# Semiparametric Estimation of Random Coefficients in Structural Economic Models

Stefan Hoderlein* — Boston College

Lars Nesheim — University College London

Anna Simoni — Università Bocconi

November 7, 2011 – PRELIMINARY –

**Abstract**

In structural economic models, individuals are usually characterized as solving a decision problem that is governed by a finite set of parameters. This paper discusses the nonparametric estimation of the density of these parameters if they are allowed to vary continuously across the population. We establish that the problem of recovering the density of random parameters falls into the class of non-linear inverse problem. This framework helps us to answer the question whether there exist densities that satisfy this relationship. It also allows us to characterize the identified set of such densities, to obtain conditions for point identification, and to establish that point identification is weak. Given this insight, we propose a consistent nonparametric estimator, and derive its asymptotic distribution. 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:** Structural Models, Heterogeneity, Nonparametric, Identification, Random Coefficients.

*Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA, email: stefan_hoderlein@yahoo.com. We have benefited from comments and discussions with Victor Aguirregabiria, Orazio Attanasio, Richard Blundell, Chris Carroll, Jeremy Fox, Dirk Krueger, Arthur Lewbel, Rosa Matzkin, Jean-Marc Robin, Yuanyuan Wan, Hal White, and seminar participants at the conferences for "Nonparametrics and Demand" at MIT, for "Revealed Preferences and Partial Identification" at Montreal University, the NBER 2011 Summer Meeting on Consumption, the CEAR 2011 conference in Denver, the Royal Economic Society Meeting 2011 in London, the ESEM 2011 in Oslo, Boston College, CalTech, Johns Hopkins University, NYU, UC San Diego, UCL, Stanford, and Toronto. All remaining errors are ours.

# 1 Introduction

**Motivation** Many structural microeconomic models postulate that individual decision makers solve complicated optimization problems which are governed by a small number of structural parameters $\theta$. These parameters are fixed for every individual, however, economic theory does not postulate that they be the same for every individual, and frequently they will differ substantially across individuals. Yet, in most empirical applications, the extent to which the individual decision makers are allowed to vary is severely constrained to depend entirely on observable variables, to involve only few discrete types, or to be monotonic in a scalar unobservable. A very natural way to make this assumption more appealing is to assume that the unobservable parameters $\theta$ in the individuals' decision problem form a continuously distributed random vector, and vary in an unrestricted, nonparametric way across the population. Proposing and analyzing such an approach is the main innovation in this paper.

To give an example, in the workhorse Euler equation models of the consumption literature, the consumption function is characterized by the first order condition

$$\partial_c u(C_t, \theta) = \mathbb{E}\left[\partial_c u(C_{t+1}, \theta) | \mathcal{I}_t\right] \tag{1.1}$$

where $u$ denotes instantaneous utility, $C_t$ consumption in period $t$, $\mathcal{I}_t$ is the information set of the economic agent in period $t$, and $\partial_x$ denotes the partial derivative with respect to $x$. The information set $\mathcal{I}_t$ consists of observable variables $Z_t$ and unobservable variables[1]. Interest centers on the distribution of (random) parameters $\theta$, but to not unduly restrict the generality of the model, we may want to allow for the possibility that there also be unobservable nuisance variables $\varepsilon_t$ in the model. These may enter $\mathcal{I}_t$, e.g., heterogeneous beliefs about the income process, or they may reflect measurement error (e.g., $C_t = C_t^* + \varepsilon_t$, where $C_t^*$ is the true consumption and $\varepsilon_t$ is a measurement error). We remark that solutions to problems of the type displayed in equation (1.1) are often only defined implicitly and have to be obtained numerically; however, our approach allows us to deal with these cases as well.

In this paper we propose a general framework to analyze a large class of such structural models. Specifically, we consider all structural economic model characterized by the following condition

$$\Psi(C, W, Z, \theta, \varepsilon) = 0 \tag{1.2}$$

where $C$ is a scalar observable outcome, $\theta$ is a $d$-vector of unobserved random parameters of interest, $\varepsilon$ is an unobservable random variable, $W$ is a $k$-vector of random variables allowed to be correlated with $\theta$ while $Z$ is a random $l$-vector of variables that are uncorrelated with $\theta$.

[1] In order to focus on the essential, we have taken the shortcut to set the interest rate equal to the discount rate.

Our analysis will cover two cases, one where there is no nuisance unobservable $\varepsilon$ and one which includes $\varepsilon$[2]. However, due to the greater generality, we largely focus on the latter. Our model is structural in that we assume that $\Psi$ is a known function coming from economic theory. Our aim is to identify and non-parametrically estimate the distribution of $\theta$ conditional on $W$. We do not require any monotonicity of $\Psi$ in $\varepsilon$ or $\theta$. While our results apply to general structural models of the form (1.2), we develop further results that apply to the specific examples in order to fix ideas and motivate the discussion.

The two key notions we pursue in this paper are heterogeneity and knowledge of the structural equation. When we lack information about the distribution of heterogeneity in the population $\left(\text{for example the density } f_{\theta|W}\right)$ but have knowledge about the structural function $\Psi$, we can use this knowledge to define a mapping from $f_{\theta|W}$ to the population distribution of observables $f_{C|WZ}$. In our setting, this mapping is provided by the equation

$$f_{C|WZ} = Tf_{\theta|W}, \quad P^W - \text{a.e.}, \tag{1.3}$$

where the operator $T$ maps the *pdf* of random parameters into the *pdf* of the observable variables $f_{C|WZ}$. When $\Psi$ can be uniquely solved for $C$ as a measurable function of the other variables, the operator $T$ can be explicitly characterized in terms of the structural economic model in (1.2). We will lagely focus on the special case in which $\Psi$ is differentiable, in which $C = \varphi(W, Z, \theta, \varepsilon)$ denotes a solution of (1.2) for $C$, given $(W, Z, \theta, \varepsilon)$. If $\varphi$ is invertible in $\varepsilon$, i.e. $\varepsilon = \varphi^{-1}(w, z, \theta, c)$, the operator takes the form

$$f_{C|WZ}(c; w, z) = \int_\Theta f_{\varepsilon|WZ\theta} \circ \varphi^{-1} \left| \frac{\partial_\varepsilon \Psi(c, w, z, \theta, \varphi^{-1}(w, z, \theta, c))}{\partial_c \Psi(c, w, z, \theta, \varphi^{-1}(w, z, \theta, c))} \right| 1_{S_{C|WZ\theta}}(c) f_{\theta|W}(\theta; w) d\theta,$$

where $f_{\varepsilon|WZ\theta}$ is the *pdf* of $\varepsilon$ conditional on $(W, Z, \theta)$ and $S_{C|WZ\theta}$ denotes the support of the conditional distribution of $C$ given $(W, Z, \theta)$. This paper focuses on this integral equation to characterize identification and obtain estimation of $f_{\theta|W}$. Given this representation, we can discuss the issues of existence, uniqueness and stability of the inverse. As a consequence of the assumption that our structural model is valid, the population value of $f_{C|WZ}$ is an element of the range of the operator. However, in a finite sample, an estimator for $f_{C|WZ}$ may not. Translated into econometric terms, existence will correspond to conditions for at least partial identification and will allow us to characterize the partially identified set, uniqueness will allow us to understand the conditions for point identification, and stability will relate to the question whether we can construct a feasible and consistent estimator, given the complex and high dimensional nature of the problem.

[2]For simplicity, we assume throughout this paper that $\varepsilon$ is a random unobserved scalar; this could be relaxed without great difficulty.

**Contributions relative to the Literature**. As already mentioned, this line of works extends the parametric structural models literature to allow for nonparametric random coefficients. This literature is vast; the consumption literature which originally motivated this research is, for instance, surveyed in Deaton (1992) and [Attanasio and Weber (2010)]. If parameter heterogeneity is to be introduced, identification becomes a crucial concern. The question is whether we can non-parametrically identify the distribution of preference parameters, and if yes, whether and how we can build an estimator on the identification principle.

To be able to answer this question, we require a nonparametric framework. The nonparametric features are not economically marginal generalizations. First, we answer the nonparametric identification question, i.e., clarify where the identifying power of the model emanates from. Second, we provide insights into when identification is only partial. Finally, we address the problems that we have in estimation even when the model is point identified. Specifically, we provide remedies that make consistent estimation feasible. All of these steps are related to contributions in the literature, and we quickly review them in the following:

First, there is a recent and innovative line of work that discusses identification of random coefficients in models that are motivated by empirical IO, see in particular Bajari, Fox, Kim and Ryan (2009), Fox and Gandhi (2010). These models are close in spirit to our approach in terms of the objectives of the analysis. The key differences are that this line of work focuses largely on a discrete support for the distribution of random coefficients, and that the identification results are not constructive. This has the effect that very high dimensional densities are identified (and estimable). Consequently, this literature lacks a notion that it is difficult to learn about the infinite dimensional objects of interest from the distribution of the data. Our approach allows to understand this effect for a large class of models and to construct consistent and stable estimators.

In terms of econometric models involving random parameters, our approach extends linear/single index nonparametric random coefficients model, as in Beran, Hall, Feuerverger (1994), [Hoderlein *et al.* (2010)], and [Gautier and Kitamura (2010)], to structural models which arise from economic theory models in a heterogeneous population and are only implicitly defined. In a model where there is no nuisance unobservable $\varepsilon_t$, our approach is related to their analysis, as outlined below.

In the case when there is nuisance heterogeneity $\varepsilon_t$, our approach resembles somwhat deconvolution approaches to modelling unobservables, a line of work that started with the seminal work of [Heckman and Singer (1984)] henceforth HS. In our notation, it is centered around the equation

$$f_{C|WZ}(c; w, z) = \int_\Theta f_{C|WZ\theta}(c; w, z, \theta) f_{\theta|W}(\theta; w) d\theta. \tag{1.4}$$

In HS's work - which is about duration analysis - the density $f_{C|WZ\theta}(c; w, z, \theta)$ is central; it is assumed to depend on a finite parameter $\sigma$ which is the main structural object of interest, while $f_{\theta|W}$ is a nuisance parameter. Closely related to HS are [Henry *et al.* (2011)], who focus on estimating $f_{C|WZ\theta}(c; w, z, \theta)$ nonparametrically while restricting $\theta$ to be finitely distributed, Kasahara and Shimotsu (2008), who also considers finitely many types, and [Bonhomme (2011)], who like HS aim at estimating a finite parameter of interest $\sigma$ when the exogenous variation comes from a Panel. This line of work is closely related to mixture models. In contrast to all of these references, in our model interest centers on $f_{\theta|W}$, and the kernel of the operator in (1.4) obtains structure from the economic primitives of the model.

The difference in the object of interest makes our work closer related to the inverse problem literature, see [Carrasco *et al.* (2007)] for an overview of inverse problems in econometrics. In particular, recovering the probability density of $\theta$ nonparametrically from (1.3) is equivalent to solving a convexly constrained integral equation of the first kind. Unconstrained integral equations of the first kind have been studied extensively in the literature on nonparametric instrumental regression, see *e.g.* [Florens (2003)], [Newey and Powell (2003)], [Darolles *et al.* (2011)] and [Hall and Horowitz (2005)]. While our object of interest is very different from a nonparametric IV regression function, we have some overlap with these references in terms of the tools we employ. In particular, we use Tikhonov regularization which was introduced by Tikhonov (1963) and introduced in econometrics by [Carrasco and Florens (2000)], [Florens (2003)] and [Darolles *et al.* (2011)], among others.

Our estimating equation is also related to the approach of [Hu and Schennach (2008)], where instead of unobserved regressors we focus on the unobserved heterogeneity parameters. Nevertheless, our model differs in many aspects from their model - not only because of the different object of interest (i.e., the distribution of random parameters), or the structural nonseparability of the model considered. Our exclusion restriction are fundamentally different from theirs (e.g., we do not assume conditional independence of $C$ and $Z$ given $\theta$) and motivated by the structural economic application. Moreover, in the case of measurement error, we do allow for correlation between the error and the unobservable of interest $\theta$. Finally, we do not require injectivity of the operator defining the estimating equation and we are able to estimate the identified set. More widely related are approaches where the nonseparable function $\Psi$ is unknown, but monotonic in at least some of its components, as in Evdokimov (2011) or Matzkin (2003), see Matzkin (2007) for a general overview.

**Structure of the Paper**. We develop our analysis of the above class of models in the following way. The next section describes our assumptions and notation and discusses our leading example: the semi-parametric Euler equation. Section 3 provides the main identification theorem. Section 4 discusses estimation by sample counterparts. Finally, we illustrate our approach with a simulation exercise in Section 5. An application using panel data on consumption panel constructed from the PSID and the CEX is transferred to a companion paper. Finally, Section 6 concludes. All the proofs are transferred to the Appendix.

# 2 The General Structural Model

## 2.1 Basic Definitions and Assumptions

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and $(C, W, Z, \theta, \varepsilon)$ be a real-valued random vector partitioned into $C \in \mathbb{R}$, $W \in \mathbb{R}^k$, $Z \in \mathbb{R}^l$, $\theta \in \mathbb{R}^d$ 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 $\sigma$-fields in $\mathbb{R}$, $\mathbb{R}^k$, $\mathbb{R}^l$, $\mathbb{R}^d$ and $\mathbb{R}$, respectively.

We use capital Latin letters for observable random variables and lowercase Latin letters for their realizations. The unobservable random variables and their realization will be denoted by lowercase Greek letters without distinction. The first assumption specifies the structural data generating process (DGP) that we are considering.

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

$$\Psi(C, W, Z, \theta, \varepsilon) = 0 \quad a.s. \tag{2.1}$$

*where $\Psi$ is a **known** real-valued function. Moreover, we assume that (2.1) has a unique global solution in terms of $C$:*

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

*where $\varphi : \mathbb{R}^{k+l+d+1} \to \mathbb{R}$ is a Borel-measurable function.*

This assumption describes how our structural model links observable variables $(C, W, Z)$ to unobservable ones $(\theta, \varepsilon)$. We distinguish between three different observable variables: $C$ is the classical dependent variable, while $W$ and $Z$ denote variables that cause $C$. The distinction between $W$ and $Z$ is made because some of these causal variables may be correlated with $\theta$, while others are independent[3]. The distinction between the unobservable variables $\theta$ and $\varepsilon$ is made to separate random parameters of interest $\theta$ from the error term $\varepsilon$. In our analysis, we are going to be completely nonparametric about the distribution of $\theta$, which is the distribution of interest, and parametric in the distribution of $\varepsilon$ in the sense that we allow unknown random parameters of finite-dimension in the distribution of $\varepsilon$.

**Remark 1** (Local versus global analysis)**.** The second part of Assumption 1, existence of a unique global solution, is quite strong. However, there are many economic models for which

[3] This obviously nests the case where all causal variables are independent of $\theta$, a situation that is probably not as common in structural modeling.

this assumption is verified. We give some examples below. In the cases in which second part of Assumption 1 is not satisfied our analysis will be valid locally, and not globally, in a neighborhood of a solution $(c_0, w_0, z_0, \theta_0, \varepsilon_0)$ of (2.1).

We do not require that the function $\varphi$ is available in closed-form. Given that it exists and is unique, it can as well be available only in a numerical form. Moreover, we do not require that $\varphi$ is globally monotone in $\varepsilon$ but it may be only piecewise monotone in $\varepsilon$. Though frequently invoked, any monotonicity condition at this stage is rather implausible, and hence this is an important weakening of assumptions. To account for piecewise monotonicity, we denote by $s$ the number of solutions in $\varepsilon$ of the equation $c = \varphi(w, z, \theta, \varepsilon)$ for given $(w, z, \theta)$. Thus, $s$ is a function of $(W, Z, \theta)$, *i.e.* $s := s(W, Z, \theta)$ and the solutions are $\varepsilon^i$, $i = 1, \dots, s$. In principle, $s$ can be either countable or uncountable but for simplicity we assume that s is a finite number. Let $\mathcal{E}_1, \dots, \mathcal{E}_s$ 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 $i = 1, \dots, s$, for given $(w, z, \theta)$. Then, we denote by $\varepsilon^i = \varphi_i^{-1}(w, z, \theta, \cdot) : \mathbb{R} \to \mathcal{E}_i$ the corresponding inverse mapping for given $(w, z, \theta)$. We write $\partial_c \varphi_i^{-1}(w, z, \theta, c)$ and $\partial_\varepsilon \varphi(w, z, \theta, \varepsilon)$ to denote the partial derivatives with respect to $C$ and $\varepsilon$ of $\varphi_i^{-1}$, $i = 1, \dots, s$ and $\varphi$, respectively, for given $(w, z, \theta)$.

We introduce a regularity assumption on $\Psi$ :

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

The next three assumptions that we introduce concern the stochastic character of the random vectors $(C, W, Z, \theta, \varepsilon)$. Under Assumption 1 the random variable $C$ has a degenerate conditional distribution, conditionally on $(W, Z, \theta, \varepsilon)$. The joint probability distribution of $(C, W, Z, \theta)$ is completely characterized by the joint probability distribution of $(W, Z, \theta, \varepsilon)$. For our analysis, all we need is knowledge, up to a finite-dimensional parameter, of the conditional distribution of $(\varepsilon | W, Z, \theta)$. This is given in the following assumption.

**Assumption 3.** *The conditional probability distribution $P_{\varepsilon|WZ\theta}$ on $\mathcal{B}_\varepsilon$ given $(W, Z, \theta)$ admits a Radon-Nikodym derivative $f_{\varepsilon|WZ\theta}$ with respect to the Lebesgue measure. This probability density function (pdf, hereafter) $f_{\varepsilon|WZ\theta}$ is known, up to a finite-dimensional parameter $\theta_\varepsilon \subset \theta$.*

This assumption allows $\varepsilon$ to depend on all variables in the model but is satisfied also in the case where $\varepsilon$ is independent of $(W, Z, \theta)$, which may be relevant in the measurement error setup. Note also that, unlike in deconvolution, $\varepsilon$ does not need to be independent of $\theta$ since we are not de-convolving a probability distribution. In general, we can allow for $\varepsilon$ and $\theta$ to be dependent even though in many applications independence is plausible. In other applications, it may be useful to split the random parameter $\theta$ into two subparameters $(\theta_1, \theta_2)$ and confine dependence between $\varepsilon$ and $\theta$ to one of the two components of $\theta$. Mathematically, this is equivalent to having the distribution of $\varepsilon$ that depends on some $\varepsilon$-specific heterogeneity parameter. For instance, we could model $P_{\varepsilon|WZ\theta}$ to be normal with mean $\mu$ and variance $\sigma^2$. Thus, $\mu$ and $\sigma^2$ may be functions of $(W, Z, \theta)$ or functions of additional parameters that may vary across the population so that $\theta$ is embedded in the vector $\tilde{\theta} = (\theta, \mu, \sigma^2)'$ and we recover the distribution of the whole $\tilde{\theta}$.

This allows a great deal of flexibility in structural modeling. By allowing $f_{\varepsilon|WZ\theta}$ to be known up to a finite dimensional random parameter (a parameter included in the vector $\theta$), we allow for cases in which very little is known about $f_{\varepsilon|WZ\theta}$; thus the specification can be very close to a nonparametric specification for $f_{\varepsilon|WZ\theta}$. To give an idea about further weakening: $f_{\varepsilon|WZ\theta}$ could also be a finite mixture of normal *pdf* where, besides the vector of means and variances, also the mixture weights could depend on $\theta$ or be part of $\theta$.

We turn now to consider the joint conditional probability distribution $P_{CZ\theta|W}$ on $\mathcal{B}_\mathcal{C} \otimes \mathcal{B}_\mathcal{Z} \otimes \mathcal{B}_\Theta$ conditional on $W$. We assume that this distribution is absolutely continuous with respect to the Lebesgue measure with Radon-Nikodym derivative $f_{CZ\theta|W}$. The conditional probability distribution $P_{\theta|W}$ on $\mathcal{B}_\Theta$ (resp. $P_{CZ|W}$ on $\mathcal{B}_\mathcal{C} \otimes \mathcal{B}_\mathcal{Z}$) conditional on $W$ admits a *pdf* $f_{\theta|W}$ (resp. $f_{CZ|W}$) with respect to the Lebesgue measure.

We denote by $\mathcal{C} \subset \mathbb{R}$, $\mathcal{Z} \subset \mathbb{R}^l$ and $\Theta \subset \mathbb{R}^d$ the supports of $P_{C|ZW}$, $P_{Z|W}$ and $P_{\Theta|W}$, respectively, where $P_{Z|W}$ is the conditional distribution on $\mathcal{B}_\mathcal{Z}$ given $W$. We consider $C$, $Z$ and $\theta$ to be jointly distributed while $W$ can be either continuous or discrete. The marginal distribution on $\mathcal{B}_\mathcal{W}$ is denoted by $P_W$ and has support $\mathcal{W} \subset \mathbb{R}^k$. The joint distribution on $\mathcal{B}_\mathcal{Z} \otimes \mathcal{B}_\mathcal{W}$ is denoted by $P_{WZ}$. We assume the following:

**Assumption 4.** *The pdf $f_{C|WZ}$ and $f_{\theta|W}$ are strictly positive and bounded on their supports $P_{WZ}$-a.s., i.e. $0 < m_c < f_{C|WZ} < M_c$ and $0 < m_\theta < f_{\theta|W} < M_\theta$ for constants $m_c, M_c, m_\theta, M_\theta$. Moreover, the support $\Theta$ of $f_{\theta|W}$ does not depend on $W$.*

We conjecture that a setup where some or all of the variables is discrete could be tackled by similar arguments, however, it is beyond the scope of this analysis as it stands.

The last assumption we introduce is an independence condition and is important for point identification of the *pdf* $f_{\theta|W}$ of the structural parameters of interest.

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

Under this assumption we can decompose $f_{CZ\theta|W}$ as $f_{C|WZ\theta} f_{\theta|W} f_{Z|W}$, where $f_{Z|W}$ is the marginal *pdf* of $Z$ given $W$. Why do we invoke this assumption, and not a marginal independence condition? First, note that assumption 5 would be satisfied if $(Z, W)$ are independent of

$\theta$ - in this case, all variables would be part of $Z$ and there would be no $W$. However, in many structural models, this is too strong an assumption, and only some variables are independent of the random coefficients, while other cannot be.

In the following, we show that our framework allows to consider various important examples:

## 2.2 Examples

**Example 1** (Linear endogenous random coefficient model). *Consider the linear model on individual level, where a true outcome $C^*$ is determined by a random vector of exogenous explanatory variables $Z_1$, and a scalar endogenous explanatory variable $X$. Suppose we have in addition access to another random variable $Z_2$ which acts as an instrument for $X$ in the sense that it is excluded from the outcome equation, but related to $X$ through a general function $\phi$ that is nonparametric in $Z_1, Z_2$ and a scalar monotonically entering unobservable $W$. More formally, our model is*

$$\begin{aligned} C^* &= \theta_0 + Z_1'\theta_1 + X\theta_2 \\ X &= \phi(Z, W) \end{aligned}$$

*where $\theta = (\theta_0, \theta_1', \theta_2)'$ is a random vector of parameters, $\phi$ is a strictly monotone function in $W$, and $Z = (Z_1, Z_2)$. In addition, we allow for a classical scalar measurement error $\eta$, i.e. $C = C^* + \eta$. We assume that the driving exogenous sources of variation $Z$ are instruments in the sense that $Z \perp (\theta, \eta, W)$, and we impose the normalization $W \sim \mathcal{U}[0,1]$, and assume for the moment that $\eta|\theta W Z \sim \mathcal{N}[0,1]$.*

*Substituting all elements into the outcome equation, we obtain*

$$C = Z_1'\theta_1 + \phi(Z_2, W)\theta_2 + \varepsilon,$$

*where $\varepsilon = \eta + \theta_0$ has $\varepsilon|\theta W Z \sim \mathcal{N}[\theta_0, 1]$. In this model, we are interested in recovering the conditional density function of $\theta$ given $W$ which, by integrating out $W$, allows to recover the density of $\theta$. Note that this model fits precisely into the framework since the function $\phi$ is known from the first stage equation alone, and hence the entire functional relationship is known. Note, moreover, that the distribution of $\varepsilon|\theta W Z$ is known as well. At this point, it is obvious that we could have let the density of $\eta$ depend on the entire vector $\theta$, or on $(W, Z)$. On top of this, the density of $\eta$ may have been coming from a mixture of normals with mixing weights that vary across the population. In addition, if a validation sample is available, a nonparametric pilot estimate of $f_{\eta|ZW}$ may be plugged in as well. Finally, a modified version of our approach also allows for the case without $\eta$, as detailed below.*

**Example 2** (Intertemporal Consumption Model)**.** *Consider the absolute risk aversion finite-horizon intertemporal utility maximization problem with constant interest rates and with income following a random walk. Specifically, let $A_t$ be a consumer's assets and $Y_t$ be his/her income. More precisely, $A_t$ is the beginning of period assets after having received all interest payments. Define $R = (1 + r)$ where $r$ is the interest rate which we assume to be non-stochastic. Let $s_t = (A_t, Y_t)$ be the state vector and let $v_t(s_t)$ be the consumer's value function at date $t$. Denote the coefficient of risk aversion by $\gamma$ and let $\beta$ represent the discount factor. At each date $t \leq T$, the consumer's value function is defined by*

$$v_t(s_t) = \max_{\{C_t^*\}} \left\{ \begin{array}{c} -\frac{e^{-\gamma c_t}}{\gamma} + \beta \mathbb{E}\left[v_{t+1}(S_{t+1}) \,|I_t(\theta_2)\right] \\ \textit{subject to} \\ A_{t+1} = R(A_t + Y_t - C_t^*) \\ Y_{t+1} = Y_t + \eta_{t+1} \end{array} \right\}$$

*where $\eta_t \sim N\left(0, \sigma_\eta^2\right)$. Here, the parameters are $\theta_1 = (\gamma, \beta)$ and $\theta_2 = \sigma_\eta^2$. At time $t$, the consumer's information set $I_t(\theta_2)$ consists of $\{\eta_s\}$ for all $s \leq t$. Let $C_t^*$ be actual consumption and suppose consumption is measured with error so that observed consumption is $C_t = C_t^* + \varepsilon_t$. Let $W_t = (A_t, Y_{t-1})$ and $Z_t = Y_t - Y_{t-1}$. Then this example can be cast precisely into the framework described above. In terms of the variables $(C_t, W_t, Z_t)$, the Euler equation is*

$$e^{-\gamma(C_t-\varepsilon_t)} - \beta \mathbb{E}\left[\partial_A v_{t+1}\left(R\left(W_t^1 + W_t^2 + Z_t - C_t + \varepsilon_t\right), W_t^2 + Z_t\right) |I_t(\theta_2)\right] = 0.$$

*The consumption function (with measurement error) takes the form*

$$C_t = \phi_{1t} W_t^1 + \phi_{2t}\left(W_t^2 + Z_t\right) + m_t(\gamma, \beta, \theta_2) + \varepsilon_t \tag{2.2}$$

*with*

$$m_t(\gamma, \beta, \theta_2) = \phi_{3t} + \phi_{4t}\gamma + \phi_{5t}(\theta_2)\frac{\ln \beta}{\gamma}.$$

*The vector $\phi_t = (\phi_{1t}, \phi_{2t}, \phi_{3t}, \phi_{4t}, \phi_{5t})$ consists of parameters that depend only on $R$, $t$ and $\theta_2$ (see, e.g., Caballero (1990)). The vector $\theta = (\theta_1, \theta_2) = (\gamma, \beta, \sigma_\eta^2)$ is assumed to be a time-invariant random coefficient vector, heterogeneously distributed in the population. We assume that the income process $(Y_t)_{t=1,..,T} \perp \theta_1$ and that $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$*

*Because $\theta$ is time invariant and determines both past and current consumption and savings decisions, in general, it will be correlated with $W_t$. Note also, that $Y_t$ cannot be used as an exogenous variable. While it is assumed to be marginally independent of $\theta$. It is not independent of $\theta$ conditional on $A_t$. Observe, however, that $Y_t - Y_{t-1}$ is assumed to be independent of the entire history so that our choice $W_t = (A_t, Y_{t-1})$ and $Z_t = Y_t - Y_{t-1}$ implies that $\theta \perp Z_t | W_t$, where $W_t = (W_t^1, Z_{t-1})$.*

**Example 3.** *Consider the independent private value* first-price sealed-bid auction model *with risk averse bidders considered in Campo, Guerre, Perrigne and Voung (2009). Let* $I \geq 2$ *denote the number of potential bidders and* $\{v_i\}_{i=1,..,I}$ *be the bidders' private values which are drawn independently from an unknown distribution. This distribution may depend on observed characteristics* $Z$ *of the auctioned objects, and hence we write* $F(\cdot|Z, I)$*. We assume that* $F(\cdot|I)$ *is absolutely continuous with density* $f(\cdot|I)$ *on a compact support* $[\underline{v}(I), \bar{v}(I)] \subset \mathbb{R}_+$*. Let* $U(\cdot)$ *be the bidders' von Neuman Morgenstern utility function with* $U(0) = 0$*,* $U'(\cdot) > 0$ *and* $U''(\cdot) \leq 0$ *because of potential risk aversion. Denote by* $s(\cdot) \equiv s(\cdot; U, F, I)$ *the Bayesian Nash equilibrium bidding strategy. Then, from equation (1) in Campo et al. (2009), the bidder i's optimal bid* $b_i = s(v_i)$ *solves the following differential equation:*

$$s'(v_i) = (I-1)\frac{f(v_i|I)}{F(v_I|I)}\lambda(v_i - b_i)$$

*for all* $v_i \in [\underline{v}(I), \bar{v}(I)]$*, where* $\lambda(\cdot) = U(\cdot)/U'(\cdot)$*. The boundary condition is* $s(\underline{v}(I)) = \underline{v}(I)$*.*

*Assume that* $U(\cdot)$ *is of CRRA type with* $U(x) = x^{1-\theta_1}$ *for* $0 \leq \theta_1 < 1$*. We assume that all bidders are symmetic within an auction, but that the preferences may vary across auctions, e.g., in response to some unobserved auction specific factor that influences the utility. Then,* $\lambda(v-b) = (1-\theta_1)^{-1}(v-b)$ *and the bidder i's optimal bid which solves the previous differential equation is*

$$b_i = v_i - [F(v_i)|I]^{-\frac{I-1}{1-\theta_1}}\left(\int [F(v_i|I)]^{\frac{I-1}{1-\theta_1}} dv_i + \int [F(v_i|I)]^{\frac{I-1}{1-\theta_1}} dv_i\Big|_{v_i=\underline{v}(I)}\right). \tag{2.3}$$

*In this example,* $b_i$ *and* $v_i$ *play the roles of* $C$ *and* $\varepsilon$ *in Assumption 1 and the function on the right hand side of (2.3) plays the role of the function* $\varphi$ *of our general model. By assuming a parametric form for the pdf of* $v_i$ *given* $I$*, we can recover the pdf of* $b_i$ *conditional on* $\theta$ *and* $Z$*. For instance, if* $v_i$ *is uniform over* $[\underline{v}(I), \bar{v}(I)]$*, then* $f_{b|\theta_1}(b; \theta_1) = \frac{1}{\bar{b}(I)-\underline{b}(I)}\frac{I-\theta_1}{I-1}$*. Moreover, note that we could let the pdf of* $v_i$ *depend on additional parameters* $\theta_2$ *that could be heterogeneous across the population. Another variable that may be unobserved and varying across the population of auctions is the actual number of bidders.*

Obviously, this list may be greatly extended: For instance, we may consider the original Heckman and Singer (1984) work on duration, where $\varepsilon$ would be an extreme value distribution. Alternatively, we may apply our framework to structural labor models of the form studied in Keane and Wolpin (1997), or to Berry, Levinsohn and Pakes type IO models in the spirit of Fox and Gandhi (2009). Instead of elaborating on the details, we leave additional applications of this framework for future research, and turn now to the completion of the definition of the formal setting of our model.

### 2.3 A Hilbert-Space Setting

The natural space for probability density functions is the $L^1$ space with respect to the Lebesgue measure endowed with either the $L^1$- or the Hellinger- metric. Despite this fact, we develop our analysis in $L^2$ spaces with respect to some suitable measures to exploit may of the desirable properties of Hilbert space operators.

For this purpose, we introduce two non-negative weighting functions on $\Theta$ and $\mathcal{C}\times\mathcal{Z}$ that we denote by $\pi_\theta$ and $\pi_{cz}$, respectively. Define the space $L^2_{\pi_\theta}(\Theta)$ (resp. $L^2_{\pi_{cz}}(\mathcal{C}\times\mathcal{Z})$) of real-valued functions defined on $\Theta$ (resp. on $\mathcal{C}\times\mathcal{Z}$), and indexed by the random variable $W$, which are $P^W$-a.s. square integrable with respect to $\pi_\theta$ (resp. $\pi_{cz}$), i.e.,

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

For brevity, we denote $L^2_{\pi_\theta}(\Theta)$ by $L^2_{\pi_\theta}$ and $L^2_{\pi_{cz}}(\mathcal{C}\times\mathcal{Z})$ by $L^2_{\pi_{cz}}$. Further, we denote the scalar product by $<\cdot,\cdot>$ and the induced norm by $||\cdot||$ in both $L^2_{\pi_\theta}$ and $L^2_{\pi_{cz}}$ without distinction. That is $\forall h_1,h_2\in L^2_{\pi_\theta}$, $<h_1,h_2>=\int h_1(\theta;W)h_2(\theta;W)\pi_\theta(\theta)d\theta$ and $\forall\psi_1,\psi_2\in L^2_{\pi_{cz}}$, $<\psi_1,\psi_2>=\int\psi_1(c,z;W)\psi_2(c;z,W)\pi_{cz}(c,z)d\theta$.

Since our analysis is conditional on $W$, we allow the weighting functions $\pi_\theta$ and $\pi_{cz}$ to be indexed by $W$ too. The set of conditional probability density functions relevant for our analysis are denoted and defined as follows

$$\begin{aligned} \mathcal{F}_{\theta|W} &:= \left\{f\in L^2_{\pi_\theta};\ f(\cdot;\cdot)\text{ is a transition probability from }(\mathbb{R}^k,\mathcal{B}_\mathcal{W})\text{ to }(\mathbb{R}^d,\mathcal{B}_\Theta)\right\} \\ \mathcal{F}_{C|WZ} &:= \left\{f\in L^2_{\pi_\theta};\ f(\cdot;\cdot,\cdot)\text{ is a transition probability from }(\mathbb{R}^{k+l},\mathcal{B}_\mathcal{W}\otimes\mathcal{B}_\mathcal{Z})\text{ to }(\mathbb{R},\mathcal{B}_\mathcal{C})\right\} \\ \mathcal{F}_{C|WZ\theta} &:= \left\{\text{ transition probabilities from }(\mathbb{R}^{k+l+d},\mathcal{B}_\mathcal{W}\otimes\mathcal{B}_\mathcal{Z}\otimes\mathcal{B}_\Theta)\text{ to }(\mathbb{R},\mathcal{B}_\mathcal{C})\right\}. \end{aligned}$$

While $\mathcal{F}_{\theta|W}\subset L^2_{\pi_\theta}$ and $\mathcal{F}_{C|WZ}\subset L^2_{\pi_{cz}}$, we do not assume that $\mathcal{F}_{C|WZ\theta}\subset L_{\pi_{cz}}\times L^2_{\pi_\theta}$. When this last condition is satisfied, we can provide a simple characterization of the identified set. However, our estimation method can be used to estimate one element of the identified set even when this last condition is not satisfied.

# 3 Identification of the Distribution of Parameters

In this section we use the structural relationship given in 1 to characterize the conditional distribution of $C$ given $(W,Z)$ ($i.e.$ observable variables) as a transformation of the conditional distribution of $\theta$ given $W$ ($i.e.$ unobservable variables). In other words, the structural economic model describes the direct mapping from the distribution of unobservables to the distribution of observables.

### 3.1 Linear Integral Equation and Non-Linear Inverse Problem

Given the the direct mapping described above, the econometrician is interested in the *inverse problem* of recovering the conditional *pdf* $f_{\theta|W}$ of $(\theta|W)$ (*i.e.* the cause) from the distribution of observables. The following theorem characterizes the structural direct/inverse problem.

**Theorem 1.** *Let Assumption 4 and 5 be satisfied and assume* $f_{C|WZ} \in \mathcal{F}_{C|WZ}$. *Then, the structural probability density function* $f_{\theta|W}$ *is a solution of*

$$f_{C|WZ} = T f_{\theta|W} \quad \wedge \quad f_{\theta|W} \in \mathcal{F}_{\theta|W}, \quad P^W - a.s. \tag{3.1}$$

*where* $T$ *is a linear operator defined as* $Th = \int_\Theta f_{C|WZ\theta}(c; w, z, \theta) h(\theta; w) d\theta$, $\forall h \in L^2_{\pi_\theta}$ *and* $f_{C|WZ\theta} \in \mathcal{F}_{C|WZ\theta}$.
*If in addition assumptions 1, 2 and 3 hold then, the operator* $T$ *can be rewritten as:* $\forall h \in L^2_{\pi_\theta}$,

$$Th = \int_\Theta \sum_{i=1}^{s} (f_{\varepsilon|WZ\theta} \circ \varphi_i^{-1})(c, w, z, \theta) \left| \frac{\partial_c \Psi(c, w, z, \theta, \varphi_i^{-1}(w, z, \theta, c))}{\partial_\varepsilon \Psi(c, w, z, \theta, \varphi_i^{-1}(w, z, \theta, c))} \right| 1_{S_{C|WZ\theta}}(c) h(\theta; w) d\theta \tag{3.2}$$

*where* $c = \varphi(w, z, \theta, \varepsilon)$ *is the a.s.-explicit solution of* $\Psi(c, w, z, \theta, \varepsilon) = 0$ *and* $S_{C|WZ\theta}$ *denotes the support of* $f_{C|WZ\theta}$.

Note that although the operator $T$ depends on $W$, we use the short-hand notation $T$ and leave implicit the dependence on $W$.

The operator $T$ in equation (3.1) is a mixing operator and the theorem characterizes the object of interest $f_{\theta|W}$ as the solution of a *convexly constrained Fredholm integral equation of I kind*. Equation (3.1) states that $f_{C|WZ}$ is a $P^W$-a.s. $f_{\theta|W}$-mixture of $\mathcal{F}_{C|WZ\theta}$ or, in other words, that $f_{C|WZ}$ belongs $P^W$-a.s. to the convex hull $\mathbb{M}_\Theta$ of $\mathcal{F}_{C|WZ\theta}$:

$$f_{C|WZ} \in \mathbb{M}_\Theta := \{h = \int_\Theta f_{C|WZ\theta}(\cdot; w, \cdot, \theta) f_{\theta|W} d\theta;\ f_{\theta|W} \in \mathcal{F}_{\theta|W}\}.$$

Therefore, recovering $f_{\theta|W}$ from (3.1) is an (ill-posed inverse problem). The main contribution of the theorem is the characterization (3.2) of the operator $T$ in terms of the structural quantities of the economic problem. This nice characterization is possible since $C$ is determined by $\theta$ through a structural relation.

While the functional form of $f_{\varepsilon|WZ\theta}$ must be specified, the specification may allow for *uncertainty* about the functional form of $f_{\varepsilon|WZ\theta}$ by parameterizing this distribution with unknown random parameters (*i.e.* hyperparameters). These parameters can then be incorporated in the vector $\theta$. The larger the number of random parameters in $f_{\varepsilon|WZ\theta}$, the less stringent our assumption about the conditional distribution of $\varepsilon$ and the more close to a nonparametric approach we are.

The function $\frac{f_{C|WZ\theta}}{\pi_\theta}$ is called the *kernel* of the operator $T$. If $\frac{f_{C|WZ\theta}}{\pi_\theta} \in L^2_{\pi_{cz}} \times L^2_{\pi_\theta}$ then the range of the operator $T$ is contained in $L^2_{\pi_{cz}}$. In fact, $\mathcal{R}(T) \subset L^2_{\pi_{cz}}$ if and only if $\forall h \in L^2_{\pi_\theta}$, $||Th|| < \infty$. By using the Cauchy-Schwarz inequality we obtain

$$\begin{aligned} ||Th||^2 &= \int_{\mathcal{C}}\int_{\mathcal{Z}} \left\langle \frac{f_{C|WZ\theta}}{\pi_\theta}, h\right\rangle^2 \pi_{cz}(c,z)dcdz \leq \int_{\mathcal{C}}\int_{\mathcal{Z}} \left|\left|\frac{f_{C|WZ\theta}}{\pi_\theta}\right|\right|^2 ||h||^2 \pi_{cz}(c,z)dcdz \\ &= ||h||^2 \left|\left|\frac{f_{C|WZ\theta}}{\pi_\theta}\right|\right|^2 \end{aligned} \tag{3.3}$$

where the last norm is the norm in $L^2_{\pi_{cz}} \times L^2_{\pi_\theta}$-space. Thus, $\frac{f_{C|WZ\theta}}{\pi_\theta} \in L^2_{\pi_{cz}} \times L^2_{\pi_\theta}$ is sufficient for $T$ to be bounded and compact. In the rest of the paper, we assume that $\pi_\theta$ and $\pi_{cz}$ are suitably chosen so that this condition is satisfied. In practice, the econometrician specifies $f_{\varepsilon|WZ\theta}$ and $\pi_\theta$ instead of $f_{C|WZ\theta}$. Point *(i)* of the next proposition gives sufficient conditions for compactness and boundedness of the operator $T$ in terms of $f_{\varepsilon|WZ\theta}$.

**Proposition 1.** *Let $T$ be the operator defined in (3.2) with domain $L^2_{\pi_\theta}$. The following holds:*

*(i) if $s$ is finite, $|\partial_\varepsilon \varphi(w,z,\theta,\varepsilon)|$ is uniformly bounded from below $P^W$-a.s. and $f_{\varepsilon|WZ\theta}$ is square integrable in $(\varepsilon, Z, \theta)$ with respect to $\frac{\pi_{cz}}{\pi_\theta}\Big|_{c=\varphi(w,z,\theta,\varepsilon)}$, $P^W$-a.s. then, $T$ is a $P^W$-a.s. bounded and compact operator with range included in $L^2_{\pi_{cz}}$;*

*(ii) if $f_{C|WZ\theta} \in \mathcal{F}_{C|WZ\theta}$, $\mathcal{F}_{C|WZ\theta} \subset L^2_{\pi_{cz}} \times L^2_{\pi_\theta}$, $P^W$-a.s. and $\pi_\theta$ is strictly positive then, $T$ is a $P^W$-a.s. bounded and compact operator with range included in $L^2_{\pi_{cz}}$.*

This proposition shows that compactness of $T$ depends both on the structural model (characterized by the structural function $\varphi$ and the density $f_{\varepsilon|WZ\theta}$) and on the weights $\pi_{cz}$ and $\pi_\theta$. Compact operators on Hilbert spaces are quite easy to work with because they have many similarities with linear operators on finite dimensional spaces. On top of that, they have appealing spectral properties and can be approximated by operators with finite-dimensional range that are norm convergent.

**Remark 2** (Singular value decomposition - SVD)**.** When the operator $T$ is compact with infinite-dimensional range (*i.e.* its kernel is not degenerate) then $T^*T$ is characterized by a countable number of eigenvalues which accumulate only at zero. Moreover, it admits the following *singular value decomposition*:

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

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 singular values are the nonnegative square roots of the eigenvalues of $T^*T$ (and also of $TT^*$). 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 $T^*$.

**Proposition 2** (Adjoint of $T$)**.** *Let $T : L^2_{\pi_\theta} \to L^2_{\pi_{cz}}$ be the operator defined in (3.2). The operator $T^*$ defined as*

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

*$\forall \psi \in L^2_{\pi_{cz}}$ and $f_{C|WZ\theta} \in \mathcal{F}_{C|WZ\theta}$, is the adjoint of $T$. The operator $T^* : L^2_{\pi_{cz}} \to L^2_{\pi_\theta}$ is bounded and linear.*

From now on we denote by $T|_{\mathcal{F}_{\theta|W}}$ the operator $T$ restricted to $\mathcal{F}_{\theta|W}$. Thus, if $\mathcal{D}(\cdot)$ and $\mathcal{R}(\cdot)$ denote the domain and the range of an operator, respectively, we have $\mathcal{D}(T|_{\mathcal{F}_{\theta|W}}) = \mathcal{D}(T) \cap \mathcal{F}_{\theta|W} = \mathcal{F}_{\theta|W}$ and $\mathcal{R}(T|_{\mathcal{F}_{\theta|W}}) \subset \mathcal{F}_{C|WZ} \subset \mathcal{R}(T) \subset L^2_{\pi_{cz}}$. Following Example 2.4 in [Carrasco *et al.* (2007)], the adjoint $T|^*_{\mathcal{F}_{\theta|W}}$ of $T|_{\mathcal{F}_{\theta|W}}$ is given by $T|^*_{\mathcal{F}_{\theta|W}} = \mathcal{P}_c T^*$ where $\mathcal{P}_c$ denotes the metric projector onto $\mathcal{F}_{\theta|W}$.

Notice that $\mathcal{F}_{\theta|W}$ is a convex set and hence the operator $T|_{\mathcal{F}_{\theta|W}}$ is an *affine* operator, *i.e.* $T|_{\mathcal{F}_{\theta|W}} : \mathcal{F}_{\theta|W} \to L^2_{\pi_{cz}}$ satisfies: $T|_{\mathcal{F}_{\theta|W}}((1-\lambda)f_1 + \lambda f_2) = (1-\lambda)T|_{\mathcal{F}_{\theta|W}}(f_1) + \lambda T|_{\mathcal{F}_{\theta|W}}(f_2)$ whenever $f_1, f_2 \in \mathcal{F}_{\theta|W}$, $0 < \lambda < 1$.

**Remark 3** (Nonlinear inverse problem)**.** Despite the linearity of $T$, the inverse problem in (3.1) is potentially non-linear because of the constraint $f_{\theta|W} \in \mathcal{F}_{\theta|W}$. By nonlinear inverse problem we mean that the $\mathcal{F}_{\theta|W}$-constrained pseudoinverse of $T$ is a non-linear operator. As we explain in more detail in section 4.3, this operator defines the constrained-best-approximate solution of (3.1).

**Remark 4** (Existence of a solution)**.** The inclusion of $\mathcal{R}(T|_{\mathcal{F}_{\theta|W}})$ in $\mathcal{F}_{C|WZ}$ guarantees the existence of at least one solution to (3.1). By using the terminology of inverse problems this mean that $f_{C|WZ}$ is "attainable".

**Remark 5** (Deconvolution)**.** When $f_{C|WZ\theta}(c; w, z, \theta)$ belongs to the location family with location parameter $\theta$, *i.e.* $f_{C|WZ\theta}(c; w, z, \theta) = f_{C|WZ}(c - \theta; w, z)$, then the operator $T$ can be written as $Th = \int_\Theta f(c - \theta; w, z)h(\theta; w)d\theta$, $\forall h \in L^2_{\pi_\theta}$. In this case, inversion of this operator is the classical deconvolution problem. Because of this particular structure, one can use special techniques to solve the inversion problem that are not available for general mixture models, see *e.g.* Carrasco and Florens (2010).

## 3.2 The identified set

In this section, we discuss point and set identification of $f_{\theta|W}$. The distribution $f_{\theta|W} \in \mathcal{F}_{\theta|W}$ will be called *identified* (with respect to the class $\mathcal{F}_{\theta|W}$) if

$$T|_{\mathcal{F}_{\theta|W}}(f_{\theta|W}) = T|_{\mathcal{F}_{\theta|W}}(\tilde{f}_{\theta|W}) \quad \Rightarrow \quad f_{\theta|W} = \tilde{f}_{\theta|W}, \tag{3.4}$$

for all $\tilde{f}_{\theta|W} \in \mathcal{F}_{\theta|W}$.

Therefore, point identification of $f_{\theta|W}$ is equivalent to requiring that the operator $T|_{\mathcal{F}_{\theta|W}}$ is injective. The injectivity of $T|_{\mathcal{F}_{\theta|W}}$ depends on the injectivity of $T$ but it is not equivalent. In fact, if $T$ is injective, *i.e.* $\mathcal{N}(T) = \{0\}$ where $\mathcal{N}(\cdot)$ denotes the null space of an operator, then $T|_{\mathcal{F}_{\theta|W}}$ is injective. However, when $T$ is non-injective the restricted operator $T|_{\mathcal{F}_{\theta|W}}$ may be injective. This is possible when the domain of $T|_{\mathcal{F}_{\theta|W}}$ is sufficiently restricted. We remark that the element 0 is not an element of $\mathcal{D}(T|_{\mathcal{F}_{\theta|W}})$ even if it is an element of $\mathcal{N}(T)$.

The following proposition characterizes the identified set for the operator $T$, denoted by $\Lambda$. We denote by $f^{\dagger}_{\theta|W}$ the solution of the linear inverse problem $f_{C|WZ\theta} = Tf_{\theta|W}$ which has minimal norm.

**Proposition 3.** *The identified set is the set of all the solutions of (3.1):*

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

*If $T$ is compact then*

$$\Lambda = \left\{h = f^{\dagger}_{\theta|W} : + \sum_{\{j;\lambda_j=0\}} \zeta_j\varphi_j,\ \textit{for}\ \{\zeta_j\}\ \textit{such that}\ \sum_{\{j;\lambda_j=0\}} \zeta_j \int_{\Theta} \varphi_j d\theta = 1-\kappa_w \wedge \sum_{\{j;\lambda_j=0\}} \zeta_j\varphi_j \geq -f^{\dagger}_{\theta|W}\right\}$$

*where $\kappa_w = \int_{\Theta} f^{\dagger}_{\theta|W}(\theta; w)d\theta$.*

This proposition characterizes the identified set in terms of quantities that depends on the SVD of $T$, which is known, and on $f_{C|WZ}$ which can be easily estimated.

The model is point-identified when $\Lambda$ is a singleton. 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 (3.1);

(ii) the operator $T$ is not injective, *i.e.* $\mathcal{N}(T) \neq \{0\}$, but $T|_{\mathcal{F}_{\theta|W}}$ is injective, *i.e.* (3.4) holds. In this case we have $\Lambda = (f^{+}_{\theta|W} + h_{\theta|W})$ where $h_{\theta|W} \in \mathcal{N}(T)$ is such that $\int_{\Theta}(f^{\dagger}_{\theta|W}(\theta; w) + h(\theta; W))d\theta = 1$ and $(f^{\dagger}_{\theta|W} + h_{\theta|W})$ is non-negative *a.e.* on $\Theta$, $P^W$-a.s. It can also be the case that $\Lambda = f^{\dagger}_{\theta|W}$.

The injectivity condition of $T$ characterizes the strength of statistical dependence among $C$ and $\theta$ conditionally on $(W, Z)$. Identification can be obtained even without injectivity of $T$. This shows that identification in our framework is a weaker concept than in the nonparametric instrumental variable (IV) literature. In fact, identification in nonparametric IV models is guaranteed only when the operator characterizing the estimating equation is injective, which corresponds to our case (i). On the contrary, we have identification also in case (ii) and this is due to the presence of the constraint in (3.1). This constraint makes the estimation problem more difficult because the estimation problem is nonlinear when the constraint is binding. On the other hand, it can help reduce the size of the identified set.

The injectivity of the operator $T$ is equivalent to the requirement that the conditional distribution $f_{\theta|CWZ}$ of $\theta$ given $(C,W,Z)$ is $L^2$-complete, provided that $f_{C|WZ}$ and $f_{\theta|W}$ are bounded away from zero and infinity. The following argument shows this fact. The operator $T$ is injective if and only if the only solution in $L^2_{\pi_\theta}$ of $0=\int_\Theta f_{C|WZ\theta}(c;w,z,\theta)h(\theta;w)d\theta$ is $h(\theta;w)=0$, $P^W$-a.s. We can rewrite $0=\int_\Theta f_{C|WZ\theta}(c;w,z,\theta)h(\theta;w)d\theta=\int_\Theta f_{\theta|CWZ}(\theta;c,w,z)\frac{f_{C|WZ}(c;w,z)}{f_{\theta|W}(\theta;w)}h(\theta;w)d\theta$ so that, by definition of completeness, this implies that $\frac{f_{C|WZ}(c;w,z)}{f_{\theta|W}(\theta;w)}h(\theta;w)=0$, $P^W$-a.s. Under Assumption 4 this in turn implies that $h(\theta;w)=0$, $P^W$-a.s. since $\inf_{c\in\mathcal{C}} f_{C|WZ}(c;w,z)>0$ and $\sup_{\theta\in\Theta} f_{\theta|W}(\theta;w)<\infty$. The equivalence between completeness and injectivity was already remarked in Florens *et al.* (1990), Florens (2003), Newey and Powell (2003) and Hu and Schennach (2008).

In our framework, $L^2$-completeness is too strong for identification, since the solution to the integral equation is constrained to be a *pdf*. In this case, the following proposition gives the weakest concept of completeness of $f_{\theta|CWZ}$ that is sufficient for identification.

**Proposition 4** ($\mathcal{T}$-completeness)**.** *Let $\mathcal{T}\subset L^2_{\pi_\theta}\times L^2_{\pi_{cz}}$ be the subset defined as $\mathcal{T}:=\Big\{h(\theta,c;w,z)\in L^2_{\pi_\theta}\times L^2_{\pi_{cz}}$ for a given $w\in\mathcal{W}$; $\int_{\mathcal{C}}\int_\Theta hd\theta dc=0,\ P^{WZ}-a.s.\Big\}$. Under Assumption 4 and if $f_{\theta|CWZ}$ is* $\mathcal{T}$-complete *then (3.4) holds.*

We refer to Mandelbaum (1987) for a definition of *completeness* of a probability distribution with respect to a family of functions $\mathcal{T}$.

In the following, we present some classes of distribution functions $\mathcal{F}_{C|WZ\theta}$ for which problem (3.1) is point-identified.

Additively-closed one-parameter family of distributions. Let $\Theta=\mathbb{R}_+$ and $\mathcal{F}_{C|WZ\theta}$ be *additively closed, i.e.* $\forall f_{C|WZ\theta}\in\mathcal{F}_{C|WZ\theta}$ and $\forall\theta_1,\theta_2\in\Theta$, $f_{C|WZ\theta}(c;w,z,\theta_1)*h_{C|WZ\theta}(c;w,z,\theta_2)=f_{C|WZ\theta}(c;w,z,\theta_1+\theta_2)$, where $*$ denotes the convolution operation. Then, $\mathcal{F}_{\theta|W}$ is identifiable. Additive closedness of $\mathcal{F}_{C|WZ\theta}$ depends on the functional form of $f_{\varepsilon|WZ\theta}$ and of the structural function $\varphi$ and can be easily checked. In particular, some distributions that belongs to the additively-closed one-parameter family, and that are relevant for our application, are the following, see Teicher (1961).

- Type III distributions: $f_{C|WZ\theta}=\frac{z^\theta}{\Gamma(\theta)}c^{\theta-1}e^{-zc}$, $c>0$, $z>0$, $\theta>0$ or $f_{C|WZ\theta}=\frac{\theta^z}{\Gamma(z)}c^{z-1}e^{-\theta c}$, $c>0$, $z>0$, $\theta>0$. The role of $W$ and $Z$ can be interchanged.
- Uniform distributions: $f_{C|WZ\theta}=\mathcal{U}[\theta-g(Z,W),\theta+g(Z,W)]$, where $g(\cdot,\cdot)$ is some function of $(Z,W)$. Therefore, $f_{C|WZ\theta}=\frac{1}{2g(Z,W)}1\{[\theta-g(Z,W)]<c<[\theta+g(Z,W)]\}$. However, for uniform distributions for which the support does not depend on $\theta$ we have no identification of $\mathcal{F}_{\theta|W}$

Location-scale one-parameter family of distributions. Let $\Theta = \mathbb{R}_+$ and $\mathcal{F}_{C|WZ\theta}$ be the one-parameter family induced by $f_{C|WZ}$ via location or scale changes, *i.e.* $\forall f_{C|WZ\theta} \in \mathcal{F}_{C|WZ\theta}$, $f_{C|WZ\theta}(c; w, z, \theta) = f_{C|WZ}(c - \theta; w, z)$ or $f_{C|WZ\theta}(c; w, z, \theta) = f_{C|WZ}(c\theta; w, z)$. For the location (resp. scale) family: if the conditional characteristic function of $C$ (resp. $\log C$), given $(W, Z)$, does not vanish $P^{WZ}$-a.s. in some non-degenerate real interval, then the class $\mathcal{F}_{\theta|W}$ is identifiable.

Exponential family of distributions.

**Lemma 3.1.** *Let us assume that* $\forall i = 1, \dots, s$, $\partial_c \varphi_i^{-1}(w, z, \cdot, c) 1_{S_{C|WZ\theta}}(c)$ *is bounded away from zero and infinity for every* $(c, w, z) \in \mathcal{C} \times \mathcal{W} \times \mathcal{Z}$, *and* $f_{\varepsilon|\theta WZ} \circ \varphi_i^{-1}(c, w, z, \theta)$ *is of the form*

$$\exp\{\tau_i(c, w, z)' m_i(\theta)\} h_i(\theta) k_i(c, w, z), \qquad i = 1, \dots, s$$

*where for every* $i = 1, \dots, s$, $h_i(\cdot)$ *is a positive function depending only on* $\theta$, $m_i(\cdot)$ *is a vector-valued function whose image has dimension equal to the dimension of* $\theta$ *and each component is an increasing function depending only on* $\theta$. *The functions* $\tau_i$ *and* $k_i$ *do not depend on* $\theta$. *Then, if the support of* $(C, Z, W)$ *has a nonempty interior,* $\mathcal{F}_{\theta|W}$ *is identifiable.*

We analyze now identification in the two previous examples that we have considered.

**Example 1** (Continued)**.** *Suppose that* $\varepsilon \sim Exp(\lambda)$ *with* $\lambda > 0$ *a known parameter. Therefore, the functional equation that identifies* $f_{\theta|W}$ *is*

$$f_{C|WZ}(c, w, z) = \int_\Theta \lambda \exp\{-\lambda(c - z_1'\theta_1 - \phi(z_2, w)\theta_2)\} 1\{c \geq z_1'\theta_1 + \phi(z_2, w)\theta_2\} f_{\theta|W} d\theta$$

*where* $z := (z_1', z_2)'$. *The function* $f_{\varepsilon|\theta WZ}(\varphi_i^{-1}(w, z, \theta, c); \theta, w, z)$ *which characterizes the kernel of the operator can be rewritten as:*

$$\begin{aligned} f_{\varepsilon|\theta WZ}(\varphi_i^{-1}(w, z, \theta, c); \theta, w, z) &= \lambda \exp\{-\lambda(c - z_1'\theta_1 - \phi(z_2, w)\theta_2)\} \\ &= \lambda \exp\{-\lambda c\} \exp\{\lambda [z_1', \phi(z_2, w)]\theta]\} \end{aligned}$$

*and satisfies the assumptions of Lemma 3.1 with* $h(\theta) = \lambda$, $m(\theta) = \theta = (\theta_1', \theta_2')'$ *is the identity function,* $\tau(c, w, z) = (z_1', \phi(z_2, w))'$ *and* $k(c, w, z) = 1$. *Then, if the support of* $(C, W, Z)$ *is nonempty,* $f_{\theta|W}$ *is point-identified.*

**Example 2** (Continued)**.** *For simplicity, assume* $\theta_2$ *is not random, eliminate the index* $t$ *and assume classical measurement error. We make use only of cross-section data for the estimation. In this example,*

$$f_{C|WZ}\left(c, w, z\right) = \int_\Theta \frac{\exp\left(-\frac{1}{2}\left(\frac{c - \phi_1 w_1 - \phi_2(w_2 + z) - m(\gamma, \beta, \theta_2)}{\sigma_\varepsilon^2}\right)^2\right)}{\sqrt{2\pi\sigma_\varepsilon^2}} f_{\gamma\beta|W}\left(\gamma, \beta; w\right) d\gamma d\beta. \qquad (3.5)$$

*Denote by $M$ the support of $m(\gamma,\beta,\theta_2)$ and by $\Gamma$ the support of $\gamma$. Define $m = m(\gamma,\beta,\theta_2)$. After a change of variable, this integral equation is can be rewritten*

$$\begin{aligned} f_{C|WZ}(c,w,z) &= \int_M\int_\Gamma \frac{\exp\left(-\frac{1}{2}\left(\frac{c-\phi_1 w_1-\phi_2(w_2+z)-m}{\sigma_\varepsilon^2}\right)^2\right)}{\sqrt{2\pi\sigma_\varepsilon^2}}\widetilde{f}_{\gamma M|W}(\gamma,m;w)\,d\gamma dm \qquad (3.6)\\ &= \int_M \frac{\exp\left(-\frac{1}{2}\left(\frac{c-\phi_1 w_1-\phi_2(w_2+z)-m}{\sigma_\varepsilon^2}\right)^2\right)}{\sqrt{2\pi\sigma_\varepsilon^2}}\widetilde{f}_M(m;w)\left(\int_\Gamma \widetilde{f}_{\gamma|WM}(\gamma,m;w)\,d\gamma\right)dm\\ &= \int_M \frac{\exp\left(-\frac{1}{2}\left(\frac{c-\phi_1 w_1-\phi_2(w_1+z)-m}{\sigma_\varepsilon^2}\right)^2\right)}{\sqrt{2\pi\sigma_\varepsilon^2}}\widetilde{f}_M(m;w)\,dm \end{aligned}$$

*where*

$$\widetilde{f}(\gamma,m;w) = f\left(\gamma,m^{-1}(\gamma,m,\theta_2)\right)\left|\frac{\partial m^{-1}(\gamma,m,\theta_2)}{\partial m}\right|.$$

*The joint density of $(\gamma,m)$ is not point-identified because any proper conditional density $\widetilde{f}_{\gamma|WM}(\gamma;m,w)$ is consistent with the data. The marginal density $\widetilde{f}_{M|W}(m;w)$ is point-identified. The identified set of densities $f_{\gamma\beta|W}$ is the set*

$$f_{\gamma\beta|W}(\gamma,\beta;w) = \widetilde{f}_{M|W}(m(\gamma,\beta,\theta_2);w)\cdot\widetilde{f}_{\gamma|WM}(\gamma,m(\gamma,\beta);w)\left|\frac{\partial m}{\partial\gamma}\right|$$

*for some conditional density $\widetilde{f}_{\gamma|WM}$.*

## 3.3 Local Identification

We discuss in this section local identification of $f_{\theta|W}$ by making reference to the results in [Chen *et al.* (2011)]. Following [Chen *et al.* (2011)] (see definition page 4), $f_{\theta|W}$ is locally identified for $\mathcal{F}\subseteq\mathcal{F}_{\theta|W}$ with $f_{\theta|W}\in\mathcal{F}$, if for all $f\in\mathcal{F}$, $||T_{\mathcal{F}_{\theta|W}}(f)||=0$ implies $||f-f_{\theta|W}||=0$. First of all, we stress that point-identification of $f_{\theta|W}$ in the sense of point *(ii)* of the previous section, *i.e.* when $T$ is not injective but $T|_{\mathcal{F}_{\theta|W}}$ does, is equivalent to local identification for $\mathcal{F}_{\theta|W}$.

However, it may be that $f_{\theta|W}$ is not-identified for $\mathcal{F}_{\theta|W}$ but there exists a set $\mathcal{F}\subset\mathcal{F}_{\theta|W}$ containing $f_{\theta|W}$ for which $f_{\theta|W}$ is identified. To see this, it is sufficient to apply Theorem 2 in [Chen *et al.* (2011)]. Let $\Delta$ be the linear space generated by differences $(f_1-f_2)$ of members of $\mathcal{F}_{\theta|W}$. Remark that $T|_{\mathcal{F}_{\theta|W}}$ has a Fréchet derivative at the point $f_{\theta|W}\in\mathcal{F}_{\theta|W}$ with respect to the norm $||\cdot||$ (in $L^2_{\pi_\theta}$) which is given by the functional $T(f_{\theta|W};d)$ defined on $d\in\Delta$ and linear in $d$ such that

$$T|_{\mathcal{F}_{\theta|W}}(f)-T|_{\mathcal{F}_{\theta|W}}(f_{\theta|W})-T(f_{\theta|W};f-f_{\theta|W})=o(||f-f_{\theta|W}||)$$

as $||f - f_{\theta|W}|| \to 0$. Moreover, $T(f_{\theta|W}; d)$ is equal to the Gâteaux derivative of $T|_{\mathcal{F}_{\theta|W}}$ at $f_{\theta|W}$ in the direction of $f \in \mathcal{F}$, *i.e.* $T(f_{\theta|W}; f - f_{\theta|W}) = T|_{\mathcal{F}_{\theta|W}}(f - f_{\theta|W})$. Thus, Theorem 2 in [Chen *et al.* (2011)] applies if there exists a $\mathcal{F} \subseteq \mathcal{F}_{\theta|W}$ containing $f_{\theta|W}$ such that $T(f_{\theta|W}; (f - f_{\theta|W})) = 0$ implies $f - f_{\theta|W} = 0$, $\forall f \in \mathcal{F}_{\theta|W}$.

The meaning of local-identification in our setting is that when $T|_{\mathcal{F}_{\theta|W}}$ is not injective then we can obtain point-identification by restricting further the set in which we are looking for the solution.

### 3.4 Economic Discussion

In the Euler equation example, it is important to note that our analysis only requires information on $f_{C|WZ}$. As long as we condition on age (i.e. as long as age is an element of $W$), and as long parameters governing aggregate uncertainty have been pre-estimated on aggregate time series data, it does not matter that the cross-sectional distribution of $C$ does not contain any cross-sectional variation in aggregate variables. In fact, even in the presence of aggregate shocks, the distribution of $f_{\theta|W}$ may be point identified if there is enough variation in idiosyncratic shocks.

### 3.5 Identification Without Nuisance Unobservables

In this section we briefly describe the case where we do not have $\varepsilon$ so that $f_{C|WZ\theta}$ cannot be recovered as in Section 3. This is relevant in models where all the unobservable variables are of interest so, $\varepsilon$ is included in $\theta$. In our setup, this implies that the general structural model (1.2) reduces to

$$\Psi(C, W, Z, \theta) = 0. \tag{3.7}$$

In this case our approach is completely non-parametric and we have two methods for estimating $f_{\theta|W}$.

The first estimation method can be applied only when $\theta$ is univariate and $\varphi$ is strictly monotonic in $\theta$. In this case, a sample of "indirect observations" of $\theta$ can be recovered by inverting $\varphi$ with respect to $\theta$: $\theta_i = \varphi^{-1}(w_i, z_i, c_i)$, $i = 1, \ldots, n$ where $(w_i, z_i, c_i; : i = 1, \ldots, n)$ is an $n$-sample of observations of $(W, Z, C)$. The *pdf* $f_{\theta|W}$ can be estimated by using any nonparametric method.

On the contrary, when $\varphi$ is not strictly monotonic in $\theta$, or $\theta$ is multivariate, we can characterize the structural *pdf* $f_{\theta|W}$ as a solution of a slightly different constrained functional equation. Let $F_{C|WZ}$ be the cumulative distribution function of $C$ conditioned on $(W, Z)$ then,

$$F_{C|WZ}(c; w, z) = Sf_{\theta|W}(\theta; w) \quad \wedge \quad f_{\theta|W} \in \mathcal{F}_{\theta|W}, \qquad P^W - a.s. \tag{3.8}$$

where $S : L^2_\pi \to L^2_{\pi_{cz}}$ is a bounded linear operator defined as

$$Sh = \int_{\{\theta;:\varphi(w,z,\theta)\leq c\}} h(\theta;w)d\theta \equiv \int_\Theta 1\{\varphi(w,z,\theta) \leq c\} h(\theta;w)d\theta, \quad \forall h \in L^2_{\pi_\theta} and F_{C|WZ} \in L^2_{\pi_{cz}}.$$

The kernel of the new operator is $g(w, z\theta, c) = \frac{1\{\varphi(w,z,\theta)\leq c\}}{\pi_\theta(\theta;w)}$ and the adjoint $S^*$ is:

$$S^*h = \int_{\mathcal{C}\times\mathcal{Z}} 1\{\varphi(w,z,\theta) \leq c\} \frac{\pi_{cz}(c,z;w)}{\pi_\theta(\theta;w)} h(c;w,z)dcdz, \quad \forall h \in L_{\pi_{cz}}(c,z).$$

For this new problem Assumptions 1 and 5 are still required while Assumption 4 can be weakened to

**Assumption 4'.** *The random variable $(C, Z, \theta|W)$ has a joint continuous distribution characterized by its cumulative distribution function $F_{C\theta Z|W}$ that is absolutely continuous with respect to the Lebesgue measure with Radon-Nikodym density $f_{C\theta Z|W}$. Moreover, the support of $f_{\theta|W}$ does not depend on $W$.*

A sufficient condition for compactness of the operator $S$ is that $\pi_\theta$ and $\pi_{cz}$ are chosen such that $\int_\Theta \frac{1}{\pi_\theta} d\theta < \infty$ and $\int_{\mathcal{C}\times\mathcal{Z}} \pi_{cz} dcdz < \infty$. When there are nuisance unobservables, the estimating equation (3.1) can be trivially recovered from (3.8) by differentiating with respect to $c$.

The estimation procedure for this case is the same as that proposed in Section 4. Our estimator defined in (4.2) is valid with $T$ replaced by $S$. We conjecture that the rate of the MISE will improve since $F_{C|WZ}$ can be estimated at a better rate than $f_{C|WZ}$. Moreover, the degree of ill-posedness will not be as severe as in the case where the kernel of $T$ is exponential.

A complete analysis of this case is beyond the scope of this paper and we leave it for future research. This short discussion simply shows that our estimation approach is quite general and can be extended to cover the case where $\varepsilon$ is not part of the structural model.

# 4 Estimation

In this section we develop the estimation theory for $f_{\theta|W}$ based on solution of the nonlinear problem (3.1). We call our estimation method *Indirect Estimation* and our estimator an *Indirect Regularized Density Estimator.*

Equation (3.1) cannot be solved directly since $f_{C|WZ}$ is unknown. In addition, the (generalized) inverse of $T$ is unbounded. Therefore, we replace $f_{C|WZ}$ in (3.1) by a (nonparametric) estimate denoted by $\hat{f}_{C|WZ}$. The maintained assumption is that we use a cross section of observations to construct the consistent estimate $\hat{f}_{C|WZ}$.

**Assumption 6.** *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* $\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2$ *converges to* $0$ *as* $n \uparrow \infty$.

In subsection 4.2.1, we analyze the asymptotic properties of our estimator when $f_{C|WZ}$ is estimated using a kernel estimator. The general procedure can be followed and similar asymptotic results can be derived for any consistent nonparametric estimator of $f_{C|WZ}$.

## 4.1 Existence of an Estimated Solution

When one replaces $f_{C|WZ}$ in (3.1) with a consistent estimator $\hat{f}_{C|WZ}$, it is no longer guaranteed that $\hat{f}_{C|WZ} \in \mathcal{R}(T)$ nor that $\hat{f}_{C|WZ} \in \mathcal{R}(T|_{\mathcal{F}_{\theta|W}})$ even though, under mild conditions, $\hat{f}_{C|WZ} \in L^2_{\pi_{cz}}$. If $\hat{f}_{C|WZ} \notin \mathcal{R}(T|_{\mathcal{F}_{\theta|W}})$ then, a solution to (3.1) does not exist. Nevertheless, we may define a generalized approximate solution that solves (3.1) approximately and that also accounts for two additional issues: the possibility that $T$ is not injective and the fact that the solution must be constrained to belong to the convex and closed subset $\mathcal{F}_{\theta|W}$. The solution concept that we adopt is the *$\mathcal{C}$-best-approximate solution*, see Neubauer (1988), denoted by $\hat{f}^{\dagger c}_{\theta|W}$ and defined in the following.

**Definition 2.** *The* $\mathcal{C}$-best-approximate solution $\hat{f}^{\dagger c}_{\theta|W}$ *is an element of* $\mathcal{F}_{\theta|W} \subset L^2_{\pi_\theta}$ *such that*

$$||T\hat{f}^{\dagger c}_{\theta|W} - \hat{f}_{C|WZ}|| = \inf\Big\{||Th - \hat{f}_{C|WZ}||;\ h \in \mathcal{F}_{\theta|W}\Big\} \tag{4.1}$$

*and*

$$||\hat{f}^{\dagger c}_{\theta|W}|| = \min\Big\{||h||;\ ||Th - \hat{f}_{C|WZ}|| = ||T\hat{f}^{\dagger c}_{\theta|W} - \hat{f}_{C|WZ}|| \text{ and } \hat{f}^{\dagger c}_{\theta|W} \in \mathcal{F}_{\theta|W}\Big\}.$$

Therefore, the *$\mathcal{C}$-best-approximate solution* is the least-squares solution on $\mathcal{F}_{\theta|W}$ that has minimal norm among all minimizers. It can be shown that $\hat{f}^{\dagger c}_{\theta|W} = T|^{\dagger}_{\mathcal{F}_{\theta|W}}\hat{f}_{C|WZ}$ where $T|^{\dagger}_{\mathcal{F}_{\theta|W}}$ denotes the $\mathcal{F}_{\theta|W}$-constrained Moore-Penrose generalized inverse. We denote $f^{\dagger c}_{\theta|W} = T|^{\dagger}_{\mathcal{F}_{\theta|W}} f_{C|WZ}$ and our estimator is an estimator of $f^{\dagger c}_{\theta|W}$.

When $\mathcal{F}_{\theta|W} = L^2_{\pi_\theta}$, *i.e.* we have no prior information on the solution, the *$\mathcal{C}$-best-approximate solution* is simply the *best-approximate solution* $\hat{f}^{\dagger}_{\theta|W}$ defined as $\hat{f}^{\dagger}_{\theta|W} = T^{\dagger} f_{C|WZ}$, where $T^{\dagger}$ denotes the Moore-Penrose generalized inverse of $T$. The *$\mathcal{C}$-best approximate solution* $\hat{f}^{\dagger c}_{\theta|W}$ exists and is unique if and only if $\mathcal{Q}_c\hat{f}_{C|WZ} \in \mathcal{R}(T|_{\mathcal{F}_{\theta|W}})$, where $\mathcal{Q}_c$ is the metric projector onto $\overline{\mathcal{R}(T|_{\mathcal{F}_{\theta|W}})}$, see proposition 5.14 in Engl *et al.* (2000). As we are considering infinite dimensional spaces, these conditions are not trivial and may be not satisfied.

While the *$\mathcal{C}$-best approximate solution* exists and is unique, it does not depend continuously on $\hat{f}_{C|WZ}$ since, in general, $\mathcal{R}(T|_{\mathcal{F}_{\theta|W}})$ is non-closed (even if it is convex because $\mathcal{F}_{\theta|W}$ is closed and convex). As a result, the inverse problem (3.1) is ill-posed. From a practical point of view, when $f_{C|WZ}$ is replaced by a consistent estimator $\hat{f}_{C|WZ}$, the corresponding estimator $\hat{f}^{\dagger c}_{\theta|W}$ is an inconsistent estimator for $f^{\dagger c}_{\theta|W}$ despite the consistency of $\hat{f}_{C|WZ}$.

One might think that the constraint makes the problem stable. This would be the case if $\mathcal{F}_{\theta|W}$ were a compact set. Unfortunately this is not the case here. Thus, a regularization procedure must be used to compute a consistent estimator of $f^{\dagger c}_{\theta|W}$. Because $f^{\dagger c}_{\theta|W} \in \mathcal{F}_{\theta|W}$, it is natural to require that the regularized estimator is in $\mathcal{F}_{\theta|W}$ too.

In order to obtain a regularized estimator of $f^{\dagger c}_{\theta|W}$ belonging to $\mathcal{F}_{\theta|W}$ we use a Tikhonov-type estimator defined as the minimizer, with respect to $h$, of

$$||Th - \hat{f}_{C|WZ}||^2 + \alpha||h||_s^2, \qquad h \in \mathcal{F}_{\theta|W} \tag{4.2}$$

where $||\cdot||_s$ denotes a norm to be specified indexed by the parameter $s$ and $\alpha > 0$ is a parameter that decreases to 0 at a suitable rate. If $s = 0$, we have the classical norm in $L^2_{\pi_\theta}$ and the estimator is the classical constrained Tikhonov regularized solution.

We call our estimator an *Indirect Regularized Density Estimator*. Problem (4.2) is nonlinear and, in general, it does not allow a solution in closed-form except in one case. Thus, in the estimation procedure we treat these two cases separately: *(i)* the case in which $f^{\dagger c}_{\theta|W} = f^{\dagger}_{\theta|W}$, for which a closed-form solution exists and *(ii)* the case in which $f^{\dagger c}_{\theta|W} \neq f^{\dagger}_{\theta|W}$. We point out that the estimators proposed below are estimators of $f^{\dagger c}_{\theta|W}$. This means that in the non-identified case our procedure gives estimators for only one element of the identified set. After that, the identified set can be estimating by using the characterization given in section 3.2.

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

In this section we consider the case $f^{\dagger c}_{\theta|W} = f^{\dagger}_{\theta|W}$, that is, the best-approximate solution belongs to $\mathcal{F}_{\theta|W}$. This is possible either when $T$ is injective or when $T$ is not injective but $T|_{\mathcal{F}_{\theta|W}}$ is and $f^{\dagger}_{\theta|W} \in \mathcal{F}_{\theta|W}$. In this particular case we can use a two- step approach to compute a regularized solution in $\mathcal{F}_{\theta|W}$ that in many cases can be faster than directly solving the nonlinear problem in (4.2).

*First step:* compute the regularized solution of the unconstrained problem:

$$\min_{h \in L^2_{\pi_\theta}} \left\{ ||Th - \hat{f}_{C|WZ}||^2 + \alpha||h||_s^2 \right\}. \tag{4.3}$$

For $s = 0$, we denote the estimator by $\hat{f}^{\alpha}_{\theta|W}$ while for $s > 0$, we use $\hat{f}^{s}_{\theta|W}$. In subsections 4.2.1 and 4.2.2 we develop these two well-known estimators, see *e.g.* Engl, Hanke and Neubauer (2000) or Florens, Johannes and Van Bellegem (2010).

*Second step:* we compute the metric projection of $\hat{f}^{\alpha}_{\theta|W}$ (resp. of $\hat{f}^{s}_{\theta|W}$) onto the set $\mathcal{F}_{\theta|W}$. We denote with $\mathcal{P}_c$ this metric projector. Thus, the *indirect Tikhonov regularized estimator* of $f_{\theta|W}$ is

$$\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W} := \max\left\{0, \hat{f}^{\alpha}_{\theta|W} - \frac{c}{\pi_\theta}\right\}. \tag{4.4}$$

where $c$ is such that $\int_{\Theta}\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W}d\theta = 1$. In practice, the constant $c$ cannot be explicitly computed. For that, one can use the following algorithm proposed by Gajek (1986):

$\mathcal{P}_c$**−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 $|C_{k+1} - 1| < \varepsilon$, for $\varepsilon > 0$.

The same algorithm works for $\hat{f}^{s}_{\theta|W}$. Gajek (1986) shows that $\hat{f}^{\alpha(k+1)}_{\theta|W}$ is the orthogonal projection of $\hat{f}^{\alpha(k)}_{\theta|W}$ onto $\mathcal{F}^{+}_{\theta|W}$ and $\hat{f}^{\alpha(k+2)}_{\theta|W}$ is the orthogonal projection of $\hat{f}^{\alpha(k+1)}_{\theta|W}$ onto $\mathcal{F}^{1}_{\theta|W}$, where $\mathcal{F}^{+}_{\theta|W}$ and $\mathcal{F}^{1}_{\theta|W}$ are the subsets of all functions in $L^2_{\pi_\theta}$ which are positive *a.e.* on $\Theta$ and which integrates to 1, $P^W$-a.s., respectively. Step 3 is well defined for each iteration if $\int_{\Theta}\frac{1}{\pi_\theta}d\theta < \infty$ and $C_{k+1} < \infty$. In particular, if $\int_{\Theta}\frac{1}{\pi_\theta}d\theta < \infty$ holds, then there exists a unique real number $c$ such that the $\mathcal{P}_c$-algorithm converges pointwise and in norm to $\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W}$ defined in (4.4).

Condition $\int_{\Theta}\frac{1}{\pi_\theta}d\theta < \infty$ is trivially satisfied if $\Theta$ is compact and $\pi_\theta$ is continuous. For the case where $\Theta$ is not compact, the condition $\int_{\Theta}\frac{1}{\pi_\theta}d\theta < \infty$ means that $\pi_\theta$ must assign high weights to arguments in the tail of the distribution of $\theta$.

### 4.2.1 Tikhonov regularized solution

The minimizer of (4.3) for $s = 0$ is the classical Tikhonov regularized solution:

$$\begin{aligned}\hat{f}^{\alpha}_{\theta|W}(\theta; w) &= (\alpha I + T^*T)^{-1}T^*\hat{f}_{C|WZ} &(4.5)\\ &= \sum_{j=1}^{\infty}\frac{\lambda_j}{\alpha+\lambda_j^2} < \hat{f}_{C|WZ}, \psi_j > &(4.6)\end{aligned}$$

where the second expression is valid when $T$ is compact. This is our first estimator for $f^{\dagger}_{\theta|W}$. The parameter $\alpha$ is a *regularization parameter* that converges to zero as the estimation error $(\hat{f}_{C|WZ} - f_{C|WZ})$ converges to zero. The Tikhonov regularization method is very well-known and developed in econometric theory so that we do not detail it here. The interested reader can refer to Kress (1999, Chapter 15) or Carrasco, Florens and Renault (2007, Section 3).

In the following, we derive the rate of the Mean Integrated Square Error (MISE) associated with the projected estimator $\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W}$. We remark that it is a weighted MISE since the norm is the norm in $L^2_{\pi_\theta}$. We introduce the following regularity assumption on $f^{\dagger}_{\theta|W}$.

**Assumption 7.** *For some $\beta > 0$ and $0 < M < \infty$ the structural density $f^{\dagger}_{\theta|W}$ is an element of the $\beta$-regularity space $\Phi_\beta(M)$ defined as*

$$\Phi_\beta(M) = \Big\{ f \in \mathcal{N}(T)^{\perp}; \quad \sum_j \frac{< f, \varphi_j >^2}{\lambda_j^{2\beta}} < M \Big\}.$$

When $M = \infty$ then, $\Phi_\beta = \mathcal{R}[(T^*T)^{\frac{\beta}{2}}]$. Assumption 7 is quite standard in inverse problem theory and is known as a *Source Condition.* It expresses the smoothness of the function $f^{\dagger}_{\theta|W}$ according with smoothing properties of the operator $T$. As $T$ is a smoothing operator, the requirement that $f^{\dagger}_{\theta|W} \in \Phi_\beta(M)$ can be considered as an abstract smoothness condition, see Engl *et al.* (2000 Section 3.2) and Carrasco, Florens and Renault (2007). We state in the following theorem the rate of the MISE.

**Theorem 3.** *Let the Assumptions 4, 3, 6, 7 be satisfied. Then, the MISE associated with $\mathcal{P}_c \hat{f}^{\alpha}_{\theta|W}$ is*

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

*Moreover, if $\alpha \asymp (\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2)^{-\frac{1}{\beta \wedge 2 + 1}}$ then,*

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

It is clear from the MISE-rate that a $\beta$ greater than 2 is useless. This is known as *saturation effect* and it represents a drawback of the classical Tikhonov regularization scheme. A high regularity of the object of interest $f^{\dagger}_{\theta|W}$ (quantified by a large value of $\beta$) cannot be exploited in order to improve the rate of the MISE. This is the price to pay for having a regularization method that is easy to implement and intuitive. When an analyst is willing to assume that the source $f_{\theta|W}$ has a higher degree of smoothness, other regularization methods can be used to exploit this smoothness. We present one of these methods that exploit higher smoothness in section 4.2.2.

In the rest of this subsection we provide the rates of Theorem 3 in the case in which $\hat{f}_{C|WZ}$ is a kernel estimate. Without loss of generality we assume that $\mathcal{C} = [0,1]$, $\mathcal{W} = [0,1]^k$, $\mathcal{Z} = [0,1]^l$. Let $K(\cdot,\cdot)$ denote a generalized kernel function of order $r$ which will be used in order to avoid boundary effects (we refer to Hall and Horowitz (2005), Darolles *et al.* (2010) and references therein for an explicit definition of $K(\cdot,\cdot)$). By abuse of notation, we use the same notation for the kernels for each of $c, w$ and $z$ though we allow that they may be different. The order of the generalize kernel is $r = \rho_c = \rho_z = \rho_w$, where $\rho_c$, $\rho_z$, and $\rho_w$ are the number of derivatives in $c$ (respectively in $z$ and in $w$) of $f_{C|WZ}$. By abuse of notation, we use $h$ to denote the bandwidth for $c$, $z$ and $w$. Define $K_h(\cdot,\cdot) = K(\frac{\cdot,\cdot}{h})$. Then, the kernel estimator of $f_{C|WZ}$ is

$$\hat{f}_{C|WZ}(c; w, z) = \frac{\frac{1}{nh^{1+k+l}} \sum_{i=1}^n K_h(c_i - c, c) K_h(w_i - w, w) K_h(z_i - z, z)}{\frac{1}{nh^{k+l}} \sum_{l=1}^n K_h(w_l - w, w) K_h(z_l - z, z)}. \tag{4.7}$$

**Assumption 8.** *Let $g_j(c_1, z_1; w) = \psi_j(c_1, z_1; w)\pi_{cz}(c_1, z_1; w)$. We assume that*

*(i)* $\sum_{j=1}^{\infty} \int \frac{1}{f_{WZ}(w,z_1)} Var(g_j(c_1, z_1; w)|w, z_1)dz_1 < \infty$

*(ii)* $\sum_{j=1}^{\infty} \left( \int \int \frac{(g_j(c_1,z_1;w) - \mathbb{E}(g_j(c_1,z_1;w)|w,z_1))}{f_{WZ}(w,z_1)} tr\left[\frac{\partial^2 f_{CWZ}(c_1,w,z_1)}{\partial w \partial w'} \int uu'K(u)du\right] dc_1 dz_1 \right)^2 < \infty$

The minimax rate of the Tikhonov estimator associated with a kernel estimator of $f_{C|WZ}$ is given in the following Corollary to Theorem 3.

**Corollary 1.** *Let $\rho := \rho_w$. Under Assumptions 4, 3, 6-8, if $h \asymp n^{-\frac{1}{2\rho+k}}$ and $\alpha \asymp n^{-\frac{2\rho}{(2\rho+k)(\beta\wedge 2+1)}}$ then,*

$$\inf_{\alpha,h} \mathbb{E}||\mathcal{P}_c \hat{f}^{\alpha}_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 \asymp n^{-\frac{2\rho(\beta\wedge 2)}{(2\rho+k)(\beta\wedge 2+1)}}.$$

Our rate is increasing in $\beta$ and $\rho$ and decreasing in the dimension $k$ of the endogenous variables $W$. We have a curse of dimensionality only in the dimension of the endogenous variables $W$. The dimension of the instruments $Z$ does not play any role in determining the rate of convergence of $\hat{f}^{\alpha}_{\theta|W}$ because of the action of $T^*$ that integrates out $(C, Z)$.

**Remark 6.** Assumption 8 is stronger than requiring $Var(g_j(c_1, z_1, w)|w, z_1) < \kappa_1$ for all $j > 0$, for some constant $\kappa_1$, as demanded *e.g.* in Carrasco and Florens (2010). This allows to us to get a better rate of convergence for the variance term of the MISE of $\hat{f}^{\alpha}_{\theta|W}$. If Assumption 8 *(i)* does not hold then, we can weaken it by assuming that $\int \frac{1}{f_{WZ}(w,z_1)} Var(g_j(c_1, z_1, w)|w, z_1)dz_1 < \kappa_1$, for all $j > 0$ and $\int \int \frac{(g_j(c_1,z_1,w) - \mathbb{E}(g_j(c_1,z_1,w)|w,z_1))}{f_{WZ}(w,z_1)} tr\left[\frac{\partial^2 f_{CWZ}(c_1,w,z_1)}{\partial w \partial w'} \int uu'K(u)du\right] dc_1 dz_1 < \kappa_2$, for all $j > 0$, for some constants $\kappa_1$ and $\kappa_2$. In this case the rate of the MISE is slower and equal to $n^{-\frac{2\rho(\beta\wedge 2)}{(2\rho+k)(\beta\wedge 2+2)}}$.

We now study pointwise asymptotic normality of the Tikhonov regularized estimator $\hat{f}^{\alpha}_{\theta|W}$ in the case where $\hat{f}_{C|WZ}$ is computed by using kernel smoothing as in (4.7). For that we introduce the following technical assumption:

**Assumption 9.** *Let $g_j(c_1, z_1; w) = \psi_j(c_1, z_1; w)\pi_{cz}(c_1, z_1; w)$ and $\tilde{\varphi}_j = \mathcal{P}_c \varphi_j$. We assume that*

$$\frac{1}{h^k} \sum_{j=1}^{\infty} \left(\frac{\lambda_j}{\alpha + \lambda_j^2}\right)^3 \int \int \mathbb{E}\left( \left| g_j(c_1, z_1; w) - \mathbb{E}[g_j(c_1, z_1; w)|w, z_1] \right|^3 |w_1, z_1 \right) |K_h(w_1 - w)|^3$$
$$\times \quad \frac{f_{WZ}(w_1, z_1)}{f^3_{WZ}(w, z_1)} dw_1 dz_1 \, \tilde{\varphi}_j^3 = \mathcal{O}(1).$$

**Lemma 4.1.** *Let Assumptions 4, 3, 6-9 hold, $\rho := \rho_w$ and $\hat{f}^{\alpha}_{\theta|W}$ be the Tikhonov regularized estimator computed by using $\hat{f}_{C|WZ}(c; w, z)$ defined in (4.7). If $\alpha \asymp n^{-\frac{2\epsilon\rho}{(2\rho+k)(\beta\wedge 2+1)}}$ for $\epsilon > 1$ and $h \asymp n^{-\frac{1}{2\rho+k}+\varepsilon_h}$, $\varepsilon_h < -\max(\frac{2(\rho-\rho_z|wedge\rho_c)}{(2\rho_w+k)(2(\rho_c\wedge\rho_z)+k)}, 0)$, then,*

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

*where*

$$
\begin{aligned}
\Omega(\theta, w) \;=\; & \sum_{j=1}^{\infty} \frac{\lambda_j^2}{(\alpha + \lambda_j^2)^2} Var\Big[\frac{(g_j(c_1, z_1; w) - \mathbb{E}[g_j(c_1, z_1; w)|w, z_1])}{f_{WZ}(w, z_1)} K_h(w_1 - w, w)\Big] : \tilde{\varphi}_j^2 \\
& + 2\sum_{j<l}^{\infty} \frac{\lambda_j \lambda_l}{(\alpha + \lambda_j^2)(\alpha + \lambda_l^2)} Cov\Big(\frac{(g_j(c_1, z_1; w) - \mathbb{E}[g_j(c_1, z_1; w)|w, z_1])}{f_{WZ}(w, z_1)} K_h(w_1 - w, w), \\
& \frac{(g_l(c_1, z_1; w) - \mathbb{E}[g_l(c_1, z_1; w)|w, z_1])}{f_{WZ}(w, z_1)} K_h(w_1 - w, w)\Big) : \tilde{\varphi}_j \tilde{\varphi}_l.
\end{aligned}
$$

In order to obtain this asymptotic normality result, we require an $\alpha$ and a bandwidth $h$ that converge at a rate that is faster than the asymptotically optimal one. This guarantees that the bias of $\hat{f}^{\alpha}_{\theta|W}(\theta; w)$ is asymptotically negligible.

### 4.2.2 Tikhonov regularization in Hilbert scales

An alternative estimator to the classical Tikhonov $\hat{f}^{\alpha}_{\theta|W}$ is obtained by minimizing (4.3) with $s > 0$. This estimator, denoted by $\hat{f}^{s}_{\theta|W}$, is particularly appealing when $f_{\theta|W}$ is highly smooth and $\beta > 2$ because it obtains a faster rate of convergence.

The estimator $\hat{f}^{s}_{\theta|W}$ is the minimizer of (4.3) for a norm $||\cdot||_s$ defined as

$$||x||_s := ||L^s x||, \quad \forall x \in \bigcap_{k \in \mathbb{R}_+} \mathcal{D}(L^k), : s \in \mathbb{R}.$$

where $L : \mathcal{D}(L) \subseteq L^2_{\pi_\theta}(\theta) \to L^2_{\pi_\theta}(\theta)$ is an unbounded, densely defined, self-adjoint, strictly positive definite operator. The inner product that induces this norm is defined as

$$< x, y >_s := < L^s x, L^s y >, \quad \forall x, y \in \bigcap_{k \in \mathbb{R}_+} \mathcal{D}(L^k), : s \in \mathbb{R}.$$

The completion of $\bigcap_{k \in \mathbb{R}_+} \mathcal{D}(L^k)$ with respect to the norm $||\cdot||_s$ is an Hilbert space usually denoted by $\mathcal{X}_s$. $(\mathcal{X}_s)_{s \in \mathbb{R}}$ is called the *Hilbert scale* induced by $L$. We refer to Engl *et al.* (2000,Chapter 8) for an exhaustive explanation of Hilbert scale. Moreover, if $s \geq 0$ we have $\mathcal{X}_s = \mathcal{D}(L^s)$. When $T$ is one-to-one we can choose $L = (T^*T)^{-1}$; the corresponding $\mathcal{X}_s$ is known as the *Canonical Hilbert Scale.*

The estimator $\hat{f}^{s}_{\theta|W}$ is defined for $s \geq 0$ and takes the form

$$\hat{f}^{s}_{\theta|W} = L^{-s}(\alpha I + L^{-s} T^* T L^{-s})^{-1} L^{-s} T^* \hat{f}_{C|WZ}. \tag{4.8}$$

For $s = 0$, $\hat{f}^{s}_{\theta|W} = \hat{f}^{\alpha}_{\theta|W}$. The constant $s$ controls the smoothness to be introduced into the regularization process since the estimator $\hat{f}^{\alpha}_{\theta|W}$ is constrained to be at least an element of $\mathcal{D}(L^s)$. When $L = (T^*T)^{-1}$, the estimator becomes $\hat{f}^{s}_{\theta|W} = (\alpha I + (T^*T)^{2s+1})^{-1}(T^*T)^{2s} T^* \hat{f}_{C|WZ}$.

For stable reconstruction of the solution $f^{\dagger}_{\theta|W}$ from an estimate $\hat{f}_{C|WZ}$ it is necessary to add additional information: *(i)* information concerning the smoothness of $f^{\dagger}_{\theta|W}$ and *(ii)* information concerning the smoothing property of the operator $T$. It is usual in *Inverse Problem theory* to formulate this additional information in terms of the operator $L$ previously introduced. We do this in the two following assumptions; Assumption 10 replaces Assumption 7.

**Assumption 10.** *For some $\tilde{\beta} > 0$ and $0 < M < \infty$ the structural density $f^{\dagger}_{\theta|W}$ is an element of the set*

$$\mathcal{X}_{\tilde{\beta}}(M) := \{f \in L^2_{\pi_\theta}(\theta); ||f||_{\tilde{\beta}} < M\}. \tag{4.9}$$

When $L = (T^*T)^{-1}$ the space $\mathcal{X}_{\tilde{\beta}}(M)$ becomes: $\mathcal{X}_{\tilde{\beta}}(M) := \{f \in \mathcal{N}(T)^{\perp};: ||f||_{\tilde{\beta}} < M\}$ which is equal to $\Phi_\beta(M)$ if $\tilde{\beta} = \frac{\beta}{2}$.

The operator $L$ defining the Hilbert Scale is usually some differential operator such that $L^{-1}$ is finitely smoothing. Hence, we can think about Assumption 10 as describing the regularity of the function $f^{\dagger}_{\theta|W}$ in terms of the number of its derivatives. That is, $f^{\dagger}_{\theta|W} \in \mathcal{X}_\beta$ is equivalent to stating that there exists $v \in L^2_{\pi_\theta}(\theta)$ such that $f^{\dagger}_{\theta|W} = L^{-\beta}v$ so that $f^{\dagger}_{\theta|W}$ has at least $\beta$ square integrable derivatives. In practice the parameter $\beta$ is unknown.

**Assumption 12.** *There exists a function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ continuous and strictly increasing with $\phi(0+) = 0$ such that:*

*(i) there exist $0 < \underline{m} \leq \overline{m} < \infty$ for which*

$$\underline{m}||[\phi(L^{-2})]^{\frac{1}{2}}f|| \leq ||Tf|| \leq \overline{m}||[\phi(L^{-2})]^{\frac{1}{2}}f|| \quad \textit{for all } f \in L^2_{\pi_\theta}(\theta),$$

*(ii) for $z = \max(s, \beta)$, the function $\psi_z : (0, c] \to \mathbb{R}^+$ defined by $\psi_z(t) := t\phi(t^{1/z})$ with $c = ||L^{-2z}||$ is convex.*

Assumption 12 characterizes the smoothing properties of the operator $T$ in terms of the smoothness of $L^{-1}$. It can be applied to cases with either finitely smooth or infinitely smooth operators. This assumption was introduced in the theory of Inverse Problems by Mair (1994) and Nair *et al.* (2005). The discussion and results on the rate of convergence that follow are along the line of these papers. In particular, we can distinguish two cases. *(I) Finitely smoothing case* (or *mildly ill-posed case*) with respect to $L$: Assumption 12 is satisfied with $\phi(t) = t^a$, where $a > 0$ can be interpreted as a degree of ill-posedness: the greater $a$ is the more ill-posed the problem is. Then, $\psi_z(t) := t^{(a+z)/z}$. *(II) Infinitely smoothing case* (or *severely ill-posed case*) with respect to $L$: Assumption 12 is satisfied with $\phi(t) = \exp(-t^{-a/2})$, where $a > 0$. Then, $\psi_z(t) := t\exp(-t^{-a/(2z)})$ which is convex on the interval $(0, ||L^{-2z}||]$ provided $||L|| \geq 1$.

When $L = (T^*T)^{-1}$, Assumption 12 is satisfied with $\phi(t) = \sqrt{t}$.

**Theorem 4.** *Under Assumptions 10 and 12, if $\alpha$ is chosen a priori by*

$$\alpha \propto \mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2 \Big[\psi_\beta^{-1}\Big(\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2\Big)\Big]^{\frac{s-\beta}{\beta}}$$

*then, we have*

$$\mathbb{E}||\mathcal{P}_c\hat{f}^s_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 \leq 4M^2\psi_\beta^{-1}\Big(\frac{\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2}{\underline{m}^2M^2}\Big).$$

It follows from Theorem 4 that:

- in the *finitely smoothing case* the optimal a priori choice for $\alpha$ is $\alpha \propto (\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2)^{\frac{a+s}{a+\beta}}$ and the corresponding rate is $\mathbb{E}||\mathcal{P}_c\hat{f}^s_{\theta|W} - f^{\dagger}_{\theta|W}||^2 = \mathcal{O}\Big((\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2)^{\frac{\beta}{a+\beta}}\Big)$;
- in the *infinitely smoothing case* the optimal a priori choice for $\alpha$ is $\alpha \propto \mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2\Big[-2\log(\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2)\Big]^{2\frac{s-\beta}{a}}$ and the corresponding rate is $\mathbb{E}||\mathcal{P}_c\hat{f}^s_{\theta|W} - f^{\dagger}_{\theta|W}||^2 = \mathcal{O}\Big(\Big[\log(\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2)^{-1}\Big]^{-\frac{2\beta}{a}}\Big)$.

Similar rates were previously given in the nonparametric instrumental regression literature by Florens *et al.* (2010), Chen and Reiss (2010) and Johannes *et al.* (2010).

Remark 4.1. In practice the regularity $\beta$ of the function $f_{\theta|W}$ is unknown, so the a-priori choice of $\alpha$ given in the previous theorem is unfeasible. On the contrary, a data-driven method for selecting $\alpha$ must be considered.

Remark 4.2. In the particular case $L := (T^*T)^{-1}$, the rate of the MISE is $(\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2)^{\frac{2\tilde{\beta}}{1+2\tilde{\beta}}}$. In order to compare this rate with the rate obtained with the classical Tikhonov regularization we have to set $\tilde{\beta} = \frac{\beta}{2}$, so that the two regularity assumptions 7 and 10 on $f^{\dagger}_{\theta|W}$ are the same. In this case, it is clear that the rate of $f^{s}_{\theta|W}$ is faster than the rate of $f^{\alpha}_{\theta|W}$ when $\beta$ is larger than 2, that is, when $f^{\dagger}_{\theta|W}$ is highly smooth.

## 4.3 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. On the contrary, we have to compute the constrained Tikhonov regularized solution by directly solving the minimization problem

$$\min_{h \in \mathcal{F}_{\theta|W}} ||Th - \hat{f}_{C|WZ}||^2 + \alpha||h||^2. \tag{4.10}$$

A closed-form solution of this problem does not exist and numerical methods must be used to compute a solution. We denote with $\check{f}^{\alpha,c}_{\theta|W}$ the estimator obtained from solution of (4.10).

In the case $f^{\dagger c}_{\theta|W} = f^{\dagger}_{\theta|W}$, *i.e.* $f^{\dagger c}_{\theta|W}$ is in the interior of $\mathcal{F}_{\theta|W}$, this procedure can be seen as an alternative to the two-step approach and gives an estimator for $f^{\dagger c}_{\theta|W}$ that has the same rate as computed in Section 4.2.1, see proposition 5.1 of Neubauer (1988).

In the case $f_{\theta|W}^{\dagger c} \in \partial\mathcal{F}_{\theta|W}$, the rate of convergence is slower and is given in Theorem (6). The next theorem, which is valid in both the cases $f_{\theta|W}^{\dagger c} = f_{\theta|W}^{\dagger}$ and $f_{\theta|W}^{\dagger c} \neq f_{\theta|W}^{\dagger}$, states the convergence of the solution of (4.10) towards $f_{\theta|W}^{\dagger c}$.

**Theorem 5.** *Let $f_{\theta|W}^{\alpha,c}$ be the solution of (4.10) when $\hat{f}_{C|WZ}$ is replaced by the true $f_{C|WZ}$ and $Q$ be the orthogonal projector of $L^2_{\pi_{cz}}$ onto $\overline{\mathcal{R}(T)}$. Then:*

*(i)* $\lim_{\alpha\to 0} f_{\theta|W}^{\alpha,c} = f_{\theta|W}^{\dagger c}$*;*

*(ii) moreover, if $\hat{f}_{C|WZ} \in L^2_{\pi_{cz}}$ is such that $||Q(\hat{f}_{C|WZ} - f_{C|WZ})|| = \mathcal{O}_p(\delta)$, for some $\delta \to 0$, if $\alpha \to 0$ and $\delta^2\alpha^{-1} \to 0$ as $\delta \to 0$ then:*

$$\lim_{\delta\to 0} \check{f}_{\theta|W}^{\alpha,c} = f_{\theta|W}^{\dagger c}.$$

The next theorem gives the rate of convergence of $\check{f}_{\theta|W}^{\dagger c}$ for the case $f_{\theta|W}^{\dagger c} \neq f_{\theta|W}^{\dagger}$.

**Theorem 6.** *Let $f_{\theta|W}^{\dagger c} \in \partial\mathcal{F}_{\theta|W}$ and $f_{\theta|W}^{\dagger c} \in \mathcal{P}_c(\mathcal{R}(T^*T)^{\nu})$ for some $\nu \in (0,1]$. If $\alpha \asymp (\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2)^{-\frac{1}{\nu\wedge 1+1}}$ then:*

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

The result of this theorem shows that in the constrained case with $f_{\theta|W}^{\dagger c} \neq f_{\theta|W}^{\dagger}$ the rate of convergence is slower than in the unconstrained case and it is at most equal to $\left(\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2\right)^{\frac{1}{2}}$. In order to get the same rate as in the unconstrained case it would necessary to add further conditions on the set $\mathcal{F}_{\theta|W}$ and on $f_{\theta|W}^{\dagger c}$. Here, we do not discuss this result further and refer to Neubauer (1986).

# 5 Monte Carlo Study

In this section we illustrate the small sample performance of our *indirect regularized density estimator* based on Tikhonov regularization for the Lifecycle savings problem described in Example 2. In our simulation, 1000 agents start at age 21, work for 45 periods, and then retire obtaining a terminal utility. Income grows until retirement. In addition, in each period each agent faces a permanent income shock $\eta_t$. These shocks are independent over time and across individuals and distributed as $\eta_t \sim \mathcal{N}(0, 0.05668)$. The initial values of income and permanent shock are set to zero. The interest rate is set to $R = 1 + r = 1.05$ and the true distribution of $\theta = (\gamma, \beta)$ has support on $[0.1, 0.700] \times [4.0, 0.999]$ and is a nonlinear transformation of a normal random variable with mean vector $(1, 0)'$ and identity variance-covariance matrix. In addition, measurement error in consumption is $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon^2$ set equal to 25% of the the cross-sectional variance of consumption.

We show in Figures **??**-**??** the quartiles of the simulated $Y_t$, $C_t^*$, $C_t$ and $A_t$ for each $t$. For the estimation, we select one cross section to be used for our estimator. The results presented in Figures refer to the period $t = 40$. Recall that in this case the joint distribution of $(\gamma, \beta)$ is not identified because they enter the kernel of the operator only as a single index. Instead we estimate the distribution of

$$m = 0.5\phi_5\gamma + \phi_3 \frac{\ln(R\beta)}{\gamma}.$$

In the figure we represent an (infeasible) kernel density estimator of the *pdf* of $m$ (in solid black line) together with our Tikhonov estimator (in dashed green line) and the MLE estimator (in dotted red line) for comparison. The regularization parameter $\alpha$ for the Tikhonov estimator has been set equal to 0.3.

# 6 Conclusion

This paper develops results on semi-parametric identification and estimation of the *pdf* of the unobserved heterogeneity in structural models. We identify the *pdf* of interest as the solution of a non-linear inverse problem. The identified set is a convex subset of $L^2_{\pi_\theta}$ that can be entirely characterized by a convex subset in $\mathbb{R}^{dim(\mathcal{N}(T))}$. The estimation methods that we propose are borrowed by the Inverse Problem literature. Finally, a simulated exercise for the Euler Equation in intertemporal consumption models shows good finite-sample properties if our procedure.

# 7 Appendix: Proofs

## 7.1 Proof of Theorem 1

We start by proving the first part of the theorem. Let us decompose the joint conditional *pdf* $f_{C\theta|WZ}$ on $\mathcal{B}_\mathcal{C} \otimes \mathcal{B}_\Theta$ conditional on $(W, Z)$ as $f_{C\theta|WZ} = f_{C|WZ\theta} f_{\theta|WZ}$ and by using Assumption 4 we obtain

$$\begin{aligned} f_{C|WZ}(c; w, z) &= \int_\Theta f_{C\theta|WZ}(c, \theta; w, z) d\theta = \int_\Theta f_{C|WZ\theta}(c; w, z, \theta) f_{\theta|WZ}(\theta; w, z) d\theta \\ &= \int_\Theta f_{C|WZ\theta}(c; w, z, \theta) f_{\theta|W}(\theta; w) d\theta \end{aligned} \tag{7.1}$$

since outside $\Theta$ the *pdf* $f_{C\theta|WZ}$ is zero. This defines an integral operator $T$ that is defined on the whole $L_{\pi_\theta}$.

Now we prove the second part of the theorem. By Assumption 1 there exists a unique $c = \varphi(w, z, \theta, \varepsilon)$ that satisfies (2.1). Thus, from $f_{\varepsilon|WZ\theta}$ specified in Assumption 3 we can characterize the *pdf* $f_{C|WZ\theta}$ of the transformation $\varphi(w, z, \theta, \cdot)$ of $\varepsilon$. 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 $(w, z, \theta)$. Let $\varepsilon^i = \varphi_i^{-1}(w, z, \theta, \cdot) : \mathbb{R} \to \mathcal{E}_i$ be the corresponding inverse mapping for given $(w, z, \theta)$. Then,

$$f_{C|WZ\theta}(c; w, z, \theta) = \sum_{i=1}^{s} f_{\varepsilon|WZ\theta}(\varphi_i^{-1}(w, z, \theta, c); \theta_2) \Big| \partial_c \varphi_i^{-1}(w, z, \theta, c) \Big| 1_{S_{C|WZ\theta}}(c) \tag{7.2}$$

Finally, since a unique solution in $C$ to (2.1) exists we have 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 using the chain rule. Therefore, $\partial_\varepsilon c = \partial_\varepsilon \varphi(w, z, \theta, \varepsilon)$ and $\partial_\varepsilon \varphi(w, z, \theta, \varepsilon) = -\frac{\partial_\varepsilon \Psi(c,w,z,\theta,\varepsilon)}{\partial_c \Psi(c,w,z,\theta,\varepsilon)}$. We conclude that

$$\begin{aligned}
\partial_c \varphi_i^{-1}(w, z, \theta, c) &= \frac{1}{\partial_\varepsilon \varphi(w, z, \theta, \varepsilon)|_{\varepsilon = \varphi_i^{-1}(w,z,\theta,c)}} = \Big[\frac{\partial_\varepsilon \Psi(\varphi(w, z, \theta, \varepsilon), w, z, \theta, \varepsilon)}{\partial_c \Psi(\varphi(w, z, \theta, \varepsilon), w, z, \theta, \varepsilon)}\Big]^{-1} \Big|_{\varepsilon = \varphi_i^{-1}(w,z,\theta,c)} \\
&= \Big[\frac{\partial_c \Psi(c, w, z, \theta, \varphi_i^{-1}(w, z, \theta, c))}{\partial_\varepsilon \Psi(c, w, z, \theta, \varphi_i^{-1}(w, z, \theta, c))}\Big]. \qquad (7.3)
\end{aligned}$$

By replacing (7.3) in (7.2) and (7.2) in (7.1) we get the result.

*(ii)* Since $f_{C|WZ} \in T\mathcal{F}_{\theta|W}$ then there exists a probability density function $h_{\theta|W} \in \mathcal{F}_{\theta|W}$ such that $f_{C|WZ} = Th_{\theta|W}$. To show the characterization (**??**) we first apply the definition of the identified set to find the first line in the formula. Then, let $\{\varphi_j\}_{j;\lambda_j=0}$ be the eigenfunctions of $T^*T$ corresponding to the 0 eigenvalue. Then, $\overline{\mathcal{N}(T)} = span\{\varphi_j; \lambda_j = 0\}$. The intersection $(f^\dagger_{\theta|W} \oplus \mathcal{N}(T)) \cap \mathcal{F}_{\theta|W}$ is given by all the functions in $(f^\dagger_{\theta|W} \oplus \mathcal{N}(T))$ that are positive a.e. on $\Theta$ and that integrate to 1 on $\Theta$. In particular integration to 1 is guaranteed by choosing $z_1$ such that $\int_\Theta h(\theta; w) = 1$. This is equivalent to

$$< f^\dagger_{\theta|W}, \frac{1}{\pi_\theta} > + \sum_{j \geq 1; \lambda_j = 0} z_j < \varphi_j, \frac{1}{\pi_\theta} >= 1$$

which gives the result.

## 7.2 Proof of Proposition 1

First, we remark that an integral operator from $L^2_{\pi_\theta}$ to $L^2_{\pi_{cz}}$ is Hilbert-Schmidt if its kernel is square integrable with respect to $\pi_\theta \times \pi_{cz}$. An Hilbert-Schmidt operator is bounded and compact. We start by proving that $\mathcal{R}(T) \subset L^2_{\pi_{cz}}$

*(i)* Under the conditions of the proposition let us compute $\int_\mathcal{C} \int_\mathcal{Z} \int_\Theta \frac{f^2_{C|WZ\theta}}{\pi^2_\theta}$:

$$\begin{aligned}
&\int_{\mathcal{C}\times\mathcal{Z}} \int_\Theta \Big[\sum_{i=1}^{s} f_{\varepsilon|WZ\theta}(\varphi_i^{-1}(w, z, \theta, c); w, z, \theta) \, |\partial_c \varphi_i^{-1}(w, z, \theta, c)|\Big]^2 1_{S_{C|WZ\theta}}(c) \frac{\pi_\theta(\theta)\pi_{cz}(c, z)}{\pi^2_\theta} d\theta dc dz \\
\leq \; & \int_\Theta \int_\mathcal{Z} \int_\mathcal{C} \sum_{i=1}^{s} f^2_{\varepsilon|WZ\theta}(\varphi_i^{-1}(w, z, \theta, c); w, z, \theta) \, |\partial_c \varphi_i^{-1}(w, z, \theta, c)|^2 \sum_{i=1}^{s} 1_{S_{C|WZ\theta}}(c) \frac{\pi_{cz}(c, z)}{\pi_\theta} dc dz d\theta \\
= \; & s \sum_{i=1}^{s} \int_\Theta \int_\mathcal{Z} \int_{S_{\varepsilon_i|WZ\theta}} f^2_{\varepsilon|WZ\theta}(\varepsilon_i; w, z, \theta) \, |\partial_{\varepsilon_i} \varphi(w, z, \theta, \varepsilon_i)|^{-1} \frac{\pi_{cz}(\varphi(w, z, \theta, \varepsilon_i), z)}{\pi_\theta(\theta)} d\varepsilon_i dz d\theta \\
= \; & s \int_\Theta \int_\mathcal{Z} \int_{S_{\varepsilon|WZ\theta}} f^2_{\varepsilon|WZ\theta}(\varepsilon; w, z, \theta) \, |\partial_\varepsilon \varphi(w, z, \theta, \varepsilon)|^{-1} \frac{\pi_{cz}(\varphi(w, z, \theta, \varepsilon), z)}{\pi_\theta(\theta)} d\varepsilon dz d\theta \\
\leq \; & \Big(\sup_{(z,\theta,\varepsilon)} |\partial_\varepsilon \varphi(w, z, \theta, \varepsilon)|^{-1}\Big) s \int_\Theta \int_\mathcal{Z} \int_{S_{\varepsilon|WZ\theta}} f^2_{\varepsilon|WZ\theta}(\varepsilon; w, z, \theta) \; \frac{\pi_{cz}(\varphi(w, z, \theta, \varepsilon), z)}{\pi_\theta(\theta)} d\varepsilon dz d\theta < \infty
\end{aligned}$$

where the first inequality follows by the Fubini theorem and the second one by the change of variable $\varphi_i^{-1}(w,z,\theta,c)=\varepsilon_i$. The notation $S_{\varepsilon_i|WZ\theta}$ means the support of the new variable $\varepsilon_i$.

By using the inequality in (3.3), this result shows that $\mathcal{R}(T)\subset L^2_{\pi_{cz}}$. Moreover, it shows that $T$ is Hilbert-Schmidt and then bounded and compact.

*(ii)* If $f_{C|WZ\theta}\in L^2_{\pi_\theta}\times L^2_{\pi_{cz}}$, $P^W$-a.s. then

$$\int_{\mathcal{C}}\int_{\mathcal{Z}}\int_{\Theta} f^2_{C|WZ\theta}(c;w,z,\theta)\frac{\pi_{cz}(c,z)\pi_\theta}{\pi_\theta^2}d\theta dzdc \leq \frac{1}{\inf_\theta \pi_\theta(\theta)}||f_{C|WZ\theta}||^2<\infty.$$

## 7.3 Proof of Proposition 2

By definition, the adjoint operator $T^*$ of the bounded linear operator $T$ satisfies: $\forall h\in L^2_{\pi_\theta}$, $\forall\psi\in L^2_{\pi_{cz}}$, $<Th,\psi>=<h,T^*\psi>$. Thus,

$$\begin{aligned}
<Th,\psi> &= \int_{\mathcal{C}}\int_{\mathcal{Z}}(Th)(c;w,z)\psi(c,z;w)\pi_{cz}(c,z)dcdz = \int_{\mathcal{C}}\int_{\mathcal{Z}}\int_{\Theta} f_{C|WZ\theta}(c;w,z,\theta)h(\theta;w)d\theta\psi(c,z;w)\pi_{cz}(c,z)dcdz\\
&= \int_{\Theta} h(\theta;w)\pi_\theta \int_{\mathcal{C}}\int_{\mathcal{Z}} f_{C|WZ\theta}(c;w,z,\theta)\psi(c,z;w)\frac{\pi_{cz}(c,z)}{\pi_\theta(\theta)}dcdzd\theta =<h,T^*\psi>
\end{aligned}$$

where the third equality follows from the Fubini theorem.

## 7.4 Proof of Lemma 3.1

For simplicity we consider the case where $\theta$ is one-dimensional (the multi-dimensional case can be recovered in a similar way). Let us suppose that $T\phi(\theta;w)=0$ $P^W$-a.e. for some function $\phi\in L^2_{\pi_\theta}(\theta)$. Then,

$$T\phi=\int_{\Theta}\sum_{i=1}^{s} f_{\varepsilon|\theta WZ}(\varphi_i^{-1}(w,z,\theta,c);\theta,w,z)|\partial_c\varphi_i^{-1}(w,z,\theta,c)|\phi(\theta;w)d\theta=0$$

implies

$$\int_{\Theta} f_{\varepsilon|\theta Wz}(\varphi_i^{-1}(w,z,\theta,c);\theta,w,z)|\partial_c\varphi_i^{-1}(w,z,\theta,c)|\phi(\theta;w)d\theta=0 \quad \forall i=1,\dots,s.$$

Then, $\forall i=1,\dots,s$ we have:

$$\begin{aligned}
0=T\phi &= \int_{\Theta}\exp\{\tau_i(c,w,z)m_i(\theta)\}h_i(\theta)k_i(c,w,z)\phi(\theta;w)|\partial_c\varphi^{-1}(w,z,\theta,c)|d\theta\\
&= \int_{\Theta}\exp\{\tau_i(c,w,z)m_i(\theta)\}h_i(\theta)k_i(c,w,z)\tilde{\phi}(\theta;w,z,c)|d\theta\\
&= \int_{\Theta}\exp\{\tau_i(c,w,z)\mu_i\}h_i(m_i^{-1}(\mu_i))k_i(c,w,z)\tilde{\phi}_i(m_i^{-1}(\mu_i);w,z,c)|dm_i^{-1}(\mu_i)=0
\end{aligned}$$

where we have used the notation $\tilde{\phi}_i(\theta;w,z,c):=\phi(\theta;w)|\partial_c\varphi_i^{-1}(w,z,\theta,c)|$ 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 $\nu_i(d\mu_i)=$

$h_i(m_i^{-1}(\mu_i))dm_i^{-1}(\mu_i)d\mu_i$. Thus,

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

where $\zeta_i(\mu_i;w,z,c) = \tilde{\phi}_i(m_i^{-1}(\mu_i);w,z,c)$, $F_i(d\mu_i;w,z,c) = \zeta_i^+(\mu_i;w,z,c)\nu_i(d\mu_i)$, $G_i(d\mu_i;w,z,c) = \zeta_i^-(\mu_i;w,z,c)\nu_i(d\mu_i)$ and , for a function $h$, $h^+$ and $h^-$ denote the posterior and negative part of it, respectively. It follows that

$$\int_\Theta \exp\{\tau_i(c,w,z)\mu_i\}F_i(d\mu_i;w,z,c) = \int_\Theta \exp\{\tau_i(c,w,z)\mu_i\}G_i(d\mu_i;w,z,c)$$

that is, $F_i$ and $G_i$ are two measures with the same Laplace transform. Then, they are equal. This implies that $\zeta_i(\mu_i;w,z,c) = 0$ and then $\phi_i(\theta;w) = 0$ for $P^W$-almost all $w \in \mathcal{W}$ since $\partial_c\varphi_i^{-1}(w,z,\cdot,c)$ is bounded away from 0 and $\infty$ $\forall(c,w,z) \in \mathcal{C}\times\mathcal{Z}\times\mathcal{W}$.

## 7.5 Proof of Theorem 3

First, since $||P_c|| \leq 1$ we have:

$$\mathbb{E}||\mathcal{P}_c\hat{f}^\alpha_{\theta|W} - f^{\dagger c}_{\theta|W}||^2 = \mathbb{E}||\mathcal{P}_c(\hat{f}^\alpha_{\theta|W} - f^\dagger_{\theta|W})||^2 \leq \mathbb{E}||\mathcal{P}_c||^2\mathbb{E}||\hat{f}^\alpha_{\theta|W} - f^\dagger_{\theta|W}||^2 \leq \mathbb{E}||\hat{f}^\alpha_{\theta|W} - f^\dagger_{\theta|W}||^2.$$

We develop here only the proof for the case with $T$ compact and we refer to Engl *et al.* (2000 Section 5.1) for a proof in the general non-compact case. Let $f^\alpha_{\theta|W} := (\alpha I + T^*T)^{-1}T^*f_{C|WZ}$, then

$$\mathbb{E}||\hat{f}^\alpha_{\theta|W} - f^\dagger_{\theta|W}||^2 \leq 2\mathbb{E}||\hat{f}^\alpha_{\theta|W} - f^\alpha_{\theta|W}||^2 + 2\mathbb{E}||f^\alpha_{\theta|W} - f^\dagger_{\theta|W}||^2 := 2(\mathcal{A}_1 + \mathcal{A}_2). \tag{7.4}$$

Term $\mathcal{A}_1$ is

$$\begin{aligned}
\mathcal{A}_1 &= \mathbb{E}||(\alpha I + T^*T)^{-1}T^*(\hat{f}_{C|WZ} - f_{C|WZ})||^2 \leq ||(\alpha I + T^*T)^{-1}T^*||^2\mathbb{E}||(\hat{f}_{C|WZ} - f_{C|WZ})||^2 \\
&= \sup_{||u||\leq 1}\sum_j \frac{\lambda_j^2}{(\alpha+\lambda_j^2)^2} < u,\psi_j >^2 \mathbb{E}||(\hat{f}_{C|WZ} - f_{C|WZ})||^2 \\
&\leq \Big(\sup_j \frac{\lambda_j}{(\alpha+\lambda_j^2)}\Big)^2\mathbb{E}||(\hat{f}_{C|WZ} - f_{C|WZ})||^2 = \frac{1}{2\alpha}\mathbb{E}||(\hat{f}_{C|WZ} - f_{C|WZ})||^2.
\end{aligned}$$

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

$$\begin{aligned}
\mathcal{A}_2 &= \mathbb{E}||(\alpha I + T^*T)^{-1}T^* f_{C|WZ} - f_{\theta|W}||^2 = ||[I - (\alpha I + T^*T)^{-1}T^*T]f_{\theta|W}||^2 \\
&= ||\alpha(\alpha I + T^*T)^{-1} f_{\theta|W}||^2 = \alpha^2 \sum_j \frac{\lambda_j^{2\beta}}{(\alpha+\lambda_j^2)^2} \frac{< f^\dagger_{\theta|W}, \varphi_j >^2}{\lambda_j^{2\beta}} \\
&\leq \alpha^2 \Big( \sup_j \frac{\lambda_j^\beta}{(\alpha+\lambda_j^2)} \Big)^2 \sum_j \frac{< f^\dagger_{\theta|W}, \varphi_j >^2}{\lambda_j^{2\beta}} \leq \alpha^\beta \frac{(2-\beta)^{2-\beta}}{4} \beta^\beta Q \quad \text{if } \beta \leq 2
\end{aligned}$$

since $f^\dagger_{\theta|W} \in \Phi_\beta(Q)$ under Assumption 7. This shows that $\mathbb{E}||\hat{f}^\alpha_{\theta|W} - f^\dagger_{\theta|W}||^2 = \mathcal{O}\Big(\alpha^{\beta\wedge 2} + \frac{1}{\alpha}\mathbb{E}||(\hat{f}_{C|WZ} - f_{C|WZ})||^2\Big)$. If we choose $\alpha$ such that $\alpha^{\beta\wedge 2} \asymp \frac{1}{\alpha}\mathbb{E}||(\hat{f}_{C|WZ} - f_{C|WZ})||^2$ then, we get the second result of the Theorem.

## 7.6 Proof of Corollary 1

Following the decomposition (7.4) in the proof of Theorem 3 we only have to deal with term $\mathcal{A}_1$:

$$\mathbb{E}||\hat{f}^\alpha_{\theta|W} - f^\alpha_{\theta|W}||^2 = \sum_{j=1}^\infty \frac{1}{(\alpha+\lambda_j^2)^2} \mathbb{E} < T^*(\hat{f}_{C|WZ} - f_{C|WZ}), \varphi_j >^2 = \sum_{j=1}^\infty \frac{1}{(\alpha+\lambda_j^2)^2} [\text{Bias}^2 + \text{Variance}]$$

where $\mathbb{E}$ is the expectation taken with respect to $f_{CWZ}$. We start by approximating $(\hat{f}_{C|WZ} - f_{C|WZ})$ by a Taylor expansion of the first order. Let $\hat{P} := \frac{1}{nh^{l+k+1}} \sum_{i=1}^n K_h(c_i - c, c) K_h(w_i - w, w) K_h(z_i - z, z)$, $P := f_{CWZ}(c, w, z)$, $\hat{Q} := \frac{1}{nh^{l+k}} \sum_{l=1}^n K_h(w_l - w, w) K_h(z_l - z, z)$ and $Q := f_{WZ}(w, z)$. Then, $\frac{\hat{P}}{\hat{Q}} = \frac{P}{Q} + \frac{1}{Q}(\hat{P} - \frac{P}{Q}\hat{Q}) + \tilde{\epsilon}_n$, where $\tilde{\epsilon}_n = \mathcal{O}_p((\hat{P} - P) + (\hat{Q} - Q))$. Then, by denoting $\mathcal{C} = [0,1]$, $\mathcal{Z} = [0,1]^l$, $\mathcal{W} = [0,1]^k$ and $\prod_{\mathcal{Z}} = \prod_{j=1}^l \left[\frac{z_j - 1}{h}, \frac{z_j}{h}\right]$, we have

$$
\begin{aligned}
\mathbb{E}||\hat{f}^{\alpha}_{\theta|W} - f^{\alpha}_{\theta|W}||^2 &= \sum_{j=1}^{\infty} \frac{1}{(\alpha+\lambda_j^2)^2}\mathbb{E}\Big( < \int_{\mathcal{Z}}\int_{\mathcal{C}} f_{C|WZ\theta}(c;w,z,\theta)\frac{1}{Q}(\hat{P} - \frac{P}{Q}\hat{Q})\frac{\pi_{CZ}(c,z;w)}{\pi_\theta(\theta;w)} dcdz, \varphi_j >^2 \Big) + \epsilon_n \\
&= \sum_{j=1}^{\infty} \frac{\lambda_j^2}{(\alpha+\lambda_j^2)^2}\mathbb{E}\Big( \int_{\mathcal{Z}} \frac{1}{nh^{l+k}} \frac{1}{f_{WZ}(w,z)} \sum_{i=1}^{n} \Big( \Big[ g_j(c_i,z;w) - \mathbb{E}(g_j(c,z;w)|w,z) \Big] \\
&\quad + \frac{h^{\rho_c}}{\rho_c!}\frac{\partial^{\rho_c}}{\partial c^{\rho_c}}\Big[g_j(\tilde{c},z;w)\Big]\sigma_u^2\Big) K_h(w_i - w, w) K_h(z_i - z, z) dz\Big)^2 + \epsilon_n \\
&= \sum_{j=1}^{\infty} \frac{\lambda_j^2}{(\alpha+\lambda_j^2)^2}\mathbb{E}\Big( \frac{1}{nh^k}\sum_{i=1}^{n} \frac{\Big[ g_j(c_i,z_i;w) - \mathbb{E}(g_j(c,z_i;w)|w,z_i)\Big]}{f_{WZ}(w,z_i)} K_h(w_i - w, w) \\
&\quad + \frac{1}{\rho_z! nh^k}\sum_{i=1}^{n}\Big[ h^{\rho_z}\int_{\prod \mathcal{Z}} \Big( \sum_{j'=1}^{l} u_{j'}\frac{\partial}{\partial z_{j'}}\Big)^{\rho_z}\Big[\frac{(g_j(c_i,\tilde{z};w) - \mathbb{E}(g_j(c,z_i;w)|w,\tilde{z}))}{f_{WZ}(w,\tilde{z})}\Big] K(u,z)du \times \\
&\quad K_h(w_i - w, w) + \frac{h^{\rho_c}}{\rho_c! f_{WZ}(w,z_i)}\frac{\partial^{\rho_c}}{\partial c^{\rho_c}}\Big[g_j(\tilde{c},z_i;w)\Big]\sigma_u^2 K_h(w_i - w, w)\Big]\Big)^2 + \epsilon_n \\
&:= \sum_{j=1}^{\infty} \frac{\lambda_j^2}{(\alpha+\lambda_j^2)^2}\mathbb{E}\Big( A_j + \varepsilon_j + C_j\Big)^2 + \epsilon_n \leq \frac{2}{\alpha}\sum_{j=1}^{\infty}\mathbb{E}\Big(A_j^2 + 2\varepsilon_j^2 + 2C_j^2\Big) + \epsilon_n
\end{aligned} \tag{7.5}
$$

where $g_j(c,z;w) = \psi_j(c,z;w)\pi_{CZ}(c,z;w)$, $\sigma_u^2 := \int_{[\frac{c-1}{h},\frac{c}{h}]} u^2 K(u,c)du$, $\tilde{c} \in [c_i, c_i + hu]$ and $\tilde{z}$ lies on the line segment joining $z_i$ and $(z_i + hu)$. Moreover, $\rho_c$ and $\rho_z$ denotes the number of derivatives of $f_{C|WZ}$ in $C$ and in $Z$, respectively. Term $\epsilon_n$ is given by $\epsilon_n = \sum_{j=1}^{\infty}\frac{1}{(\alpha+\lambda_j^2)^2}\mathbb{E} < T^*\tilde{\epsilon}_n(c,w,z), \varphi_j >^2$ and it is negligible with respect to the other terms in (7.5) under Assumption 8.

We start by computing the bias and variance of $A_j$; it will result clear that the bias and variance of the other two terms $\varepsilon_j$ and $C_j$ are negligible. The bias of $A_j$ is of the same order as $\mathbb{E}(A_j)$ where

$$
\begin{aligned}
\mathbb{E}(A_j) &= \frac{1}{nh^k}\sum_{i=1}^{n}\int_{\mathcal{W}}\int_{\mathcal{Z}}\int_{\mathcal{C}} \frac{g_j(c_i,z_i;w) - \mathbb{E}(g_j(c,z_i;w)|w,z_i)}{f_{WZ}(w,z_i)} K_h(w_i - w, w) f_{CWZ}(c_i,w_i,z_i) dc_i dz_i dw_i \\
&= \frac{h^\rho}{\rho!}\int_{\mathcal{Z}}\int_{\mathcal{C}} \frac{g_j(c_i,z_i;w) - \mathbb{E}(g_j(c_i,z_i;w)|w,z_i)}{f_{WZ}(w,z_i)} \\
&\quad \times \int_{\prod_{j=1}^{k}\Big[\frac{w_j - 1}{h},\frac{w_j}{h}\Big]} \Big( \sum_{j'=1}^{k} u_{j'}\frac{\partial}{\partial w_{j'}}\Big)^{\rho} f_{CWZ}(c_i,w,z_i) K(u,w) du dc_i dz_i + \mathcal{O}(h^\rho)
\end{aligned}
$$

after having applied a Taylor expansion at the order $\rho$. The first $(\rho - 1)$ terms in the expansion are equal to 0. The Variance of $T^*\hat{f}_{C|WZ}$ is approximately equal to $Var(A_j)$ which is of the order

$$
\begin{aligned}
Var(A_j) &\leq \frac{1}{n^2h^{2k}}\sum_{i=1}^{n}\int_{\mathcal{W}}\int_{\mathcal{Z}} \frac{1}{f_{WZ}^2(w,z_i)} Var(g_j(c_i,z_i;w)|w_i,z_i) K_h^2(w_i - w, w) f_{WZ}(w_i,z_i) dz_i dw_i \\
&= \frac{1}{nh^k}\int_{\mathcal{Z}} \frac{1}{f_{WZ}(w,z_i)}\sigma_j^2(w,z_i)dz_i \int_{\prod_{j=1}^{k}\Big[\frac{w_j - 1}{h},\frac{w_j}{h}\Big]} K^2(u,w)du + \mathcal{O}\Big(\frac{1}{nh^k}\Big).
\end{aligned}
$$

where $\sigma_j^2(w,z_i) = Var[g_j(c_i,z_i;w)|w,z_i]$. Going back to formula (7.5), under Assumption 8 the first part $2\sum_{j=1}^{\infty}\frac{\lambda_j^2}{(\alpha+\lambda_j^2)^2}\mathbb{E}(A_j^2)$ of this expression is bounded above by

$$\begin{aligned}
\leq\;& \frac{2}{\alpha}\sum_{j=1}^{\infty}(Var(A_j)+\mathbf{E}^2(A_j)) = \frac{2}{\alpha nh^k}\sum_{j=1}^{\infty}\int\frac{\sigma_j^2(w,z_i)}{f_{WZ}(w,z_i)}dz_i\int_{\prod_{j=1}^k\left[\frac{w_j-1}{h},\frac{w_j}{h}\right]}K^2(u,w)du+\\
& \frac{2h^{2\rho}}{\alpha(\rho!)^2}\sum_{j=1}^{\infty}\Big(\int_{\mathcal{C}}\int_{\mathcal{Z}}\frac{1}{f_{WZ}(w,z_i)}[g_j(c_i,z_i;w)-\mathbb{E}(g_j(c_i,z_i;w)|w,z_i)]\times\\
& \int_{\prod_{j=1}^k\left[\frac{w_j-1}{h},\frac{w_j}{h}\right]}\Big(\sum_{j'=1}^{k}u_{j'}\frac{\partial}{\partial w_{j'}}\Big)^{\rho}f_{CWZ}(c_i,w,z_i)K(u,w)dudc_idz_i\Big)^2+\mathcal{O}\Big(\frac{1}{\alpha nh^k}+\frac{h^{2\rho}}{\alpha}\Big)\\
\asymp\;& \frac{1}{\alpha}\Big(\frac{1}{nh^k}+h^{2\rho}\Big)
\end{aligned}$$

where the last line is true under Assumption 8. From this result it is clear that the terms in (7.5) depending on $\varepsilon_j$ and $C_j$ are negligible with respect to $\mathbb{E}(A_j^2)$. Finally,

$$\mathbb{E}||\hat{f}^{\alpha}_{\theta|W}-f^{\alpha}_{\theta|W}||^2\leq\frac{1}{\alpha}\sum_{j=1}^{\infty}(\kappa_{j1}^2\frac{1}{nh^k}+\kappa_{j2}^2h^{2\rho})+\mathcal{O}\Big(\frac{1}{\alpha nh^k}+\frac{h^{2\rho}}{\alpha}\Big)=\mathcal{O}\Big(\frac{1}{\alpha}\Big(\frac{1}{nh^k}+h^{2\rho}\Big)\Big).$$

It can be shown that this rate is uniform over $\mathcal{F}$ and it is minimized for $h\asymp n^{-\frac{1}{2\rho+k}}$, so that $\sup_{\mathcal{F}}\inf_h\mathbb{E}||\hat{f}^{\alpha}_{\theta|W}-f^{\alpha}_{\theta|W}||^2=\mathcal{O}(\frac{1}{\alpha}n^{-\frac{2\rho}{2\rho+k}})$. >From the proof of Theorem 3 we know that the bias is of order $\alpha^{\beta}$. Hence, $\sup_{\mathcal{F}}\inf_h\mathbb{E}||\hat{f}^{\alpha}_{\theta|W}-f^{\dagger}_{\theta|W}||^2=\mathcal{O}(\alpha^{\beta\wedge 2}\frac{1}{\alpha}n^{-\frac{2\rho}{2\rho+k}})$ and the value of $\alpha$ that minimizes $\mathbb{E}||\hat{f}^{\alpha}_{\theta|W}-f^{\dagger}_{\theta|W}||^2$ is $\alpha\asymp n^{-\frac{2\rho}{(2\rho+k)(\beta\wedge 2+1)}}$. This concludes the proof.

## 7.7 Proof of Lemma 4.1

We want to determine the rate $\varepsilon_n$ and the function $\Omega(\theta;w)$ such that $\varepsilon_n(\mathcal{P}_c\hat{f}^{\alpha}_{\theta|W}-f^{\dagger c}_{\theta|W})\Rightarrow\mathcal{N}(0,\Omega(\theta;w))$ for a fixed value of $(\theta;w)$. We use formula (4.6) and the decomposition $\mathcal{P}_c(\hat{f}^{\alpha}_{\theta|W}-f_{\theta|W})=\mathcal{P}_c(\hat{f}^{\alpha}_{\theta|W}-f^{\alpha}_{\theta|W})+\mathcal{P}_c(f^{\alpha}_{\theta|W}-f^{\dagger}_{\theta|W})$ where $f^{\alpha}_{\theta|W}:=(\alpha I+T^*T)^{-1}T^*f_{C|WZ}$. First, we make the following developments:

$$\begin{aligned}
<T^*f_{C|WZ},\varphi_j> &= \int_{\Theta}\int_{\mathcal{Z}}\int_{\mathcal{C}}f_{C|WZ\theta}(c;w,z,\theta)f_{C|WZ}(c;w,z)\varphi_j(\theta)\pi_{cz}(c;z)dcdzd\theta\\
&= \lambda_j\int_{\mathcal{Z}}\int_{\mathcal{C}}f_{C|WZ}(c;w,z)g_j(c,z;w)dcdz := A_{1j}
\end{aligned}$$

where $g_j(c,z;w):=\psi_j(c,z;w)\pi_{cz}(c,z;w)$ and

$$\begin{aligned}
<T^*\hat{f}_{C|WZ},\varphi_j> &= \int_{\Theta}\int_{\mathcal{Z}}\int_{\mathcal{C}}\frac{\frac{1}{nh^{l+k+1}}\sum_{i=1}^n K_h(c_i-c,c)K_h(w_i-w,w)K_h(z_i-z,z)}{\frac{1}{nh^{l+k}}\sum_{l=1}^n K_h(w_l-w,w)K_h(z_l-z,z)}\\
&\quad\times f_{C|WZ\theta}(c;w,z,\theta)\varphi_j(\theta)\pi_{cz}(c;z)dcdzd\theta\\
&= \lambda_j\int_{\mathcal{Z}}\frac{\frac{1}{nh^{l+k}}\sum_{i=1}^n g_j(c_i,z;w)K_h(w_i-w,w)K_h(z_i-z,z)}{\frac{1}{nh^{l+k}}\sum_{l=1}^n K_h(w_l-w,w)K_h(z_l-z,z)}dz+\\
&\quad\frac{h^{\rho_c}}{h^{\rho_c}!}\lambda_j\int_{\mathcal{Z}}\frac{\frac{1}{nh^{l+k}}\sum_{i=1}^n\sigma_u^2\frac{\partial^{\rho_c}}{\partial c^{\rho_c}}[g_j(\tilde{c};w,z)]K_h(w_i-w)K_h(z_i-z)}{\frac{1}{nh^{l+k}}\sum_{l=1}^n K_h(w_l-w)K_h(z_l-z)}dz\\
:=&\quad \hat{A}_{1j}+\hat{A}_{2j}
\end{aligned}$$

where $\sigma_u^2 := \int_{[\frac{c-1}{h},\frac{c}{h}]} u^{\rho_c} K(u,c)du$, $\tilde{c} \in [c_i, c_i + hu]$ and $\rho_c$ denotes the number of derivatives of $f_{C|WZ}$ in $C$. Therefore,

$$\mathcal{P}_c(\hat{f}^{\alpha}_{\theta|W} - f^{\alpha}_{\theta|W}) = \sum_{j=1}^{\infty} \frac{1}{\alpha + \lambda_j^2}(\hat{A}_{1j} - A_{1j})\tilde{\varphi}_j + \sum_{j=1}^{\infty} \frac{1}{\alpha + \lambda_j^2} \hat{A}_{2j}\tilde{\varphi}_j := A_1 + A_2 \tag{7.6}$$

where $\tilde{\varphi}_j = \mathcal{P}_c\varphi_j$. Let $\hat{P} := \frac{1}{nh^{l+k}}\sum_{i=1}^n g_j(c_i, z; w)K_h(w_i - w, w)K_h(z_i - z, z)$, $P := \int_{\mathcal{C}} f_{CWZ}(c, w, z)g_j(c, z; w)dc$, $\hat{Q} = \frac{1}{nh^{l+k}}\sum_{l=1}^n K_h(w_l - w, w)K_h(z_l - z, z)$ and $Q := f_{WZ}(w, z)$. By a Taylor expansion of the first order, we have: $\frac{\hat{P}}{\hat{Q}} = \frac{P}{Q} + \frac{1}{Q}(\hat{P} - \frac{P}{Q}\hat{Q}) + \tilde{\epsilon}_n$, where $\tilde{\epsilon}_n = o_p((\hat{P} - P) + (\hat{Q} - Q))$. Then,

$$\begin{aligned} A_1 &= \sum_{j=1}^{\infty} \frac{\lambda_j}{\alpha + \lambda_j^2}\Big(\int_{\mathcal{Z}} (\frac{\hat{P}}{\hat{Q}} - \frac{P}{Q})dz\Big)\tilde{\varphi}_j \\ &= \sum_{j=1}^{\infty} \frac{\lambda_j}{\alpha + \lambda_j^2}\Big(\int_{\mathcal{Z}} \frac{1}{Q}(\hat{P} - \frac{P}{Q}\hat{Q})dz\Big)\tilde{\varphi}_j + \epsilon_n := \sum_{j=1}^{\infty} \frac{\lambda_j}{\alpha + \lambda_j^2}\Big(\varepsilon_{1j} - \varepsilon_{2j}\Big)\tilde{\varphi}_j + \epsilon_n. \end{aligned}$$

where $\varepsilon_{1j}$ and $\varepsilon_{2j}$ can be further developed as:

$$\begin{aligned} \varepsilon_{1j} &= \frac{1}{nh^k}\sum_{i=1}^n \frac{1}{f_{WZ}(w, z_i)} g_j(c_i, z; w)K_h(w_i - w, w) + \\ &\quad \frac{h^{\rho_z}}{\rho_z!}\frac{1}{nh^k}\sum_{i=1}^n \int_{\prod_{\mathcal{Z}}} \Big(\sum_{j'=1}^{l} u_{j'}\frac{\partial}{\partial z_{j'}}\Big)^{\rho_z}\Big[\frac{(g_j(c_i, \tilde{z}; w) - \mathbb{E}(g_j(c, z_i; w)|w, \tilde{z}))}{f_{WZ}(w, \tilde{z})}\Big]K(u, z)duK_h(w_i - w, w) \\ := &\quad \varepsilon_{3j} + \varepsilon_{4j}; \\ \varepsilon_{2j} &= \frac{1}{nh^k}\sum_{i=1}^n \frac{\mathbb{E}[g_j(c_i, z_i; w)|z_i, w]}{f_{WZ}(w, z_i)}K_h(w_i - w, w) + \\ &\quad \frac{h^{\rho_z}}{\rho_z!}\frac{1}{nh^k}\sum_{i=1}^n \int_{\prod_{\mathcal{Z}}} \Big(\sum_{j'=1}^{l} u_{j'}\frac{\partial}{\partial z_{j'}}\Big)^{\rho_z}\Big[\frac{(g_j(c_i, \tilde{z}; w) - \mathbb{E}(g_j(c, z_i; w)|w, \tilde{z}))}{f_{WZ}(w, \tilde{z})}\Big]K(u, z)duK_h(w_i - w, w) \\ := &\quad \varepsilon_{5j} + \varepsilon_{6j} \end{aligned}$$

where $\prod_{\mathcal{Z}} = \prod_{j=1}^l \left[\frac{z_j - 1}{h}, \frac{z_j}{h}\right]$, $\tilde{z}$ lies on the line segment joining $z_i$ and $(z_i + hu)$ and $\rho_z$ denotes the number of derivatives of $f_{C|WZ}$ in $Z$. Then,

$$\begin{aligned} \mathcal{P}_c(\hat{f}^{\alpha}_{\theta|W} - f^{\alpha}_{\theta|W}) &= \sum_{j=1}^{\infty} \frac{\lambda_j}{\alpha + \lambda_j^2}(\varepsilon_{3j} + \varepsilon_{4j} - \varepsilon_{5j} - \varepsilon_{6j})\tilde{\varphi}_j + \sum_{j=1}^{\infty} \frac{1}{\alpha + \lambda_j^2}\hat{A}_{2j}\tilde{\varphi}_j + o_p((nh^k)^{-1/2} + h^{\rho \wedge \rho_z}) \\ &= \sum_{j=1}^{\infty} \frac{\lambda_j}{\alpha + \lambda_j^2}(\varepsilon_{3j} - \varepsilon_{5j} - \mathbb{E}(\varepsilon_{3j} - \varepsilon_{5j}))\tilde{\varphi}_j + \qquad (7.7) \\ &\quad \sum_{j=1}^{\infty} \frac{\lambda_j}{\alpha + \lambda_j^2}(\mathbb{E}(\varepsilon_{3j} - \varepsilon_{5j}) + \varepsilon_{4j} - \varepsilon_{6j})\tilde{\varphi}_j + \sum_{j=1}^{\infty} \frac{1}{\alpha + \lambda_j^2}\hat{A}_{2j}\tilde{\varphi}_j + o_p((nh^k)^{-1/2} + h^{\rho \wedge \rho_z}). \end{aligned} \tag{7.8}$$

It is easy to see that the terms in (7.8) are asymptotically irrelevant under Assumption 8 if $\sqrt{nh^k}h^{\rho_c} \rightarrow 0$ and $\sqrt{nh^k}h^{\rho_z} \rightarrow 0$. Therefore, we only have to prove that the term in (7.7) is asymptotically normal. For that, let us notice that we have a triangular array:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} Z_{ni} &= \frac{1}{n}\sum_{i=1}^{n}(\tilde{Z}_{ni} - \mathbb{E}(\tilde{Z}_{ni})) = \sum_{j=1}^{\infty}\frac{\lambda_j}{\alpha+\lambda_j^2}(\varepsilon_{3j} - \varepsilon_{5j} - \mathbb{E}(\varepsilon_{3j} - \varepsilon_{5j}))\tilde{\varphi}_j \qquad (7.9)\\
\text{with } : \tilde{Z}_{ni} &= \sum_{j=1}^{\infty}\frac{\lambda_j}{\alpha+\lambda_j^2}\frac{1}{h^k f_{WZ}(w,z_i)}\Big[g_j(c_i,z_i;w) - \mathbb{E}[g_j(c_i,z_i;w)|w,z_i]\Big]K_h(w_i-w,w)\tilde{\varphi}_j.
\end{aligned}$$

We have to check the Liapounov condition: $\sum_{i=1}^{n}\mathbb{E}|L_{ni}|^{2+\delta} \rightarrow 0$ as $n \rightarrow \infty$ for some $\delta > 0$ where $L_{ni} = \frac{Z_{ni}}{[nVar(Z_{ni})]^{\frac{1}{2}}}$. By using the $C_r$-inequality we obtain

$$\begin{aligned}
\sum_{i=1}^{n}\mathbb{E}|L_{ni}|^{2+\delta} &= [nVar(Z_{n1})]^{-(1+\delta/2)}n\mathbb{E}|Z_{n1}|^{2+\delta}\\
&\leq [nVar(Z_{n1})]^{-(1+\delta/2)}n2^{\delta+1}(\mathbb{E}|\tilde{Z}_{n1}|^{2+\delta} + |\mathbb{E}\tilde{Z}_{n1}|^{2+\delta})\\
&\leq [nVar(Z_{n1})]^{-(1+\delta/2)}n2^{\delta+1}(\mathbb{E}|\tilde{Z}_{n1}|^{2+\delta} + \mathbb{E}|\tilde{Z}_{n1}|^{2+\delta})\\
&= [Var(Z_{n1})]^{-(1+\delta/2)}n^{-\delta/2}2^{\delta+2}h^{-k(2+\delta)}\mathbb{E}|h^k\tilde{Z}_{n1}|^{2+\delta}\\
&= [h^kVar(Z_{n1})]^{-(1+\delta/2)}(nh^k)^{-\delta/2}2^{\delta+2}\mathbb{E}\Big(\frac{1}{h^k}|h^q\tilde{Z}_{n1}|^{2+\delta}\Big).
\end{aligned}$$

Since $nh^k \rightarrow \infty$ the Liapounov condition is satisfied if we show that $h^kVar(Z_{n1})$ is bounded from below and $\mathbb{E}\Big(\frac{1}{h^k}|h^k\tilde{Z}_{n1}|^{2+\delta}\Big) = \mathcal{O}(1)$ for some $\delta > 0$.

$$Var(Z_{n1}) = \frac{1}{h^k}\sum_{j=1}^{\infty}\frac{\lambda_j^2}{(\alpha+\lambda_j^2)^2}\sigma_{jj}\tilde{\varphi}_j^2 + \frac{2}{h^k}\sum_{j<l}\frac{\lambda_j\lambda_l}{(\alpha+\lambda_j^2)(\alpha+\lambda_l^2)}\sigma_{jl}\tilde{\varphi}_j\tilde{\varphi}_l$$

with

$$\begin{aligned}
\sigma_{jj} &= \int\frac{1}{f_{WZ}(w,z_1)}\omega_{jj}(w,z_1)dz_1\sigma_{uw} + \mathcal{O}(h^k)\\
\sigma_{jl} &= \int\frac{1}{f_{WZ}(w,z_1)}\omega_{jl}(w,z_1)dz_1\sigma_{uw} + \mathcal{O}(h^k)
\end{aligned}$$

where $\sigma_{uw} = \int_{\prod_{j=1}^{k}\left[\frac{w_j-1}{h},\frac{w_j}{h}\right]}K^2(u,w)du$, $\omega_{jj}(w,z_1) = Var[g_j(c_1,z_1;w)|w,z_1]$ and $\omega_{jl}(w,z_1) = Cov[g_j(c_1,z_1;w),g_l(c_1,z_1;w)|w,z_1]$. Because $h^kVar(Z_{n1})$ is a sum of positive terms it is bounded from below. Next, we develop $\mathbb{E}\Big(\frac{1}{h^k}|h^k\tilde{Z}_{n1}|^{2+\delta}\Big)$ for $\delta = 1$:

$$\begin{aligned}
\mathbb{E}\Big(\frac{1}{h^k}|h^k\tilde{Z}_{n1}|^3\Big) &= \mathbb{E}\Big(\frac{1}{h^k}\Big|\sum_{j=1}^{\infty}\frac{\lambda_j}{\alpha+\lambda_j^2}\frac{[g_j(c_1,z_1;w)-\mathbb{E}(g_j(c_1,z_1;w)|w,z_1)]}{f_{WZ}(w,z_1)}K_h(w_1-w,w)\tilde{\varphi}_j\Big|^3\Big)\\
&\leq \frac{1}{h^k}\sum_{j=1}^{\infty}\Big(\frac{\lambda_j}{\alpha+\lambda_j^2}\Big)^3\int_{\mathcal{W}}\int_{\mathcal{Z}}\int_{\mathcal{C}}\Big|g_j(c_1,z_1;w)-\mathbb{E}(g_j(c_1,z_1;w)|w,z_1)\Big|^3|K_h(w_1-w,w)|^3\\
&\quad\times\frac{f_{CWZ}(c_1,w_1,z_1)}{f_{WZ}^3(w,z_1)}dc_1dw_1dz_1\,\tilde{\varphi}_j^3+\text{cross product}\\
&= \frac{1}{h^k}\sum_{j=1}^{\infty}\Big(\frac{\lambda_j}{\alpha+\lambda_j^2}\Big)^3\int_{\mathcal{W}}\int_{\mathcal{Z}}\mathbb{E}\Big(\Big|g_j(c_1,z_1;w)-\mathbb{E}(g_j(c_1,z_1;w)|w,z_1)\Big|^3|w_1,z_1\Big)\times\\
&\quad|K_h(w_1-w,w)|^3\frac{f_{WZ}(w_1,z_1)}{f_{WZ}^3(w,z_1)}dw_1dz_1\,\tilde{\varphi}_j^3+\text{cross products.}
\end{aligned}$$

The cross products are dominated by the first term, then from Assumption 9 it follows that the Liapunov condition is satisfied for $\delta = 1$. By the Liapounov Central Limit Theorem (see *e.g.* Theorem 2.7.2 page 101 in Lehmann (2004))

$$\sqrt{nh^k}\frac{\frac{1}{n}\sum_{i=1}^n Z_{ni}}{\Omega(\theta,w)} \quad\Rightarrow\quad \mathcal{N}(0,1)$$

where

$$\begin{aligned}
\Omega(\theta,w) &= \sum_{j=1}^{\infty}\frac{\lambda_j^2}{(\alpha+\lambda_j^2)^2}Var\Big[\frac{(g_j(c_1,z_1;w)-\mathbb{E}[g_j(c_1,z_1;w)|w,z_1])}{f_{WZ}(w,z_1)}K_h(w_1-w,w)\Big]\,\tilde{\varphi}_j^2\\
&\quad+2\sum_{j<l}^{\infty}\frac{\lambda_j\lambda_l}{(\alpha+\lambda_j^2)(\alpha+\lambda_l^2)}Cov\Big(\frac{(g_j(c_1,z_1;w)-\mathbb{E}[g_j(c_1,z_1;w)|w,z_1])}{f_{WZ}(w,z_1)}K_h(w_1-w,w),\\
&\quad\frac{(g_l(c_1,z_1;w)-\mathbb{E}[g_l(c_1,z_1;w)|w,z_1])}{f_{WZ}(w,z_1)}K_h(w_1-w,w)\Big)\,\tilde{\varphi}_j\tilde{\varphi}_l.
\end{aligned}$$

Finally, we have to consider the bias term $\mathcal{P}_c f_{\theta|W}^{\alpha} - f_{\theta|W}^{\dagger c}$. This term is of order $\alpha^{\beta/2}$ and disappears after multiplication by $\sqrt{nh^k}$ if $\alpha$ and $h$ are properly chosen. Moreover, $h$ must be selected so that $\sqrt{nh^k}h^{\rho_c}\to 0$ and $\sqrt{nh^k}h^{\rho_z}\to 0$. The first result is guaranteed if $\alpha \asymp n^{-\frac{2\rho\epsilon}{(\beta+1)(2\rho+k)}}$ while the second result is guaranteed if $h \asymp n^{-\frac{1}{2\rho+k}+\varepsilon_h}$, with $\varepsilon_h < 0$ if $(\rho_z\wedge\rho_c) > \rho$ and with $\varepsilon_h < 0$ chosen such that $-\varepsilon_h > \frac{2(\rho-(\rho_z\wedge\rho_c))}{(2\rho+k)(2(\rho_c\wedge\rho_z)+k)}$.

## 7.8 Proof of Theorem 4

Since $||P_c|| \leq 1$ we have:

$$\mathbb{E}||\mathcal{P}_c\hat{f}_{\theta|W}^s - f_{\theta|W}^{\dagger c}||^2 = \mathbb{E}||\mathcal{P}_c(\hat{f}_{\theta|W}^s - f_{\theta|W}^{\dagger})||^2 \leq \mathbb{E}||\mathcal{P}_c||^2\mathbb{E}||\hat{f}_{\theta|W}^s - f_{\theta|W}^{\dagger}||^2 \leq \mathbb{E}||\hat{f}_{\theta|W}^s - f_{\theta|W}^{\dagger}||^2.$$

Then, we apply Theorems 4.3 and 5.2 in Nair *et al.* (2005) with $\mathbb{E}||\hat{f}_{C|WZ} - f_{C|WZ}||^2 = \delta$ in their notation. We refer to this paper for the proof.

### 7.9 Proof of Theorem 5

*(i)* See proof of Theorem 2.4 in Neubauer (1988).

*(ii)* See proof of Theorem 2.7 in Neubauer (1988).

# References


[Attanasio and Weber (2010)] Attanasio and Weber (2010), "Consumption and saving: models of intertemporal allocation and their implications for public policy," NBER Working Paper.

[Blundell *et al.* (2007)] Blundell, R., Chen, X. and Kristensen, D. (2007). *Semi-nonparamtric IV estimation of shape-invariant engel curves.* Econometrica, Vol. 75, **6**, 1613-1669.

[Bonhomme (2011)] Bonhomme, S. (2011). Functional differencing. Preprint.

[Campo *et al.* (2009)] Campo, S., Guerre, E., Perrigne, I. and Q., Vuong (2009). Semiparametric estimation of first-price quctions with risk averse bidders. *Review of Economic Studies*, forthcoming.

[Carrasco and Florens (2000)] Carrasco, M. and J.P., Florens (2000), 'Generalization of GMM to a continuum of moment conditions', *Econometric Theory*, **16**, 797-834.

[Carrasco and Florens (2010)] Carrasco, M. and J.P., Florens (2010). *Spectral Method for Deconvolving a Density*, Econometric Theory, **27**, 546-581.

[Carrasco *et al.* (2007)] Carrasco, M., Florens, J.P., and E., Renault (2007). *Linear inverse problems in structural econometrics: estimation based on spectral decomposition and regularization*, in: J., Heckman and E., Leamer, (Eds.), Handbook of Econometrics, Vol.6B, 5633-5751. Elsevier, North Holland.

[Chen and Reiss (2010)] Chen, X. and M., Reiss, (2010). *On rate optimality for ill-posed inverse problems in econometrics, working paper.* Econometric Theory, **27**, 497-521.

[Chen *et al.* (2011)] Chen, X., Chernozhukov, V., Lee, S. and W.K., Newey (2011). Local identification of nonparametric and semiparametric models, Cowles foundation discussion paper No. 1795.

[Darolles *et al.* (2011)] Darolles, S., Fan, Y., Florens, J.P., and E., Renault (2011). *Nonparametric instrumental regression.* Econometrica, **79**, 1541-1565.

[1] Deaton, A., (1993), Understanding Consumption.

[Engl *et al.* (2000)] Engl, H.W., Hanke, M. and Neubauer, A. (2000). *Regularization of inverse problems*, Kluwer Academic, Dordrecht.

[Escancianoy and Hoderlein (2010)] Escanciano, Juan Carlos, and Hoderlein, Stefan (2010), "Nonparametric Identification of Euler Equations", unpublished manuscript.

[2] Evdokimov, K. (2009). ""Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity", Yale University, Working paper.

[Florens (2003)] Florens, J.P., 2003, Inverse problems and structural econometrics: the example of instrumental variables. Invited Lectures to the World Congress of the Econometric Society, Seattle 2000. In: M., Dewatripont, L.-P., Hansen, and S.J., Turnovsky, (Eds.), Advances in Economics end econometrics: theory and applications, Vol.II, pp. 284-311. Cambridge University Press.

[Florens *et al.* (2010)] Florens, J.P., Johannes, J. and S., Van Bellegem (2010). *Identification and estimation by penalization in nonparametric instrumental regression.* Econometric Theory, **27**, 472-496.

[Florens *et al.* (1990)] Florens, J.P., Mouchart, M. and J.M., Rolin (1990). *Elements of Bayesian statistics.* Dekker, New York.

[Gajek (1986)] Gajek, L. (1986). *On improving density estimators which are not bona fide functions.* Annals of Statistics.

[Gautier and Kitamura (2010)] Gautier, E. and Y. Kitamura (2010). Nonparametric estimation in random coefficients binary choice models. Preprint.

[Hall (1978)] Hall, Robert E, 1978. "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, 86(6), 971-87.

[Hall and Horowitz (2005)] Hall, P. and J., Horowitz (2005). *Nonparametric methods for inference in the presence of instrumental variables.* Annals of Statistics **33**, 2904-2929.

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

[Henry *et al.* (2011)] Henry, M., Kitamura, Y. and B. Salanié (2011). Identifying finite mixtures in econometric models. Preprint.

[Hoderlein *et al.* (2010)] Hoderlein, S., Klemelae, J. and Mammen, E. (2010), "Analyzing the Random Coefficient Model Nonparametrically," forthcoming *Econometric Theory.*

[Hong and Shum (2009)] Hong, H., and Shum, M. (2009) "Pairwise-Difference Estimation of a Dynamic Optimization Model," *Review of Economic Studies*, 77(1): 273-304.

[Hu and Schennach (2008)] Hu, Y. and S., Schennach (2008). 'Instrumental variable treatment of nonclassical measurement error models'. *Econometrica*, **76**, 195-216.

[Johannes *et al.* (2010)] Johannes, J., Van Bellegem, S. and Vanhems, A. (2010). *Convergence Rates for ill-posed inverse Problems with an unknown operator.* Econometric Theory, **27**, 522-545.

[Kato (1995)] Kato, T. (1995), 'Perturbation theory for linear operators', Springer.

[3] Keane, Michael P & Wolpin, Kenneth I, (1997). 'The Career Decisions of Young Men,' *Journal of Political Economy*, 105, 473-522.

[Kress (1999)] Kress, R. (1999). *Linear integral equation*, Springer.

[Lehmann (2004)] Lehmann, E.L. (2004). *Elements of Large-Sample Theory*, Springer.

[Mair (1994)] Mair, B.A. (1994). *Tikhonov regularization for finitely and infinitely smoothing operators*, SIAM J. Math.Anal., **25**, 135-47.

[Mandelbaum (1987)] Mandelbaum, A. and Rüschendorf, L. (1987). *Complete and symmetrically complete families of distributions*, The Annals of Statistics, Vol. 14, No.3, 1229-1244.

[Matzkin (2003)] Matzkin, R. (2003). *Nonparametric Estimation of Nonadditive Random Functions*, Econometrica.

[Nair *et al.* (2005)] Nair, M.T.,Pereverzev, S.V. and Tautenhahn, U. (2005). "Regularization in Hilbert scales under general smoothing conditions", *Inverse Problems*, 21, 1851-1869.

[Neubauer (1986)] Neubauer, A. (1986). *Tikhonov regularization of ill-posed linear operator equations on closed convex sets.* PhD thesis, Johannes Kepler Universität Linz, 1985, appeard in Verlag der Wissenschaftlichen Gesellschaften Österreichs, Wien.

[Newey and Powell (2003)] Newey, W.K. and J.L., Powell, (2003). *Instrumental variable estimation of nonparametric models.* Econometrica, Vol.71, **5**, 1565-1578.

[Serfling (1980)] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley.

[Teicher (1961)] Teicher, H. (1961). 'Identifiability of Mistures', *Ann. Math. Stat.*, **32**, 244-248.

[Tikhonov(1963)] Tikhonov, A.N. (1963). *Regularization of incorrectly posed problems*, Soviet Math. Dokl. 4, 1624–1627.