j(\theta, w)$ where $\{\lambda_j, \psi_j, \varphi_j\}_{j \in \mathbb{N}}$ denotes the SVD of T.

Second step. Compute the orthogonal projection, denoted by $\mathcal{P}_c \hat{f}^{\alpha}_{\theta|W}$, of $\hat{f}^{\alpha}_{\theta|W}$ onto the set $\mathcal{F}_{\theta|W}$:

$$\mathcal{P}_c \hat{f}^{\alpha}_{\theta|W} = \max \left\{ 0, \hat{f}^{\alpha}_{\theta|W} - \frac{c}{\pi_\theta} \right\} \tag{5.5}$$

where c is such that $\int_\Theta \mathcal{P}_c \hat{f}^{\alpha}_{\theta|W} d\theta = 1$.

We call $\mathcal{P}_c \hat{f}^{\alpha}_{\theta|W}$ the *indirect Tikhonov regularized estimator* of $f^{\dagger c}_{\theta|W}$. Gajek (1986) shows that the projection can be computed simply using the following iterative algorithm.

$\mathcal{P}_c$**−algorithm:**

1. Set $\hat{f}^{\alpha(0)}_{\theta|W} = \hat{f}^{\alpha}_{\theta|W}$ and $k = 0$.
2. Set $\hat{f}^{\alpha(k+1)}_{\theta|W} = \max\{0, \hat{f}^{\alpha(k)}_{\theta|W}\}$ and check $C_{k+1} = \int_\Theta \hat{f}^{\alpha(k+1)}_{\theta|W}(\theta, w) d\theta$. If $C_{k+1} = 1$ stop. Otherwise,
3. Set $\hat{f}^{\alpha(k+2)}_{\theta|W} = \hat{f}^{\alpha(k+1)}_{\theta|W} - \frac{(C_{k+1} - 1)}{\pi_\theta \int_\Theta \frac{1}{\pi_\theta} d\theta}$.
4. Set $k = k + 2$ and repeat 2 - 4 until $|C_{k+1} - 1| < \epsilon$, for a small $\epsilon > 0$.

While other projection methods exist, Gajek (1986) shows that this algorithm converges pointwise and in norm to $\mathcal{P}_c \hat{f}^{\alpha}_{\theta|W}$ and that $\mathcal{P}_c \hat{f}^{\alpha}_{\theta|W}$ minimizes the weighted MISE $\mathbb{E}|| \cdot ||^2$.

Several procedures for choosing α have been proposed in the literature. These include: cross-validation, procedures based on extensions of the discrepancy principle of Morozov (1993) (see Darolles et al., 2011; Florens & Simoni, 2012), or the Bayesian-type procedures proposed in Florens & Simoni (2016) and Johannes, Simoni, & Schenk (2015). In the simulations in Section 6, we use cross-validation. To the best of our knowledge, among these procedures, the only procedures that have been proved to be adaptive are the last two.

5.1.1 Rates of convergence

The two-step estimator is consistent when $f^{\dagger c}_{\theta|W} = f^{\dagger}_{\theta|W}$, that is, when $f^{\dagger}_{\theta|W} \in \mathcal{F}_{\theta|W}$. This is possible for instance when either T or $T|_{\mathfrak{D}}$ are injective. Theorem 3 below provides the rate of the (weighted) Mean Integrated Square Error (MISE) associated with the two-step estimator $\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W}$. The rate depends on the smoothness of $f^{\dagger c}_{\theta|W}$ and on the smoothness of T. The next assumption (a type of *source condition*[10]) quantifies the smoothness of $f^{\dagger c}_{\theta|W}$ relative to the smoothness of T. It is only required to derive the rate of convergence of the estimator.

Assumption 8. *Let $\phi : [0, \infty) \to [0, \infty)$ be a continuous, strictly increasing function with $\phi(0) = 0$. Let $T : L^2_{\pi_\theta} \to L^2_{\pi_{cz}}$ be as defined in (4.2) and bounded. There exists a source $\nu \in L^2_{\pi_\theta}$ such that for some $0 < M < \infty$,*

$$f^{\dagger c}_{\theta|W} = \phi(T^*T)\nu \qquad \textit{and} \qquad ||\nu|| \leq M.$$

When the operator T is finitely smooth (mildly ill-posed case) and $f^{\dagger c}_{\theta|W}$ belongs to a Sobolev class of functions, then an appropriate choice of ϕ is $\phi(t) = t^{\beta/2}$ for some $\beta > 0$. For example, in the intertemporal consumption model, this choice of ϕ is appropriate if $f^{\dagger c}_{\theta|W}$ is infinitely differentiable. In contrast, when T is infinitely smooth (severely ill-posed case) and $f^{\dagger c}_{\theta|W}$ is not analytic, then an appropriate choice of ϕ is $\phi(t) = (-\log(t))^{-\beta/2}$ for some $\beta > 0$. In this latter case, the rate of convergence is much slower.

The following theorem states the rate of convergence:

Theorem 3. *Let Assumptions 1-5 and 7-8 be satisfied, and $f^{\dagger c}_{\theta|W} = f^{\dagger}_{\theta|W} \in \mathcal{F}_{\theta|W}$. Assume that there exists a constant γ_ϕ such that*

$$\sup_{t \in \sigma(T^*T)} \left|\phi(t)\alpha(\alpha + t)^{-1}\right| \leq \gamma_\phi \phi(\alpha), \qquad \alpha \to 0 \tag{5.6}$$

where $\sigma(T^*T)$ denotes the spectrum of T^*T. Then, the weighted MISE associated with $\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W}$ is $\mathbb{E}||\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 = \mathcal{O}(\phi^2(\alpha) + \alpha^{-1}\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2)$. If $\phi(t) = t^{\beta/2}$ with $\beta > 0$ then, (5.6) is satisfied for $\beta \leq 2$ and

$$\inf_{\alpha>0} \mathbb{E}||\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 = \mathcal{O}\left([\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2]^{\frac{\beta\wedge 2}{\beta\wedge 2+1}}\right).$$

If $\phi(t) = (-\log(t))^{-\beta/2}$ with $\beta > 0$ and (5.6) is satisfied then

$$\inf_{\alpha>0} \mathbb{E}||\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 = \mathcal{O}\left(\left[-\log\left(\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2\right)\right]^{-\beta}\right).$$

In the case $\phi(t) = (-\log(t))^{-\beta/2}$, Hohage (2000) shows that (5.6) holds automatically for $0 < \beta \leq 2$. The rate in the theorem is at most of order $[\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2]^{\frac{2}{3}}$ and hence is slower than the minimax rate for density function estimation because we use indirect observations of θ to estimate $f_{\theta|W}$. Let $\hat{f}^{\alpha(k)}_{\theta|W}$ be the two-step estimator obtained by using the $\mathcal{P}_c$-algorithm. It is possible to show that $\mathbb{E}||\hat{f}^{\alpha(k)}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 \leq \mathbb{E}||\hat{f}^{\alpha}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2$. Therefore, this theorem and its proof also provide the rate of convergence for the approximation of $\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W}$ obtained from the $\mathcal{P}_c$-algorithm.

The rate of Theorem 3 can be made explicit if the rate of convergence of $\hat{f}_{C|WZ}$ is known. Consider the case where $\hat{f}_{C|WZ}$ is a kernel estimator. Suppose

$$\hat{f}_{C|WZ}(c,w,z) = \frac{\frac{1}{nh_n^{1+k+l}}\sum_{i=1}^n K_h(c_i - c, c)K_h(w_i - w, w)K_h(z_i - z, z)}{\frac{1}{nh_d^{k+l}}\sum_{l=1}^n K_h(w_l - w, w)K_h(z_l - z, z)}, \tag{5.7}$$

where h_n and h_d denote bandwidths and $K_h(\cdot,\cdot)$ is a generalized kernel function[11] of order $r = 2$. Assume without loss of generality that $\mathcal{C} = [0,1]$, $\mathcal{W} = [0,1]^k$, $\mathcal{Z} = [0,1]^l$. By standard Taylor series arguments, as in Rosenblatt (1969), it is easy to show that $\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2 = \mathcal{O}\left(\frac{1}{n\min\{h_n,h_d\}^{k+l+1}} + \max\{h_n^4, h_d^4\}\right)$. If $h_n = h_d = h$ is chosen such that $\frac{1}{nh^{k+l+1}} \asymp h^4$ then $\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2 = \mathcal{O}(n^{-4/(k+l+1+4)})$. Plugging this into the expression from Theorem 3, for the case $\phi(t) = t^{\beta/2}$, we obtain

$$\inf_{\alpha>0}\mathbb{E}||\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 = \mathcal{O}\left(n^{-\frac{4(\beta\wedge 2)}{(k+l+1+4)(\beta\wedge 2+1)}}\right) \tag{5.8}$$

and for the case $\phi(t) = (-\log(t))^{-\beta/2}$, we obtain $\inf_{\alpha>0}\mathbb{E}||\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 = \mathcal{O}((-\log(1/n))^{-\beta})$.

We show now that this rate can be improved and made independent of the dimension of Z. This is possible since applying the operator T^* to the error term $(\hat{f}_{C|WZ} - f_{C|WZ})$ has a smoothing effect, integrating out (C,Z), so that the dimension of (C,Z) does not play any role in the rate. The following corollary to Theorem 3 gives the new rate.

Corollary 1. *Let Assumptions 1-5, 7-8 and* (5.6) *be satisfied, and* $f^{\dagger c}_{\theta|W} = f^{\dagger}_{\theta|W} \in \mathcal{F}_{\theta|W}$ *and* $\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W}$ *be the two-step estimator computed by using* $\hat{f}_{C|WZ}(c,w,z)$ *defined in (5.7). Then,* $\mathbb{E}||\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 = \mathcal{O}(\phi^2(\alpha) + \alpha^{-2}\left(\max\{h_n^4, h_d^4\} + n^{-1}(\min\{h_n, h_d\})^{-k}\right))$. *Moreover, if* $h_n = h_d \asymp n^{-1/(4+k)}$ *and* $\phi(t) = t^{(\beta\wedge 2)/2}$ *we have* $\inf_{\alpha}\mathbb{E}||\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 = \mathcal{O}\left(n^{-\frac{4(\beta\wedge 2)}{(4+k)((\beta\wedge 2)+2)}}\right)$.

The rate in Corollary 1 is faster than the rate in (5.8) if $(l+1)(\beta\wedge 2+1) > 4+k$. Under the conditions of the corollary, if we have no W and if $h_n = h_d \asymp n^{-1/4}$ then $\mathbb{E}||T^*(\hat{f}_{C|WZ} - f_{C|WZ})||^2 = \mathcal{O}(n^{-1})$. The rate is increasing in β and decreasing in the dimension of W. So, there is a curse of dimensionality only in the dimension of the endogenous variables W and not in the dimension of the instruments Z. This is due to the action of the operator T^* that integrates out (C,Z).

In some applications, the covariate W may need to be estimated. If this is the case, the estimation error in the generated covariate $\hat{W}$ would affect the rates of convergence of $\hat{f}_{C|\hat{W}Z}$ (see Mammen, Rothe, & Schienle (2012)) and then of our estimator. This can be seen from Theorem 3. Moreover, in such applications the operator T would be estimated. Thus, the rate of convergence of nonparametric estimators of T and T^* would have to be taken into account when computing the rate of the estimator of $f_{\theta|W}$. Nevertheless, under assumptions similar to Assumptions A.3 and A.4 in Darolles et al. (2011) and with appropriate choice of tuning parameters, the error of the estimated operator can be made negligible so that it does not affect the rate of convergence of the estimator of $f_{\theta|W}$ (cf. also footnote 4).

5.1.2 Asymptotic normality

We now study pointwise asymptotic normality of the two-step estimator $\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W}$ in the case where $\hat{f}_{C|WZ}$ is estimated using kernel estimator (5.7). For that we introduce the following technical assumption which uses the SVD of T introduced in (4.4).

Assumption 9. *Let $\mathbb{E}_{WZ}$ denote the conditional expectation given (W,Z) and $\hat{f}_{WZ}$ denote the kernel estimator of the joint pdf of (W,Z). We assume that for every $\theta \in \Theta$ and $w \in \mathcal{W}$: (i)*

$$\mathbb{E}\left|\sum_{j=1}^{\infty}\frac{\lambda_j}{\alpha+\lambda_j^2}\left\langle (K_h(c_i-c,c)-\mathbb{E}_{WZ}(K_h(c_i-c,c)))\frac{K_h(z_i-z,z)K_h(w_i-w,w)}{h_n^{k+l+1}\hat{f}_{WZ}},\psi_j\right\rangle\varphi_j(\theta,w)\right|^3 = \mathcal{O}\left(\alpha^{-3/2}h_n^{-2k}\right)$$

and (ii) there exists a constant $\kappa > 0$ such that

$$Var\left(\sum_{j=1}^{\infty}\frac{\lambda_j}{\alpha+\lambda_j^2}\left\langle (K_h(c_i-c,c)-\mathbb{E}_{WZ}(K_h(c_i-c,c)))\frac{K_h(z_i-z,z)K_h(w_i-w,w)}{h_n^{k+l+1}\hat{f}_{WZ}},\psi_j\right\rangle\varphi_j(\theta,w)\right) > \kappa\alpha^{-2}h_n^{-k}.$$

In the following theorem we use the notation '⇒' to denote pointwise convergence in distribution.

Theorem 4. *Let Assumptions 1-9 and (5.6) hold, $f^{\dagger c}_{\theta|W} = f^{\dagger}_{\theta|W} \in \mathcal{F}_{\theta|W}$ and $\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W}$ be the two-step estimator computed by using $\hat{f}_{C|WZ}(c,w,z)$ defined in (5.7). Let $\mathbb{E}_{WZ}$ denote the conditional expectation given (W,Z) and $\hat{f}_{WZ}$ denote the kernel estimator of the joint pdf of (W,Z). If $n\alpha h_n^{k+4} \to 0$, $\alpha^3/(h_n^k n) \to 0$ and if $n\alpha^2 h_n^k\phi^2(\alpha) \to 0$ as $n \uparrow \infty$, then for every $\theta \in \Theta$ and $w \in \mathcal{W}$:*

$$\frac{\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W}(\theta,w) - f^{\dagger c}_{\theta|W}(\theta,w)}{\sqrt{V_c(\theta,w)}} \Rightarrow N(0,1)$$

where

$$V_c(\theta,w) \quad = \quad \frac{1}{n}Var\left(\mathcal{P}_c^{\dagger}\sum_{j=1}^{\infty}\frac{\lambda_j}{(\alpha+\lambda_j^2)}\left\langle (K_h(c_i-c,c)-\mathbb{E}_{WZ}(K_h(c_i-c,c)))\frac{K_{h,i}(z,w)}{h^{k+l+1}\hat{f}_{WZ}},\psi_j\right\rangle\varphi_j(\theta,w)\right),$$

$K_{h,i}(z,w) = K_h(z_i-z,z)K_h(w_i-w,w)$ and $\mathcal{P}_c^{\dagger}$ denotes the projection on the tangent cone of $\mathcal{F}_{\theta|W}$ at $f^{\dagger c}_{\theta|W}$ defined as $\overline{\{\lambda(f - f^{\dagger c}_{\theta|W}) : \lambda \geq 0,\ f \in \mathcal{F}_{\theta|W}\}}$.

In order to obtain this asymptotic normality result, we require a regularization parameter α that converges to 0 at a faster rate than the asymptotically optimal one. This guarantees that the bias of $\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W}(\theta,w)$ is asymptotically negligible.

5.2 Estimation of $f^{\dagger c}_{\theta|W}$: constrained Tikhonov regularization

When $f^{\dagger c}_{\theta|W} \neq f^{\dagger}_{\theta|W}$ the two-step procedure can no longer be applied. Instead, we have to compute the constrained Tikhonov regularized solution by directly solving the minimization problem

$$\min_{h\in\mathcal{F}_{\theta|W}} \left\{||Th - \hat{f}_{C|WZ}||^2 + \alpha||h||^2\right\}. \tag{5.9}$$

The existence of a unique solution to problem (5.9) is proved in Neubauer (1988). A closed-form solution of this problem does not exist and numerical methods must be used to compute a solution. We denote by $\check{f}^{\alpha,c}_{\theta|W}$ the estimator obtained by solving (5.9) and by $P_{\mathcal{F}_{\theta|W}}$ the orthogonal projector of $L^2_{\pi_\theta}$ onto $\mathcal{F}_{\theta|W}$. The next theorem states consistency of the estimator $\check{f}^{\alpha,c}_{\theta|W}$.

Theorem 5. *Let Assumptions 1-5 and 7 hold. Let T be a bounded operator and let $f^{\dagger c}_{\theta|W} \in \mathcal{R}(P_{\mathcal{F}_{\theta|W}}T^*)$. Then, if $\alpha \to 0$ and $\alpha^{-1}\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2 \to 0$ as $n \uparrow \infty$, then $\mathbb{E}||\check{f}^{\alpha,c}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 \to 0$.*

Under a smoothness assumption about $f^{\dagger c}_{\theta|W}$, it is also possible to recover the convergence rate for the constrained estimator. To derive this rate, a regularity condition for $f^{\dagger c}_{\theta|W}$ different from Assumption 8 is required. This regularity condition is stated as follows.

Define $M = \{ f \in \mathcal{N}(T)^{\perp} \big| P_{\mathcal{F}_{\theta|W}} f = f^{\dagger c}_{\theta|W}\}$ and, $\forall f \in M$, $\tilde{L}_f = \{h \in L^2_{\pi_\theta}; \langle f^{\dagger c}_{\theta|W} - f, h\rangle = 0\}$. Define $\tilde{P}_f$ to be the orthogonal projector of $L^2_{\pi_\theta}$ onto $\tilde{L}_f$. Finally, for $\beta > 0$ define $N_\beta = \{f \in M : \tilde{P}_f f^{\dagger c}_{\theta|W} \in \mathcal{R}\left((\tilde{P}_f T^* T \tilde{P}_f)^{\beta/2}\right)\}$. The regularity condition is stated in terms of the set N_β.

Theorem 6. *Let Assumptions 1-5 and 7 hold and let T be a bounded operator. Suppose $f^{\dagger c}_{\theta|W} \neq f^{\dagger}_{\theta|W}$ and $N_\beta \neq \varnothing$. Then, $\mathbb{E}||\check{f}^{\alpha,c}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 = \mathcal{O}\left(\alpha^{\beta\wedge 2} + \alpha^{-1}\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2\right)$ and if $\alpha \asymp (\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2)^{\frac{1}{(\beta\wedge 2)+1}}$: $\mathbb{E}||\check{f}^{\alpha,c}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 = \mathcal{O}\left([\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2]^{\frac{\beta\wedge 2}{\beta\wedge 2+1}}\right)$.*

To understand the condition $N_\beta \neq \emptyset$, consider Figure 1, adapted from Neubauer (1986). The figure shows a finite dimensional version of the result. The set N_β is the set of all functions $f \in \mathcal{N}(T)^{\perp}$ for which both the orthogonal projection onto $\mathcal{F}_{\theta|W}$ equals $f^{\dagger c}_{\theta|W}$ and the orthogonal projection of $f^{\dagger c}_{\theta|W}$ onto the hyperplane $\tilde{L}_f$ is "smooth". The regularity condition on $f^{\dagger c}_{\theta|W}$ is imposed via the smoothness of its projection $\tilde{P}_f f^{\dagger c}_{\theta|W}$. Smoothness of the projection is measured by smoothness of the operator $(\tilde{P}_f T^* T \tilde{P}_f)^{\beta/2}$. When N_β is not empty, there exists at least one $f \in M$ such that the projection of $f^{\dagger c}_{\theta|W}$ on the corresponding hyperplane $\tilde{L}_f$ has smoothness β. This regularity condition is analogous to an approximation condition in classical nonparametric estimation. In that case, regularity of the function is implied by the assumption that the function can be well approximated by a smooth function. Here, the approximating function is the projection $\tilde{P}_f f^{\dagger c}_{\theta|W}$.

5.3 Case with non-random parameters

Suppose $\theta = (\theta_1, \theta_2)$ where θ_1 is deterministic. Let θ_1 belong to a compact set $\Theta_1 \subset \mathbb{R}^{d_1}$ and let θ_2 be a vector of random parameters (with dimension d_2) distributed according to a probability

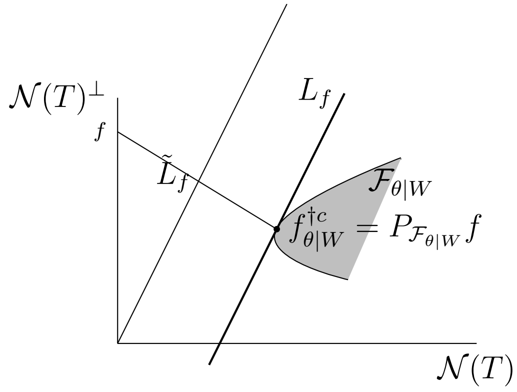


Figure 1: Representation of (part of) the set $\mathcal{F}_{\theta|W}$ (grey area), the supporting hyperplane $L_f = \{h \in L^2_{\pi_\theta} : \langle f^{\dagger c}_{\theta|W} - f, h - f^{\dagger c}_{\theta|W} \rangle = 0\}$ of $\mathcal{F}_{\theta|W}$ in $f^{\dagger c}_{\theta|W}$, the hyperplane $\tilde{L}_f$ and an element f of M.

distribution $P_{\theta_2|W}$ satisfying Assumptions 4 and 5. In this case, we can use either of the two estimation procedures we have proposed, after some minor modifications.

Here, we focus on the constrained Tikhonov regularization procedure.[12] The minimization problem (5.9) should be replaced by

$$\min_{\theta_1 \in \Theta_1,\, h \in \mathcal{F}_{\theta_2|W}} \left\{ ||T_{\theta_1} h - \hat{f}_{C|WZ}||^2 + \alpha ||h||^2 \right\} \tag{5.10}$$

where we write T_{θ_1} to make explicit the dependence of the operator on θ_1. The kernel of T_{θ_1} is (up to the factor $\frac{1}{\pi_\theta}$) equal to $f_{C|WZ\theta}(c, w, z, \theta_1, \theta_2)$ which has the same expression as in (4.2).

Let the parameter space be $\mathcal{G} = \Theta_1 \times \mathcal{F}_{\theta_2|W}$. Denote by $g = (\theta_1, h)$ a generic element of $\mathcal{G}$, with $\theta_1 \in \Theta_1$ and $h \in \mathcal{F}_{\theta_2|W}$. Let $g^0 = (\theta_1^0, f^0_{\theta_2|W})$ be the true value of g. Moreover, denote $\widehat{Q}_n(\theta_1, h) = ||T_{\theta_1} h - \hat{f}_{C|WZ}||^2$ and $Q(\theta_1, h) = ||T_{\theta_1} h - f_{C|WZ}||^2$. The estimator computed by solving (5.10) will be denoted by $\hat{g} = (\hat{\theta}_1, \check{f}^{\alpha,c}_{\theta_2|W})$ and belongs to $\mathcal{G}$. Define $\|\hat{g} - g^0\| = \|\hat{\theta}_1 - \theta_1^0\|_E + \|\check{f}^{\alpha,c}_{\theta_2|W} - f^0_{\theta|W}\|$, where $\|\cdot\|_E$ denotes the Euclidean norm induced by the scalar product $\langle \cdot, \cdot \rangle_E$ in $\mathbb{R}^{d_1}$. Theorem 7 below states consistency of $\hat{g}$. We introduce the following assumption.

Assumption 10. *Let the following statements hold:*

i. *The subset $\Theta_1 \subset \mathbb{R}^{d_1}$ is compact.*

ii. *The family of functions $\{\widehat{Q}_n(\cdot, h) + \alpha \|h\|^2\}_{h \in \mathcal{F}_{\theta_2|W}}$ is equidifferentiable at every $\theta_1 \in \Theta_1$.*

iii. *Let $\widehat{Q}_{n,1}(\theta_1, h)$ denote the first derivative of $\widehat{Q}_n(\theta_1, h)$ with respect to θ_1 evaluated at θ_1. We assume that $\sup_{h \in \mathcal{F}_{\theta_2|W}} |\widehat{Q}_{n,1}(\theta_1, h)| < \infty$ for every $\theta_1 \in \Theta_1$.*

iv. *$Q(g^0) = 0$ and any $(\theta_1, h) \in \mathcal{G}$ that satisfies $Q(\theta_1, h) = 0$ also satisfies $\theta_1 = \theta_1^0$ and $h = f^0_{\theta_2|W}$ almost everywhere.*

v. *The function $f_{C|WZ\theta}(c, w, z, \theta_1, \theta_2)$ is continuous in θ_1 a.e..*

vi. *The criterion $\hat{Q}_n$ satisfies: $|\hat{Q}_n(g^0) - Q(g^0)| = O_p(\delta_n)$, where $\delta_n = o(1)$.*

Theorem 7. *Let Assumptions 1-5, 7 and 10 hold. Then: (i) a solution to* (5.10) *exists and (ii), if* $\delta_n = O(\alpha)$*:* $\|\hat{g} - g^0\| \to 0$ *in probability as* $n \uparrow \infty$*.*

6 Monte Carlo simulation

6.1 Simulation 1: Linear endogenous random coefficient model

Consider model (2.1). Assume that $g(Z_2, W) = Z_2 W$. Then equation (2.1) becomes

$$C = \theta_1 Z_1 + \theta_2 Z_2 W + \varepsilon. \tag{6.1}$$

Assume that $\varepsilon \sim N(0, 0.1)$, $W \sim U[1,2]$, $Z \sim N(0, \Sigma_z)$ with $Z = (Z_1, Z_2)$ and Σ_z is equal to the identity matrix. Finally, assume that $\theta|W \sim N(\mu_\theta, \Sigma_\theta)$ with $\theta = (\theta_1, \theta_2)$, $\mu_\theta(1) = \mu_\theta(2) = \beta_0 + \beta_1 W$, $(\beta_0, \beta_1) = (1,1)$ and Σ_θ equal to σ_θ^2 times the identity matrix, where $\sigma_\theta^2 > 0$.

We simulate 2000 Monte Carlo datasets from this model, 500 for each sample size of 100, 500, 1000 and 5000. For each dataset, we first estimate $\widehat{f_{C|WZ}}$ using a Gaussian product kernel density estimator with bandwidth chosen as discussed below. Then we compute $\hat{f}^{\alpha}_{\theta|W}$ using (5.4). Finally, we compute $\mathcal{P}_c \widehat{f^{\alpha}_{\theta|W}}$ as in (5.5), at the 30th, 50th, and 70th percentiles of the distribution of W.

To facilitate accurate numerical integration, we first make a change of variable, mapping (C, Z_1, Z_2) into the region $[-1,1]^3$. Specifically, we define $U_c = 2\Phi\left(\frac{C-\mu_c}{\sigma_c}\right) - 1$, $U_1 = 2\Phi\left(\frac{Z_1 - \mu_{z_1}}{\sigma_{z_1}}\right) - 1, U_2 = 2\Phi\left(\frac{Z_2 - \mu_{z_2}}{\sigma_{z_2}}\right) - 1$, where Φ is the standard normal CDF and (μ_c, σ_c), $(\mu_{z_i}, \sigma_{z_i})$, $i = 1,2$ are the empirical mean and standard deviation of C, Z_1 and Z_2, respectively. Substituting these new variables into (6.1) , and solving for ε, the structural function $\varepsilon = \varphi^{-1}(W, Z, \theta, \varepsilon)$ can be written as

$$\varepsilon = \mu_c + \sigma_c \Phi^{-1}\left(\frac{U_c + 1}{2}\right) - \theta_1\left(\mu_{z_1} + \sigma_{z_1}\Phi^{-1}\left(\frac{U_{z_1}+1}{2}\right)\right) - \theta_2\left(\mu_{z_2} + \sigma_{z_2}\Phi^{-1}\left(\frac{U_{z_2}+1}{2}\right)\right) W.$$

Next, let w_{30}, w_{50}, w_{70} denote the 30th, 50th and 70th percentile of W, respectively. Using the weight functions $\pi_{cz} = 1$ and $\pi_\theta = 1$, for each $w \in \{w_{30}, w_{50}, w_{70}\}$, we then compute $\hat{f}^{\alpha}_{\theta|W}$ to solve

$$\min_h \left\{ \int \left(\hat{f}_{U_c|WU_{z_1}U_{z_2}} - Th\right)^2 dcdz_1dz_2 + \alpha \int h(\theta)^2 d\theta \right\}. \tag{6.2}$$

Figure 2 displays contour plots of the true density and of the estimated density of θ given W, for the three different quantiles of W, obtained from one of our Monte Carlo datasets (with $n = 1000$ and $\sigma_\theta^2 = 0.1$). In each panel of the figure, the top panel shows the true density and the bottom panel shows the estimate. In all cases both the shape and location of the estimate track the true density quite closely. In particular, the unimodality of the density is well covered, and the location of the mode almost exactly coincides with the true mode. Moreover, the spread and the elliptic shape of the density also very much coincides in every dimension with the true spread and elliptic shape of the density of random coefficients.

Results are obtained using bandwidths $h_n = h_d = 0.05$ and the Tikhonov regularization parameter $\alpha = 0.01$. Bandwidths are chosen to minimize the average of the square root of the density weighted mean squared error:

$$AMSE = \frac{1}{3*500}\sum_{r,q}\left(\int\left[\mathcal{P}_c\widehat{f}^{\alpha}_{\theta|W,r}\left(\theta,w_q\right)-f_{\theta|W}\left(\theta,w_q\right)\right]^2 f_{\theta|W}\left(\theta,w_q\right)d\theta\right)^{0.5} \tag{6.3}$$

where r indexes Monte Carlo replications and $w_q \in \{w_{30}, w_{50}, w_{70}\}$. That is, the average is calculated as the empirical average across 500 Monte Carlo replications and three quantiles of the distribution of W.

Table 1 compares the performance of our estimator to that of the oracle estimator (i.e., the infeasible kernel density estimator). It displays the AMSE of the estimators for different sample sizes and for different values of the variance of the random coefficients σ^2_θ.

Table 1: AMSE as a function of sample size

Sample Size		Tikhonov	Oracle	Ratio
$\sigma^2_\theta = 0.1$	100	0.631	0.418	1.51
	500	0.534	0.216	2.47
	1000	0.493	0.170	3.47
	5000	0.447	0.129	3.47
$\sigma^2_\theta = 0.5$	100	0.089	0.052	1.71
	500	0.064	0.041	1.56
	1000	0.056	0.039	1.44
	5000	0.050	0.028	1.79
$\sigma^2_\theta = 1.0$	100	0.041	0.024	1.71
	500	0.030	0.017	1.76
	1000	0.027	0.014	1.93
	5000	0.023	0.012	1.92

Several features are noteworthy: First, observe that the ratio is approximately twofold, which is not very large if one considers the small sample size and the complexity of the procedure. Second, note the absolute value of the AMSE decreases as the sample size increases, showing consistency. Third, the ratio of the two averages increases slightly. This is to be expected given the fact that the unfeasible oracle estimator converges faster. Nevertheless, the ratio is almost constant, largely between 1.5 and 2, suggesting that the theoretical large sample differences may slightly overstate the small sample differences. Finally, we see that when σ^2_θ increases the AMSE of both estimators decreases.

6.2 Simulation 2: Intertemporal consumption model

To analyse the CARA model of Example 2, we simulated $n = 1000$ agents. Each agent starts at age $t = 21$, works for 45 periods and then obtains a terminal retirement utility. In each period each agent faces a permanent i.i.d. income shock η_t distributed as $\eta_t \sim N(0, 0.01668)$. The initial value of income is set to 0.2 (scaled so that 0.2 equals \$20,000) and the initial permanent shock is set to zero. The interest rate is set to $R = 1 + r = 1.05$ and the random parameters γ and β have support on $(0.5, 4.0)$ and $(0.7, 0.999)$ respectively, covering a range of values suggested in the literature. The joint distribution of (γ, β) is generated as follows. We define $x \sim N\left(\mu_x, \sigma_x^2 I\right)$ with $x = (x_1, x_2)$, $\mu_x = (1, 0)'$, $\sigma_x^2 > 0$ and generate $\gamma = 0.5 + 3.5\,\Phi\left(x_1\right)$, and $\beta = 0.7 + 0.299\,\Phi\left(x_2\right)$, where Φ is the standard normal CDF. In addition, measurement error in consumption is $\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$ with σ_ε^2 set equal to 25% of the cross-sectional variance of consumption. We select one cross section, age 31, as the sample for our estimator. While the precise estimates vary with age as the model predicts, the qualitative results are quite similar for other cross sections.

As discussed in Section 4.1, the joint distribution of (γ, β) is not identified because the variables enter the kernel of the operator only through a single index. Instead we estimate the distribution of

$$\delta = \varphi_3 + \varphi_4 \gamma + \varphi_5 \frac{\ln\left(R\beta\right)}{\gamma} \tag{6.4}$$

where φ_3, φ_4, and φ_5 are parameters that depend only on the interest rate R and the time period t.

For the estimation, we use a Gaussian kernel with bandwidths $h_n = h_d = 0.3$ and with Tikhonov regularization parameter $\alpha = 0.01$. For the infeasible kernel density estimator we set the bandwidth to $h_\theta = 0.3$. While tuning parameters may be chosen using least-squares cross-validation, for the purposes of illustration, we chose tuning parameters to minimize the square root of density weighted mean squared error computed across the 500 Monte Carlo replications.

The distribution of δ conditional on W is difficult to compute because it is endogenously determined. Therefore we compute the following average of the square root of the density weighted mean squared error averaged across 500 Monte Carlo replications and 2 quantiles of the W distribution:

$$AMSE = \frac{1}{2 * 500} \sum_{r,q} \left(\int \left(\left[\mathcal{P}_c \widehat{f}^{\alpha}_{\delta|W,r}\left(\delta, w_q\right) - \widehat{f}^{Ker}_{\delta|W}\left(\delta, w_q\right) \right]^2 \right) \left(\widehat{f}^{Ker}_{\delta|W}\left(\delta, w_q\right) \right) d\delta \right)^{0.5}, \tag{6.5}$$

where r indexes Monte Carlo replications, $w_q \in \{w_{0.25}, w_{0.75}\}$, and $\widehat{f}^{Ker}_{\delta|W}$ denotes the kernel density estimator of $f_{\delta|W}$.

In Figures 3- 4 we show an (infeasible) kernel density estimator of the *pdf* of δ (the dotted black line) together with our Tikhonov estimator (the solid black line) and pointwise 95% confidence intervals obtained using the 1000 bootstrap replications. Each figure shows the estimation results for a different value of the variance σ_x^2 of the random coefficients: $\sigma_x^2 = 1$ and $\sigma_x^2 = 5$. In each figure, the estimate is conditional on fixed levels of assets and income. "Low" levels of each variable correspond to the 25th percentile and "high" levels correspond to the 75th percentile. As the results reveal, the unfeasible oracle estimator which we take in place of the true density is, for every value w of W we consider, within the confidence intervals. This suggests that our estimator is reasonably accurate in spite of the only moderate sample size of $n = 1000$.

Varying σ_x^2 has two effects on the results. First, the estimated density is more spread out in the higher variance case. Second, the pointwise confidence regions are slightly larger though similar in the high variance case. Note that the kernel of T in the operator equation, is the same in both cases. The confidence regions are slightly wider in the high variance case because high variance in θ results in higher estimation error in the estimate of the density of C conditional on W and Z. In contrast to Simulation 1, the way in which this higher variance affects the outcome c is not linear.

These results characterize the density of δ conditional on W. They also place constraints on the joint distribution of (β, γ) given W. Let δ_q be a quantile of the distribution of δ. For each quantile δ_q we can draw a curve representing the values of (β, γ) satisfying (6.4) when $\delta = \delta_q$. This curve is a "quantile level set". It represents the set of values of (β, γ) that keep δ_q constant. Moreover, since (6.4) is monotonic in β, it must be the case that with probability q, (β, γ) lie below this level set and with probability $1 - q$ they lie above this level set.

Figure 5 shows these level set curves conditional on various values of $W_t = (A_{t=31}, Y_{t=30})$ for the case with $\sigma_x^2 = 1$. For example, the solid line in Figure 5 shows the 0.1 quantile level set. With probability 0.1, (β, γ) lie below this curve. The quantile-level-sets partition the (β, γ) space into convex regions. The convex region in Figure 5 bounded by the 0.1 and 0.9 quantile level sets shows that people with low assets and low income are likely to be very impatient ($\beta < 0.9$) if they are risk averse ($\gamma > 3.5$) but are likely to be patient if they have low risk aversion. The other panels in the figure show that this convex region shifts upward for people with higher assets or income. As theory predicts, individuals with higher asset holdings are on average more patient and risk averse, but there is some evidence of a trade off between patience and risk aversion.

7 Conclusion

A key challenge in structural empirical work is to develop models that allow for complex multi-dimensional heterogeneity by allowing parameters to vary across the population while retaining tractability. In this paper, we develop methods that allow for a nonparametric distribution of these heterogeneity parameters by separating estimation of these models from computational aspects. We develop results on identification and estimation for the class of models we study and show through simulations how to apply these methods to two examples, a linear random coefficients demand model with endogeneity and a semiparametric dynamic consumption model. The nonlinear random coefficients framework we develop, however, is very rich. Beyond these examples, it can be applied to a wide range of applications. Indeed, we believe that it can serve as a blueprint on how to estimate complex structural models with heterogeneous parameters.

Notes

[1] In some applications, the distribution of θ may be time-varying or endogenous. Our approach captures this time-variation through dependence on W. Simulation 2 in Section 6 provides an illustration of this.

[2] An online Supplement is provided at Cambridge Journals Online. Readers may cite the supplementary material associated with this article as "Supplementary Material on 'Semiparametric Estimation of Random Coefficients in Structural Economic Models'", available at Cambridge Journals Online (journals.cambridge.org/ect).

[3] The weighting functions can be chosen to ensure that the operator T defined below is compact and bounded (see Proposition 1 for an example) and to reflect the researcher's loss function (see discussion after equation (5.2)).

[4] If Ψ is not known but is estimated in a first-stage, it affects neither our procedure nor the rate of convergence as long as the first-stage estimator converges at a faster rate. In this case, the asymptotic normality result we provide still holds under further assumptions similar to Assumption 6 in Horowitz (2007).

[5] The analysis can be extended to the multivariate ε case without great difficulty.

[6] This finite dimensional parameter may be either included in the vector θ and treated as a random parameter or estimated as a fixed parameter.

[7] Hoderlein et al. (2012) develop a richer application where φ must be computed numerically.

[8] An exception is Carrasco & Florens (2011).

[9] In our case, the generalized Fourier coefficient $\langle f_{C|WZ}, \psi_j \rangle$, cannot be simplified as in Carrasco & Florens (2011). Therefore, $f_{C|WZ}$ must be estimated nonparametrically and plugged-in allowing us to obtain a convergence rate that is in general faster.

[10] Chen & Reiss (2011) discuss different types of source conditions in inverse problems.

[11] See Hall & Horowitz (2005) or Darolles et al. (2011) for an explicit definition of $K_h(\cdot, \cdot)$. For simplicity, we use the same second order kernel K_h and the same bandwidths for all variables.

[12] The two-step procedure is described in Appendix E in the online Supplement.

References

Ai, C. & X. Chen (2003) Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* 71(6), 1795–1843.

Attanasio, O. P. & G. Weber (2010) Consumption and saving: Models of intertemporal allocation and their implications for public policy. *Journal of Economic Literature* 48(3), 693–751.

Beran, R., A. Feuerverger, & P. Hall (1996) On nonparametric estimation of intercept and slope distributions in random coefficient regression. *The Annals of Statistics* 24(6), 2569–2592.

Beran, R. & P. W. Millar (1994) Minimum distance estimation in random coefficient regression models. *The Annals of Statistics* 22(4), 1976–1992.

Blundell, R., X. Chen, & D. Kristensen (2007) Semi-nonparametric IV estimation of shape-invariant engel curves. *Econometrica* 75(6), pp. 1613–1669.

Canay, I. A., A. Santos, & A. M. Shaikh (2013) On the testability of identification in some nonparametric models with endogeneity. *Econometrica* 81(6), 2535–2559.

Carrasco, M. & J.-P. Florens (2000) Generalization of GMM to a continuum of moment conditions. *Econometric Theory* 16(6), 797–834.

Carrasco, M. & J.-P. Florens (2011) A spectral method for deconvolving a density. *Econometric Theory* 27, 546–581.

Carrasco, M., J.-P. Florens, & E. Renault (2007) Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. In J. J. Heckman & E. E. Leamer, eds., *Handbook of Econometrics*, volume 6, Part B, pages 5633 – 5751. Elsevier.

Chen, X. & D. Pouzo (2012) Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica* 80(1), 277–321.

Chen, X. & M. Reiss (2011) On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory* 27(6), 497–521.

Darolles, S., Y. Fan, J. P. Florens, & E. Renault (2011) Nonparametric instrumental regression. *Econometrica* 79(5), 1541–1565.

Dunker, F., S. Hoderlein, & K. Kaido (2013) Random coefficients in static games of complete information. Technical report, cemmap working paper, CWP12/13.

Engl, H., M. Hanke, & A. Neubauer (2000) *Regularization of inverse problems.* Kluwer Academic, Dordrecht.

Florens, J., M. Mouchart, & J.-M. Rolin (1990) *Elements of Bayesian statistics.* Dekker, New York.

Florens, J.-P. (2003) Inverse problems and structural econometrics: The example of instrumental variables. In M. Dewatripont, L. P. Hansen, & S. J. Turnovsky, eds., *Advances in Economics and Econometrics*, volume 2, pages 284–311. Cambridge University Press.

Florens, J.-P. & A. Simoni (2012) Nonparametric estimation of an instrumental regression: A quasi-bayesian approach based on regularized posterior. *Journal of Econometrics* 170(2), 458 – 475.

Florens, J.-P. & A. Simoni (2016) Regularizing priors for linear inverse problems. *Econometric Theory* 32, 71–121.

Gajek, L. (1986) On improving density estimators which are not bona fide functions. *The Annals of Statistics* 14(4), 1612–1618.

Gautier, E. & S. Hoderlein (2015) A triangular treatment eect model with random coefficients in the selection equation. Technical report, Toulouse School of Economics.

Gautier, E. & Y. Kitamura (2013) Nonparametric estimation in random coefficients binary choice models. *Econometrica* 81(2), 581–607.

Hall, P. & J. L. Horowitz (2005) Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics* 33(6), 2904–2929.

Hansen, P. (1988) Computation of the singular value expansion. *Computing* 40(3), 185–199.

Heckman, J. & B. Singer (1984) A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52(2), 271–320.

Henry, M., Y. Kitamura, & B. Salanié (2014) Partial identification of finite mixtures in econometric models. *Quantitative Economics* 5(1), 123–144.

Hoderlein, S. (2011) How many consumers are rational? *Journal of Econometrics* 164(2), 294 – 309.

Hoderlein, S., J. Klemelä, & E. Mammen (2010) Analyzing the random coefficient model nonparametrically. *Econometric Theory* 26, 804–837.

Hoderlein, S., L. Nesheim, & A. Simoni (2012) Heterogeneous euler equations: a semiparametric structural approach. Technical report, University College London.

Hohage, T. (2000) Regularization of exponentially ill-posed problems. *Numerical Functional Analysis and Optimization* 21(3-4), 439–464.

Horowitz, J. L. (2007) Asymptotic normality of a nonparametric instrumental variables estimator. *International Economic Review* 48(4), 1329–1349.

Hu, Y. & S. M. Schennach (2008) Instrumental variable treatment of nonclassical measurement error models. *Econometrica* 76(1), 195–216.

Ichimura, H. & T. Thompson (1998) Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution. *Journal of Econometrics* 86(2), 269–295.

Johannes, J., A. Simoni, & R. Schenk (2015) Adaptive Bayesian estimation in indirect Gaussian sequence space models. *ArXiv e-prints 1502.00184* .

Kasahara, H. & K. Shimotsu (2009) Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica* 77(1), 135–175.

Lewbel, A. & K. Pendakur (2016) Unobserved preference heterogeneity in demand using generalized random coefficients. *Journal of Political Economy* forthcoming.

Mammen, E., C. Rothe, & M. Schienle (2012) Nonparametric regression with nonparametrically generated covariates. *The Annals of Statistics* 40(2), 1132–1170.

Mandelbaum, A. & L. Ruschendorf (1987) Complete and symmetrically complete families of distributions. *The Annals of Statistics* 15(3), 1229–1244.

Masten, M. (2014) Random coefficients on endogenous variables in simultaneous equations models. Technical report, cemmap working paper, CWP01/14.

Matzkin, R. L. (2007) Nonparametric identification. In J. J. Heckman & E. E. Leamer, eds., *Handbook of Econometrics*, volume 6, Part B, pages 5307–5368. Elsevier.

Morozov, V. (1993) *Regularization Methods for Ill-Posed Problems.* FL: CRC Press.

Neubauer, A. (1986) *Tikhonov regularization of ill-posed linear operator equations on closed convex sets.* Johannes Kepler-Universität Linz edition.

Neubauer, A. (1988) Tikhonov-regularization of ill-posed linear operator equations on closed convex sets. *Journal of Approximation Theory* 53(3), 304–320.

Newey, W. K. & J. L. Powell (2003) Instrumental variable estimation of nonparametric models. *Econometrica* 71(5), 1565–1578.

Rosenblatt, M. (1969) Conditional probability density and regression estimators. In P. Shnaiah, ed., *Multivariate Analysis II*, pages 25–31. Academic Press, New York,.

van der Vaart, A. W. (1998) *Asymptotic Statistics.* Cambridge University Press.

A Proofs

A.1 Proof of Theorem 1

By Assumption 1, there exists a unique $c = \varphi(w, z, \theta, \varepsilon)$ that satisfies (3.1). Thus, using the transformation $\varphi(w, z, \theta, \cdot)$ mapping ε to c, the density of ε, $f_{\varepsilon|WZ\theta}$, specified in Assumption 3, and $f_{\theta|W}$ specified in Assumption 4, we can characterize the *pdf* of $f_{C\theta|WZ}$. Let $\mathcal{E}_1, \ldots, \mathcal{E}_s$ be a partition of $\mathbb{R}$ such that $\varphi(w, z, \theta, \cdot) : \mathcal{E}_i \to \mathbb{R}$ is one-to-one for each $i = 1, \ldots, s$, for given (w, z, θ) and $s \in \mathbb{N}_+$. Let $\varphi_i^{-1}(w, z, \theta, \cdot) : \operatorname{Im}(\mathcal{E}_i\, |w, z, \theta) \to \mathcal{E}_i$ be the corresponding inverse mapping for given (w, z, θ). Then,

$$f_{C|WZ\theta}(c, w, z, \theta) = \sum_{i=1}^{s} f_{\varepsilon|WZ\theta}\left[\varphi_i^{-1}(w, z, \theta, c), w, z, \theta\right] \left|\partial_c \varphi_i^{-1}(w, z, \theta, c)\right| 1_{\mathcal{C}_i}(c). \tag{A.1}$$

Further, using Assumption 5 we have $f_{C\theta|WZ} = f_{C|WZ\theta} f_{\theta|W}$. This implies that

$$f_{C|WZ}(c, w, z) = \int_{\Theta} f_{C|WZ\theta}(c, w, z, \theta) f_{\theta|W}(\theta, w) d\theta. \tag{A.2}$$

Finally, since a unique solution in C to (3.1) exists, the chain rule implies that: $\partial_\varepsilon \Psi(c, w, z, \theta, \varepsilon) = \partial_c \Psi(c, w, z, \theta, \varepsilon) \partial_\varepsilon c + \partial_\varepsilon \Psi(c, w, z, \theta, \varepsilon) = 0$, by abuse of notation. Since $\partial_\varepsilon c = \partial_\varepsilon \varphi(w, z, \theta, \varepsilon)$ we have $\partial_\varepsilon \varphi(w, z, \theta, \varepsilon) = -\frac{\partial_\varepsilon \Psi(c,w,z,\theta,\varepsilon)}{\partial_c \Psi(c,w,z,\theta,\varepsilon)}$. We conclude that

$$\begin{aligned} \partial_c \varphi_i^{-1}(w, z, \theta, c) &= \frac{1}{\partial_\varepsilon \varphi(w, z, \theta, \varepsilon)|_{\varepsilon=\varphi_i^{-1}(w,z,\theta,c)}} = -\left[\frac{\partial_\varepsilon \Psi(c, w, z, \theta, \varepsilon)}{\partial_c \Psi(c, w, z, \theta, \varepsilon)}\right]^{-1} \Big|_{\varepsilon=\varphi_i^{-1}(w,z,\theta,c)} \\ &= -\left[\frac{\partial_c \Psi(c, w, z, \theta, \varphi_i^{-1}(w, z, \theta, c))}{\partial_\varepsilon \Psi(c, w, z, \theta, \varphi_i^{-1}(w, z, \theta, c))}\right]. \end{aligned} \tag{A.3}$$

By replacing (A.3) in (A.1) and (A.1) in (A.2) we get the result.

A.2 Proof of Theorem 2

The first characterization of Λ follows trivially from (4.3) since every element obtained as the sum of $f_{\theta|W}^{\dagger}$ and an element of $\mathcal{N}(T)$ is solution of the unconstrained inverse problem. Then, to obtain the set of solutions to the constrained problem, we only have to take the intersection with $\mathcal{F}_{\theta|W}$.

To obtain the second characterization of Λ note that we can always write a generic element of $f_{\theta|W}^{\dagger} \oplus \mathcal{N}(T)$ in terms of the orthonormal basis $\{\{\varphi_j\}_{j\in\mathbb{N}}, \{\tilde{\varphi}_l\}_{l\in J_0}\}$ which exists under Assumption 6. Then, we obtain the identified set by considering only $h \in \{f^{\dagger}_{\theta|W} \oplus \mathcal{N}(T)\}$ that satisfy three restrictions. First, $\int_{\Theta} h(\theta, w)d\theta = 1$, a.s.. Second, $\int_{\Theta} h^2(\theta, w)\pi_{\theta}(\theta)d\theta < \infty$, a.s.. Third, $h(\theta, w) \geq 0$, a.s.. The first constraint is automatically satisfied since for every (w, z, θ), $\int_{\mathcal{C}} f_{C|WZ\theta}(c, w, z, \theta)dc = 1$ and, by using Fubini's theorem: $\int_{\Theta} h(\theta, w)d\theta = \int_{\mathcal{C}} \int_{\Theta} f_{C|WZ\theta}d\theta h(\theta, w)dc = \int_{\mathcal{C}} f_{C|WZ}dc = 1$ (where we have used the fact that $f_{C|WZ\theta}$ integrates to 1 and $Th = f_{C|WZ}$). The second constraint is equivalent to $||f^{\dagger}_{\theta|W}||^2 + \sum_{l \in J_0} \zeta_l^2 < \infty$ for some $\zeta_l \in \mathbb{R}$ and, by definition of $f^{\dagger}_{\theta|W}$, $||f^{\dagger}_{\theta|W}||^2 < \infty$. Finally, the third constraint is equivalent to requiring that the negative part of every function in Λ is equal to 0.

A.3 Proof of Proposition 2

Suppose that $\mathcal{F}_{\theta|CWZ}$ is $\mathcal{T}$-complete and that for $f^1_{\theta|W}, f^2_{\theta|W} \in \mathcal{F}_{\theta|W}$, $T|_{\mathcal{F}_{\theta|W}}(f^1_{\theta|W}) = T|_{\mathcal{F}_{\theta|W}}(f^2_{\theta|W})$ holds. By using the decomposition $f_{C|WZ\theta} = f_{\theta|CWZ}f_{C|WZ}/f_{\theta|W}$ this equality can be rewritten as

$$\begin{aligned} 0 &= T|_{\mathcal{F}_{\theta|W}}(f^1_{\theta|W}) - T|_{\mathcal{F}_{\theta|W}}(f^2_{\theta|W}) = \int_{\Theta} f_{C|WZ\theta}(c, w, z, \theta) \left[f^1_{\theta|W}(\theta, w) - f^2_{\theta|W}(\theta, w) \right] d\theta \\ &= \int_{\Theta} f_{\theta|CWZ}(\theta, c, w, z) \frac{f_{C|WZ}(c, w, z)}{f_{\theta|W}(\theta, w)} \left[f^1_{\theta|W}(\theta, w) - f^2_{\theta|W}(\theta, w) \right] d\theta \end{aligned} \tag{A.4}$$

which is equivalent to

$$0 = \int_{\Theta} f_{\theta|CWZ}(\theta, c, w, z) \frac{1}{f_{\theta|W}(\theta, w)} \left[f^1_{\theta|W}(\theta, w) - f^2_{\theta|W}(\theta, w) \right] d\theta \tag{A.5}$$

because, by Assumptions 2 and 3, $0 < f_{C|WZ} < \infty$. Moreover, $\frac{1}{f_{\theta|W}(\theta, w)} \left[f^1_{\theta|W}(\theta, w) - f^2_{\theta|W}(\theta, w) \right] \in \mathcal{T}$ so that (A.5) implies $\frac{1}{f_{\theta|W}(\theta, w)} \left[f^1_{\theta|W}(\theta, w) - f^2_{\theta|W}(\theta, w) \right] = 0$ which in turns implies $f^1_{\theta|W}(\theta, w) = f^2_{\theta|W}(\theta, w)$ under Assumption 4.

On the other hand, if (4.5) holds, then $0 = \int_{\Theta} f_{\theta|CWZ}(\theta, c, w, z) \frac{1}{f_{\theta|W}(\theta, w)} \left[f^1_{\theta|W}(\theta, w) - f^2_{\theta|W}(\theta, w) \right] d\theta$ implies that $\frac{1}{f_{\theta|W}(\theta, w)} \left[f^1_{\theta|W}(\theta, w) - f^2_{\theta|W}(\theta, w) \right] = 0$ because, by Assumptions 2, 3 and 4, $0 < f_{C|WZ} < \infty$ and $0 < f_{\theta|W} < \infty$. This concludes the proof.

A.4 Proof of Proposition 3

For simplicity we consider the case where θ is one-dimensional (the multi-dimensional case can be recovered in a similar way). Let us suppose that $T\phi(\theta, w) = 0$, a.s. for some function $\phi \in \mathfrak{D}$. Then, $\forall (c, z) \in \mathcal{C} \times \mathcal{Z}$

$$T\phi = \int_{\Theta} \sum_{i=1}^{s} f_{\varepsilon|\theta WZ} \left[\varphi_i^{-1}(w, z, \theta, c), \theta, w, z \right] \left| \partial_c \varphi_i^{-1}(w, z, \theta, c) \right| 1_{\mathcal{C}_i}(c) \phi(\theta, w) d\theta = 0 \qquad a.s.$$

implies that $\forall (c, z) \in \mathcal{C}_i \times \mathcal{Z}$

$$\int_{\Theta} f_{\varepsilon|\theta WZ} \left[\varphi_i^{-1}(w, z, \theta, c), \theta, w, z \right] \left| \partial_c \varphi_i^{-1}(w, z, \theta, c) \right| \phi(\theta, w) d\theta = 0 \quad a.s. \quad \forall i = 1, \ldots, s.$$

Then, $\forall (c,z) \in \mathcal{C}_i \times \mathcal{Z}$ and $\forall i = 1, \dots, s$, we have:

$$\begin{aligned} 0 &= \int_\Theta \exp\left[\tau_i(c,w,z)m_i(\theta)\right] h_i(\theta) k_i(c,w,z)\phi(\theta,w) \left|\partial_c \varphi_i^{-1}(w,z,\theta,c)\right| d\theta \\ &= \int_\Theta \exp\left[\tau_i(c,w,z)\mu_i\right] h_i\left[m_i^{-1}(\mu_i)\right] k_i(c,w,z)\tilde{\phi}_i\left[m_i^{-1}(\mu_i), w, z, c\right] dm_i^{-1}(\mu_i) \quad \text{a.s.} \end{aligned}$$

where we have used the notation $\tilde{\phi}_i(\theta,w,z,c) = \phi(\theta,w)\left|\partial_c \varphi_i^{-1}(w,z,\theta,c)\right|$ and the change of variable $m_i(\theta) = \mu_i$. Moreover, since $dm_i^{-1}(\mu_i)$ and h_i are positive functions, we can define a measure $\nu_i(d\mu_i) = h_i\left[m_i^{-1}(\mu_i)\right] dm_i^{-1}(\mu_i)$. Thus, $\forall (c,z) \in \mathcal{C}_i \times \mathcal{Z}$ and $\forall i = 1, \dots, s$,

$$\begin{aligned} 0 &= k_i(c,w,z) \int_\Theta \exp\left[\tau_i(c,w,z)\mu_i\right] \tilde{\phi}_i\left[m_i^{-1}(\mu_i), w, z, c\right] \nu_i(d\mu_i) \\ &= k_i(c,w,z) \int_\Theta \exp\left[\tau_i(c,w,z)\mu_i\right] \zeta_i(\mu_i, w, z, c)\nu_i(d\mu_i) \\ &= k_i(c,w,z) \int_\Theta \exp\left[\tau_i(c,w,z)\mu_i\right] \left[\zeta_i^+(\mu_i,w,z,c) - \zeta_i^-(\mu_i,w,z,c)\right] \nu_i(d\mu_i) \\ &= k_i(c,w,z) \left\{\int_\Theta \exp\left[\tau_i(c,w,z)\mu_i\right] F_i(d\mu_i,w,z,c) - \int_\Theta \exp\left[\tau_i(c,w,z)\mu_i\right] G_i(d\mu_i,w,z,c)\right\} \end{aligned}$$

a.s. where

$$\begin{aligned} \zeta_i(\mu_i,w,z,c) &= \tilde{\phi}_i \circ m_i^{-1} \\ F_i(d\mu_i,w,z,c) &= \zeta_i^+(\mu_i,w,z,c)\nu_i(d\mu_i) \\ G_i(d\mu_i,w,z,c) &= \zeta_i^-(\mu_i,w,z,c)\nu_i(d\mu_i) \end{aligned}$$

and, for a function h, h^+ and h^- denote the positive and negative part of it, respectively. It follows that

$$\int_\Theta \exp\left[\tau_i(c,w,z)\mu_i\right] F_i(d\mu_i,w,z,c) = \int_\Theta \exp\left[\tau_i(c,w,z)\mu_i\right] G_i(d\mu_i,w,z,c),$$

that is, F_i and G_i are two measures with the same Laplace transform. Then, they are equal since $\tau_i(c,w,z)$ vary over $\mathbb{R}$. This implies that $\zeta_i(\mu_i,w,z,c) = 0$ and then $\phi_i(\theta,w) = 0$, a.s. since $\partial_c \varphi_i^{-1}(w,z,\theta,c)$ is bounded away from 0 and ∞ by Assumption 2, $\forall (c,w,\theta,z) \in \mathcal{C} \times \mathcal{W} \times \Theta \times \mathcal{Z}$.

A.5 Proof of Theorem 3

First, since $||\mathcal{P}_c|| \leq 1$ we have: $\mathbb{E}||\mathcal{P}_c \hat{f}^\alpha_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 = \mathbb{E}||\mathcal{P}_c(\hat{f}^\alpha_{\theta|W} - f^\dagger_{\theta|W})||^2 \leq ||\mathcal{P}_c||^2 \mathbb{E}||\hat{f}^\alpha_{\theta|W} - f^\dagger_{\theta|W}||^2 \leq \mathbb{E}||\hat{f}^\alpha_{\theta|W} - f^\dagger_{\theta|W}||^2$. Let $f^\alpha_{\theta|W} = (\alpha I + T^*T)^{-1}T^* f_{C|WZ}$, then

$$\mathbb{E}||\hat{f}^\alpha_{\theta|W} - f^\dagger_{\theta|W}||^2 \leq 2\mathbb{E}||\hat{f}^\alpha_{\theta|W} - f^\alpha_{\theta|W}||^2 + 2\mathbb{E}||f^\alpha_{\theta|W} - f^\dagger_{\theta|W}||^2 = 2(\mathcal{A}_1 + \mathcal{A}_2). \qquad \text{(A.6)}$$

Because the operator T^*T is a bounded self-adjoint linear operator, its spectrum, denoted by $\sigma(T^*T)$, is a nonempty set of real numbers. Hence we can analyze the two terms $\mathcal{A}_1$ and $\mathcal{A}_2$ as follows. Term $\mathcal{A}_1$ is

$$\begin{aligned}\mathcal{A}_1 &= \mathbb{E}||(\alpha I + T^*T)^{-1}T^*(\hat{f}_{C|WZ} - f_{C|WZ})||^2 \leq ||(\alpha I + T^*T)^{-1}T^*||^2\mathbb{E}||(\hat{f}_{C|WZ} - f_{C|WZ})||^2 \\ &\leq \sup_{t\in\sigma(T^*T)} |(\alpha+t)^{-1}\sqrt{t}|^2\mathbb{E}||(\hat{f}_{C|WZ} - f_{C|WZ})||^2 = \mathcal{O}\left(\frac{1}{\alpha}\mathbb{E}||(\hat{f}_{C|WZ} - f_{C|WZ})||^2\right). \qquad \text{(A.7)}\end{aligned}$$

Next, we develop term $\mathcal{A}_2$:

$$\begin{aligned}\mathcal{A}_2 &= \mathbb{E}||(\alpha I + T^*T)^{-1}T^*f_{C|WZ} - f^{\dagger}_{\theta|W}||^2 = ||[I - (\alpha I + T^*T)^{-1}T^*T]f^{\dagger}_{\theta|W}||^2 \\ &= ||\alpha(\alpha I + T^*T)^{-1}f^{\dagger}_{\theta|W}||^2 = ||\alpha(\alpha I + T^*T)^{-1}\phi(T^*T)\nu||^2 \\ &\leq \sup_{t\in\sigma(T^*T)} |\phi(t)\alpha(\alpha+t)^{-1}|^2\nu^2 = \mathcal{O}(\phi^2(\alpha)) \qquad \text{(A.8)}\end{aligned}$$

where we have used Assumption 8 and the last inequality follows from (5.6). This shows that

$$\mathbb{E}||\hat{f}^{\alpha}_{\theta|W} - f^{\dagger}_{\theta|W}||^2 = \mathcal{O}\left(\phi^2(\alpha) + \frac{1}{\alpha}\mathbb{E}||(\hat{f}_{C|WZ} - f_{C|WZ})||^2\right).$$

Next, consider the case $\phi(t) = t^{\beta/2}$. Then,

$$\sup_{t\in\sigma(T^*T)} |t^{\beta/2}\alpha(\alpha+t)^{-1}| = \frac{1}{2}\alpha^{\beta/2}$$

if $\beta < 2$ and $\sup_{t\in\sigma(T^*T)} |\phi(t)\alpha(\alpha+t)^{-1}| = \alpha$ if $\beta = 2$. Hence, (5.6) is satisfied and by choosing $\alpha \asymp (\mathbb{E}||(\hat{f}_{C|WZ} - f_{C|WZ})||^2)^{1/(\beta\wedge 2+1)}$ we get the result. Finally, consider the case $\phi(t) = (-\log(t))^{-\beta/2}$. If we choose $\alpha \asymp (\mathbb{E}||(\hat{f}_{C|WZ} - f_{C|WZ})||^2)^{\epsilon}$ for $0 < \epsilon < 1$ we get the final result of the theorem.

A.6 Proof of Corollary 1

Following the decomposition (A.6) in the proof of Theorem 3, the upper bound for $\mathcal{A}_2$ remains unchanged while term $\mathcal{A}_1$ is now bounded above by $\mathcal{A}_1 \leq ||(\alpha I + T^*T)^{-1}||^2\mathbb{E}||T^*(\hat{f}_{C|WZ} - f_{C|WZ})||^2$ so that $\mathcal{A}_1 = \mathcal{O}\left(\alpha^{-2}\mathbb{E}||T^*(\hat{f}_{C|WZ} - f_{C|WZ})||^2\right)$. We have to compute the rate of $\mathbb{E}||T^*(\hat{f}_{C|WZ} - f_{C|WZ})||^2$. Remark that $\mathbb{E}||T^*(\hat{f}_{C|WZ} - f_{C|WZ})||^2 = \int_{\Theta}\left(Var(T^*\hat{f}_{C|WZ}) + (\mathbb{E}(T^*\hat{f}_{C|WZ}) - T^*f_{C|WZ})^2\right)\pi_{\theta}(\theta)d\theta$. By using standard Taylor series approximations, it is easy to show (see Lemma 3 in the online Supplement) that the squared bias term is of order $\left(\mathbb{E}(T^*\hat{f}_{C|WZ} - T^*f_{C|WZ})\right)^2 = \mathcal{O}\left(\max\{h_n^4, h_d^4\}\right)$ and the variance term is $Var(T^*\hat{f}_{C|WZ}) = \mathcal{O}\left(n^{-1}(\min\{h_n, h_d\})^{-k}\right)$. Therefore, the rate of the MISE is:

$$\mathbb{E}||\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 = \mathcal{O}\left[\phi^2(\alpha) + \frac{1}{\alpha^2}\left(\max\{h_n^4, h_d^4\} + \frac{1}{n(\min\{h_n, h_d\})^k}\right)\right].$$

A.7 Proof of Theorem 4

Denote by $\mathbb{E}_{WZ}$ the conditional expectation given (W, Z). Let us consider the decomposition

$$\begin{aligned}(\hat{f}^{\alpha}_{\theta|W} - f^{\dagger c}_{\theta|W})(\theta, w) &= \left[\hat{f}^{\alpha}_{\theta|W} - (\alpha I + T^*T)^{-1}T^*\mathbb{E}_{WZ}\left(\hat{f}_{C|WZ}\right)\right](\theta, w) \\ &\quad + \left[(\alpha I + T^*T)^{-1}T^*\mathbb{E}_{WZ}\left(\hat{f}_{C|WZ}\right) - f^{\dagger}_{\theta|W}\right](\theta, w) \\ &= A + B.\end{aligned}$$

The result of Lemma 4 follows from proving that $\frac{\mathcal{P}_c A}{\sqrt{V_c(\theta,w)}} \to^d N(0,1)$ and $\frac{\mathcal{P}_c B}{\sqrt{V_c(\theta,w)}} = o_p(1)$. We start by proving that $\frac{A}{\sqrt{V(A)}} \to^d N(0,1)$ where $V(A) = Var(A)$. Let $\{\lambda_j, \varphi_j, \psi_j\}_{j\in\mathbb{N}}$ denote the SVD of T, $\hat{f}_{WZ}$ denote the kernel estimator of the joint *pdf* of (W, Z) and $K_{h,i}(z, w) = K_h(z_i - z, z)K_h(w_i - w, w)$, then

$$\begin{aligned}A &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{\infty}\frac{1}{\alpha + \lambda_j^2}\left\langle T^*\left\{K_h\left(c_i - c, c\right) - \mathbb{E}_{WZ}\left[K_h\left(c_i - c, c\right)\right]\right\}\frac{K_{h,i}(z, w)}{h_n^{k+l+1}\hat{f}_{WZ}}, \varphi_j\right\rangle\varphi_j(\theta, w) \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{\infty}\frac{\lambda_j}{\alpha + \lambda_j^2}\left\langle\left\{K_h\left(c_i - c, c\right) - \mathbb{E}_{WZ}\left[K_h\left(c_i - c, c\right)\right]\right\}\frac{K_{h,i}(z, w)}{h_n^{k+l+1}\hat{f}_{WZ}}, \psi_j\right\rangle\varphi_j(\theta, w) = \frac{1}{n}\sum_{i=1}^{n}Z_{ni}.\end{aligned}$$

By a triangular array version of the Liapounov's central limit theorem it follows that

$$\frac{A}{\sqrt{V(A)}} = \frac{1}{n}\sum_{i=1}^{n}Z_{ni}/\sqrt{n^{-1}Var(Z_{ni})} \to^d N(0,1)$$

if $\sum_{i=1}^{n}\mathbb{E}\left|Z_{ni}/\sqrt{nVar(Z_{ni})}\right|^3 \to 0$ as $n \to \infty$. Theorem 4 in the online Supplement shows that this latter convergence holds if $\alpha^3/(nh_n^k) \to 0$. To prove $\frac{\mathcal{P}_c A}{\sqrt{V_c(\theta,w)}} \to^d N(0,1)$ we use the functional delta method (see e.g. van der Vaart (1998) Theorem 20.8). This requires that the projection operator $\mathcal{P}_c$ is Hadamard differentiable. The (one-sided) Hadamard derivative of $\mathcal{P}_c$ in $f^{\dagger c}_{\theta|W}$ is a projection as well, denoted by $\mathcal{P}_c^{\dagger}$, that projects on the tangent cone of $\mathcal{F}_{\theta|W}$ at $f^{\dagger c}_{\theta|W}$ defined as in the statement of Lemma 4. Moreover, $V_c(\theta, w) = Var(\mathcal{P}_c^{\dagger}A)$ and $V(A)$ and $V_c(\theta, w)$ have the same rate.

To prove the second result we follow the strategy in the proof of Proposition 6 in Carrasco & Florens (2011) and prove that $\frac{B^2}{V(A)} \to 0$. Let us decompose B as

$$B = (\alpha I + T^*T)^{-1}T^*\left[\mathbb{E}\left(\hat{f}_{C|WZ}\right) - f_{C|WZ}\right](\theta, w) - \left[(\alpha I + T^*T)^{-1}T^*f_{C|WZ} - f^{\dagger}_{\theta|W}\right](\theta, w).$$

Then

$$\begin{aligned}B^2 &\leq 2\left|(\alpha I + T^*T)^{-1}T^*\left[\mathbb{E}_{WZ}\left(\hat{f}_{C|WZ}\right) - f_{C|WZ}\right](\theta, w)\right|^2 \\ &\quad +2\left|\left[(\alpha I + T^*T)^{-1}T^*f_{C|WZ} - f^{\dagger}_{\theta|W}\right](\theta, w)\right|^2.\end{aligned}$$

Note that $\mathbb{E}_{WZ}(\hat{f}_{C|WZ}) = f_{C|WZ} + \mathcal{O}(h_n^2)$. Then, by using (A.7) and (A.8), under Assumption 8 we conclude that $\frac{B^2}{V(A)} = \mathcal{O}_p\left(n\alpha h_n^{k+4} + n\alpha^2\phi^2(\alpha)h_n^k\right)$ which converges to zero under the conditions of the theorem. Since $\mathcal{P}_c$ is a nonexpansive map, these rates are not affected by replacing B with $\mathcal{P}_c B$ and $V(A)$ with $V_c(\theta, w)$ since $V(A)$ and $V_c(\theta, w)$ have the same rate.

A.8 Proof of Theorem 5

The functional $J_\alpha(h) = ||Th - \hat{f}_{C|WZ}||^2 + \alpha||h||^2$ is a strictly convex and Fréchet differentiable functional with Fréchet derivative $2\left((T^*T + \alpha I)h - T^*\hat{f}_{C|WZ}\right)$. Hence, the convex problem (5.9) has a unique solution $\check{f}^{\alpha,c}_{\theta|W}$ that is characterized as the unique element in $\mathcal{F}_{C|WZ}$ such that the following variational inequality holds:

$$\langle (\alpha I + T^*T)\check{f}^{\alpha,c}_{\theta|W} - T^*\hat{f}_{C|WZ}, f - \check{f}^{\alpha,c}_{\theta|W}\rangle \geq 0, \quad \forall f \in \mathcal{F}_{\theta|W}. \tag{A.9}$$

For every $\alpha > 0$ define the inner product $\langle f_1, f_2\rangle_\alpha = \langle (\alpha I + T^*T)f_1, f_2\rangle$ on $L^2_{\pi_\theta}$. Then (A.9) is equivalent to

$$\langle \check{f}^{\alpha,c}_{\theta|W} - (\alpha I + T^*T)^{-1}T^*\hat{f}_{C|WZ}, f - \check{f}^{\alpha,c}_{\theta|W}\rangle_\alpha \geq 0, \quad \forall f \in \mathcal{F}_{\theta|W}. \tag{A.10}$$

Thus, $\check{f}^{\alpha,c}_{\theta|W} = \mathcal{P}^\alpha_c(\alpha I + T^*T)^{-1}T^*\hat{f}_{C|WZ}$ where $\mathcal{P}^\alpha_c$ denotes the projector onto $\mathcal{F}_{\theta|W}$ with respect to $\langle \cdot, \cdot\rangle_\alpha$. By denoting $f^{\alpha,c}_{\theta|W} = \mathcal{P}^\alpha_c(\alpha I + T^*T)^{-1}T^*f_{C|WZ}$ we can write

$$\begin{aligned}\mathbb{E}||\check{f}^{\alpha,c}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 &\leq 2\mathbb{E}\|\mathcal{P}^\alpha_c(\alpha I + T^*T)^{-1}T^*(\hat{f}_{C|WZ} - f_{C|WZ})\|^2 + 2||f^{\alpha,c}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 \\ &= \mathcal{O}\left(\alpha^{-1}\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2\right) + 2||f^{\alpha,c}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2.\end{aligned} \tag{A.11}$$

It remains to show that $||f^{\alpha,c}_{\theta|W} - f^{\dagger c}_{\theta|W}||$ converges to 0. By definition of $f^{\alpha,c}_{\theta|W}$:

$$\langle (\alpha I + T^*T)f^{\alpha,c}_{\theta|W} - T^*f_{C|WZ}, f^{\dagger c}_{\theta|W} - f^{\alpha,c}_{\theta|W}\rangle \geq 0. \tag{A.12}$$

Define the closed and convex set $U = \{u \in \overline{\mathcal{R}(T)} : P_{\mathcal{F}_{\theta|W}}T^*u = f^{\dagger c}_{\theta|W}\}$ and let $\bar{u}$ be the element of U with minimal norm. It follows that $\langle f^{\dagger c}_{\theta|W} - T^*\bar{u}, f^{\alpha,c}_{\theta|W} - f^{\dagger c}_{\theta|W}\rangle \geq 0$ or, equivalently,

$$\langle T^*\bar{u} - f^{\dagger c}_{\theta|W}, f^{\dagger c}_{\theta|W} - f^{\alpha,c}_{\theta|W}\rangle \geq 0. \tag{A.13}$$

By summing (A.12), with $f_{C|WZ}$ replaced by $Tf^{\dagger c}_{\theta|W}$, and (A.13), multiplied by $\alpha > 0$, we obtain:

$$\langle (\alpha I + T^*T)(f^{\alpha,c}_{\theta|W} - f^{\dagger c}_{\theta|W}) + \alpha T^*\bar{u}, f^{\dagger c}_{\theta|W} - f^{\alpha,c}_{\theta|W}\rangle \geq 0$$

which is equivalent to $||T(f^{\alpha,c}_{\theta|W} - f^{\dagger c}_{\theta|W})||^2 + \alpha||f^{\alpha,c}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 \leq \alpha\langle T^*\bar{u}, f^{\dagger c}_{\theta|W} - f^{\alpha,c}_{\theta|W}\rangle$. Then, since $\alpha\langle T^*\bar{u}, f^{\dagger c}_{\theta|W} - f^{\alpha,c}_{\theta|W}\rangle \leq \alpha||\bar{u}||||T(f^{\alpha,c}_{\theta|W} - f^{\dagger c}_{\theta|W})||$, it follows that $||T(f^{\alpha,c}_{\theta|W} - f^{\dagger c}_{\theta|W})|| \leq \alpha||\bar{u}||$ and hence, $||f^{\alpha,c}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 \leq \alpha^2||\bar{u}||^2$ which converges to 0 as $\alpha \to 0$. From (A.11) and this result, we conclude that: $\mathbb{E}||\check{f}^{\alpha,c}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 \to 0$.

A.9 Proof of Theorem 6

The first part of the proof is the same as the proof of Theorem 5. Thus, (A.11) still holds and we only have to determine the rate of $||f^{\alpha,c}_{\theta|W} - f^{\dagger c}_{\theta|W}||$. To do this we slightly modify the proof of Lemma 3.9 in Neubauer

(1986). For every $f \in \mathcal{N}(T)^{\perp}$, let $f_{\theta|W}^{\alpha,L_f}$ be the solution of (5.9) with $\hat{f}_{C|WZ}$ replaced by $f_{C|WZ}$ and $\mathcal{F}_{\theta|W}$ replaced by $L_f = \{h \in L^2_{\pi_\theta} : \langle f_{\theta|W}^{\dagger c} - f, h - f_{\theta|W}^{\dagger c}\rangle = 0\}$. Note that L_f is the supporting hyperplane of $\mathcal{F}_{\theta|W}$ in $f_{\theta|W}^{\dagger c}$. By the triangular inequality:

$$||f_{\theta|W}^{\alpha,c} - f_{\theta|W}^{\dagger c}|| \leq ||f_{\theta|W}^{\alpha,c} - f_{\theta|W}^{\alpha,L_f}|| + ||f_{\theta|W}^{\alpha,L_f} - f_{\theta|W}^{\dagger c}||. \tag{A.14}$$

We start by analyzing the second term on the right in (A.14). Since $f_{\theta|W}^{\alpha,L_f}, f_{\theta|W}^{\dagger c} \in L_f$ then $\tilde{P}_f(f_{\theta|W}^{\alpha,L_f} - f_{\theta|W}^{\dagger c}) = f_{\theta|W}^{\alpha,L_f} - f_{\theta|W}^{\dagger c}$. Therefore, the rate of the second term can be easily determined if we show that $\tilde{P}_f f_{\theta|W}^{\alpha,L_f}$ is an unconstrained Tikhonov-regularized solution of the form $\tilde{P}_f f_{\theta|W}^{\alpha,L_f} = (\tilde{P}_f T^*T\tilde{P}_f + \alpha I)^{-1}\tilde{P}_f T^*T\tilde{P}_f f_{\theta|W}^{\dagger c}$. In order to show this, we start by showing that

$$\tilde{P}_f \left(T^*T f_{\theta|W}^{\alpha,L_f} + \alpha f_{\theta|W}^{\alpha,L_f} - T^* f_{C|WZ}\right) = 0. \tag{A.15}$$

This is equivalent to showing that $\langle T^*T f_{\theta|W}^{\alpha,L_f} + \alpha f_{\theta|W}^{\alpha,L_f} - T^* f_{C|WZ}, h\rangle = 0$ for every $h \in \tilde{L}_f$. Let $h \in \tilde{L}_f$, then $h + f_{\theta|W}^{\dagger c} \in L_f$. By definition of $f_{\theta|W}^{\alpha,L_f}$, the variational inequality

$$\langle (\alpha I + T^*T) f_{\theta|W}^{\alpha,L_f} - T^* f_{C|WZ}, h' - f_{\theta|W}^{\alpha,L_f}\rangle \geq 0, \qquad \forall h' \in L_f \tag{A.16}$$

holds with equality. Therefore, for $h \in \tilde{L}_f$,

$$\langle (\alpha I + T^*T) f_{\theta|W}^{\alpha,L_f} - T^* f_{C|WZ}, h\rangle + \langle (\alpha I + T^*T) f_{\theta|W}^{\alpha,L_f} - T^* f_{C|WZ}, f_{\theta|W}^{\dagger c} - f_{\theta|W}^{\alpha,L_f}\rangle = 0 \tag{A.17}$$

where the second term is equal to 0 by applying (A.16) with equality and since $f_{\theta|W}^{\dagger c} \in L_f$. We conclude that $\langle (\alpha I + T^*T) f_{\theta|W}^{\alpha,L_f} - T^* f_{C|WZ}, h\rangle = 0$ for every $h \in \tilde{L}_f$. This proves (A.15).

By using the result of Lemma 5 in the online Supplement and by rearranging terms we get: $\tilde{P}_f f_{\theta|W}^{\alpha,L_f} = (\tilde{P}_f T^*T\tilde{P}_f + \alpha I)^{-1}\tilde{P}_f T^*T\tilde{P}_f f_{\theta|W}^{\dagger c}$. By Lemma 3.5 (a) in Neubauer (1986): $f_{\theta|W}^{\dagger c} = f_{\theta|W}^{\dagger L_f}$, where $f_{\theta|W}^{\dagger L_f}$ satisfies:

$$\begin{aligned} \|T f_{\theta|W}^{\dagger L_f} - f_{C|WZ}\| &= \inf\left\{\|Th - f_{C|WZ}\| : h \in L_f\right\} \\ \|f_{\theta|W}^{\dagger L_f}\| &= \min\left\{\|h\| : h \in L_f \text{ and } \|Th - f_{C|WZ}\| = \|T f_{\theta|W}^{\dagger L_f} - f_{C|WZ}\|\right\}. \end{aligned}$$

This implies $||f_{\theta|W}^{\alpha,L_f} - f_{\theta|W}^{\dagger c}|| = ||f_{\theta|W}^{\alpha,L_f} - f_{\theta|W}^{\dagger L_f}||$. So, by using the regularity condition $N_\beta \neq \varnothing$ we conclude that

$$\begin{aligned} ||f_{\theta|W}^{\alpha,L_f} - f_{\theta|W}^{\dagger c}||^2 &= ||\tilde{P}_f(f_{\theta|W}^{\alpha,L_f} - f_{\theta|W}^{\dagger c})||^2 = ||(\tilde{P}_f T^*T\tilde{P}_f + \alpha I)^{-1}\tilde{P}_f T^*T\tilde{P}_f f_{\theta|W}^{\dagger c} - \tilde{P}_f f_{\theta|W}^{\dagger c}||^2 \\ &= ||\alpha(\tilde{P}_f T^*T\tilde{P}_f + \alpha I)^{-1}\tilde{P}_f f_{\theta|W}^{\dagger c}||^2 = \mathcal{O}(\alpha^{\beta\wedge 2}). \end{aligned} \tag{A.18}$$

We now consider the first term on the right in (A.14). By Lemma 3.6 (b) in Neubauer (1986) the following inequality holds for $\alpha > 0$ sufficiently small and $f \in M$: $||T(f_{\theta|W}^{\alpha,c} - f_{\theta|W}^{\alpha,L_f})||^2 + \alpha||f_{\theta|W}^{\alpha,c} - f_{\theta|W}^{\alpha,L_f}||^2 \leq ||T(f_{\theta|W}^{\dagger c} - f_{\theta|W}^{\alpha,L_f})||^2 + \alpha||f_{\theta|W}^{\dagger c} - f_{\theta|W}^{\alpha,L_f}||^2$. This implies that $||f_{\theta|W}^{\alpha,c} - f_{\theta|W}^{\alpha,L_f}||^2 \leq \frac{1}{\alpha}||T(f_{\theta|W}^{\dagger c} - f_{\theta|W}^{\alpha,L_f})||^2 + ||f_{\theta|W}^{\dagger c} - f_{\theta|W}^{\alpha,L_f}||^2$. From (A.18) and the fact that $||T(f_{\theta|W}^{\dagger c} - f_{\theta|W}^{\alpha,L_f})||^2 = \mathcal{O}(\alpha^{\beta\wedge 2+1})$ (which also follows from (A.18)) we conclude that $||f_{\theta|W}^{\alpha,c} - f_{\theta|W}^{\alpha,L_f}||^2 = \mathcal{O}(\alpha^{\beta\wedge 2})$. By putting all these results together we have proved the results of the theorem.

A.10 Proof of Theorem 7

Part *(i)* follows from Lemma 6 in the online Supplement. Hence, we prove *(ii)*. Let $\mathcal{U}_w(g^0)$ denote an open neighborhood in $\mathcal{G}$ in the weak topology around g^0 and $\mathcal{U}_w(\theta_1^0)$ denote its projection onto Θ_1, that is, $\mathcal{U}_w(\theta_1^0) = \{\theta_1 \in \Theta_1 : \exists h \in \mathcal{F}_{\theta_2|W} \text{ such that } (\theta_1, h) \in \mathcal{U}_w(g^0)\}$. Because $\Theta_1 \subset \mathbb{R}^d$ and because the weak and norm topologies coincide on finite dimensional spaces, then $\|\hat{\theta}_1\|_E \to 0$ in probability if and only if $P(\hat{\theta}_1 \in \mathcal{U}_w(\theta_1^0)) \to 1$. This last result follows from Lemma 9 in the online Supplement and the inequality $P(\hat{\theta}_1 \in \mathcal{U}_w(\theta_1^0)) \geq P(\hat{g} \in \mathcal{U}_w(g^0))$.

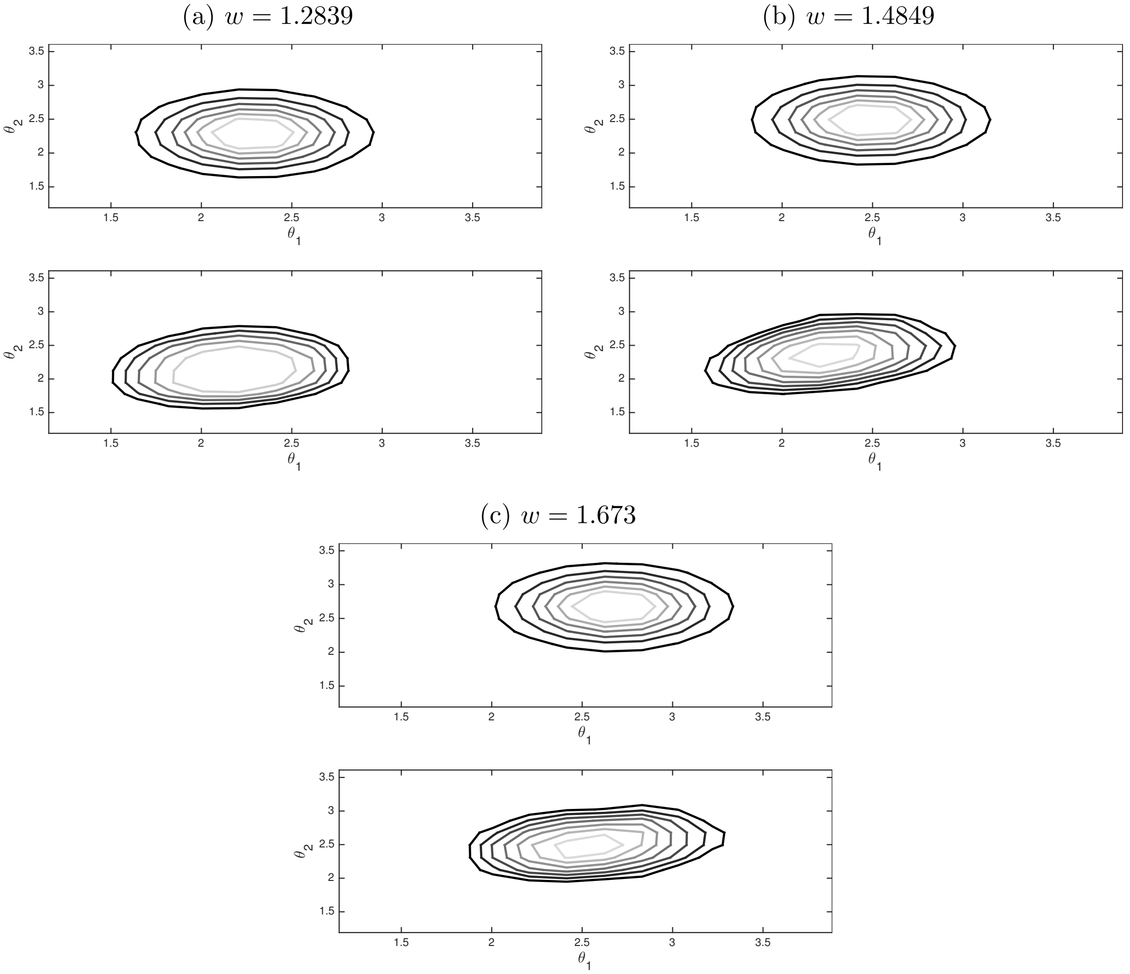
Next, we show $\|\hat{f}_{\theta_2|W} - f_{\theta_2|W}^0\| \to 0$ in probability. Lemma 9 shows consistency under the weak topology which implies $\langle \tilde{g}, \hat{f}_{\theta_2|W} - f_{\theta_2|W}^0\rangle$ for every $\tilde{g} \in \Theta_1 \times \mathcal{F}_{\theta|W}$. From Lemma 7

$$\begin{aligned} \|\hat{f}_{\theta_2|W}\|^2 - \|f_{\theta_2|W}^0\|^2 &\geq \langle f_{\theta_2|W}^0, \hat{f}_{\theta_2|W} - f_{\theta_2|W}^0\rangle + c\|\hat{f}_{\theta_2|W} - f_{\theta_2|W}^0\|^2 \\ &= \langle g^0, \hat{g} - g^0\rangle - \langle \theta_1^0, \hat{\theta}_1 - \theta_1^0\rangle + c\|\hat{f}_{\theta_2|W} - f_{\theta_2|W}^0\|^2. \end{aligned}$$

The first two terms of the right hand side converge to zero in probability by Lemma 9 and the left hand side converges to zero in probability by Lemma 8 *(i)*. Hence, $\|\hat{f}_{\theta|W} - f_{\theta|W}^0\|^2 \to 0$ in probability.

B Figures

Figure 2: Log linear demand ($n = 1000$, $\sigma_\theta^2 = 0.1$)
true $f_{\theta|W}$ (upper) vs. $\mathcal{P}_c\widehat{f}_{\theta|W}$ (lower)



(a) $w = 1.2839$

(b) $w = 1.4849$

(c) $w = 1.673$

Figure 3:
CARA example: Case 1 ($\sigma_x^2 = 1.0$)
(density of δ)

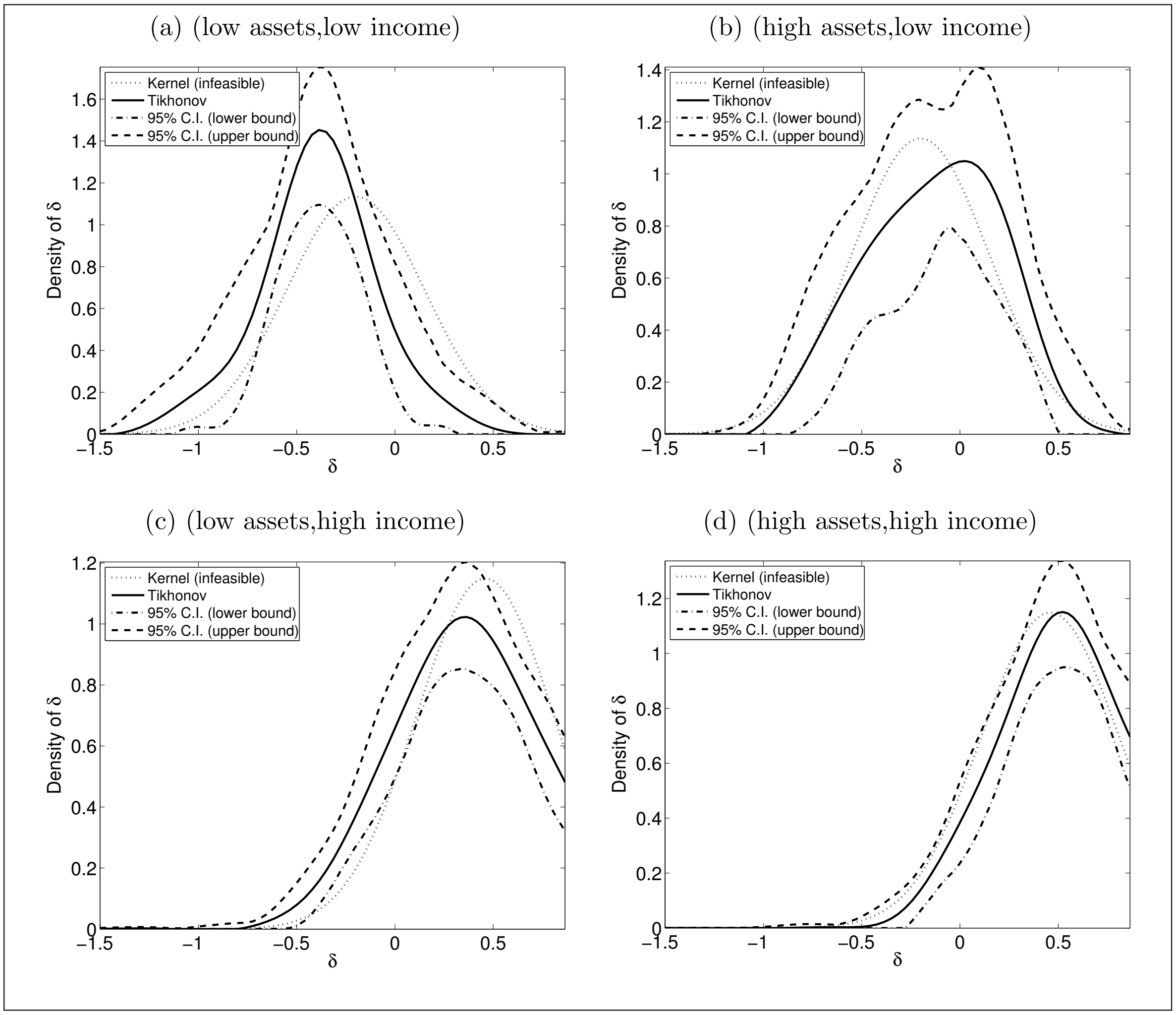


Figure 4:
CARA example: Case 2 ($\sigma_x^2 = 5.0$)
(density of δ)

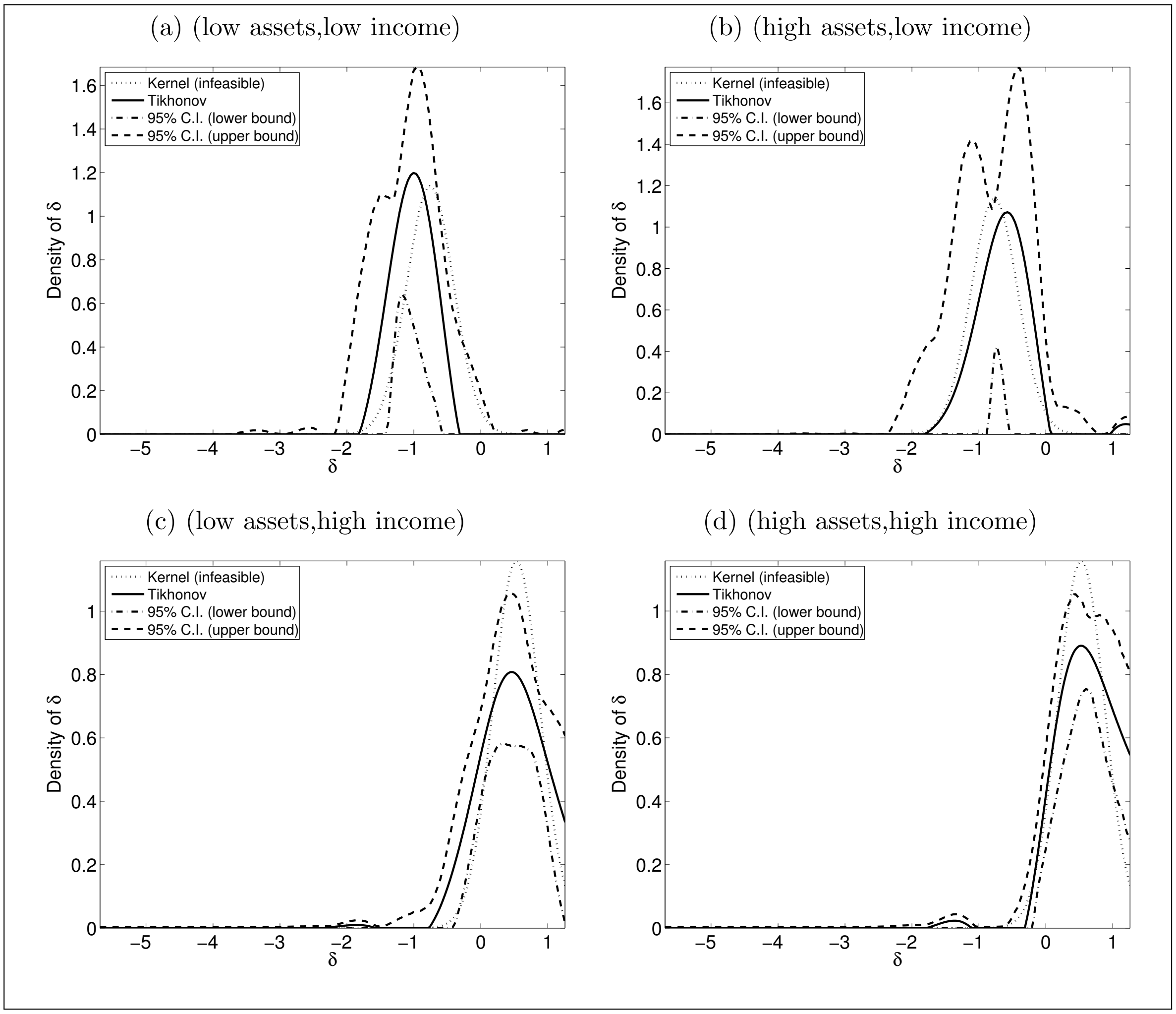
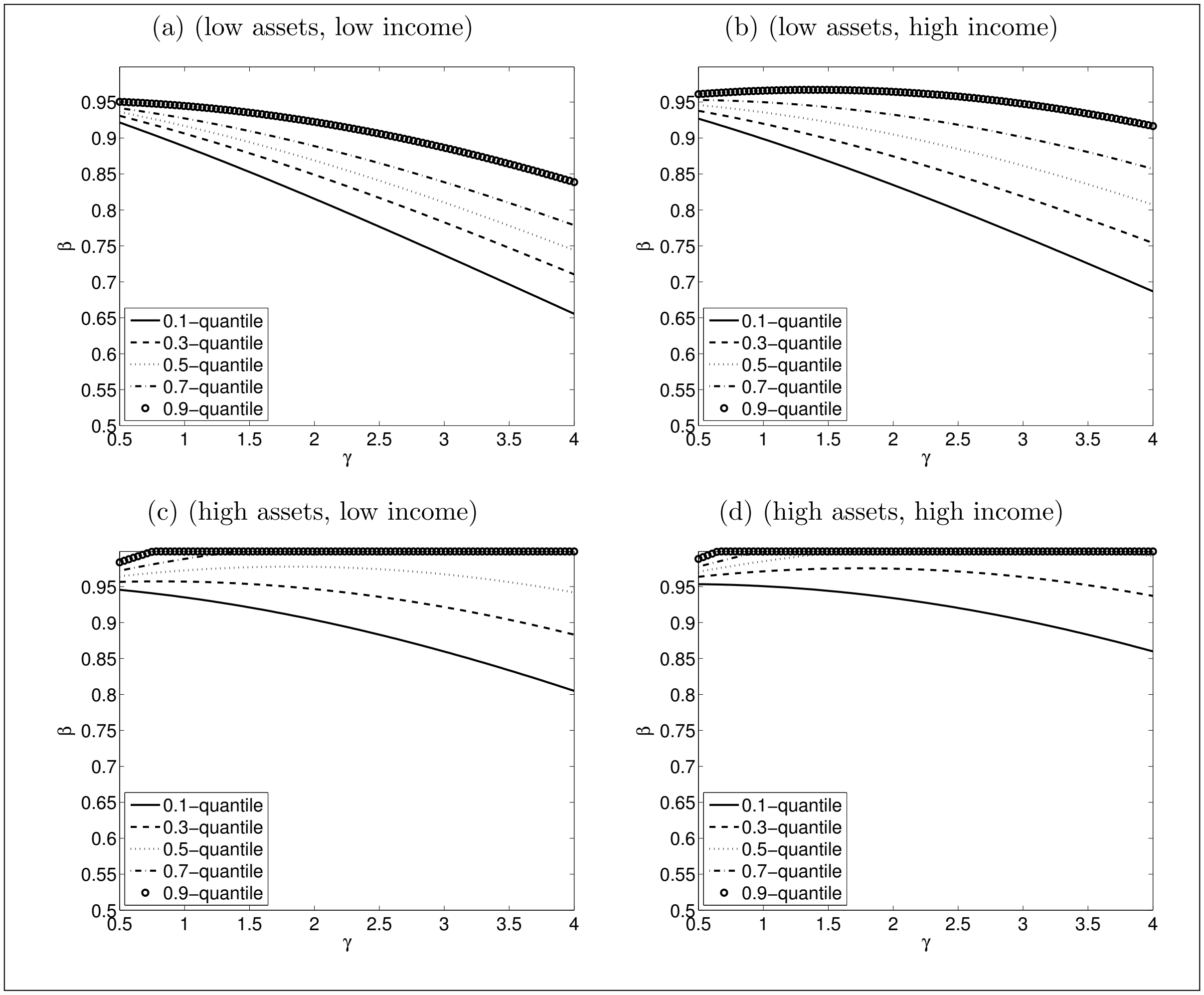


Figure 5:
CARA example: Case 1 ($\sigma_x^2 = 1.0$)
(quantile level sets of δ)


(a) (low assets, low income)
(b) (low assets, high income)
(c) (high assets, low income)
(d) (high assets, high income)
β
γ
0.1–quantile
0.3–quantile
0.5–quantile
0.7–quantile
0.9–quantile