Irregular Identification of Structural Models with Nonparametric Unobserved Heterogeneity

Juan Carlos Escanciano*

Universidad Carlos III de Madrid

September 8th, 2020

Abstract

One of the most important empirical findings in microeconometrics is the pervasiveness of heterogeneity in economic behaviour (cf. Heckman 2001). This paper shows that distribution functions and quantiles of the nonparametric unobserved heterogeneity have an infinite efficiency bound in many structural economic models of interest. The paper presents a novel and relatively simple check of this fact. The usefulness of the theory is demonstrated by showing irregular identification in several relevant examples in economics, including, among others, the proportion of individuals with severe long term unemployment duration, Average Marginal Effects (AME) in a correlated random coefficient model, bounds on average equivalent variation under endogeneity, and the distribution and quantiles of random coefficients in linear, binary and the semiparametric Mixed Logit models. In particular, it is shown that the commonly used monotonicity assumption is necessary for regular identification of the AME in a model with heterogenous effects.

Keywords: Irregular Identification; Semiparametric Models; Nonparametric Unobserved Heterogeneity.

JEL classification: C14; C31; C33; C35

*Department of Economics, Universidad Carlos III de Madrid, Calle Madrid 126, Getafe 28907, Madrid, Spain. E-mail: jescanci@eco.uc3m.es. Web Page: https://sites.google.com/view/juancarlosescanciano. Research funded by the Spanish Grant PGC2018-096732-B-I00.

1 Introduction

A tenet in empirical microeconometrics research is the pervasiveness of heterogeneity in behaviour of otherwise observationally equivalent individuals (cf. Heckman 2001). This paper shows that, for a large class of structural economic models, regular identification of functionals of nonparametric unobserved heterogeneity (UH), that is, identification of these functionals with a finite efficiency bound, implies certain *necessary* smoothness conditions on the functional, leading to a novel and practically simple check for regularity (or lack thereof). In particular, this paper uses these implications to show that cumulative distribution functions (CDFs) and quantiles of UH often have infinite efficiency bounds in many empirically relevant economic models with nonparametric UH. These results have important practical implications, as these parameters are relevant for policy analysis, and they explain why any inferences on such parameters are expected to be unstable in empirical work. In particular, if a parameter is irregularly identified, then no regular estimator with a parametric rate of convergence exists (see Chamberlain 1986).

The parameters (functionals) we consider are of interest in their own. For example, labour economists are interested in the proportion of individuals at risk of severe long term unemployment, and more generally, social scientists are interested in evaluating the effects of treatments and policy interventions (e.g. average marginal effects and average signs). The functionals that we entertain, such as CDFs and quantiles of UH, are also used as inputs in subsequent counterfactual exercises. Our research limits the kind of inferences that are attainable on these parameters by *any* method in models where UH is nonparametric.

These observations are applicable to a wide class of models with nonparametric UH. We consider first continuous mixtures, which have been commonly employed as a modeling device to account for UH in a variety of economic settings ranging from labour to industrial organization; see Compiani and Kitamura (2016) for a recent review. The canonical example is a tightly specified structural *parametric* model that is made flexible by allowing all (or a subset) of parameters to be individual specific, thereby accounting for UH. We show that if the mapping from the individual specific parameters to the conditional likelihood is smooth, then there will be many functionals of UH that will not be regularly identified. Heuristically, smoothness of the conditional likelihood translates into a multicollinearity problem, as we further explain below. There are important economic applications that fall under this setting, see, e.g., Heckman and Singer (1984a, 1984b) for the study of unemployment duration. We illustrate the usefulness of these results in the context of duration data by establishing an infinite efficiency bound for the distribution and quantiles of UH in the structural model of unemployment duration with two spells and nonparametric UH recently proposed by Alvarez, Borovicková and Shimer (2016). Duration models with nonparametric UH are often specified with conditional likelihoods that are very smooth as a function of UH (see, e.g., mixed proportional hazard models). The results on the model in Alvarez et al. (2016) are thus illustrative of a wide class of problems giving rise to “smooth” models for which our impossibility results apply.

Extending the efficiency bound results from “smooth” models to Random Coefficients (RC) models poses some significant technical challenges, because these models have discontinuous conditional likelihoods given UH. The most interesting result we provide is for the linear RC model, which forms the basis for the irregular identification of the Average Marginal Effect (AME) and the proportion of individuals with a positive marginal effect in a correlated RC model. We show that the well-known monotonicity assumption of first stages is necessary for the regular identification of the AME with nonparametric UH. These results are expected to hold in more complex models, such as simultaneous equation models, using the same methods proposed here.

The models treated up to this point are indexed by the distribution of UH, and only by that distribution. However, a simple and powerful observation of this paper is that our analysis can be trivially extended to more complex semiparametric models indexed by UH and additional (possibly infinite-dimensional) parameters. We illustrate this point with several examples, including semiparametric mixture models where some parameters are fixed and others are random. A leading example is the popular RC Logit or Mixed Logit model, which is one of the most commonly used models in applied choice analysis. This model was introduced by Boyd and Mellman (1980) and Cardell and Dunbar (1980), and it is widely used in environmental economics, industrial economics, marketing, public economics, transportation economics and other fields. Applying our results to this model we obtain an infinite efficiency bound for CDFs and quantiles of the RC. The Mixed Logit example nicely illustrates the simplicity of our method of proof. This should be contrasted with direct efficiency bounds calculations, which are the standard approach in the literature and are particularly challenging for this model (or for any of the models we consider for that matter). These results have practical implications for proposed estimators of the Mixed Logit model (see, e.g., Bajari, Fox and Ryan, 2007).

Another example of semiparametric model that we provide is to a bound on average equivalent variation, as in Hausman and Newey (2016). Existing results have focused on regular identification for this functional, see, e.g., Santos (2011) and Chernozhukov, Escanciano, Ichimura, Newey, and Robins (2018). When prices are endogenous and the demand is estimated by nonparametric instrumental variables methods, we show that the bound on average equivalent variation becomes an irregular functional under mild conditions. Further illustrations demonstrating the utility of our results in semiparametric settings are gathered in an Appendix and include a canonical model of infectious diseases with UH and measurement error models with two measurements identified by means of Kotlarski’s lemma.

What can be done to obtain regular identification of CDFs and quantiles of UH in these models? We show that functional form assumptions that restrict the conditional likelihood of observables given heterogeneity do not generally help for the purpose of achieving regularity of quantiles and CDFs if UH is still nonparametric and the conditional likelihood is smooth in UH. Thus, our results show that restricting UH is somewhat necessary to attain finite efficiency bounds for the distribution and quantiles of UH in many of the aforementioned models. Commonly used strategies in practice, such as the use of parametric distributions for UH or considering discrete heterogeneity, indeed restore the regular identification of functionals of UH but they rely on assumptions that can be deemed too strong. Semiparametric restrictions are preferable, and we find necessary conditions for regular identification under semiparametric restrictions on UH, although we recognize that giving general primitive assumptions for regularity seems difficult. Our recommendation for inference on CDFs and quantiles of UH is to use flexible semiparametric specifications such as sieve methods; see, e.g., Shen (1997), Chen (2007), Bajari, Fox and Ryan (2007), Hu and Schennach (2008), Bester and Hansen (2007), Chen and Liao (2014), Fox, Kim and Yang (2016) and references therein, coupled with regularization (penalization) to reduce the high variance of estimates of functionals of UH when the conditional likelihood is a very smooth function of UH (as in e.g. the Mixed logit model).

The rest of the paper is organized as follows. After a literature review, Section 3 sets notation and considers the class of continuous mixtures, where the method is most transparent. This section illustrates the theoretical results in the structural model of Alvarez, Borovicková and Shimer (2016). Section 4 extends the analysis to several classes of RC models. Section 5 extends further the analysis to semiparametric models, illustrating the theory with the Mixed Logit model and the bound on average equivalent variation. Section 6 discusses different strategies, some of them considered in the literature, to regularize the estimation of CDFs and quantiles of UH. Section 7 concludes. An Appendix contains proofs of the main results, further results on nonlinear RC models and further examples.

2 Literature Review

Our paper relates to a number of studies providing sufficient conditions for nonparametric identification for the distribution of UH in the aforementioned models. See, among many others, Elbers and Ridder (1982), Heckman and Singer (1984a, 1984b) and Alvarez, Borovicková and Shimer (2016) for structural models of unemployment duration, Beran and Hall (1992), Beran, Feuerverger and Hall (1996), and Hoderlein, Klemela and Mammen (2010) for linear RC, Ichimura and Thompson (1998), Gautier and Kitamura (2013) and Hoderlein and Sherman (2015) for binary RC, Briesch, Chintagunta and Matzkin (2010) and Fox, Kim, Ryan and Bajari (2012) for RC multinomial choice models, Hoderlein, Holzmann and Meister (2017) for triangular RC models, Masten (2017) for simultaneous RC models, and Lewbel and Pendakur (2017) for nonlinear RC models. For a review of nonparametric identification results see Matzkin (2007, 2013) and Lewbel (2019). What differentiates our paper from these and other related studies is our focus on establishing whether identification is regular or not.

Establishing an infinite efficiency bound for functionals of UH in these models is a priori a rather challenging task and was much unexplored prior to this paper, with the exception of the classical deconvolution problem for which rates (e.g. Fan 1991, Dattner et al. 2011) and information bounds (Trabs 2015) have been derived. Parametric rates of convergence for CDFs, and even functional weak convergence, have been established in Söhl and Trabs (2012). As shown in Dattner et al. (2011), the attainability of parametric rates depends on a delicate trade off between smoothness of the error's distribution and the smoothness of the parameter space. Nonparametric rates for CDFs are often obtained when the error's distribution is unknown, see e.g. Adusumilli et al. (2020), although presumably, by limiting the smoothness of the error's distribution parametric rates might be attainable. Beyond the classical deconvolution not much is known about efficiency bounds for CDFs and quantiles of UH in RC and related models.

The standard approach in the literature for obtaining efficiency bounds consists in characterizing the so-called tangent space of the model, the efficient score and the Fisher information, see Newey (1990) for an accessible review of efficiency bounds and some of the related concepts, and see, e.g., Chamberlain (1986), Severini and Tripathi (2001) and Khan and Tamer (2010) for further illustrations. Fisher informations are hard to compute in RC and related models, which explains the lack of theoretical work on semiparametric efficiency bounds in these models. Our method of proof avoids the complications in directly computing the tangent space and the Fisher information, and it is relatively much simpler to apply. The basic tool is a dominated convergence theorem, with regularity conditions that are easy to check in many models (although not in all models).

The starting point of our research is a fundamental result by van der Vaart (1991), who found a necessary condition for regular estimation of a parameter. The main observation and novel contribution of our paper consists in systematically exploiting the implications that van der Vaart's (1991) necessary condition has on the smoothness of certain influence functions. Prior to our paper, van der Vaart (1991), Groeneboom and Wellner (1992) and Bickel, Klassen, Ritov and Wellner (1998) have used the necessary condition of van der Vaart (1991) to show that CDFs are irregularly identified in some specific *univariate* exponential and uniform mixture models. See Section 3.2 for a detailed comparison with the irregularity result in van der Vaart (1991). Relative to this work, our contribution is to derive sufficient conditions for a general method of proof, thereby extending the scope of applications to models of economic interest. In particular, we allow for multidimensional UH, semiparametric models and non-smooth conditional likelihoods such as those that arise with RC models. The general method of proof and the applicability to RC models are novel to this paper, and have not been studied in the aforementioned references.

Although not the focus of this paper, a large class of models for which our results are applicable are panel data models with fixed effects. Within this setting, Chamberlain (1992) established regular identification of the AME in a linear RC panel data model, while Arellano and Bonhomme (2012) showed the identification of the full distribution of UH in a model with limited serial dependence in errors. Graham and Powell (2012) pointed out the irregular identification of the AME when regressors exhibit little variation across periods, while Bonhomme (2011) derived conditions for regular and irregular identification of moments of UH in nonlinear panel data. Our research is highly complementary to these papers, as we consider different models and our approach for proving irregular identification is different and exploits the smoothness implications of regular identification.

Also related are the impossibility results obtained in Hirano and Porter (2012) for non-differentiable functionals. The settings and problems investigated are, however, quite different. They study some important limitations on inference derived from non-differentiability, while we show infinite efficiency bounds implied by the lack of smoothness in influence functions for functionals of nonparametric UH. Both papers point out to important limitations on inference for structural parameters of interest.

3 Basic Setting and Results

Let $\{(Z_i, \alpha_i)\}_{i=1}^n$ denote an independent and identically distributed (iid) sample with the same distribution as (Z, α). The observed data is $Z_1, ..., Z_n$, while α_i denotes the i-th individual's UH. Assume each observation Z_i has a probability $\mathbb{P}$ and a density with respect to (wrt) a $\sigma-$finite measure μ given by

$$f_{\eta_0}(z) = \int_{\mathcal{A}} f_{z/\alpha}(z) d\eta_0(\alpha), \tag{1}$$

where $f_{z/\alpha}(z)$ denotes the known conditional density of Z given α, and η_0 is the unknown distribution of α with support on $\mathcal{A} \subseteq \mathbb{R}^{d_\alpha}$ (the results can potentially be extended to abstract heterogeneity spaces, but for simplicity of exposition we focus on the Euclidean case). The assumption of known conditional density $f_{z/\alpha}(z)$ is relaxed in Section 5.

Suppose we are interested in estimating a moment of UH,

$$\phi(\eta_0) = \mathbb{E}_{\eta_0}[r(\alpha)],$$

for a measurable function $r(\cdot) \in L_2(\eta_0)$, where, henceforth, $\mathbb{E}_{\eta_0}$ denotes the expectation under the distribution η_0 and $L_p(\nu)$ denotes the space of (equivalence classes of) real-valued measurable functions h such that $\int |h|^p \, d\nu < \infty$, for a generic measure ν. Henceforth, we drop the sets of integration in integrals and the qualification $\nu-$almost surely for simplicity of notation. So, for example, a function in $L_2(\nu)$ is discontinuous when there is no continuous function in its equivalence class. Also, we drop the reference to the measure ν in $L_2(\nu)$ when $\nu = \mathbb{P}$, and write simply L_2. We will be concerned with regular identification of $\phi(\eta_0)$, i.e. identification of $\phi(\eta_0)$ with a finite efficiency bound, when UH is nonparametric as formally defined below.

The basic message of this paper is based on two observations. First, from a general result in van der Vaart (1991), we prove that a necessary condition for regular identification of $\phi(\eta_0)$ when UH is nonparametric is the existence of a measurable function $s(Z)$ with zero mean and finite variance such that

$$r(\alpha) - \phi(\eta_0) = \int s(z) f_{z/\alpha}(z) d\mu(z). \tag{2}$$

Second, if the mapping $\alpha \to f_{z/\alpha}$ is continuous (smooth), then under mild regularity conditions, (2) implies that $r(\cdot)$ must be also continuous (smooth). The main contribution of this paper is a formalization of the second observation and its application to some economic models of interest.

The precise sense of UH being nonparametric is the usual one, formalized as follows. Let H denote a class of distributions on $\mathcal{A}$, and assume $\eta_0 \in H$. Let $\eta_t \in H$ be a parametric submodel indexed by $t \in [0, \varepsilon)$, for some $\varepsilon > 0$, such that for a $b \in L_2(\eta_0)$ the classical mean square differentiability condition holds,

$$\int \left[\frac{d\eta_t^{1/2} - d\eta_0^{1/2}}{t} - \frac{1}{2} b d\eta_0^{1/2} \right]^2 \to 0 \text{ as } t \downarrow 0. \tag{3}$$

Often the score function b can be simply computed as $b = \partial \log d\eta_t / \partial t$, i.e. the score associated to the parametric submodel η_t at the "truth" $t = 0$ (corresponding to η_0), where, henceforth, derivatives wrt to t are one-sided and evaluated at zero. Denote by $T(\eta_0)$ the linear span of the scores $b's$ in (3) and let $L_2^0(\nu)$ denote the subspace of functions in $L_2(\nu)$ with zero $\nu-$mean. For further discussion on the tangent set $T(\eta_0)$ see Newey (1990). Then, a formal definition of nonparametric UH is given as follows.

Definition 3.1 *UH is nonparametric if $T(\eta_0)$ is dense in $L_2^0(\eta_0)$.*

This definition is a formalization of the standard assumption in the literature that UH is nonparametric. Heuristically, it means that UH is essentially unrestricted. Henceforth, we assume, unless otherwise stated, that UH is nonparametric. The first result in this section shows that in the presence of nonparametric UH in model (1), regular identification of $\mathbb{E}_{\eta_0}[r(\alpha)]$ requires necessarily that (2) holds.

Lemma 3.1 *If UH is nonparametric, then (2) is necessary for regular identification of $\phi(\eta_0)$.*

Severini and Tripathi (2006, 2012) and Bonhomme (2011) have found related results in the context of nonparametric instrumental variables and nonlinear panel data models, respectively. These results and Lemma 3.1 are special cases of a more general result in van der Vaart (1991). Additionally, Escanciano (2020) has shown that (2) is not only necessary but also sufficient for regular identification of $\phi(\eta_0)$ in model (1). Note that we are not assuming here that η_0 or s in (2) are identified. This level of generality is important because these functions may not be identified in many structural economics models under weak assumptions, which does not prevent us from identifying and estimating certain functionals of them (cf. Hurwicz 1950).[1]

We now proceed with the main insight of this paper, which is that if the mapping $\alpha \rightarrow f_{z/\alpha}$ is continuous (smooth), then, under regularity conditions, $r(\cdot)$ must be also continuous (smooth). This simple observation follows by dominated convergence, and it implies non-regularity of CDFs, signs, quantiles, and other functionals of UH in "smooth models" satisfying the following assumption. Let N denote an open subset of $\mathcal{A} \subset \mathbb{R}^{d_\alpha}$.

Assumption 1 *(i) $\alpha \rightarrow f_{z/\alpha}(z)$ is continuous on N a.e-μ; (ii) for all $\alpha \in N$ there exists a neighborhood of α, say $\Gamma_0 \subset N$, such that for all s satisfying (2),*

$$\int |s(z)| \sup_{\alpha \in \Gamma_0} f_{z/\alpha}(z) d\mu(z) < \infty. \tag{4}$$

Assumption 1(i) is easy to check. Assumption 1(ii) is a dominance condition. The main complication in checking Assumption 1(ii) is that s belongs to $L_2(\mathbb{P})$ but not necessarily to $L_1(\mu)$ or $L_2(\mu)$. We verify these conditions in a number of examples below.

Lemma 3.2 *Let the conditional density $f_{z/\alpha}(z)$ satisfy Assumption 1. Then, $r(\alpha)$ in (2) is continuous in α on N.*

The following corollary is a direct consequence of the previous two lemmas.

Corollary 3.1 *Let Assumption 1 hold. The CDF $\phi(\eta_0) = \mathbb{E}_{\eta_0}[1(\alpha \leq \alpha_r)]$, for $\alpha_r \in N$, is not regularly identified.*

Moments in general and CDFs in particular are examples of linear functionals. Quantiles of UH are, in contrast, nonlinear functionals, and are thus not covered by the previous results. To extend the theory to a more general setting including nonlinear functionals we need to introduce some notation. A functional $\phi(\eta_0) : H \rightarrow \mathbb{R}$ is said to be differentiable if there exists an $r_\phi \in L_2^0(\eta_0)$ such that for all paths satisfying (3), it holds

$$\lim_{t \rightarrow 0} \frac{\phi(\eta_t) - \phi(\eta_0)}{t} = \mathbb{E}_{\eta_0}[r_\phi(\alpha) b(\alpha)]. \tag{5}$$

Under nonparametric UH such r_ϕ is unique, as in Newey (1994). This function r_ϕ is called the influence function of $\phi(\eta_0)$ and plays the role of the preceding moment function r.

[1] Of course, if η_0 is identified, so is $\phi(\eta_0)$ (since r is known). Identification of $\phi(\eta_0)$ follows from (2) because we can find an identified function $\tilde{s}(Z)$, depending only on $f_{z/\alpha}$ and r, such that $r(\alpha) = \mathbb{E}[\tilde{s}(Z)|\,\alpha]$ holds, and thus by iterated expectations $\phi(\eta_0) = \mathbb{E}_{\eta_0}[r(\alpha)] = \mathbb{E}_{\eta_0}[\mathbb{E}[\tilde{s}(Z)|\,\alpha]] = \mathbb{E}[\tilde{s}(Z)]$.

To illustrate with an example, consider the scalar UH case and assume η_0 is absolute continuous with a strictly positive Lebesgue density in a neighborhood of $\phi(\eta_0)$, where $\phi(\eta_0)$ is such that

$$\int_{-\infty}^{\phi(\eta_0)} d\eta_0(\alpha) = \tau, \ \ \tau \in (0,1). \tag{6}$$

That is, $\phi(\eta_0)$ is the τ-quantile of η_0. It is well-known that the quantile functional is differentiable under the conditions above with influence function

$$r_\phi(\alpha) = \frac{-\left\{1(\alpha < \phi(\eta_0)) - \tau\right\}}{\dot{\eta}_0(\phi(\eta_0))},$$

where $\dot{\eta}_0$ is the density pertaining to η_0. It follows from our results that the discontinuity of the influence function $r_\phi(\cdot)$ implies irregular identification. Next result, formalizes this finding.

Corollary 3.2 *Let Assumption 1 hold. Assume η_0 is absolute continuous with a strictly positive Lebesgue density in a neighborhood of $\phi(\eta_0)$ satisfying (6). If $\phi(\eta_0) \in N$, then the τ-quantile of the nonparametric UH distribution is not regularly identified.*

Remark 3.1 *Henceforth, whenever we discuss identification of quantiles, we implicitly assume that the components of UH have densities that satisfy the conditions in Corollary 3.2.*

Remark 3.2 *The quantile example illustrates that our results are also applicable to nonlinear differentiable functionals. For other nonlinear functionals the researcher needs to find r_ϕ satisfying (5) and replace r by r_ϕ in our results below.*

We discuss now the technical complications of the more standard approach in the literature of directly computing the Fisher Information or the efficiency bound. Define the so-called tangent space of scores $\mathcal{S} := \{s \in L_2^0 : s(z) = \mathbb{E}\left[b(\alpha)|\, Z\right]$ for some $b \in T(\eta_0)\}$. Then, a standard result in linear inverse problems is that all solutions s of equation (2) have the same orthogonal projection onto the closure of $\mathcal{S}$ (see Engl, Hanke and Nuebauer, 1996). Denote by s^* such orthogonal projection, the so-called efficient score. The efficiency bound is given by the variance of $s^*(Z)$ (see e.g. Newey 1990, van der Vaart 1998, Bickel et al. 1998, and Escanciano 2020). Thus, an alternative to our approach is to compute $s^*(Z)$ and check that it has infinite variance. However, computing $s^*(Z)$ can be cumbersome, particularly because characterizing the mean squared closure of $\mathcal{S}$ can be a rather difficult task in the models we analyze here. In fact, to the best of our knowledge, the analytical expression for s^* remains unknown for the functionals and models we study. In passing, we note that these arguments show that it suffices to check the dominance condition (4) for s in the closure of $\mathcal{S}$. This additional information will turn out to be quite useful in some of our applications, such as the linear RC model.

3.1 An Application To A Structural Model of Unemployment

We illustrate the applicability of the previous results in the context of a structural model of unemployment with nonparametric UH. Nonparametric heterogeneity has played a critical role in rationalizing unemployment duration ever since the seminal contributions by Elbers and Ridder (1982) and Heckman and Singer (1984a, 1984b). Recent work by Alvarez et al. (2016) is motivated from this perspective. These authors have shown nonparametric identification of the distribution of UH in their nonparametric structural model for unemployment with two spells. Specifically, Alvarez, Borovicková and Shimer (2016) propose a structural model for transitions in and out of employment that implies a duration of unemployment given by the first passage time of a Brownian motion with drift, a random variable with an inverse Gaussian distribution. The parameters of the inverse Gaussian distribution are allowed to vary in arbitrary ways to account for UH in workers. These authors investigate nonparametric identification of the distribution of UH, η_0, when two unemployment spells $Z_i = (t_{i1}, t_{i2})$ are observed on the set $\mathcal{T}^2$, $\mathcal{T} \subseteq [0, \infty)$. The reduced form parameters $\alpha = (\alpha_1, \alpha_2)' \in \mathbb{R} \times [0, \infty)$ are functions of structural parameters. The distribution of Z_i is absolutely continuous with Lebesgue density $f_{\eta_0}(t_1, t_2)$ given, up to a normalizing constant, by

$$f_{\eta_0}(t_1, t_2) = \int_{\mathbb{R}\times[0,\infty)} \frac{\alpha_2^2}{t_1^{3/2} t_2^{3/2}} e^{-\frac{(\alpha_1 t_1 - \alpha_2)^2}{2t_1} - \frac{(\alpha_1 t_2 - \alpha_2)^2}{2t_2}} d\eta_0(\alpha_1, \alpha_2). \tag{7}$$

Alvarez, Borovicková and Shimer (2016) show that η_0 is nonparametrically identified up to the sign of α_1, but they do not investigate if specific functionals of this distribution are regularly or irregularly identified, which is the focus of study here. Specifically, we show that the CDF of η_0 at a point, and other functionals of η_0 with discontinuous influence functions, such as quantiles, have infinite efficiency bounds. These functionals are important parameters. For example, $\phi(\eta_0) = \mathbb{E}_{\eta_0}[1(\alpha_1 \leq \alpha_{10}) 1(\alpha_2 \leq \alpha_{20})]$, for a fixed $\alpha_{10} < 0 < \alpha_{20}$ and large absolute values of α_{10} and α_{20}, quantifies the proportion of individuals at risk of severe long term unemployment (an individual with parameters α_1 and α_2, $\alpha_1 \leq \alpha_{10}$ and $\alpha_2 \leq \alpha_{20}$, has a probability larger or equal than $1 - \exp(2\alpha_{10}\alpha_{20})$ of remaining unemployed forever). We apply our previous results to this example for a generic moment $\phi(\eta_0) = \mathbb{E}_{\eta_0}[r(\alpha_1, \alpha_2)]$, under the following mild condition.

Assumption 2 *(i) Let the set $\mathcal{T} \subseteq [0, \infty)$ be a convex set with a non-empty interior; (ii) the moment function r is locally bounded.*

Proposition 3.1 *Under Assumption 2, if $\phi(\eta_0) = \mathbb{E}_{\eta_0}[r(\alpha_1, \alpha_2)]$ is regularly identified, then*

$$r(\cdot) \in \left\{ b(\alpha_1, \alpha_2) \in L_2^0(\eta_0) : b(\alpha_1, \alpha_2) = C_1 + C_2 \alpha_2^2 e^{2\alpha_1\alpha_2} h(\alpha_1^2, \alpha_2^2) \right\},$$

for constants C_1 and C_2 and a continuous function $h(u, v)$ defined on $(0, \infty)^2$ that, if $\mathcal{T}$ is bounded, is an infinite number of times differentiable at $u \in (0, \infty)$, for all $v \in (0, \infty)$.

For the purpose of proving an infinite efficiency bound for CDFs and quantiles only the continuity part of Proposition 3.1 suffices. Thus, an implication of Proposition 3.1 is that the CDF of UH at the fixed point $(\alpha_{10}, \alpha_{20})$, i.e. $\phi(\eta_0) = \mathbb{E}\left[1(\alpha_1 \leq \alpha_{10})1(\alpha_2 \leq \alpha_{20})\right]$, is not regularly identified because $r_\phi(\alpha_1, \alpha_2) = 1(\alpha_1 \leq \alpha_{10})1(\alpha_2 \leq \alpha_{20})$ is not continuous when $(\alpha_{10}, \alpha_{20})$ is in the interior of the support of η_0.

Corollary 3.3 *Under Assumption 2(i), the CDFs and quantiles of UH in the model (7) are not regularly identified.*

The high smoothness of the mapping $\alpha \to f_{z/\alpha}(z)$ in the model of Alvarez et al. (2016) makes inference on functionals of UH hard, and Corollary 3.3 is one way to formalize this statement for CDFs and quantiles of UH.

3.2 A comparison with van der Vaart (1991)

It is useful to compare these irregularity results for CDFs with those obtained in van der Vaart (1991). He specifically studied regular estimation of the CDF in the univariate exponential mixture model with conditional density

$$f_{z/\alpha}(z) = h(z)c(\alpha)e^{z\alpha}, \tag{8}$$

for a scalar α. He argued that "from the completeness of the exponential family, it follows readily that (with our notation)

$$\int s(z) f_{z/\alpha}(z) dz$$

cannot be constant in an open interval, unless it is constant everywhere" (cf. van der Vaart, 1991, p. 191). Since the indicator function is constant in an open interval, it follows that (2) is not satisfied, and thus, the CDF of η_0 is not regularly estimable.

The argument used in van der Vaart (1991) to establish irregularity is thus quite different from the argument used in our paper. We exploit the lack of smoothness of the influence function, rather than the completeness of the exponential family. Additionally, it should be noted that van der Vaart's (1991) irregularity argument is specific to the univariate exponential mixture model in (8) and it does not apply to the models we focus here, including the application to Alvarez et al. (2016) or the RC models that we study next.

4 Random Coefficient Models

Random coefficient models have long been used in economics to model nonparametric UH. There is by now an extensive literature on nonparametric identification of UH in these models, see, e.g., Lewbel (2019). In this paper we focus on establishing irregular identification of CDFs and quantiles of the distributions of RC (or more generally of any other functional with a discontinuous influence function). To the best of our knowledge, this is the first paper to do so.

A general class of random coefficient models, including nonlinear models, is given by

$$Y_i = m\left(X_i, \alpha_i\right), \tag{9}$$

where $Z_i = (Y_i, X_i)$ are observed, but α_i is unobserved and independent of X_i with support $\mathcal{A}$. Assume $m : \mathcal{X} \times \mathcal{A} \to \mathbb{R}^r$ is a measurable map, where $\mathcal{X}$ is the support of X. The functional form of m is known, and the nonparametric part is given by the distribution of α_i. The assumptions of known m and the independence of α_i and X_i are relaxed below. The density of the data is

$$f_{\eta_0}(y, x) = \int_{\mathcal{A}} 1\left(y = m(x, \alpha)\right) d\eta_0(\alpha),$$

where $1(A)$ denotes the indicator function of the event A. In this setting, the dominating measure μ is defined on $\mathcal{Z} = \mathcal{Y} \times \mathcal{X}$ as $\mu\left(B_1 \times B_2\right) = \nu_Y\left(B_1\right) \nu_X(B_2)$, where B_1 and B_2 are Borel sets of $\mathcal{Y}$ and $\mathcal{X}$, respectively, ν_Y is either the counting measure for discrete outcomes or the Lebesgue measure $\lambda(\cdot)$ for continuous outcomes, and $\nu_X(\cdot)$ is the probability measure for X. The main challenge we face with RC models is that $f_{z/\alpha}(z) = 1\left(y = m(x, \alpha)\right)$ is not continuous, and thus the previous results need to be generalized. The generalization is non-trivial, particularly so for continuous outcomes. The discontinuity of $\alpha \to f_{z/\alpha}$ makes regularity of functionals of UH more likely. With a decreasing level of "smoothness", and hence an increasing level of difficulty, we consider first the binary choice RC model and next the linear RC. Section 9.1 in the Appendix contains some generic results for nonlinear RC.

4.1 Binary Choice Random Coefficient

The binary choice random coefficient model is given by

$$Y_i = 1\left(X_i' \alpha_i \geq 0\right),$$

where we observe $Z_i = (Y_i, X_i)$ but α_i is unobservable. The random vector α_i is independent of X_i, normalized to $|\alpha_i| = 1$ and satisfies $\mathbb{P}\left(\alpha_i = 0\right) = 0$. As in the existing literature, we assume η_0 is absolutely continuous wrt the uniform spherical measure $\sigma(\cdot)$ in $\mathbb{S}^{d_\alpha - 1}$, where $\mathbb{S}^{d_\alpha - 1} = \{b \in \mathbb{R}^{d_\alpha} : |b| = 1\}$ denotes the unit sphere in $\mathbb{R}^{d_\alpha}$. The density of the data for a positive outcome (i.e. the choice probability function) is given by

$$f_{\eta_0}(x) = \int_{\mathbb{S}^{d_\alpha - 1}} 1\left(x' s \geq 0\right) d\eta_0(s). \tag{10}$$

Ichimura and Thompson (1998) and Gautier and Kitamura (2013) have found sufficient conditions for nonparametric identification of η_0. These authors, however, have not investigated whether identification is regular or irregular, which is the focus here.

By (10) and Lemma 3.1 a necessary condition for regular identification of $\phi(\eta_0) = \mathbb{E}_{\eta_0}[r(\alpha)]$ under nonparametric UH is

$$r(\alpha) - \phi(\eta_0) = \int 1\,(x'\alpha \geq 0)\, s(1,x) dv_X(x), \tag{11}$$

for some $s \in L_2^0$. The following result provides necessary conditions for regular identification. Write $\alpha = (\alpha_1, \alpha_2')'$.

Proposition 4.1 *If the distribution of $X/|X|$ is absolutely continuous, then $r(\cdot)$ in (11) must be uniformly continuous on $\mathbb{S}^{d_\alpha - 1}$. If $X = (1, \tilde{X})$ and $\alpha_2'\tilde{X}$ is absolutely continuous, then $r(\alpha_1, \alpha_2)$ must be an absolutely continuous function of α_1.*

An implication of this proposition is that functionals such as the CDF and quantiles of random coefficients are not regularly identified in the binary RC model. To the best of our knowledge, this result is new in the literature.

Corollary 4.1 *Under the conditions of Proposition 4.1, the CDFs and quantiles of UH in the binary RC model are not regularly identified.*

This result is hardly surprising given the fact that in the simpler model $Y_i = 1\,(X_i'\beta_0 - \alpha_i \geq 0)$ with a scalar α_i independent of X_i, the distribution of α_i is identified as $\eta_0(\alpha) = \mathbb{E}\left[Y_i | \, X_i'\beta_0 = \alpha\right].$ The full independence of α_i of X_i provide further overidentifying restrictions on η_0, but they do not lead to regular identification of η_0.

The situation in the linear RC is different and cannot be reduced to well-known irregular functionals, such as conditional means. For one thing, simpler specifications such as $Y_i = X_i'\beta_0 + \alpha_i$ do lead to regular identification of η_0. As we will see, the linear RC model is less "smooth" than the binary RC model, in a precise sense defined below, which makes regularity of functionals of UH more likely in the linear model than in the binary choice model.

4.2 Linear Random Coefficient

The linear RC model has a long history in econometrics, see, e.g., Hildreth and Huock (1968) and Swamy (1970). This model is given by

$$Y_i = X_i'\alpha_i,$$

where we observe a d_z-dimensional vector $Z_i = (Y_i, X_i)$, but α_i is unobservable and independent of X_i. The dimension of X_i and α_i is d_α, so $d_z = d_\alpha + 1$. Like in Hoderlein, Klemelä and Mammen (2010), we normalize X_i so that $|X_i| = 1$. The density of the data is

$$f_{\eta_0}(z) = \int_{\mathbb{R}^{d_\alpha}} 1\,(y = x'\alpha)\, d\eta_0(\alpha). \tag{12}$$

Nonparametric identification and estimation of η_0 has been studied by Beran and Hall (1992), Beran, Feuerverger and Hall (1996), and Hoderlein, Klemelä and Mammen (2010), among others. These authors exploit the relation between (12) and the Radon transform. In this paper we study necessary conditions for regular identification of $\phi(\eta_0) = \mathbb{E}_{\eta_0}[r(\alpha)]$, for a measurable function $r(\cdot)$ with $\mathbb{E}_{\eta_0}[r^2(\alpha)] < \infty$, and regular identification of quantiles of the components of α. To the best of our knowledge, no efficiency bounds calculations for functionals of UH are available in the literature for this model.

The discontinuity of $1(y = x'\alpha)$ may a priori suggest that many functionals of UH are regularly identified in this model. By Lemma 3.1 a necessary condition for regular identification of $\phi(\eta_0) = \mathbb{E}_{\eta_0}[r(\alpha)]$ under nonparametric UH is

$$r(\alpha) - \phi(\eta_0) = \int s(x'\alpha, x)dv_X(x), \tag{13}$$

for some $s \in L_2^0$. Under suitable conditions scores in the tangent space $\mathcal{S} = \{s \in L_2^0 : s(z) = \mathbb{E}[b(\alpha)|\, Z]$ for some $b \in T(\eta_0)\}$ are continuous, but providing conditions under which elements of the *closure* of $\mathcal{S}$ are continuous is a much harder task. In fact, without additional restrictions, elements in the closure of $\mathcal{S}$ can be potentially very discontinuous (cf. Smith, Solmon and Wagner 1977). We shall provide regularity conditions below that guarantee that any element of the closure of $\mathcal{S}$ can be written as

$$s(z) = \frac{g(z)}{f_{\eta_0}(z)}, \tag{14}$$

where $g(z)$ has an squared integrable weak derivative with respect to the first argument y in $z = (y, x)$. As we show below, the representation in (14) will be instrumental for checking the sufficient conditions for the dominated convergence theorem in Lemma 3.2.

Let $\eta_{0,x}$ denote the Lebesgue density of $x'\alpha$ when α has distribution η_0. The set $\eta_0 T(\eta_0)$ is defined as $\eta_0 T(\eta_0) := \{\eta_0 b : b \in T(\eta_0)\}$, while the definition of a Sobolev space $H^{\rho_0}(\mathcal{A})$ is provided after (26) in the Appendix. The index ρ_0 quantifies the degree of smoothness (with higher ρ_0 corresponding to higher smoothness).

Assumption 3 *For $d_\alpha > 1$ and N as in Assumption 1: (i) the density of the distribution η_0 is bounded, has bounded support, with a corresponding density $\eta_{0,x}$ that is continuous and satisfies $\inf_{\alpha \in N} \eta_{0,x}(x'\alpha) \geq 1/l(x)$ for a positive measurable function $l(\cdot)$ such that $\mathbb{E}_X[l^2(X)] < \infty$; (ii) X is absolutely continuous with a bounded density $f_X(\cdot)$; (iii) $\eta_0 T(\eta_0) \subseteq H^{\rho_0}(\mathcal{A})$, where $\rho_0 + (d_\alpha - 1)/2 > 2$; (iv) r belongs to the closure of $T(\eta_0)$.*

The bounded support of Assumption 3(i) is often considered in the literature, see, e.g., Hoderlein, Klemelä and Mammen (2010). If the infinite efficiency bound holds in a model with bounded support of α it also holds in the more general model where the support is unrestricted. A sufficient condition for the continuity of $\eta_{0,x}$ is that the Fourier transform of the density of η_0

is integrable, which was also assumed in Hoderlein, Klemelä and Mammen (2010). Assumptions 3(i-ii) establish a link between the tails of η_0 and $f_X(\cdot)$. Assumption 3(iii) imposes a mild smoothness condition on the tangent space of UH. This assumption and Assumption 3(iv) allow but do not require nonparametric UH.

Proposition 4.2 *Under Assumption 3 and if r satisfies (13), then it must be continuous on N.*

Corollary 4.2 *Under the conditions of Proposition 4.2, the CDFs and quantiles of UH are not regularly identified in the linear RC model.*

Without exploiting the specific structure of the closure of the tangent space (i.e. 14) Proposition 4.2 and Corollary 4.2 may not hold. This stands in contrast to other RC models, such as the binary RC model, where the continuity of the integral in (2) holds for any s in L_2.

4.3 Correlated Random Coefficients: AME

The independence assumption between regressors and UH rules out important models and parameters in economics, such as the Average Marginal Effect (AME) $\phi(\eta_0) = \mathbb{E}_{\eta_0}[\gamma_i]$ and the Proportion of individuals with a Positive AME (PPAME), $\phi(\eta_0) = \mathbb{E}_{\eta_0}[1(\gamma_i > 0)]$, where γ_i is the coefficient of an endogenous continuous variable in a RC triangular system. We extend our previous results to these cases. We will show that under nonparametric UH these important parameters are not regularly identified. These results appear to be new in the literature under this generality. For simplicity, we focus on a triangular model, but the same arguments are potentially applicable to a wide class of random coefficient models, including simultaneous equation models, nonlinear models with endogeneity, or variations of these models that include covariates, multiple endogenous variables, and mixed random and non-random coefficients.

Consider the triangular model:

$$Y_1 = \gamma Y_2 + U_1, \qquad Y_2 = \delta X + U_2, \tag{15}$$

where γ, U_1, δ and U_2 are RC, and we observe $Z = (Y_1, Y_2, X)'$. The variable Y_2 is a continuous treatment variable, possibly endogenous, in the sense that U_1 and U_2 are correlated, and X is an instrument, independent of all the random coefficients. Suppose, the researcher is interested in the AME $\phi(\eta_0) = \mathbb{E}_{\eta_0}[\gamma]$ or the PPAME $\phi(\eta_0) = \mathbb{E}_{\eta_0}[1(\gamma > 0)]$. We will provide conditions under which both parameters have an infinite efficiency bound. To see this, we obtain the reduced forms

$$Y_1 = \gamma\delta X + \gamma U_2 + U_1 \equiv \pi_1 X + \pi_0,$$
$$Y_2 = \delta X + U_2,$$

which, with some abuse of notation, are jointly written as $Y = \alpha_0 + \alpha_1 X$, where $Y = (Y_1, Y_2)'$, $\alpha = (\alpha_0, \alpha_1)$, $\alpha_0 = (\pi_0, U_2)'$ and $\alpha_1 = (\pi_1, \delta)'$. Proposition 4.2 can then be applied to the reduced form. Because the corresponding influence functions for the AME and PPAME are $r_{AME}(\alpha) = \pi_1/\delta$ and $r_{PPAME}(\alpha) = 1(\pi_1 > 0)1(\delta > 0) + 1(\pi_1 < 0)1(\delta < 0)$, respectively, and they are discontinuous functions of $\alpha_1 = (\pi_1, \delta)'$, non-regularity follows from Proposition 4.2. Consider the following assumption. Let N be an open set in the interior of $\mathcal{A}$, the support of the vector of reduced form random coefficients α.

Assumption 4 *(i) Assumption 3 holds for the reduced form $Y = \alpha_0 + \alpha_1 X$; (ii) X is independent of the random coefficients $(\gamma, U_1, \delta, U_2)$; (iii) $(p_0, u_2, 0, d_0) \in N$ for some (p_0, u_2, d_0); (iv) $(p_0, u_2, p_1, 0) \in N$ for some (p_0, u_2, p_1).*

Assumption 4(iv) means that zero is an interior point in the support for the first stage effect δ, i.e. a lack of monotonicity.

Proposition 4.3 *Suppose (15) and Assumption 4(i-ii) holds. If in addition Assumption 4(iii) or Assumption 4(iv) holds, then the PPAME is not regularly identified. If Assumption 4(iv) holds and $\mathbb{E}[\gamma^2] < \infty$, then the AME is not regularly identified.*

Proposition 4.3 proves non-regularity for the AME and the PPAME. The condition $\mathbb{E}[\gamma^2] < \infty$ ensures that the AME is a continuous functional in $L_2(\eta_0)$. If f_{δ^2} denotes the (Lebesgue) density of δ^2 and $h(u) = \mathbb{E}[\pi_1^2 | \delta^2 = u] f_{\delta^2}(u)$, then a sufficient condition for $\mathbb{E}[\gamma^2] < \infty$ is $\lim_{u \to 0^+} h(u)/u^\rho < \infty$ for some $\rho > 0$ and $\mathbb{E}[\pi_1^2] < \infty$; see Khuri and Casella (2002, pg. 45).

Intuitively, non-regularity of the AME comes from the presence of a set of individuals with near-zero first-stage effects (Assumption 4(iv)), although $\mathbb{P}(\delta = 0) = 0$. When the instrument satisfies a monotonicity restriction, in the sense that $\mathbb{P}(\delta > 0) = 1$ or $\mathbb{P}(\delta < 0) = 1$, then regular identification of the AME might be possible. An interpretation of Proposition 4.3 is that the monotonicity condition is necessary for regular identification of the AME with nonparametric UH. Indeed, Heckman and Vytlacil (1998) and Wooldridge (1997, 2003, 2008) show that with homogenous first-stage effects regular estimation by IV methods holds. See also Florens et al. (2008), Masten and Torgovitsky (2016), and the extensive literature following the seminal contributions by Imbens and Angrist (1994) and Heckman and Vytlacil (2005) for identification results on conditional and weighted AME or their discrete versions.

The PPAME is non-regular under more general conditions than the AME. The difficulty of identification of the PPAME is well recognized in the literature. Heckman, Smith and Clements (1997) provide bounds for the analog to PPAME in the binary treatment case, and identification when gains are not anticipated at the time of the program.

The literature on nonparametric identification in the model (15) is relatively scarce. In important work, Masten (2017, Proposition 4) gives conditions for nonparametric identification of the distribution of γ, but he did not discuss efficiency bounds for the AME or the PPAME under his conditions. For different models, Khan and Tamer (2010) and Graham and Powell (2012) show irregularity of the AME when $\mathbb{E}[\gamma^2] = \infty$. The nature of irregularity of the AME documented here is different and it holds in a setting where $\mathbb{E}[\gamma^2] < \infty$. The lack of monotonicity of first stages results in a discontinuity at zero for the influence function of the AME. This result formalizes the importance of monotonicity assumptions in reliable estimation of the AME in models with nonparametric heterogenous effects.

4.4 Are CDFs and quantiles of UH always irregular in RC models?

Whether CDFs and quantiles of UH are regular or irregularly identified depends in a very delicate way on the model m and the distribution of regressors. To illustrate these issues, we discuss two examples where regular identification for CDFs is expected. Consider the canonical monotonic, possibly nonseparable, model

$$Y_i = m(X_i, \alpha_i)$$

with scalar UH α_i and where $\alpha \to m(x, \alpha)$ is strictly increasing with inverse $m^{-1}(y, x)$. Then, if we define $s(Y_i, X_i) = 1(m^{-1}(Y_i, X_i) \leq 0)$, then the regularity condition of Lemma 3.1 is satisfied with $r(\alpha) = 1(\alpha \leq 0)$, provided both s and r have certain finite moments, proving that the necessary condition for regular identification of the CDF at 0 (or at any other point in fact) holds. In invertible models like this, regular identification of CDFs and quantiles is satisfied under mild conditions. See the Appendix for a specific example showing which one of our sufficient conditions for irregularity are not satisfied in this setting.

Alternatively, if the regressors are discrete, then identification will be regular. By way of example, consider the binary choice RC model. We note that the necessary condition for regular estimation (11) is valid whether or not the distribution of X is continuous, discrete or mixed, although Proposition 4.1 focuses on the continuous case. Suppose now that X is discrete with finite support $\mathcal{X} = \{x_1, ..., x_J\}$. Then, $\phi(\eta_0) = \mathbb{E}_{\eta_0}[r(\alpha)]$ is identified in the binary choice RC if $r(\alpha)$ belongs to the span of $\left\{1\left(x_j'\alpha \geq 0\right)\right\}_{j=1}^{J}$, and if so, identification is regular. For example, if the the $k-th$ canonical vector $e_k = (0, ..., 1, ...0)'$ with a 1 in the $k-th$ component belongs to the support of $\mathcal{X}$, then the average survival function of the $k-th$ component of α at zero, i.e. $\mathbb{E}_{\eta_0}[1(\alpha_k \geq 0)]$, will be regularly identified (indeed $\mathbb{E}_{\eta_0}[1(\alpha_k \geq 0)] = \mathbb{E}[Y_i | X_i = e_k]$). Thus, discrete regressors make regular identification of CDFs and quantiles more likely.

On the contrary, if X is continuous and the mapping $\alpha \to f_{z/\alpha}(z)$ is sufficiently smooth, regular identification of CDFs and quantiles may not be possible unless we impose strong assumptions on UH. For example, for the RC binary choice model, Gautier and Kitamura (2013, Proposition 3.1) have shown that the right hand side of (2) necessarily belongs to $H^\rho(\mathcal{A})$ for $\rho = d_\alpha/2$, and by Sobolev embedding (since $\rho > (d_\alpha - 1)/2$), the moment function $r(\cdot)$ must be a continuous function. They use this feature to show that the RC binary model imposes restrictions on the density of observables, while we use it here to show irregularity of CDFs and quantiles.

More generally, regular identification of CDFs and quantiles can be obtained when the mapping $\alpha \to f_{z/\alpha}(z)$ is not very smooth and the distribution of UH satisfies a certain level of smoothness, as illustrated by Dattner et al. (2011) and Söhl and Trabs (2012) for the classical convolution problem. The situation is more involved for other models such as the linear RC model, as we discuss now. Smith, Solmon and Wagner (1977) have shown that the closure of the range of the Radom transform operator

$$Rb(z) = \int 1\left(y = x'\alpha\right) b(\alpha) d\alpha,$$

as a mapping in $L_2(\lambda)$ contains any even function of the Sobolev space $H^{\rho}(\mathcal{Z})$, where $\mathcal{Z}$ is the support of $Z = (Y, X)$ and $\rho = (d_\alpha - 1)/2$. For the binary RC model $\rho = d_\alpha/2$ for densities on a $(d_\alpha - 1)-$dimensional space (cf. Gautier and Kitamura, 2013), while for the linear RC $\rho = (d_\alpha - 1)/2$, for densities on a $d_\alpha-$dimensional argument Z. It is in this precise sense that we say the binary RC model is smoother than the linear RC model. These arguments imply fundamental differences between the binary and linear RC models when it comes to regular identification of functionals of UH. In Proposition 4.2 we provide sufficient conditions for irregular identification, but we leave open the possibility of regular identification of CDFs and quantiles when these conditions are not satisfied. Such analysis is beyond the scope of this paper and is deferred to future research.

5 Extension to Semiparametric Models

This section extends our results to semiparametric models. The main point is as follows, if a functional is irregular in a model, it will be irregular in a larger model that nests the original model as a special case. Information can only decrease (or remain the same) when we know less. This basic observation has important implications, and it widens substantially the applicability of our previous results, as we illustrate with two examples and with further examples in the Appendix.

5.1 The Mixed Logit Model

Consider first a conditional semiparametric mixture model with density

$$f_{\eta_0,\theta_0}(y, x) = \int f_{y/x,\alpha}(y; \theta_0) d\eta_0(\alpha),$$

where θ_0 is an additional unknown parameter, finite or infinite-dimensional. The basic idea here is that irregularity of $\phi(\eta_0) = \mathbb{E}_{\eta_0}[r(\alpha)]$ in the model where θ_0 is known implies irregularity in the model where θ_0 is unknown.

We illustrate our point with the random coefficients Logit model, also known as the Mixed Logit—one of the most commonly used models in applied choice analysis. Fox, Kim, Ryan and Bajari (2012) have recently shown nonparametric identification for the semiparametric Mixed Logit model. Here, we show that the identification of the CDF and quantiles of the distribution of RC is necessarily irregular when UH is nonparametric. The CDF and quantiles of this distribution are important parameters in applications of discrete choice.

The data $Z_i = (Y_i, X_i)$ is a random sample from the density (wrt μ below),

$$f_{\lambda_0}(y,x) = \int f_{y/x,\alpha}(y;\theta_0)d\eta_0(\alpha),$$

where $\lambda_0 = (\theta_0, \eta_0) \in \Theta \times H$, $\theta_0 = (\theta_{01}, ..., \theta_{0J})'$,

$$f_{y/x,\alpha}(y;\theta_0) = \frac{\exp\left(\theta_{0y} + x_y'\alpha\right)}{1 + \sum_{j=1}^{J} \exp\left(\theta_{0j} + x_j'\alpha\right)},$$

for $x = (x_0, x_1, ..., x_J) \in \mathcal{X}$ and $y \in \mathcal{Y} = \{0, 1, ..., J\}$. The consumer can choose between $j = 1, ..., J$, $J < \infty$, mutually exclusive inside goods and one outside good ($y = 0$). The utility for the inside good is normalized so that $\theta_{00} = 0$ and $x_0 = 0$. The random coefficients α are independent of the regressors X, and have a distribution η_0. The main result below also applies to the correlated random coefficient case. Moreover, non-regular identification for CDFs and quantiles is proved even when θ_0 is known. This will imply non-regularity when θ_0 is unknown and/or when random coefficients are dependent of the characteristics.

The measure μ is defined on $\mathcal{Z} = \mathcal{Y} \times \mathcal{X}$ as $\mu(B_1 \times B_2) = \tau(B_1)\nu_X(B_2)$, where $B_1 \subset \mathcal{Y}$, B_2 is a Borel set of $\mathcal{X}$, $\tau(\cdot)$ is the counting measure and $\nu_X(\cdot)$ is the probability measure for X. The vector α and covariates x_y are $K-$dimensional. The parameter space Θ is an open set of $\mathbb{R}^J$. The set H consists of measurable functions $\eta : \mathbb{R}^K \to \mathbb{R}$ whose support $\mathcal{A}$ has a non-empty interior and $\int_{\mathcal{A}} d\eta(\alpha) = 1$.

Applying the necessary condition for regular identification to a continuous linear functional $\phi(\eta) \in \mathbb{R}$ with influence function r_ϕ in the model where θ_0 is known, it must be true that for some $s \in L_2$,

$$r_\phi(\alpha) - \phi(\eta_0) = \int f_{y/x,\alpha}(y;\theta_0)s(y,x)d\mu(y,x). \tag{16}$$

It is straightforward to show that the right hand side in (16) is continuous in α in the interior of its support. In fact, more is true in general: it is an analytic function of α (a function that is infinitely differentiable with a convergent power series expansion). But continuity suffices for proving the non-regularity of CDFs and quantiles of η_0. This follows without computing least favorable distributions, simply by dominated convergence. We gather the proof here to illustrate the simplicity of our method of proof.

Proposition 5.1 *r_ϕ in (16) is continuous in the interior of $\mathcal{A}$.*

Proof of Proposition 5.1: Write

$$\int f_{y/x,\alpha}(y;\theta_0)s(y,x)d\mu(y,x) = \sum_{j=0}^{J}\int f_{y/x,\alpha}(j;\theta_0)s(j,x)\upsilon_X(dx).$$

Each of the summands in the last expression is continuous in α in the interior of its support, by continuity and boundedness of $f_{y/x,\alpha}(j;\theta_0)$ and the dominated convergence theorem. ■

Proposition 5.1 implies that identification of the CDF and quantiles of the distribution of η_0 under the conditions specified in Fox et al. (2012) must be irregular. Bajari, Fox and Ryan (2007) propose a simple estimator of the CDF of η_0, and Fox, Kim and Yang (2016) show its consistency (in the weak topology) and obtain its rates of convergence. Proposition 5.1 implies that the estimator in Fox et al. (2016), or any other estimator for that matter, cannot achieve regular parametric rates of convergence. The lack of regularity is not evident from the rates established in Fox et al. (2016). Let F_0 be the CDF pertaining to η_0 and $\widehat{F}_\eta$ the "fixed grid" estimator of Bajari et al. (2007), Fox et al. (2011) and Fox et al. (2016) based on D grid points ($D \equiv D(n)$, where n is the sample size). The order of the bias established in Fox et al. (2016) is $D^{-\bar{s}/K}$ where $\bar{s}$ is the smoothness of the mapping $\alpha \to f_{y/x,\alpha}$ (here $\bar{s} = \infty$). This suggests that parametric rates might be attainable, but our results show that this is not possible (at least in a local uniform sense). The order of the variance for $\widehat{F}_\eta$ is inversely related to the minimum eigenvalue of the $D \times D$ matrix Ψ_D with $(d_1, d_2) - th$ element, $1 \leq d_1, d_2 \leq D$, given by

$$\mathbb{E}\left[g'(X,\alpha_{d_1})g(X,\alpha_{d_2})\right], \tag{17}$$

where $g(x,\alpha_d) = (f_{y/x,\alpha_d}(0;\theta_0), ..., f_{y/x,\alpha_d}(J;\theta_0))'$ are conditional choice probabilities when UH is evaluated at the $d - th$ grid point α_d, $d = 1, ..., D$. This minimum eigenvalue quantifies the level of multicollinearity in the least squares regression of Fox et al. (2016), and we conjecture that given the high smoothness of the mapping $\alpha \to f_{y/x,\alpha}$ this term will go to zero exponentially fast, so it will be the main determinant in the (slow) rate of convergence of $\widehat{F}_\eta$. A detailed theoretical analysis of this issue is beyond the scope of this paper, but see the discussion in Section 6. We stress that these arguments are not a criticism of "fixed grid" estimators per se, but rather represent a limitation that *any* estimation method would have due to the statistical difficulty of the problem.

5.2 Bound on of exact consumer surplus

Another new example of irregular identification in a semiparametric setting is for a weighted average over income values of an average (across heterogenous individuals) of exact consumer surplus bounds, as in Hausman and Newey (2016). Here Y is quantity consumed, $X = (X_1, X_2)'$, X_1 is price, X_2 is income, $\eta_0(x)$ is the demand function, price is changing between $\breve{x}_1$ and $\bar{x}_1$, and B is a bound on the income effect. Let $w(x_2)$ be some weight function and $v(x_1) = 1(\breve{x}_1 \leq x_1 \leq \bar{x}_1)e^{-B(x_1 - \breve{x}_1)}$. Consider the functional

$$\phi(\eta_0) = \mathbb{E}_{\eta_0}\left[w(X_2)\int v(u)\eta_0(u, X_2)du\right],$$

which corresponds to a bound on the average of equivalent variation over unobserved individual heterogeneity and income (see Hausman and Newey, 2016). Santos (2011) and Chernozhukov et al. (2016) have investigated regular identification and estimation of related quantities when the demand function satisfies the instrumental variables restriction

$$\mathbb{E}_{\eta_0}[Y_i - \eta_0(X_i)|W_i] = 0,$$

for an instrument W_i. In this paper we will discuss conditions under which $\phi(\eta_0)$ is irregularly identified. To the best of our knowledge, this result is new in the literature.

Let $f_0(x_1|x_2)$ denote the conditional pdf of X_1 given X_2, and let

$$r(x) = f_0(x_1|x_2)^{-1}v(x_1)w(x_2).$$

Consider a submodel where $r(\cdot)$ is known and has finite variance, and note the moment representation

$$\phi(\eta_0) = \mathbb{E}_{\eta_0}\left[\eta_0(X)r(X)\right].$$

Severini and Tripathi (2012, Lemma 4.1) have shown that a necessary condition for regular identification of moments such as $\phi(\eta_0)$ is the existence of $s(W)$ with finite variance such that

$$r(x) = \mathbb{E}[s(W_i)|X_i = x]. \tag{18}$$

We apply our results to obtain irregular identification under the following mild assumptions. Let $f(w|x)$ denote the conditional pdf of W given X wrt μ. Fix x_2 in the support of X_2.

Assumption 5 *(i) The conditions of Severini and Tripathi (2012, Lemma 4.1) hold; (ii) the densities $f_0(x_1|x_2)$ and $f(w|x)$ and continuous in a neighborhood of $\bar{x}_1$, say Γ_0; (ii)*

$$\int |s(w)| \sup_{x_1 \in \Gamma_0} f(w|x_1, x_2)d\mu(w) < \infty. \tag{19}$$

The conditions of Severini and Tripathi (2012, Lemma 4.1) are standard in the literature of nonparametric instrumental variables.

Proposition 5.2 *Suppose Assumption 5 holds. Then, $\phi(\eta_0)$ is irregularly identified.*

The source of irregularity is the discontinuity of $v(x_1)$ at $\bar{x}_1$. The continuity of $f_0(x_1|x_2)$ at $\bar{x}_1$ is just a simplifying assumption, while the continuity $f(w|x)$ is important for the result. We only need this continuity to hold for a submodel of the larger semiparametric model. The dominance condition will hold if, for example, the mapping $x_1 \to f(w|x_1, x_2)$ is quasi-convex, since in that case and for $\Gamma_0 = (\alpha_1, \alpha_2)$, the left hand side of (19) is bounded by

$$\max\left\{\int |s(w)|\, f(w|\alpha_1, x_2) d\mu(w), \int |s(w)|\, f(w|\alpha_2, x_2) d\mu(w)\right\} < \infty.$$

Alternative identification strategies that assume exogeneity of prices or a control function approach lead to regular identification under mild conditions, as shown in Hausman and Newey (1995, 2016). Our results have important implications for the analysis in Santos (2011) and Chernozhukov et al. (2016).

6 Regularization

The previous examples show that regular identification of CDFs and quantiles of UH in the models considered may require restricting the nature of heterogeneity. In this section we investigate how common approaches considered in the literature address the lack of regularity of these functionals. Additionally, we provide a necessary condition for CDFs and quantiles to be regularly identified when UH is semiparametric and a discussion on how smoothness of $\alpha \to f_{z/\alpha}$ translates into a multicollinearity problem for sieve and related estimators.

Our first observation is derived from the main idea in the previous section: functional form assumptions that restrict the conditional likelihood may not help with the irregular identification of CDFs and quantiles if still the mapping $\alpha \to f_{z/\alpha}$ is smooth, while UH is nonparametric. For example, knowing the finite dimensional parameters of a semiparametric mixture, knowing the functional forms of the idiosyncratic error terms in Kotlarski's lemma, or knowing the functional form of the baseline hazard in the mixed proportional hazard model do not help in restoring regular identification of CDFs and quantiles of UH when UH is nonparametric.

We discuss how restrictions on UH translate into regularity of functionals of UH. Denote by $\overline{T(\eta_0)}$ the mean squared closure of $T(\eta_0)$ in $L_2(\eta_0)$. That UH is semiparametric (rather than nonparametric) formally means that $\overline{T(\eta_0)}$ is a strict subset of $L_2^0(\eta_0)$. The extension of the necessary condition for regular identification of $\phi(\eta_0) = \mathbb{E}_{\eta_0}[r(\alpha)]$, for a measurable function $r(\cdot)$ with $\mathbb{E}_{\eta_0}[r^2(\alpha)] < \infty$, is given in the following lemma. Let $\Pi_{\overline{V}}$ denote the orthogonal projection operator onto $\overline{V}$, where $\overline{V}$ denotes the closure of V in the norm topology.

Lemma 6.1 *The necessary condition for regular identification of $\phi(\eta_0) = \mathbb{E}_{\eta_0}[r(\alpha)]$ when UH is semiparametric is*

$$\Pi_{\overline{T(\eta_0)}} r(\alpha) = \Pi_{\overline{T(\eta_0)}} \mathbb{E}\left[\, s(Z) | \, \alpha\right], \text{ for some } s \in L_2^0. \tag{20}$$

The mismatch in smoothness between $r(\alpha)$ and $\mathbb{E}\left[s(Z)|\,\alpha\right]$, which was the source of irregularity when UH was nonparametric, may now be restored by the projection onto $\overline{T(\eta_0)}$. The left hand side of (20) is the influence function for $\phi(\eta_0)$ with semiparametric UH. We briefly discuss how different restrictions on UH translate into regularity of CDFs and quantiles in view of this general characterization.

A popular approach in practice is to consider a parametric distribution for the UH. A leading example of parametric model is a finite mixture with known and finite support points. Parametric heterogeneity leads to a finite dimensional tangent space $T(\eta_0)$, which is then closed $T(\eta_0) = \overline{T(\eta_0)}$, and which is generated by the scores of the specified distribution. Denote by l_η the score of UH, i.e. $\overline{T(\eta_0)} = T(\eta_0) = span(l_\eta)$, assume $\mathbb{E}_{\eta_0}\left[l_\eta(\alpha)l_\eta'(\alpha)\right]$ is non-singular, and define the projected score $s_0(Z) = \mathbb{E}\left[l_\eta(\alpha)|\,Z\right]$. Often, $l_\eta = \partial \log d\eta_\theta/\partial\theta$, where θ is the parameter indexing UH, and $s_0(Z) = \partial \log f_\theta/\partial\theta$, where

$$f_\theta(z) = \int f_{z/\alpha}(z)d\eta_\theta(s).$$

Then, simple algebra shows that a solution to (20) in s is given by s_r defined by

$$s_r(Z) = \lambda_r' s_0(Z),$$

where λ_r is a solution to

$$\mathbb{E}\left[s_0(Z)s_0'(Z)\right]\lambda_r = \mathbb{E}\left[r(\alpha)l_\eta'(\alpha)\right]. \tag{21}$$

If the Fisher information for η_0 is positive definite, which means $\mathbb{E}\left[s_0(Z)s_0'(Z)\right]$ is non-singular, then there is a unique solution λ_r of (21), and $\phi(\eta_0)$ is regularly identified. More generally, $\phi(\eta_0)$ may be regularly identified even when η_0 is not, and this corresponds to the system in (21) having some (non-unique) solution in λ_r. The drawback of the parametric approach is the high misspecification risk, which can be quantified by the dimension and form of the model's tangent space. If the dimension of $T(\eta_0)$ is D, then the tangent space of the model is at most $D-$dimensional and given by $\mathcal{S} := \{s \in L_2^0 : s(z) = \lambda' s_0(z) \text{ for some } \lambda \in \mathbb{R}^D\}$. Estimators for functionals of UH will be in general inconsistent when the model is misspecified.

As usual, a semiparametric approach is more robust to misspecification than a parametric one. In Lemma 6.1 we have derived the necessary condition for regular identification of moments when UH is semiparametric, so $\overline{T(\eta_0)}$ is a strict subset of $L_2^0(\eta_0)$ of infinite dimension. Examples of semiparametric models include finite mixtures with unknown support points and sieve methods with incomplete sieve basis. Existing rate results for finite mixtures with unknown support points suggest irregularity of the CDFs in general (see, e.g., Chen 1995 and Heinrich and Kahn 2018), although we are not aware of any paper investigating semiparametric efficiency bounds for finite mixtures with unknown support points. We recognize that, although the sufficient condition for semiparametric restrictions in Lemma 6.1 is general, it may be hard to find primitive conditions for it, as computing the closure of $T(\eta_0)$ and the projections onto it may not be straightforward in applications.

As a practical approach, we recommend a sieve method where the span of $\{l_\eta(\alpha)\}$ increases with the sample size, i.e. $D \to \infty$ as $n \to \infty$. Without loss of generality normalize l_η so that $\mathbb{E}_{\eta_0}\left[l_\eta(\alpha)l'_\eta(\alpha)\right]$ is the identity matrix. A key quantity for sieve estimation is the minimum eigenvalue of the Fisher information matrix $\mathbb{E}\left[s_0(Z)s'_0(Z)\right]$, denoted by $\xi_{\min} \equiv \xi_{\min}(D)$; see Fox, Kim and Yang (2016) and (21). We provide a useful bound for $\xi_{\min}$. To that end, we assume the score operator $Ab = \mathbb{E}\left[b(\alpha)|\, Z\right]$ from $L_2(\eta_0)$ to L_2 is compact. A well known sufficient condition for this is

$$\int \frac{f^2_{z/\alpha}(z)}{f_{\eta_0}(z)} d\eta_0(\alpha)d\mu(z) < \infty. \tag{22}$$

Under this condition, A has a sequence of singular values $\{\mu_d\}_{d=1}^{\infty}$ (see Engl, Hanke and Nuebauer, 1996).

Lemma 6.2 *If (22) holds, then* $\xi_{\min}(D) \leq \mu_D^2$.

It is well known that condition (22) yields $\mu_D \to 0$ as $D \to \infty$. Thus, Lemma 6.2 implies that also $\xi_{\min}(D) \to 0$. This is the multicollinearity problem referred to above. Furthermore, the score operator A is an integral operator with kernel $K(z,\alpha) = f_{z/\alpha}(z)/f_{\eta_0}(z)$, and it is well known that the smoother the mapping $\alpha \to K(z,\alpha)$, the faster the singular values μ_D go to zero. In particular, for analytical kernels the singular values decay exponentially fast to zero (Hille and Tamarkin 1931). The minimum eigenvalue $\xi_{\min}(D)$ is also closely related to the sieve measure of ill-posedness τ_D proposed in econometrics (see Chen 2007 and Blundell, Chen and Kristensen 2007) through the relation

$$\tau_D^2 = \frac{1}{\xi_{\min}(D)}.$$

Prior to this paper, Blundell, Chen and Kristensen (2007, Lemma 1) obtained the bound $\tau_D \geq 1/\mu_D$ in a nonparametric IV setting. Thus, the modest contribution here is the interpretation in terms of the minimum eigenvalue of the Fisher information matrix. For applications of sieve estimators along this line and the important role of τ_D (or $\xi_{\min}(D)$) see, e.g., Chen (2007), Bajari, Fox and Ryan (2007), Hu and Schennach (2008), Bester and Hansen (2007), Chen and Liao (2014), Fox, Kim and Yang (2016) and references therein. These arguments formalize the idea that the smoother the mapping $\alpha \to f_{z/\alpha}$, the more difficult estimation of functionals of UH is.

7 Conclusions

We have studied irregular identification of CDFs and quantiles (or more generally, functionals with discontinuous influence functions) of nonparametric UH in some structural economic models. Example applications include the structural model of unemployment with two spells in Alvarez et al. (2015), the binary and linear RC models (possibly with correlated effects), the AME in a triangular model with near zero first-stage effects, bounds on average equivalent variation, and the distribution and quantiles of UH in the Mixed Logit model. These are only some applications, but the results are applicable more widely. Further examples in the Appendix include a canonical model for heterogenous infectious diseases, and measurement error models with two measurements identified by means of Kotlarski's lemma. Furthermore, as we discuss in the Appendix, we expect our approach to be applicable to the many situations where the so-called Information Operator (see e.g. Begun, Hall, Huang and Wellner (1983)) is a smoothing operator.

The most appealing feature of our method of proof is its simplicity, relative to alternative approaches that directly compute efficiency bounds, which are particularly difficult to compute in the models we have studied. Instead, we exploit some necessary smoothness conditions that the influence function of a regularly identified functional must satisfy. The Mixed Logit example is illustrative of the simplicity of our method of proof. In contrast, directly computing the Fisher information and the efficiency bound in this model is rather challenging (and were unknown prior to this paper).

8 Appendix A: Proofs of Main Results

Proof of Lemma 3.1: First, the functional $\eta_0 \rightarrow \phi(\eta_0) = \mathbb{E}_{\eta_0}[r(\alpha)]$ is differentiable with influence function

$$\chi(\alpha) = \Pi_{\overline{T(\eta_0)}} r(\alpha),$$

where $\Pi_{\overline{V}}$ denotes the orthogonal projection operator onto the closure of V, $\overline{V}$. To see this, note that by linearity of $\eta_0 \rightarrow \phi(\eta_0)$, for all $b \in T(\eta_0)$,

$$\begin{aligned}\lim_{t\rightarrow 0} \frac{\phi(\eta_t) - \phi(\eta_0)}{t} &= \mathbb{E}_{\eta_0}[r(\alpha)b(\alpha)] \\ &= \mathbb{E}_{\eta_0}[\left(\Pi_{\overline{T(\eta_0)}} r(\alpha)\right) b(\alpha)].\end{aligned}$$

Since UH is nonparametric $\Pi_{\overline{T(\eta_0)}} r(\alpha) = r(\alpha) - \phi(\eta_0)$. On the other hand, by Lemma 25.34 in van der Vaart (1998) the adjoint of the score operator is given by

$$A^* s = \mathbb{E}\left[s(Z)|\, \alpha\right] - \mathbb{E}\left[s(Z)\right].$$

The lemma then follows from Theorem 3.1 and Theorem 4.1 in van der Vaart (1991), which establish that a necessary condition for positive Fisher information for $\phi(\eta_0)$ is

$$r(\alpha) - \phi(\eta_0) = \mathbb{E}\left[s(Z)|\, \alpha\right],$$

since $\mathbb{E}\left[s(Z)\right] = 0$. ■

Proof of Lemma 3.2: Let $\alpha_n, \alpha \in N$ such that $\alpha_n \rightarrow \alpha$, and define $h_n(z) = s(z) f_{z/\alpha_n}(z)$. Note (i) implies $h_n(z) \rightarrow h(z) := s(z) f_{z/\alpha}(z)$ a.e-μ. Also, by the dominance condition, for a sufficiently large n,

$$\int |h_n(z)|\, d\mu(z) < \infty.$$

We conclude by dominated convergence that

$$\int s(z) f_{z/\alpha_n}(z) d\mu(z) \rightarrow \int s(z) f_{z/\alpha}(z) d\mu(z).$$

■

Proof of Corollary 3.1: By Lemma 3.2 if the influence function of the functional is discontinuous then the functional is not regularly identified. Since the indicator is not continuous, this proves the lemma. ■

Proof of Corollary 3.2: Lemma 21.3 in van der Vaart (1998) shows the pathwise differentiability of the quantile functional with an influence function

$$r_\phi(\alpha) = \frac{-\left\{1(\alpha < \phi(\eta_0)) - \tau\right\}}{\dot{\eta}_0(\phi(\eta_0))}.$$

That is, under the regularity conditions of the corollary, the quantile functional $\eta_0 \to \phi(\eta_0)$ satisfies, for all $b \in T(\eta_0)$,

$$\lim_{t\to 0} \frac{\phi(\eta_t) - \phi(\eta_0)}{t} = \mathbb{E}_{\eta_0}[r_\phi(\alpha) b(\alpha)].$$

From Van der Vaart (1991) it follows that a necessary condition for the quantile functional to be differentiable is

$$r_\phi(\alpha) - \phi(\eta_0) = \int s(z) f_{z/\alpha}(z) d\mu(z).$$

By Lemma 3.2 if the influence function of the functional is discontinuous then the functional is not regularly identified. Since the influence function of the quantile is not continuous, this proves the lemma. ■

Proof of Proposition 3.1: By substitution of $f_{z/\alpha}(t_1, t_2)$ we obtain

$$\begin{aligned}\mathbb{E}\left[s(Z)|\,\alpha\right] &= \int_{\mathcal{T}^2} s(t_1, t_2) f_{z/\alpha}(t_1, t_2) dt_1 dt_2 \\ &= C\beta^2 e^{2\alpha\beta} h(\alpha_1^2, \alpha_2^2),\end{aligned}$$

where

$$h(u, v) = \int_{\mathcal{T}^2} s(t_1, t_2) \frac{1}{t_1^{3/2} t_2^{3/2}} s(u, v; t_1) s(u, v; t_2) dt_1 dt_2$$

and

$$s(u, v; t) = \exp\left(-\frac{ut}{2} - \frac{v}{2t}\right),\ t \in \mathcal{T},\ (u, v) \in (0, \infty).$$

We check that the conditions for an application of the Leibniz's rule hold. These conditions are

1. The partial derivative $\partial^m s(u, v; t_1) s(u, v; t_2)/\partial^m u$ exists and is a continuous function on an open neighborhood B of (u, v), for a.s. $(t_1, t_2) \in \mathcal{T}^2$.

2. There is a positive function $h_m(t_1, t_2)$ such that

$$\sup_{(u,v)\in B} \left| \frac{\partial^m s(u, v; t_1) s(u, v; t_2)}{\partial^m u} \right| \leq h_m(t_1, t_2) \tag{23}$$

and

$$\int_{\mathcal{T}^2} s(t_1, t_2) \frac{1}{t_1^{3/2} t_2^{3/2}} h_m(t_1, t_2) dt_1 dt_2 < \infty. \tag{24}$$

Simple differentiation and induction show that for any integer $m \geq 0$

$$\frac{\partial^m s(u, v; t_1) s(u, v; t_2)}{\partial^m u} = 2^{-m}(-1)^m (t_1 + t_2)^m s(u, v; t_1) s(u, v; t_2).$$

Therefore, by monotonicity we can find u^* and v^* such that (23) holds with

$$h_m(t_1, t_2) = 2^{-m}(t_1 + t_2)^m s(u^*, v^*; t_1) s(u^*, v^*; t_2).$$

Furthermore, by $\mathbb{E}\left[s(Z)|\, \alpha\right] < \infty$ for all α in a local neighborhood (by local boundedness of r), and the boundedness of $\mathcal{T}$, condition (24) holds. The continuity of $h(u, v)$ is a special case of the previous arguments with $m = 0$ (note the term $(t_1 + t_2)^m$ is one and the boundedness of $\mathcal{T}$ is not needed in this case). ■

Proof of Proposition 4.1: Define

$$\begin{aligned} b(\alpha) &= \mathbb{E}\left[s(Y_i = 1, X_i)|\, \alpha_i = \alpha\right] \\ &= \int 1\left(x'\alpha \geq 0\right) s(1, x) dv_X(x). \end{aligned}$$

We prove that b is continuous and by compactness of the sphere is therefore uniformly continuous. Since the halfspaces $1\left(x'\alpha \geq 0\right)$ and $1\left(x'\alpha_0 \geq 0\right)$ intersect in sets having surface measure of order $|\alpha - \alpha_0|$, it follows from the absolutely continuity of the angular component of X that

$$|b(\alpha) - b(\alpha_0)| = O\left(|\alpha - \alpha_0|\right).$$

When $x = (1, \tilde{x})$, then

$$\begin{aligned} b(\alpha) &= \int 1\left(\tilde{x}'\alpha_2 \geq -\alpha_1\right) s(1, 1, \tilde{x}) dv_X(\tilde{x}), \\ &= \int 1\left(u \geq -\alpha_1\right) s_{\alpha_2}(u) f_{\alpha_2}(u) du, \end{aligned}$$

where $s_{\alpha_2}(u) = \mathbb{E}\left[s(Y_i = 1, 1, \tilde{X}_i) \middle|\, \alpha_2'\tilde{X}_i = u\right]$ and f_{α_2} denotes the density of $\alpha_2'\tilde{X}_i$. The absolute continuity in α_1 follows from the integrability of $s_{\alpha_2}(u) f_{\alpha_2}(u)$ and Royden (1968, Chapter 5). ■

Proof of Corollary 4.1: The proof follows as in Corollaries 3.1 and 3.2. ■

For a function $a \in L_1(\lambda) \cap L_2(\lambda)$, define the Fourier transform $\hat{a}(t) = \int e^{it'\alpha} a(\alpha) d\alpha$, where $i = \sqrt{-1}$. Use the notation

$$\tilde{g}(p, x) = \int e^{ipy} g(y, x) dy,$$

for the Fourier transform with respect to just the first argument (for $g(\cdot, x) \in L_1(\lambda) \cap L_2(\lambda)$). Define the norms

$$|g|_{1,\rho}^2 = \int_{\mathbb{S}^{d_\alpha - 1}} \int_{\mathbb{R}} |\tilde{g}(p, x)|^2 (1 + |p|^2)^\rho dp dx \tag{25}$$

and

$$|g|_\rho^2 = \int |\hat{g}(t)|^2 (1 + |t|^2)^\rho dt. \tag{26}$$

The Sobolev space $H^\rho(\mathcal{A})$ is defined as the set of measurable functions g such that $|g|_\rho < \infty$.

Proof of Proposition 4.2: Define the score operator $A : T(\eta_0) \to L_2$

$$Ab(z) = \frac{Rb\eta_0(z)}{f_{\eta_0}(z)} 1(f_{\eta_0}(z) > 0),$$

where R denotes the Radon transform

$$Ra(y,x) = \int a(\alpha)1(y = x'\alpha)d\alpha.$$

Define $g(z) = s(z)f_{\eta_0}(z)$ and $a(\alpha) = b(\alpha)\eta_0(\alpha)$. Since $f_{\eta_0}(z)$ and η_0 are bounded, it follows that g and a are in $L_1(\lambda) \cap L_2(\lambda)$. From the definition of $Ra(y,x)$

$$\sup_{y,x} |Ra(y,x)| \leq \int |a(\alpha)|\, d\alpha < \infty, \tag{27}$$

and since the supports of α and X are bounded, the support of Y is also bounded and $Ra \in L_2(\lambda)$, so we can view $R : L_2(\lambda) \to L_2(\lambda)$.

First, we show that if s belongs to the closure of the range of A, then $g(z) = s(z)f_{\eta_0}(z)$ belongs to the closure of the range of R. Indeed, if s_n is a sequence in the range of A converging to s in L_2, then $g_n = s_n f_{\eta_0}(z) \equiv Ra_n$ and clearly

$$\int |g_n(z) - g(z)|^2\, dz \leq \int |s_n(z) - s(z)|^2\, f_{\eta_0}(z)dz \to 0.$$

Next, we shall show that any function g in the closure of the range of R will have an squared integrable weak derivative with respect to the first argument (in y). By Theorem 2.4.1 in Ramm and Katsevich (1996) and Assumption 3(iii) it follows that $|g|_{1,\rho} < \infty$ for $\rho = \rho_0 + (d_\alpha - 1)/2$. While by well known results in Fourier analysis, with $\partial_y g$ denoting the weak derivative with respect to y

$$\begin{aligned}\int_{\mathbb{S}^{d_\alpha-1}} \int \left|\widetilde{\partial_y g}(p,x)\right|^2 dpdx &\leq \int_{\mathbb{S}^{d_\alpha-1}} \int |p|^2 \left|\widetilde{g}(p,x)\right|^2 dpdx \\ &\leq \int_{\mathbb{S}^{d_\alpha-1}} \int |\tilde{g}(p,x)|^2 (1+|p|^2)^\rho dpdx \\ &< \infty,\end{aligned}$$

and similarly, by Cauchy-Schwarz

$$\begin{aligned}\int_{\mathbb{S}^{d_\alpha-1}} \int \left|\widetilde{\partial_y g}(p,x)\right| dpdx &\leq \int_{\mathbb{S}^{d_\alpha-1}} \int \left(1+|p|^2\right)^{1/2} |\widetilde{g}(p,x)|\, dpdx \\ &\leq C\left(\int_{\mathbb{S}^{d_\alpha-1}} \int (1+|p|^2)^{1-\rho} dpdx\right)^{1/2} \\ &< \infty, \text{ because } \rho > 2.\end{aligned}$$

Thus $\widetilde{\partial_y g}(p,x) \in L_1(\lambda) \cap L_2(\lambda)$ and by Plancherell's theorem $\partial_y g(\cdot) \in L_2(\lambda)$, as we claimed.

Define $\varphi(\cdot) = \partial_y g(\cdot) \in L_2(\lambda)$. We proceed to verify the conditions of the dominated convergence theorem, see Lemma 3.2. First, we show that $g(y,x)$ is continuous in y. Indeed, by the bounded support assumption

$$g(y,x) = \int_{-\infty}^{y} \varphi(u,x)dx$$

is absolutely continuous in y (see Royden 1968, Chapter 5).

Next, by independence of α_i and X_i,

$$\mathbb{P}\left[Y_i \leq y \middle| X_i = x\right] = \mathbb{P}\left[x'\alpha_i \leq y\right],$$

and taking derivatives we conclude $f_{\eta_0}(z) = \eta_{0,x}(y)$. Thus, $f_{\eta_0}(z)$ is also continuous in y by Assumption 3(i). Moreover,

$$\inf_{\alpha \in N} \eta_{0,x}(x'\alpha) \geq 1/l(x) > 0,$$

which yields the continuity of $\alpha \to s(x'\alpha, x)$ in N. Furthermore, by Cauchy-Schwarz and

$$\begin{aligned}
\int \sup_{\alpha \in \Gamma_0} |s(x'\alpha, x)| \, f_X(x)dx &= \int \sup_{\alpha \in \Gamma_0} |g(x'\alpha, x)| \sup_{\alpha \in \Gamma_0} \left| \frac{f_X(x)}{f_{\eta_0}(x'\alpha, x)} \right| dx \\
&\leq \left(\int |\varphi(u, x)|^2 \, dudx \right)^{1/2} \left(\int \sup_{\alpha \in \Gamma_0} \left| \frac{f_X(x)}{f_{\eta_0}(x'\alpha, x)} \right|^2 dx \right)^{1/2} \\
&\leq C \left(\int l^2(x) f_X(x) dx \right)^{1/2} \\
&\leq C.
\end{aligned}$$

Thus, by dominated convergence r must be continuous in N. ■

Proof of Corollary 4.2: The proof follows as in Corollaries 3.1 and 3.2. ■

Proof of Proposition 4.3: A necessary condition for a reduced form functional $\phi(\eta_0) = \mathbb{E}_{\eta_0}[r(\alpha)]$ to be regularly identified is

$$r(\alpha) - \phi(\eta_0) = \int s(\alpha_0 + \alpha_1 x, x) dv_X(x), \;\; \alpha = (\alpha_0', \alpha_1') = (\pi_0, U_2, \pi_1, \delta)'.$$

Thus, by Proposition 4.2 $r(\alpha)$ must be continuous in N. However, the influence function for the PPAME

$$r_{PPAME}(\alpha) = 1(\pi_1 > 0)1(\delta > 0) + 1(\pi_1 < 0)1(\delta < 0)$$

is discontinuous at the points $(p_0, u_2, 0, d_0)$ or $(p_0, u_2, p_1, 0)$. Conclude that the PPAME is not regularly identified. As for AME, by $\mathbb{E}\left[\gamma^2\right] < \infty$ this functional is differentiable in the sense of van der Vaart (1991) with an influence function $r_{AME}(\beta) = \pi_1/\delta$. Since there is no continuous function that is η_0-a.s equal to $r_{AME}(\beta) = \pi_1/\delta$ when $(p_0, u_2, p_1, 0)$ is a point in the interior of the support, we conclude that the AME is not regularly identified. ■

Proof of Proposition 5.2: By Severini and Tripathi (2012, Lemma 4.1) condition (18) must hold under regular identification. The dominated convergence theorem, as in Lemma 6, then implies that r must be continuous at $\bar{x}_1$. This yields irregularity for functionals with discontinuous r, such as the bound on average equivalent variation. ■

Proof of Lemma 6.1: By Lemma 25.34 in van der Vaart (1998) the so-called score operator is given by

$$Ab(z) = \mathbb{E}\left[b(\alpha) \middle| Z \right], \; b \in T(\eta_0)$$

Thus, by the law of iterated expectations

$$\begin{aligned}
\mathbb{E}\left[Ab(Z)s(Z)\right] &= \mathbb{E}\left[b(\alpha)s(Z)\right] \\
&= \mathbb{E}\left[b(\alpha)\mathbb{E}\left[s(Z) \middle| \alpha\right]\right] \\
&= \mathbb{E}\left[b(\alpha)\Pi_{\overline{T(\eta_0)}}\mathbb{E}\left[s(Z) \middle| \alpha\right]\right].
\end{aligned}$$

In Lemma 3.2 we have shown that the functional $\eta_0 \to \phi(\eta_0) = \mathbb{E}_{\eta_0}[r(\alpha)]$ is differentiable with influence function

$$\chi(\alpha) = \Pi_{\overline{T(\eta_0)}} r(\alpha).$$

The lemma then follows from Theorem 3.1 in van der Vaart (1991). ■

Proof of Lemma 6.2: The sieve measure of ill-posedness (cf. Blundell, Chen and Kristensen 2007) is

$$\tau_D = \sup_{b \in T(\eta_0), b \neq 0} \frac{\|b\|}{\|Ab\|}.$$

Since $T(\eta_0) = span(l_\eta)$ and $\mathbb{E}_{\eta_0}\left[l_\eta(\alpha)l_\eta'(\alpha)\right]$ is the identity then $b = \lambda' l_\eta$ and $\|b\|^2 = \lambda'\lambda = |\lambda|^2$, while $\|Ab\|^2 = \lambda'\mathbb{E}\left[s_0(Z)s_0'(Z)\right]\lambda$. Thus,

$$\begin{aligned}
\tau_D^2 &= \sup_{\lambda \in \mathbb{R}^D, \lambda \neq 0} \frac{|\lambda|^2}{\lambda'\mathbb{E}\left[s_0(Z)s_0'(Z)\right]\lambda} \\
&= \frac{1}{\inf_{\lambda \in \mathbb{R}^D, |\lambda|=1} \lambda'\mathbb{E}\left[s_0(Z)s_0'(Z)\right]\lambda} \\
&= \frac{1}{\xi_{\min}(D)}.
\end{aligned}$$

The bound then follows from Lemma 1 in Blundell, Chen and Kristensen (2007). ■

9 Appendix B: Further Results

9.1 Nonlinear RC

In this section we describe a generic approach that can be used for generic nonlinear RC models with continuous outcomes. We also illustrate how certain invertible RC models are ruled out by our conditions. For the generic RC model in (9), the regularity condition reads

$$r(\alpha) - \phi(\eta_0) = \mathbb{E}\left[s(m(X_i, \alpha), X_i)\right]. \tag{28}$$

Again, the main difficulty in proving that the right hand side of (28) is continuous is that the score function $s(\cdot)$ is only known to be in L_2 (thus, s is potentially very discontinuous). To overcome this difficulty, we resort to Fourier analysis and use the so-called Parseval's identity (see Rudin 1987, pg. 187). To describe the method, assume X is absolutely continuous with density $f_X(x)$, and define

$$g(z) = s(z) f_{\eta_0}(z) \qquad \text{and} \qquad w(z, \alpha) = \frac{1\left(y = m(x, \alpha)\right) f_X(x)}{f_{\eta_0}(z)} 1(f_{\eta_0}(z) > 0).$$

Note that $g \in L_1(\lambda)$, and since f_{η_0} is bounded, also $g \in L_2(\lambda)$. Let $\eta_{m,x}$ denote the density of $m(x, \alpha)$ when α has density η_0. Under our conditions below, $w(\cdot, \alpha) \in L_1(\lambda) \cap L_2(\lambda)$, and by Parseval's identity, if r satisfies (28) then

$$r(\alpha) - \phi(\eta_0) = \int \hat{g}(t) \overline{\hat{w}(t, \alpha)} dt, \tag{29}$$

where, for a generic function $h \in L_1(\lambda)$, $\hat{h}(t) = (2\pi)^{-d_z/2} \int e^{-it'z} h(z) dz$ denotes the Fourier transform, with $i = \sqrt{-1}$, $\overline{v}$ denotes the complex conjugate of v and

$$\overline{\hat{w}(t, \alpha)} = (2\pi)^{-d_z/2} \int \frac{f_X(x)}{\eta_{m,x}(x'\alpha)} e^{i(t_1 m(x,\alpha) + t_2' x)} dx.$$

This integral representation is now amenable to our Lemma 3.2 under the following assumption.

Assumption 6 *(i) The vector X is absolutely continuous with a bounded density $f_X(\cdot)$; (ii) the density $\eta_{m,x}$ is continuous and satisfies $\inf_{\alpha \in N} \eta_{m,x}(m(x, \alpha)) > 1/l(x)$ for an a.s. positive measurable function $l(\cdot)$ such that $\mathbb{E}_X[l^2(X)] < \infty$; (iii) the function $\alpha \to m(x, \alpha)$ is continuous a.s. in x; (iv) for all $\hat{g}$ satisfying (29),*

$$\int |\hat{g}(t)| \sup_{\alpha \in \Gamma_0} \left| \overline{\hat{w}(t, \alpha)} \right| dt < \infty. \tag{30}$$

Proposition 9.1 *Under Assumption 6 and if r satisfies (13), then $r(\cdot)$ must be continuous on N.*

Proof of Proposition 9.1: First, we need to check that g and $w(z, \alpha)$ are in $L_1(\lambda) \cap L_2(\lambda)$, so we can apply Parseval's identity. From $s \in L_2$ and the definition of $g(z) = s(z) f_{\eta_0}(z)$, it is clear that $g \in L_1(\lambda)$. Next, note

$$f_{\eta_0}(z) \leq \int_{\mathbb{R}^d} d\eta_0(\alpha) = 1.$$

Thus, g also belongs to $L_2(\lambda)$. Furthermore, by independence of α_i and X_i,

$$\mathbb{P}\left[Y_i \leq y | X_i = x\right] = \mathbb{P}\left[m(x, \alpha_i) \leq y\right],$$

and taking derivatives we conclude $f_{\eta_0}(z) = \eta_{m,x}(y)$. Then, for $p = 1$ or 2,

$$\begin{aligned}\int |w(z,\alpha)|^p \, dz &= \int \left| \frac{f_X(x)}{\eta_{m,x}(x'\alpha)} \right|^p dx \\ &\leq \int l^p(x) \, |f_X(x)|^p \, dx \\ &\leq C \int l^p(x) f_X(x) dx \\ &< \infty,\end{aligned}$$

because f_X is bounded. Then, we can apply Parseval's identity and obtain

$$r(\alpha) - \phi(\eta_0) = \int \hat{g}(t)\overline{\hat{w}(t,\alpha)}dt.$$

We now proceed to verify the conditions of Lemma 3.2 with $\hat{g}(\cdot)$ playing the role of s and $\overline{\hat{w}(t,\alpha)}$ that of the conditional density. Note

$$\overline{\hat{w}(t,\alpha)} = (2\pi)^{-d_z/2} \int \frac{f_X(x)}{\eta_{m,x}(m(x,\alpha))} e^{i(t_1 m(x,\alpha)+t_2' x)} dx.$$

Under the conditions of the proposition the function $\alpha \to \overline{\hat{w}(t,\alpha)}$ is continuous on N since $\eta_{m,x}(\cdot)$ and $m(x,\cdot)$ are continuous and $\eta_{m,x}(m(x,\alpha))$ is bounded away from zero on N. Furthermore, the dominance condition holds from (30). Conclude applying one more time dominated convergence under the dominance condition Assumption 6(iii). ■

We give a specific example where the conditions above are not satisfied. Consider the model $Y_i = X_i + \alpha_i$. Then, $s(Y_i, X_i) = 1(Y_i \leq X_i)$ solves (2) with $r(\alpha) = 1(\alpha \leq 0)$, which is discontinuous at 0. This is of course an unrealistic model, but the idea is simply to illustrate which of our assumptions is key for the results to hold. In this example, Assumption 6(i-ii) is satisfied under mild conditions, since $\eta_{m,x}(m(x,\alpha)) = \eta_0(\alpha)$, but the integrability condition (30) fails, since for $s(Y_i, X_i) = 1(Y_i \leq X_i)$

$$\begin{aligned}\int |\hat{g}(t)| \sup_{\alpha\in\Gamma_0} \left|\overline{\hat{w}(t,\alpha)}\right| dt &= \inf_{\alpha\in\Gamma_0} \eta_0(\alpha) \int |\hat{g}(t_1)| \, dt_1 \\ &= \infty,\end{aligned}$$

where $\hat{g}(t_1) = \int 1(\alpha \leq 0)\eta_0(\alpha)e^{it_1\alpha}d\alpha$. Note that the discontinuity implies the lack of integrability.

9.2 Identification under Kotlarski's Assumptions

There is a growing literature in econometrics identifying the distribution of latent variables by means of Kotlarski's Lemma (see Prakasa Rao (1983) for a description of the method). In this setting we observe $Z = (Y_1, Y_2)$ satisfying

$$Y_1 = \alpha_1 + \alpha_2$$
$$Y_2 = \alpha_1 + \alpha_3,$$

where $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$ is a vector of UH with independent components, and (with some abuse of notation) Lebesgue densities η_{0j}, for $j = 1, 2, 3$. The density of the data is given by

$$\begin{aligned} f_{\eta_0}(y_1, y_2) &= \int 1(y_1 = \alpha_1 + \alpha_2) 1(y_2 = \alpha_1 + \alpha_3) \eta_{01}(\alpha_1) \eta_{02}(\alpha_2) \eta_{03}(\alpha_3) d\alpha_1 d\alpha_2 d\alpha_3 \\ &= \int \eta_{02}(y_1 - \alpha_1) \eta_{03}(y_2 - \alpha_1) \eta_{01}(\alpha_1) d\alpha_1. \end{aligned}$$

Consider a parametric submodel where η_{02} and η_{03} are known and continuous. The model reduces then to our original setting where $f_{z/\alpha}(z) = \eta_{02}(y_1 - \alpha)\eta_{03}(y_2 - \alpha)$ is known and continuous in α. If the dominance condition of Lemma 3.2 is satisfied, then the CDF and quantiles of η_{01} will be irregularly identified.

9.3 A Canonical Model of Infectious Diseases

The canonical Heterogenous Mixing model of infectious diseases, see e.g. Geoffard and Philipson (1995) and references therein, gives rise to the specification of the conditional hazard

$$h_{y/x,\alpha}(y) = \beta(x, \alpha)\psi(y),$$

where the probability that an infected individual of type (x_1, α_1) will infect a susceptible individual of class (x_2, α_2) is

$$\beta(x_1, \alpha_1) \times \beta(x_2, \alpha_2)$$

(the so called factorized matching assumption) and $\psi(y)$ is a baseline hazard function satisfying

$$\psi(y) = \int \beta(x, \alpha) I_{y/x,\alpha}(y) d\eta_0(x, \alpha), \tag{31}$$

where $I_{y/x,\alpha}$ is the proportion of infected individuals at time t in class (x, α) with distribution $\eta_0(x, \alpha)$. It is often assumed that $\beta(x, \alpha) = \phi(x)\alpha$. This model is then like a Mixed Proportional Hazard model but with restrictions on the baseline hazard. The conditional density of Y given X given by

$$f_{\eta_0}(y, x) = \int \phi(x)\psi(y)\alpha e^{-\phi(x)\Psi(y)\alpha} d\eta_0(\alpha),$$

where we follow the notation and setting of the main text. In submodel where $\phi(x)$ and $\psi(y)$ are known, with $\psi(y)$ following the restriction (31), the model fits our original formulation with

$f_{z/\alpha}(z) = \phi(x)\psi(y)\alpha e^{-\phi(x)\Psi(y)\alpha}$ known and continuous as a function of α. It remains to verify the dominance condition. By a simple argument, with $\Gamma_0 = (\alpha_1, \alpha_2)$ and $\alpha_1 > 0$,

$$\int |s(z)| \sup_{\alpha \in (\alpha_1,\alpha_2)} \phi(x)\psi(y)\alpha e^{-\phi(x)\Psi(y)\alpha} d\mu(z) \leq \frac{\alpha_2}{\alpha_1} \int |s(z)| \, \phi(x)\psi(y)\alpha_1 e^{-\phi(x)\Psi(y)\alpha_1} d\mu(z) < \infty.$$

This shows the irregularity of the CDFs and quantiles of η_0. In related research, Horowitz (1999) has established very slow rates of convergence (logarithmic) for the CDF of α in the standard Mixed Proportional Hazard model. We note the irregularity holds even when $\phi(x)$ is known and $\psi(y)$ satisfies further restrictions. Our simple proof reveals that this is a generic feature of heterogeneous mixing models where $\alpha \to \beta(x, \alpha)$ is smooth.

9.4 Anatomy of the general problem

The necessary condition for regular estimation in van der Vaart (1991) is quite general, and in its abstract form reads as

$$\tilde{\psi} \in R(A^*),$$

where $\tilde{\psi}$ is the so-called gradient, which for our original moment functional is $\tilde{\psi}(\alpha) = r(\alpha) - \phi(\eta_0)$, and A^* is the adjoint of the so-called score operator A. In many semiparametric models, A^* is a smoothing integral operator, in the sense that

$$A^* s = \int s(z) k(z, \alpha) d\mu(z)$$

is an operator from L_2 to $L_2(\eta_0)$ with a kernel function k such that $\alpha \to k(z, \alpha)$ is smooth, at least for some submodel. We expect our results to be applicable in this general setting.

References

ADUSUMILLI, K., D. KURISU, T. OTSU AND Y-J WHANG (2020): "Inference on Distribution Functions Under Measurement Error," *Journal of Econometrics*, 215, 131-164.

ALVAREZ, F., BOROVIČKOVÁ, K. AND R. SHIMER (2016): "Decomposing Duration Dependence in a Stopping Time Model," unpublished manuscript.

ARELLANO, M. AND S. BONHOMME (2012): "Identifying Distributional Characteristics in Random Coefficients Panel Data Models," *Review of Economic Studies,* 79, 987-1020.

BAJARI, P., FOX, J.T., RYAN, S., (2007): "Linear Regression Estimation of Discrete Choice Models with Nonparametric Distributions of Random Coefficients," *American Economic Review*, 72, 459-463.

BEGUN, J. M., W. J. HALL, W. M. HUANG, AND J. A. WELLNER (1983): "Information and Asymptotic Efficiency in Parametric-Nonparametric Models," *The Annals of Statistics*, 11, 432-452.

BERAN, R. AND P. HALL (1992): "Estimating Coefficient Distributions in Random Coefficient Regressions," *The Annals of Statistics*, 20, 1970-1984.

BERAN, R., FEUERVERGER, A., AND HALL, P. (1996): "On Nonparametric Estimation of Intercept and Slope Distributions in Random Coefficient Regression", *The Annals of Statistics*, 24, 2569-2592.

BESTER, A., AND C. HANSEN (2007): "Flexible Correlated Random Effects Estimation in Panel Models with Unobserved Heterogeneity," unpublished manuscript.

BICKEL, P. J., C. A. J. KLASSEN, Y. RITOV, AND J. A. WELLNER (1998): *Efficient and Adaptive Estimation for Semiparametric Models*. New York: Springer-Verlag.

BLUNDELL, R., X. CHEN, AND D. KRISTENSEN (2007): "Semi-nonparametric IV Estimation of Shape-invariant Engel Curves, " *Econometrica*, 75, 1613-1670.

BONHOMME, S. (2011): "Panel Data, Inverse Problems, and the Estimation of Policy Parameters," Unpublished manuscript.

BOYD, J.H., AND MELLMAN, R.E. (1980): "Effect of Fuel Economy Standards on the US Automotive Market: An Hedonic Demand Analysis," *Transportation Research B* 14(5), 367-378.

Briesch, R. A., P. K. Chintagunta and R. L. Matzkin (2010): "Nonparametric Discrete Choice Models With Unobserved Heterogeneity," *Journal of Business & Economic Statistics*, 28, 291-307.

Cardell, N.S. and Dunbar, F.C. (1980): "Measuring the Societal Impacts of Automobile Downsizing," *Transportation Research B* 14(5), 423-434.

Chamberlain, G. (1986): "Asymptotic Efficiency in Semi-Parametric Models with Censoring," *Journal of Econometrics*, 34, 305-334.

Chamberlain, G. (1992): "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60, 567-596.

Chen, J. (1995): "Optimal Rate of Convergence for Finite Mixture Models," *The Annals of Statistics*, 23, 221-233.

Chen, X., (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in: *Handbook of Econometrics*, vol. 7. Elsevier.

Chen, X. and Z. Liao (2014): "Sieve M-Inference of Irregular Parameters", *Journal of Econometrics*, 182, 70-86.

Chernozhukov, V., J. C. Escanciano, H. Ichimura, W. K. Newey, and J. M. Robins (2016): "Locally Robust Semiparametric Estimation," arXiv preprint arXiv:1608.00033.

Dattner, I., Goldenshluger, A., Juditsky, A. (2011): "On Deconvolution of Distribution Functions," *Annals of Statistics* 39, 2477-2501.

Elbers, C. and G. Ridder (1982): "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model," *Review of Economic Studies*, 49, 403-409.

Engl, H. W., M. Hanke, and A. Nuebauer (1996): *Regularization of Inverse Problems.* Dordrecht: Kluwer Academic Publishers.

Escanciano, J.C., (2020): "Semiparametric Identification and Fisher Information", unpublished manuscript.

Fan, J. (1991): "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems," *Annals of Statistics* 19, 1257-1272.

Florens, J.P., J.J. Heckman, C. Meghir and E. Vytlacil, (2008): "Identification of Treatment Effects Using Control Functions in Models with Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica*, 76, 1191-1206.

Fox, J. T., K.-I. Kim, S. P. Ryan, and P. Bajari (2011): "A Simple Estimator for the Distribution of Random Coefficients," *Quantitative Economics*, 2, 381-418.

Fox, J. T., K.-I. Kim, S. P. Ryan, and P. Bajari (2012): "The Random Coefficients Logit Model Is Identified," *Journal of Econometrics*, 166 (2), 204-212.

Fox, J. T., K.-I. Kim, C. Yang (2016): "A Simple Nonparametric Approach to Estimating the Distribution of Random Coefficients in Structural Models," *Journal of Econometrics*, 195, 236-254.

Gautier, E. and Y. Kitamura (2013): "Nonparametric Estimation in Random Coefficients Binary Choice Models," *Econometrica*, 81, 581-607.

Geoffard, P-Y. and T. Philipson (1995): "The Empirical Content of Canonical Models of Infectious Diseases: The Proportional Hazard Specification," *Biometrika*, 82, 101-111.

Graham, B. W., and J. L. Powell (2012): "Identification and Estimation of Average Partial Effects in 'Irregular' Correlated Random Coefficient Panel Data Models," *Econometrica*, 80, 2105-2152.

Groeneboom P. and Wellner J.A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. DMV Seminar, vol 19. Birkhäuser, Basel.

Hausman, J.A., and W.K. Newey (1995): "Nonparametric Estimation of Exact Consumer Surplus
and Deadweight Loss," *Econometrica*, 63, 1445-1476.

Hausman, J.A., and W.K. Newey (2016): "Individual Heterogeneity and Average Welfare," *Econometrica* 84, 1225-1248.

Heckman, J. J. (2001): "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture," *Journal of Political Economy*, 109(4), 673-748.

Heckman, J.J. and B. Singer (1984a): "The Identifiability of the Proportional Hazard Model", *Review of Economic Studies*, 51,231-241.

Heckman, J.J. and B. Singer (1984b): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271-320.

Heckman, J.J., Smith, J. and N. Clements (1993): "Making The Most Out of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64, 487-535.

Heckman, J.J. and E. Vytlacil (1998): "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return is Correlated with Schooling," *Journal of Human Resources*, 33, 974-987.

Heckman, J.J. and E. Vytlacil (2005): "Structural Equations, Treatment, Effects and Econometric Policy Evaluation," *Econometrica*, 73, 669-738.

Heinrich, P. and J. Kahn (2018): "Optimal Rates for Finite Mixture Estimation," *The Annals of Statistics*, 46, 2844-2870.

Hildreth, C. and J.P. Huock (1968): "Some Estimators for a Linear Model with Random Coefficients," *Journal of the American Statistical Association,* 63, 584-92.

Hille, E. and J. D. Tamarkin, (1931): "On the Characteristic Values of Linear Integral Equations," *Acta Mathematica*, 57, 1-76.

Hirano, K. and Porter, J.R. (2012): "Impossibility Results for Nondifferentiable Functionals," *Econometrica*, 80, 1769-1790.

Hoderlein, S., H. Holzmann, and A. Meister (2017): "The Triangular Model with Random Coefficients," *Journal of Econometrics*, 201, 144-169.

Hoderlein, S., S. Klemela, and E. Mammen (2010): "Analyzing the Random Coefficient Model Nonparametrically," *Econometric Theory*, 26, 804-837.

Hoderlein S. and Sherman, B. (2015): "Identification and estimation in a correlated random coefficients binary response model," *Journal of Econometrics*, 188, 135-149.

Horowitz, J. L. (1999): "Semiparametric Estimation of a Proportional HazardModel with Unobserved Heterogeneity," *Econometrica*, 67, 1001-1028.

Hu, Y., and S. M. Schennach (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76 (1), 195-216.

Hurwicz L. (1950): "Generalization of the Concept of Identification," In Statistical Inference in Dynamic Economic Models, ed. T.C. Koopmans. New York: Wiley.

Ichimura, H., and T. Thompson (1998): "Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution," *Journal of Econometrics*, 86, 269-295.

Imbens, G., and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 61, 467-476.

Khan, S. and E. Tamer (2010): "Irregular Identification, Support Conditions, and Inverse Weight Estimation," *Econometrica*, 6, 2021-2042.

Khuri, A. and G. Casella (2002): "The Existence of the First Negative Moment Revisited", *The American Statistician,* 56, 44-47.

Lewbel, A. and K. Pendakur (2017): "Unobserved Preference Heterogeneity in Demand Using Generalized Random Coefficients", *Journal of Political Economy*, 125, 1100-1148.

Lewbel, A. (2019): "The Identification Zoo – Meanings of Identification in Econometrics" *Journal of Economic Literature*, 57(4), 835-903.

Masten, M.A. (2017): "Random Coefficients on Endogenous Variables in Simultaneous Equations Models", *Review of Economic Studies*, 85, 1193 1250.

Masten, M.A. and A. Torgovitsky (2016): "Identification of Instrumental Variables Correlated Random Coefficients Models," *The Review of Economics and Statistics*, 98, 1001-1005.

Matzkin, R. L. (2007). Nonparametric Identification. In J.J. Heckman and E.E. Leamer (eds.), Handbook of Econometrics, Vol. 6b, pp. 5307-5368, Elsevier, New York.

Matzkin, R. L. (2013): "Nonparametric Identification in Structural Economic Models," *Annual Review of Economics*, 5.

Newey, W. K. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99-135.

Newey, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica,* 62, 1349-1382.

Prakasa Rao B. L. S. (1983). *Nonparametric Functional Estimation.* New York: Acad. Press.

Ramm, A. G. and A. I. Katsevich (1996). *The radon transform and local tomography.* CRC Press.

Royden, H. L. (1968). *Real Analysis.* Second Edition, The Macmillan Company, New York.

Rudin, W. (1987). *Real and complex analysis.* 3rd ed., McGraw-Hill, New York.

Santos, A. (2011): "Instrumental Variable Methods for Recovering Continuous Linear Functionals," *Journal of Econometrics*, 161(2), 129–146.

Severini, T. A., and G. Tripathi (2006): "Some Identification Issues in Nonparametric Linear Models with Endogenous Regressors," *Econometric Theory*, 22(2), 258–278.

Severini, T. A., and G. Tripathi (2001): "A Simplified Approach to Computing Effciency Bounds in Semi-
parametric Models," *Journal of Econometrics*, 102, 23-66.

Severini, T. A., and G. Tripathi (2012): "Efficency Bounds for Estimating Linear Functionals of Nonparametric Regression Models with Endogenous Regressors," *Journal of Econometrics*, 170(2), 491-498.

Shen, X. (1997): "On Methods of Sieves and Penalization," *The Annals of Statistics*, 25, 2555-2591.

Smith, K.T., D. C. Solmon and S. L. Wagner (1977): "Practical and Mathematical Aspects of the Problem of Reconstructing Objects from Radiographs," *Bulletin of the American Mathematical Society* 83, 1227-1270.

Söhl, J., Trabs, M. (2012): "A Uniform Central Limit Theorem and Efficiency for Deconvolution Estimators," *Electronic Journal of Statistics*, 6, 2486-2518.

Swamy, P.A.V.B. (1970): "Efficient Inference in a Random Coefficient Model," *Econometrica*, 38, 311-23.

Trabs, M. (2015): "Information Bounds for Inverse Problems with Application to Deconvolution and Lévy Models," *Annales de l'I.H.P. Probabilités et statistiques*, 51, 1620-1650.

van der Vaart, A. W. (1991): "On Differentiable Functionals," *The Annals of Statistics*, 19, 178-204.

van der Vaart, A. W. (1998). *Asymptotic Statistics*, vol. 3 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

Wooldridge, J. (1997): "On Two Stage Least Squares Estimation of the Average Treatment Effect in a Random Coefficient Model," *Economics Letters*, 56, 129-133.

Wooldridge, J. (2003): "Further Results on Instrumental Variables Estimation of Average Treatment Effects in the Correlated Random Coefficient Model," *Economics Letters*, 79, 185-191.

Wooldridge, J. (2008): "Instrumental Variables Estimation of the Average Treatment Effect in the Correlated Random Coefficient Model," in Modelling and Evaluating Treatment Effects in Econometrics (*Advances in Econometrics*, Volume 21), ed. by D. Millimet, J. Smith, and E. Vytlacil, Emerald Group Publishing Limited.