

# On the identification of structural linear functionals

---

Juan Carlos Escanciano  
Wei Li

The Institute for Fiscal Studies  
Department of Economics, UCL

**cemmap** working paper CWP48/13

# On the Identification of Structural Linear Functionals <sup>\*</sup>

Juan Carlos Escanciano<sup>†</sup>  
Indiana University

Wei Li <sup>‡</sup>  
Indiana University

October 2nd, 2013

## Abstract

This paper asks which aspects of a structural Nonparametric Instrumental Variables Regression (NPIVR) can be identified well and which ones cannot. It contributes to answering this question by characterizing the identified set of linear continuous functionals of the NPIVR under norm constraints. Each element of the identified set of NPIVR can be written as the sum of a common “identifiable component” and an idiosyncratic “unidentifiable component”. The identified set for any continuous linear functional is shown to be a closed interval, whose midpoint is the functional applied to the “identifiable component”. The formula for the length of the identified set extends the popular omitted variables formula of classical linear regression. Some examples illustrate the wide applicability and utility of our identification result, including bounds and a new identification condition for point-evaluation functionals. The main ideas are illustrated with an empirical application of the effect of children on labor-market outcomes for women.

**Keywords:** Instrumental Variables; Nonparametric; Semiparametric; Bounds; Linear functional.

**JEL classification:** C26; C14; C21.

---

<sup>\*</sup>We thank Joachim Freyberger and Joel Horowitz for providing the data used in our application, and Stefan Hoderlein for helpful comments.

<sup>†</sup>Department of Economics, Indiana University, 349 Wylie Hall, 100 South Woodlawn Avenue, Bloomington, IN 47405–7104, USA. E-mail: [jescanci@indiana.edu](mailto:jescanci@indiana.edu). Web Page: <http://mypage.iu.edu/~jescanci/>. Research funded by the Spanish Plan Nacional de I+D+I, reference number ECO2012-33053.

<sup>‡</sup>Department of Economics, Indiana University, 105 Wylie Hall, 100 South Woodlawn Avenue, Bloomington, IN 47405–7104, USA. E-mail: [w15@indiana.edu](mailto:w15@indiana.edu)

# 1 Introduction

Consider a nonparametric instrumental variables (IV) setting, where the dependent variable  $Y$  is related to the vector  $X$  through the equation

$$Y = g(X) + \varepsilon, \tag{1}$$

where  $E[\varepsilon|Z] = 0$ , almost surely (a.s), for a vector of instruments  $Z$ . The nonparametric IV Regression (NPIVR) model (1) is different from the classical regression model in that  $X$  is allowed to be endogenous, that is,  $E[\varepsilon|X] \neq 0$  with positive probability, a condition that is likely to arise in many economic applications. The setting in (1) allows for  $X$  and  $Z$  to have common components (i.e. some variables to be exogenous). Suppose  $g \in \mathcal{G}$ , where  $\mathcal{G}$  is a Hilbert space of measurable functions of  $X$  with values in  $\mathbb{R}$ , equipped with the norm  $\|\cdot\|_{\mathcal{G}}$ ; see examples below. We consider the parameter space  $\Theta := \{g \in \mathcal{G} : \|g\|_{\mathcal{G}} \leq B\}$ , for a positive constant  $B > 0$ . Identification of  $g$  in (1) (i.e. uniqueness of a solution to (1) in  $\Theta$ ) requires far more stringent conditions in the nonparametric case investigated here than in typical parametric settings; see [Newey and Powell \(2003\)](#).<sup>1</sup> Furthermore, even when these strong identification conditions hold, the best estimator of  $g$ , in a minimax rate sense, can have very slow rates of convergence.<sup>2</sup> These arguments motivate us to investigate identification of aspects of  $g$ , rather than  $g$  itself, in the form of  $Lg$ , where  $L$  is a generic linear continuous functional on  $(\mathcal{G}, \|\cdot\|_{\mathcal{G}})$ . Numerous examples of linear functionals are provided below, including the point-evaluation functional  $Lg = g(x_0)$ , for some  $x_0$  in the support of  $X$ , which shows the wide applicability of our approach. In this paper, we characterize the sharp identified set for  $Lg$ , where  $g$  satisfies (1) and  $g \in \Theta$ .

Parameter spaces with norm constraints have been routinely used in the context of (1) for various choices of  $\|\cdot\|_{\mathcal{G}}$ ; see [Newey and Powell \(2003\)](#), [Blundell, Chen and Kristensen \(2007\)](#), [Horowitz \(2011\)](#) and [Santos \(2012\)](#), among many others. These constraints are often applied with norms  $\|\cdot\|_{\mathcal{G}}$  that involve derivatives, so that highly variable solutions  $g$  are ruled out. Under point identification of  $g$ , norm constraints do not play any role in inference, provided the point identified function  $g$  satisfies  $\|g\|_{\mathcal{G}} < B$ . However, under partial identification, norm constraints play a fundamental role in identification and inference, as we show in this paper. In particular, we investigate the impact of  $B$  on inferences under partial identification, and show that the midpoint of the identified set for  $Lg$  is independent of  $B$ . Moreover, we find necessary and sufficient conditions under which set inferences are insensitive to  $B$ .

The nonparametric IV model in (1) has been a subject of much recent research in econometrics. The literature has mainly focused on conditions for point identification and estimation under point identification; see the aforementioned references. An exception to requiring point identification is [Santos \(2012\)](#), who discusses nonparametric inference under partial identification of  $g$ . Our focus here is on semiparametric aspects of  $g$ , rather than nonparametric ones. The papers most closely related to ours are those by [Severini and Tripathi \(2006, 2012\)](#), [Santos \(2011\)](#) and, more recently, [Freyberger and Horowitz \(2013\)](#). [Severini and Tripathi \(2006, 2012\)](#) provided necessary and sufficient conditions

---

<sup>1</sup>Essentially, it is required that any transformation of the instrument  $Z$  is a relevant instrument for *any* non-constant transformation of  $X$ ; see Lemma 2.1 in [Severini and Tripathi \(2006\)](#).

<sup>2</sup>See, for instance, [Hall and Horowitz \(2005\)](#), [Blundell, Chen and Kristensen \(2007\)](#), [Darolles, Fan, Florens and Renault \(2011\)](#), [Chen and Reiss \(2011\)](#) and [Chen and Pouzo \(2012\)](#), to mention just a few.

for identification of  $Lg$ , and a necessary condition for its  $\sqrt{n}$ -consistent estimation, where  $n$  is the sample size. Furthermore, they also established the efficiency bounds for estimating  $Lg$  when it is identified. Santos (2011) proposed a  $\sqrt{n}$ -asymptotically normal estimator for the point identified  $Lg$ . These papers allow for partial identification of  $g$ , but assume  $Lg$  to be point identified. In contrast, our paper focuses on a partially identified  $Lg$ . Recently and independently of our work, Freyberger and Horowitz (2013) have investigated partial identification and estimation of  $Lg$  when observations have a finite discrete support and when shape restrictions, such as monotonicity, hold. In this paper we allow for continuous and/or discrete variables. Moreover, we analyze for the first time in the literature the case where  $Lg$  is partially identified under norm constraints, and consider a general setting that includes new examples such as, for instance, point-evaluation functionals. Traditional Hilbert spaces considered in the literature cannot handle point-evaluation functionals because they are discontinuous in the underlying topologies. To solve this problem, we use as  $\mathcal{G}$  a Reproducing Kernel Hilbert Space (RKHS), under which point-evaluation functionals are continuous, see for instance Chapters 13 to 15 of Parzen (1967). When the theory is specialized to this example, we obtain a new sufficient condition for point identification of  $g(x_0)$ , and we relate this condition to the classical completeness assumption. For the particular case of discrete variables, our results are complementary to existing ones, as our norm constraints cannot be written as the linear inequality constraints considered by Freyberger and Horowitz (2013). Thus, our results significantly complement and extend existing results, characterizing the factors that determine the identified set of linear continuous functionals under general norm constraints, introducing new examples and shedding some light on the impact of norm constraints on inference under partial identification.

Based on our identification results, we introduce the concept of  $\mathcal{G}$ -consistent estimation of  $g$  in (1). We say a nonparametric estimator  $\hat{g}$  of  $g$  is  $\mathcal{G}$ -consistent when  $L\hat{g}$  consistently estimates the midpoint of the identified set for  $Lg$ , for *all* linear continuous functionals  $L : \mathcal{G} \rightarrow \mathbb{R}$ .<sup>3</sup>  $\mathcal{G}$ -consistent estimators are appealing because under point identification they are consistent for the parameter of interest, and under partial identification they are consistent for the mean value of the parameter without further information (i.e. with an uninformative prior on the identified set). See also Song (2013) for an optimality theory of inference based on the midpoint. In some sense,  $\mathcal{G}$ -consistent estimators are robust to lack of point identification. This consistency concept can be extended to more general situations with parameters that are identified in an interval; see Manski (2003, 2007), Tamer (2010) and references therein for numerous examples. Our arguments below lead to natural candidates for  $\mathcal{G}$ -consistent estimators; see Section 3. Nevertheless, a companion paper to this paper will deal with estimation of the identified set characterized here, and in particular, it will investigate  $\mathcal{G}$ -consistent estimators and their asymptotic properties.

The following notations are used throughout the paper. Henceforth,  $A'$ ,  $tr(A)$  and  $|A| := (tr(A'A))^{1/2}$  denote the transpose, trace and the Euclidean norm of a matrix  $A$ , respectively. The symbol  $:=$  denotes definitional relation. For a real-valued function  $h(x)$ , denote by  $h^{(j)}(x) := \partial^j h(x) / \partial x^j$  its  $j^{th}$  deriva-

---

<sup>3</sup>An extension of this concept allows for restricted classes of linear functionals. We say  $\hat{g}$  is  $\mathcal{L}$ -consistent when  $L\hat{g}$  consistently estimates the midpoint of the identified set for  $Lg$ , for linear continuous functionals  $L : \mathcal{G} \rightarrow \mathbb{R}$  in the class  $\mathcal{L}$ . Hence,  $\mathcal{G}$ -consistent estimators are  $\mathcal{L}$ -consistent estimators with  $\mathcal{L}$  equals the class of all linear continuous functionals.

tive, with  $h^{(0)}(x) := h(x)$ . For a generic random variable  $W$ ,  $L_2(W)$  denotes the space of real-valued (measurable) functions of  $W$  that are square integrable, i.e.  $f(W)$ , such that  $\|f\|^2 := E[f^2(W)] < \infty$ .  $L_2(W)$  is a Hilbert space with inner product  $\langle f, g \rangle := E[f(W)g(W)]$ . Let  $\mathcal{S}_W$  denote the support of  $W$ . Notice the abuse of notation, as  $\|f\|$  and  $\langle f, g \rangle$  depend on the distribution of  $W$ . We hope this will not cause confusion. We assume, for simplicity,  $\mathcal{G} \subset L_2(X)$  and denote by  $\langle f, g \rangle_{\mathcal{G}}$  the inner product associated to  $\|\cdot\|_{\mathcal{G}}$ . For a linear operator  $K : \mathcal{G}_1 \rightarrow \mathcal{G}_2$  between two Hilbert spaces, denote the subspaces  $\mathcal{R}(K) := \{f \in \mathcal{G}_2 : \exists s \in \mathcal{G}_1, Ks = f\}$  (range of  $K$ ) and  $\mathcal{N}(K) := \{f \in \mathcal{G}_1 : Kf = 0\}$  (kernel of  $K$ ). Let  $\mathcal{D}(K)$  denote the domain of definition of  $K$ . For a subspace  $V \subset \mathcal{G}$ ,  $V^\perp$  and  $\bar{V}$  denote, respectively, its orthogonal complement and closure, with respect to (wrt) the norm topology, in  $\mathcal{G}$ . Henceforth, for a closed subspace  $V$ ,  $P_V$  denotes its orthogonal projection operator. We will extensively use basic results from operator theory and Hilbert spaces. See [Carrasco, Florens and Renault \(2006\)](#) for an excellent review of these results.

The remainder of the paper is organized as follows. Section 2 illustrates the general applicability of our setting with some examples. Section 3 establishes our main result, the characterization of the identified set for  $Lg$ . This result is discussed in the context of previous examples in Section 4. This section also contains an empirical application of the effect of children on labor-market outcomes for women, using data from [Angrist and Evans \(1998\)](#). Finally, Section 5 concludes and discusses further extensions. Some tables and figures, as well as some asymptotic results for discrete variables, are gathered in an Appendix.

## 2 Examples

Here we illustrate our theory with several examples. Some of these examples and further examples can be found in, e.g., [Stock \(1989\)](#), [Brown and Newey \(1998\)](#), [Ai and Chen \(2003\)](#), [Chen and Pouzo \(2009, 2012\)](#), [Severini and Tripathi \(2006\)](#), [Santos \(2011\)](#) and [Severini and Tripathi \(2012\)](#).

**Example 1 (Engel Curves):** Let  $Y$  denote household's expenditure share on food, let  $X$  be the logarithm of the household's total expenditures, and let  $Z$  be the logarithm of the household's gross earnings. [Blundell, Chen and Kristensen \(2007\)](#) considered an empirical application of (1) within this setting, establishing rates of convergence for an estimator of  $g$ . They provided sufficient conditions for point identification in the form of bounded completeness assumptions. Here, we complement their analysis by discussing methods that are robust to potential identification failures. Let  $\mathcal{G}$  be a subspace of  $L_2(X)$  of differentiable functions with integrable derivatives, and take  $\|\cdot\|_{\mathcal{G}} = \|\cdot\|$ . A functional of interest is the average partial effect

$$Lg = E \left[ g^{(1)}(X) \right].$$

In this application a natural bound is  $B = 1$ . Under mild conditions (see [Powell, Stock, and Stoker \(1989\)](#), Lemma 2.1),  $L$  is a continuous functional on  $\mathcal{G}$ . We will provide below sharp bounds for  $Lg$ . ■

**Example 2 (Consumer Surplus):** Suppose a researcher aims to estimate the approximated consumer surplus for a price change from  $a$  to  $b$ , where  $0 < a < b < \infty$ . Take for example a market for fish,

as in Angrist, Graddy, and Imbens (2000). Here  $Y$  denotes demanded quantity of fish,  $X$  denotes its price and  $g(\cdot)$  represents the demand function. The instrument  $Z$  measures weather conditions at sea. Approximated consumer surplus is computed as

$$Lg = \int_a^b g(x)dx.$$

See, e.g., Newey and McFadden (1994). The operator  $L$  is linear and continuous in  $\mathcal{G}$ , the space of square integrable functions on  $[a, b]$ , and which are zero elsewhere. ■

**Example 3 (Best Linear Approximations):** A linear functional of interest in applied work is the Best Linear Approximation (BLA) functional in  $\mathcal{G} = L_2(X)$ . Here, we focus on a linear combination of the BLA coefficients

$$\beta \equiv \beta(g) = E[XX']^{-1}E[Xg(X)],$$

i.e.,  $Lg = \alpha'\beta$  for some known  $\alpha \in \mathbb{R}^p$ , where  $p$  is the dimension of  $X$ . For instance,  $\alpha = (1, 0, \dots, 0)'$  corresponds to the first component of  $\beta$ . It is well-known that  $\beta$  solves the problem

$$\min_{\gamma \in \mathbb{R}^p} E[(g(X) - \gamma'X)^2],$$

and hence, it contains important structural information. In the context of the application by Angrist and Krueger (1991) on returns to schooling, Newey (2013) shows that under Gaussianity,  $\beta'X$  is the part of  $g$  that can be estimated with certain precision, other aspects of  $g$  are subject to large sample variability. ■

**Example 4 (Point-evaluation Functionals):** To illustrate the wide applicability of our results, consider the point-evaluation functional  $Lg = g(x_0)$ , for some  $x_0$  in the support of  $X$ . This functional is not continuous in the usual Hilbert spaces considered in the literature, e.g.  $\mathcal{G} = L_2(X)$ . However, if we take as  $\mathcal{G}$  a RKHS, see for instance Parzen (1967) and Wahba (1990), then  $L$  is continuous by definition.<sup>4</sup> For example, suppose  $X$  is univariate with  $\mathcal{S}_X = [a, b]$ , and consider  $\mathcal{G}$  as the Sobolev Space

$$W_2^1 := \{g : g \text{ is absolute continuous and } g^{(1)} \in L_2[a, b]\},$$

where  $L_2[a, b]$  is the space of square integrable functions in  $[a, b]$ . The space  $W_2^1$  is endowed with the square norm

$$\|g\|_{W_2^1}^2 := \int_a^b g^2(x)dx + \int_a^b (g^{(1)}(x))^2 dx.$$

Then, it is known that  $W_2^1$  is a Hilbert space and  $Lg = g(x_0)$  a continuous functional on it; see, e.g., Adams (1975). Newey and Powell (2003), Blundell, Chen and Kristensen (2007), Horowitz (2011), Santos (2012), among many others, have used related spaces in the context of NPIVR under norm constraints. This setting can be easily modified to allow for point-evaluation functionals of derivatives

---

<sup>4</sup>RKHS are exactly defined as Hilbert spaces where the evaluation functionals are continuous. So, the concept RKHS is exactly the right tool to use to handle point-evaluation functionals by standard Hilbert space methods.

of  $g$ , multivariate covariates and unbounded supports. Readers are referred to [Berlinet and Thomas-Agnan \(2004\)](#) for numerous examples of RKHS and their properties. ■

**Example 5 (Discrete Variables):** Our results can be applied to discrete variables. To illustrate this in a specific example, consider the application of [Angrist and Evans \(1998\)](#), where  $Y$  is number of weeks a woman works in a year,  $X$  is the number of children a woman has (2,3,4 or 5), and the instrumental variable  $Z$  is binary with  $Z = 1$  indicating the woman's first two children have the same sex and  $Z = 0$  otherwise. This is also the application considered in [Freyberger and Horowitz \(2013\)](#). Suppose one is interested in the increment functional  $L_{23}g = g(3) - g(2)$  or in the probability-weighted increment functional  $L_{P23}g = P(X = 3)g(3) - P(X = 2)g(2)$ . All linear functionals have the form  $Lg = c'g$ , where here we identify  $g$  with the vector  $g = (g(2), \dots, g(5))'$  and  $c = (c_1, \dots, c_4)' \in \mathbb{R}^4$  is an arbitrary vector. In this application,  $\mathcal{G} = \mathbb{R}^4$ , endowed with the usual Euclidean norm, and we can take  $B = 104 (= 52 \times \sqrt{4})$ . Below, we provide bounds and inference for  $Lg$  incorporating the bound  $|g| \leq B$ . The example can be easily extended to any situation where the cardinality in the discrete support of  $X$ , say  $d(X)$ , is larger than the cardinality of the support of  $Z$ , say  $d(Z)$ , which is common in applications. See, for instance, [Angrist and Krueger \(1991\)](#), [Bronars and Grogger \(1994\)](#), [Card \(1995\)](#) and [Lochner and Moretti \(2004\)](#), among many others, for empirical applications where  $d(X) > d(Z)$ . In such situations  $g$  is not identified; see [Newey and Powell \(2003\)](#) and [Freyberger and Horowitz \(2013\)](#). It is straightforward to prove that the dimension of the space of linear functionals that are identified is  $d(Z)$  under strong instruments, which can be much smaller than  $d(X)$ . ■

### 3 Identification

Let  $T : \mathcal{G} \rightarrow L_2(Z)$  be the linear operator  $Tg(z) := E[g(X)|Z = z]$ . Henceforth, we assume  $T$  is continuous wrt  $\|\cdot\|_{\mathcal{G}}$ .<sup>5</sup> Define  $r(z) := E[Y|Z = z]$ , which is assumed to exist and satisfy  $r \in \mathcal{R}(T) \subset L_2(Z)$ . Then, the identified set for  $g$  is  $\mathcal{G}_0 \cap \Theta$ , where recall  $\Theta = \{g \in \mathcal{G} : \|g\|_{\mathcal{G}} \leq B\}$ , and  $\mathcal{G}_0 := \{g \in \mathcal{G} : Tg = r\}$ . That is,  $\mathcal{G}_0 \cap \Theta$  is the set of functions in the parameter space  $\Theta$  that are compatible with the exogeneity conditions of  $Z$ . We will assume that the model is correctly specified (i.e. the identified set is nonempty) and allow  $T$  to be non-injective, so partial identification of  $g$  is permitted. Therefore,  $\mathcal{G}_0$  is not a singleton if  $\mathcal{N}(T) \neq \{0\}$ . Since  $\mathcal{N}(T)$  plays a key role in the identification of  $g$  and  $Lg$ , we provide some intuitive interpretation for this set. Informally speaking,  $\mathcal{N}(T)$  consists of transformations  $h(X) \in \mathcal{G}$  that, although related to  $Y$ , are completely unrelated to the instrument  $Z$ , in the sense that variation in the instruments does not recover any information on  $h(X)$  (by definition  $E[h(X)|Z = z] = 0$  a.s, so  $h(X)$  is uncorrelated with any measurable function of  $Z$ ).

The main result of this paper characterizes the identified set  $\mathcal{L}_0 := \{Lg : g \in \mathcal{G}_0 \cap \Theta\}$ , which is assumed to be non-empty. Since  $L$  is continuous wrt  $\|\cdot\|_{\mathcal{G}}$ , by the Riesz Representation theorem,  $Lg = \langle g, \ell \rangle_{\mathcal{G}}$ , for some  $\ell \in \mathcal{G}$  and all  $g \in \mathcal{G}$ . The function  $\ell$  is called the Riesz representer of  $L$ . Define

<sup>5</sup>If  $\|\cdot\|_{\mathcal{G}} = \|\cdot\|$ , then  $T$  is clearly continuous, as  $\|Tg\|^2 = E[(E(g(X)|Z))^2] \leq E[E(g(X)^2|Z)] = \|g\|^2$ .

$g_0 := \arg \min_{g \in \mathcal{G}_0} \|g\|_{\mathcal{G}}$  and assume  $\|g_0\|_{\mathcal{G}} \leq B$ . It is well-known that  $g_0$  is uniquely defined.<sup>6</sup> It is also known that  $g_0 = T^\dagger r$ , where  $T^\dagger$  denotes the Moore-Penrose pseudoinverse of  $T$  (see Engl, Hanke and Neubauer (1996), p. 33).<sup>7</sup> The function  $g_0$  has an important structural interpretation. For any  $g \in \mathcal{G}_0$ ,  $g_0 = P_{\mathcal{N}(T)^\perp} g$ , so that

$$g = \underbrace{g_0}_{\text{Identified}} + \underbrace{P_{\mathcal{N}(T)} g}_{\text{Unidentified}}.$$

Following Severini and Tripathi (2006), we call  $g_0$  the ‘‘identifiable part’’ of  $g$  and  $P_{\mathcal{N}(T)} g$  the ‘‘unidentifiable part’’. As we will show, the function  $g_0$  plays a fundamental role in the partial identification of linear structural functionals. Figure 1 in the Appendix illustrates the geometry of the problem. The set  $\mathcal{G}_0$  is a linear variety (an affine hyperplane),  $g_0$  is the orthogonal projection of the origin onto  $\mathcal{G}_0$ . In this plot,  $\mathcal{G}_0$  intersects  $\|g\|_{\mathcal{G}} = B$  in the points  $g_{\min}$  and  $g_{\max}$ , respectively, and these points are such that  $Lg_{\min} = \min_{g \in \mathcal{G}_0, \|g\|_{\mathcal{G}} \leq B} Lg$  and  $Lg_{\max} = \max_{g \in \mathcal{G}_0, \|g\|_{\mathcal{G}} \leq B} Lg$ . The identified set for  $Lg$  is the interval  $[Lg_{\min}, Lg_{\max}]$  and  $Lg_0$  its midpoint. It is evident from this plot that the only identified functional is that where  $\ell$  is collinear with  $g_0$ , i.e. orthogonal to  $\mathcal{N}(T)$ . Any other case leads to partial identification of  $Lg$ . The following theorem formally characterizes  $\mathcal{L}_0$ .

**Theorem 3.1** *Assume  $T$  is continuous wrt  $\|\cdot\|_{\mathcal{G}}$ ,  $\|g_0\|_{\mathcal{G}} \leq B$  and that (1) holds. Then, the identified set  $\mathcal{L}_0$  is a closed interval with midpoint  $Lg_0$  and radius*

$$\rho := (B - \|g_0\|_{\mathcal{G}}) \|P_{\mathcal{N}(T)} \ell\|_{\mathcal{G}}.$$

**Proof.** Continuity of  $T$  implies that its kernel  $\mathcal{N}(T)$  is a closed subspace. In this proof, we take  $\mathcal{N}$  to mean  $\mathcal{N}(T)$ . Notice also  $\mathcal{G}_0 = g_0 + \mathcal{N}$ , so  $\mathcal{G}_0$  is a closed linear variety. The existence and uniqueness of  $g_0$  is guaranteed by Theorem 1 in Luenberger (1997) p. 64. We then can write

$$\begin{aligned} \sup_{g \in \mathcal{G}_0, \|g\|_{\mathcal{G}} \leq B} Lg &= Lg_0 + \sup_{g \in \mathcal{N}, \|g_0 + g\|_{\mathcal{G}} \leq B} Lg \\ &= Lg_0 + \sup_{g \in \mathcal{N}, \|g\|_{\mathcal{G}} \leq B - \|g_0\|_{\mathcal{G}}} Lg \\ &= Lg_0 + (B - \|g_0\|_{\mathcal{G}}) \sup_{g \in \mathcal{N}, \|g\|_{\mathcal{G}} \leq 1} Lg \\ &= Lg_0 + (B - \|g_0\|_{\mathcal{G}}) \min_{h \in \mathcal{N}^\perp} \|\ell - h\|_{\mathcal{G}}. \end{aligned}$$

The third equality holds by Pythagorean theorem and the fact that  $g_0$  orthogonal to  $\mathcal{N}$  (Luenberger (1997) p. 64). The last equality holds by Theorem 2 in Luenberger (1997) p. 121. Similarly, following the fact that

$$\inf_{g \in \mathcal{N}, \|g\|_{\mathcal{G}} \leq B} Lg = - \sup_{g \in \mathcal{N}, \|g\|_{\mathcal{G}} \leq B} L(-g)$$

<sup>6</sup>In the mathematical literature  $g_0$  is called the best-approximate solution to  $Tg = r$ , typically allowing for misspecification. See Engl, Hanke and Neubauer (1996).

<sup>7</sup>For other equivalent representations of  $g_0$  see the Table 1 in the Appendix.



we have

$$\inf_{g \in \mathcal{G}_0, \|g\|_{\mathcal{G}} \leq B} Lg = Lg_0 - (B - \|g_0\|_{\mathcal{G}}) \min_{h \in \mathcal{N}^{\perp}} \|\ell - h\|_{\mathcal{G}}.$$

Finally, we show that the supremum and infimum can be attained. Since  $T$  is continuous,  $\mathcal{G}_0$  in fact is a closed hyperplane and therefore it is convex. Together with the fact that  $\{g : \|g\| \leq B\}$  is convex, closed and bounded, we have  $\mathcal{G}_0 \cap \{g : \|g\| \leq B\}$  is convex, closed and bounded. By Mazur's Theorem, this set is also weakly closed (i.e, closed in weak topology). Since  $L$  is a continuous linear functional, it is clearly weakly (sequentially) upper and lower semicontinuous. Therefore  $L$  can attain its maximum and minimum over  $\mathcal{G}_0 \cap \{g : \|g\| \leq B\}$  (Theorem 7.3.5 in [Kurdila and Zabrankin \(2005\)](#)). ■

Theorem 3.1 emphasizes the importance of  $g_0$  in partial identification; functionals of this function correspond to the midpoint of the identified set. The length of the identified set for  $Lg$  depends on two factors: the magnitude of  $g_0$  relative to  $B$ ; and a factor that depends on the relevance of the unidentified factors  $\mathcal{N}(T)$  in predicting  $\ell$ . The first factor  $B - \|g_0\|_{\mathcal{G}}$  is independent of the functional  $L$ . In particular, in the extreme case,  $B = \|g_0\|_{\mathcal{G}}$  is a sufficient condition for point identification of *any* continuous functional, a result implicit in [Santos \(2012\)](#). This factor formalizes the intuition that the larger is the identifiable part of  $g$ , the more information we have. The presence of the second factor confirms previous results in the literature and it is specific to the functional considered. This factor is independent of the norm constraints. Moreover,  $\partial\rho/\partial B = \|P_{\mathcal{N}(T)}\ell\|_{\mathcal{G}}$ . Therefore, this second factor measures the sensitivity of the identified set (specifically, the radius) wrt  $B$ , and as such, it provides useful information in applications where there is no natural choice of  $B$ . Point identification holds if  $\ell \in \mathcal{N}(T)^{\perp}$ , a result given first by [Severini and Tripathi \(2006\)](#). Also, when  $B = \infty$ , we see that, except in the latter case, there is no information on the linear functional  $Lg$ , which is consistent with a similar finding in the discrete case by [Freyberger and Horowitz \(2013\)](#). One can also see that the condition  $\mathcal{N}(T) = \{0\}$  simply reiterates the fact that point identification of  $g$  is equivalent to point identification of  $Lg$  for *all*  $L$ , when  $\|g_0\|_{\mathcal{G}} < B$ . To provide some intuition for the second factor, write  $g = P_{\mathcal{N}(T)^{\perp}}g + P_{\mathcal{N}(T)}g \equiv g_0 + g_u$  and  $\ell = P_{\mathcal{N}(T)^{\perp}}\ell + P_{\mathcal{N}(T)}\ell$ . Then,  $Lg_u = \langle g_u, P_{\mathcal{N}(T)}\ell \rangle_{\mathcal{G}}$  represents the identification bias  $Lg - Lg_0$ , and, subject to the normalization  $\|g_u\|_{\mathcal{G}} = 1$ , this bias attains the maximum absolute value of  $\|P_{\mathcal{N}(T)}\ell\|_{\mathcal{G}}$ . This identification bias formula generalizes the omitted variable bias of classical linear regression, see Example 3 below. From these arguments, we easily obtain the bound  $|Lg - Lg_0| \leq (B - \|g_0\|_{\mathcal{G}}) \|P_{\mathcal{N}(T)}\ell\|_{\mathcal{G}}$  by Cauchy-Schwartz inequality. The equality is attained at elements in the identified set  $\mathcal{G}_0$  with  $\|g\|_{\mathcal{G}} = B$  and which are collinear with  $P_{\mathcal{N}(T)}\ell$ . The set is sharp, as shown in Theorem 3.1.

The midpoint of the interval is given by  $Lg_0$  and is independent of  $B$ . We provide an alternative characterization of  $Lg_0$  that has important implications for estimation. Assume  $\|\cdot\|_{\mathcal{G}} = \|\cdot\|$  and consider the following least squares problem where  $\ell \in \mathcal{R}(T^*) + \overline{\mathcal{R}(T^*)}^{\perp}$  and  $\theta_0$  satisfies

$$\|T^*\theta_0 - \ell\| = \inf_{\theta \in L_2(Z)} \|T^*\theta - \ell\|. \quad (2)$$

Here  $T^*$  denotes the adjoint operator of  $T$ , i.e.  $T^*\theta(x) := E[\theta(Z)|X = x]$ . It is well-known that

$T^*\theta_0 = P_{\overline{\mathcal{R}(T^*)}}\ell$ .<sup>8</sup> Define the prediction error  $u = T^*\theta_0 - \ell$ , which is non-zero under  $\ell \notin \overline{\mathcal{R}(T^*)} = \mathcal{N}(T)^\perp$ , i.e., under partial identification of  $Lg$  if  $\|g_0\|_{\mathcal{G}} < B$ . Then, we have

$$\begin{aligned} E[Y\theta_0(Z)] &= E[g_0(X)\theta_0(Z)] \\ &= E[g_0(X)T^*\theta_0(X)] \\ &= E[g_0(X)\ell(X)] + E[g_0(X)u(X)] \\ &= E[g_0(X)\ell(X)] \\ &= Lg_0, \end{aligned}$$

where the fourth equality follows from  $g_0 \in \mathcal{N}(T)^\perp$  and  $u \in \overline{\mathcal{R}(T^*)}^\perp = \mathcal{N}(T)$ . Hence, an alternative representation of  $Lg_0$  when  $\ell \in \mathcal{R}(T^*) + \overline{\mathcal{R}(T^*)}^\perp$  is  $E[Y\theta_0(Z)]$ . This representation was used first by Santos (2011) to construct an estimator for  $Lg$  under point identification, and our arguments above show that it is valid more generally for estimating  $Lg_0$  under partial identification provided  $\ell \in \mathcal{R}(T^*) + \overline{\mathcal{R}(T^*)}^\perp$ . One advantage of following this program is that in estimating  $Lg_0$  one also estimates  $\|P_{\mathcal{N}}\ell\|$ , as the latter is the objective function in (2), i.e.  $\|P_{\mathcal{N}}\ell\| = \|u\|$ .

We remark now on further implications of our Theorem 3.1 for estimation. Our theorem justifies inference based on functionals of  $g_0$ , as they correspond to the parameter of interest under point identification and the midpoint of the identified set under partial identification. We call nonparametric estimators of  $g_0$  under partial identification  $\mathcal{G}$ -consistent estimators. Note that some of the proposed procedures for estimating  $g$  under point identification are  $\mathcal{G}$ -consistent. Specifically, procedures that regularize the problem by penalization on the norm  $\|g\|_{\mathcal{G}}$  are well suited for estimation of  $g_0$  even under partial identification; see e.g. the Tikhonov regularization estimation approach in Hall and Horowitz (2005) and Darolles, Fan, Florens and Renault (2011). See also Carrasco, Florens and Renault (2006), Florens, Johannes and Van Bellegem (2011) and Gagliardini and Scaillet (2012) for further motivation. For a general estimation method in ill-posed problems with penalization see Chen and Pouzo (2012). These existing results can be used to obtain  $\mathcal{G}$ -consistent estimators. These and other issues pertaining to estimation are beyond the scope of this paper and will be investigated in a companion paper.

## 4 Examples Revisited

**Example 1 (Engel Curves, cont.):** Let  $f(x)$  denote the Lebesgue density of  $X$ . Assume that  $f$  is continuously differentiable and that it vanishes at the boundary of its support. Define the score function  $s(X) := f^{(1)}(X)/f(X)$ , and assume  $s \in L_2(X)$ . Then, for the average partial effect functional,  $\ell(x) = -s(x) = -\dot{f}(x)/f(x)$ . Then, the identified set will be small if  $\|g_0\|$  is close to 1 or if  $\|P_{\mathcal{N}(T)}s\|$  is small. If the smallest of the observationally equivalent Engel curves has sufficient variation (as measured by  $\|g_0\|$ ), then the identified set will be small. Informally speaking, we expect this to be the case if a large amount of consumers spend large shares in food.

---

<sup>8</sup>Note that in general  $\mathcal{R}(T^*)$  is not closed. If  $\ell \in \overline{\mathcal{R}(T^*)} \setminus \mathcal{R}(T^*) + \overline{\mathcal{R}(T^*)}^\perp$ , this equality does not hold.

In their application, [Blundell, Chen and Kristensen \(2007\)](#) consider norm constraints with

$$\|g\|_{\mathcal{G}}^2 := \int_0^1 g^2(x)f(x)dx + \int_0^1 \left(g^{(2)}(x)\right)^2 dx. \quad (3)$$

Under point identification of  $g$  these constraints do not play any role in the asymptotic distribution derived in [Blundell, Chen and Kristensen \(2007\)](#). However, we have shown here that they are important under partial identification. An implication of our results is that the estimator proposed by [Blundell, Chen and Kristensen \(2007\)](#) is  $\mathcal{G}$ -consistent, where  $\mathcal{G}$  is the subspace of  $g \in L_2(X)$  with  $\|g\|_{\mathcal{G}} < \infty$ , and where  $\|\cdot\|_{\mathcal{G}}^2$  is defined in (3). ■

**Example 2 (Consumer Surplus, cont.):** With  $\langle f, g \rangle_{\mathcal{G}} = \int_a^b f(x)g(x)dx$ , for the consumer surplus functional  $\ell = 1$ . The operator  $T$  is continuous wrt to  $\|\cdot\|_{\mathcal{G}}$  if, for instance, the following condition holds

$$\sup_{a \leq x \leq b} f_{X/Z=z}^2(x) \in L_2(Z),$$

where  $f_{X/Z=z}(x)$  denotes the conditional density of  $X$  given  $Z$ , and evaluated at  $x$  and  $z$ , respectively. The identified set for the consumer surplus is given by the interval

$$[\langle 1, g_0 \rangle_{\mathcal{G}} - \rho, \langle 1, g_0 \rangle_{\mathcal{G}} + \rho],$$

where  $\rho = (B - \|g_0\|_{\mathcal{G}}) \|P_{\mathcal{N}(T)}1\|_{\mathcal{G}}$ ,  $B$  is a bound on the quantity of the good (capacity constraint) and  $g_0$  is the identified part of the demand. ■

**Example 3 (Best Linear Approximations, cont.):** In this application,  $\ell(x) = \alpha' E[XX']^{-1}x$ , with  $\|\cdot\|_{\mathcal{G}} = \|\cdot\|$ . Different combinations of BLA coefficients lead to different sizes of the corresponding identified sets. Define the matrix  $Q := E[XX']^{-1}E[(P_{\mathcal{N}(T)}X)(P_{\mathcal{N}(T)}X)']E[XX']^{-1}$ . Then, the second factor in the radius,  $\|P_{\mathcal{N}(T)}\ell\|_{\mathcal{G}}$ , is the square root of  $\alpha'Q\alpha$ . Note that if the unidentifiable components are uncorrelated with  $X$ , i.e.  $P_{\mathcal{N}(T)}X = 0$ , then  $\beta$  is identified. Hence,  $P_{\mathcal{N}(T)}X = 0$  can be understood as a generalization of the classic omitted variable bias condition. Suppose we restrict the analysis to  $\alpha$  such that  $|\alpha| = 1$ . Under this constraint, the maximum and minimum values of the second factor are equal to  $\sqrt{\lambda_{\max}}$  and  $\sqrt{\lambda_{\min}}$ , respectively, where  $\lambda_{\max}$  and  $\lambda_{\min}$  are the maximum and minimum eigenvalues of  $Q$ , and the corresponding normalized eigenvectors,  $\alpha_{\max}$  and  $\alpha_{\min}$ , are the coefficients. In particular,  $\alpha'_{\min}\beta$  corresponds to the combination of coefficients with the smallest identified set. ■

**Example 4 (Point-evaluation Functionals, cont.):** By the continuity of the point-evaluation functionals and the Riesz Representation theorem, there exists a function  $k(\cdot, x_0) \in \mathcal{G}$  such that  $h(x_0) = \langle h, k(\cdot, x_0) \rangle_{\mathcal{G}}$ , for all  $h \in \mathcal{G}$  and all  $x_0 \in \mathcal{S}_X$ . This is the so-called *reproducing property*. The function  $k(x, z) = \langle k(\cdot, x), k(\cdot, z) \rangle_{\mathcal{G}}$  is called the kernel of the RKHS. It is well-known that  $\mathcal{G} = W_2^1$  is a RKHS with kernel<sup>9</sup>

$$k(x, z) = \frac{1}{2 \sinh(b-a)} [\cosh(x+z-b-a) + \cosh(|x-z|-b+a)]; \quad (4)$$

<sup>9</sup>The Hyperbolic sine and cosine are given by  $\sinh x = (e^x - e^{-x})/2$  and  $\cosh x = (e^x + e^{-x})/2$ , respectively.

see, e.g., [Du and Cui \(2008\)](#). Since  $k(\cdot, x_0) \in W_2^1$ , the reproducing property implies that point-evaluation functionals are continuous in  $W_2^1$ , with the corresponding representer  $\ell(x) = k(x, x_0)$ . The operator  $T$  is also continuous wrt to  $\|\cdot\|_{\mathcal{G}}$ . Hence, we have pointwise bounds for any element of the identified set

$$g_0(x_0) - \rho \leq g(x_0) \leq g_0(x_0) + \rho,$$

where  $\rho = (B - \|g_0\|_{\mathcal{G}}) \|P_{\mathcal{N}(T)}k(\cdot, x_0)\|_{\mathcal{G}}$ . There are several expressions for  $g_0$  available. Here, we consider one given by [Du and Cui \(2008\)](#). Let  $\{z_i\}_{i=1}^{\infty}$  denote a dense set in the support of  $Z$ . Then, define the functions

$$\psi_i(x) := E[k(X, x) | Z = z_i].$$

Then, let  $\{\bar{\psi}_i\}$  denote the Gram-Schmidt orthonormalization of  $\{\psi_i\}$ , say

$$\bar{\psi}_i(x) = \sum_{j=1}^i \beta_{ij} \psi_j(x),$$

where  $\beta_{ij}$  are the coefficients of the Gram-Schmidt orthonormalization. Then, Theorem 3.2 in [Du and Cui \(2008\)](#) shows that

$$g_0(x) = \sum_{i=1}^{\infty} \sum_{j=1}^i \beta_{ij} r(z_j) \bar{\psi}_i(x).$$

Alternative characterizations based on the Tihkonov functional

$$g_{\lambda} = \arg \min_{g \in \mathcal{G}} \|Tg - r\|^2 + \lambda \|g\|_{\mathcal{G}},$$

or empirical analogs of this are also available, see the important Representer Lemma in [Wahba \(1990\)](#). Polynomial Spline estimators were historically motivated from this representation, and the convergence from  $g_{\lambda}$  to  $g_0$  as  $\lambda \rightarrow 0$  has been extensively investigated in the mathematical literature; see, e.g., [Engl, Hanke and Neubauer \(1996\)](#). These results based on the Tihkonov functional are valid for general Hilbert spaces and not just for RKHS.

As for the second factor  $\|P_{\mathcal{N}(T)}k(\cdot, x_0)\|_{\mathcal{G}}$ , suppose  $\{h_j\}_{j=1}^{\infty}$  is an orthonormal basis of  $\mathcal{N}(T)$ . Then, by the reproducing property,

$$\begin{aligned} P_{\mathcal{N}(T)}k(\cdot, x_0) &= \sum_{j=0}^{\infty} \langle h_j, k(\cdot, x_0) \rangle_{\mathcal{G}} h_j \\ &= \sum_{j=0}^{\infty} h_j(x_0) h_j, \end{aligned}$$

and hence,

$$\|P_{\mathcal{N}(T)}k(\cdot, x_0)\|_{\mathcal{G}} = \sum_{j=0}^{\infty} h_j^2(x_0).$$

Thus, point-identification of  $g(x_0)$  holds if  $\sum_{j=0}^{\infty} h_j^2(x_0) = 0$ . Clearly this is the case under the completeness condition, which requires  $\mathcal{N}(T) = \{0\}$ , or  $\sum_{j=0}^{\infty} h_j^2(x) = 0$  for all  $x$  in the support of  $X$ . So, as

one can see here, when applied to all  $x_0$ , our identification condition is equivalent to the completeness condition. Our identification assumption is well motivated when the interest is not on the value of  $g$  at all points of its support, but rather at specific points. For instance, we might be interested in estimating a demand function at new fixed prices, say  $p_0$ , without the need to introduce assumptions on the identification of demand at other prices different from  $p_0$ .

The results presented in this example can be extended to other continuous linear functionals. A convenient property of RKHS is that the Riesz representer for a continuous linear functional  $L$  has the general expression  $\ell(x) = Lk(x, \cdot)$ . In particular, the previous example can be extended to point-evaluation functionals of derivatives. Consider, for instance, the Sobolev space

$$W_2^2 := \{g : g, g^{(1)} \text{ are absolute continuous and } g^{(2)} \in L_2(\mathbb{R})\},$$

where  $L_2(\mathbb{R})$  is the space of square integrable functions in  $\mathbb{R}$ . The space  $W_2^2$  is endowed with the square norm

$$\|g\|_{W_2^2}^2 := \int_{-\infty}^{\infty} g^2(x) dx + \frac{1}{\tau^4} \int_{-\infty}^{\infty} \left(g^{(2)}(x)\right)^2 dx,$$

where the parameter  $\tau > 0$  allows for flexible penalization (different weights to the  $L_2(X)$  and second derivative norms). For a related space see [Blundell, Chen and Kristensen \(2007\)](#). It is well-known that  $\mathcal{G} = W_2^2$  is a RKHS with kernel

$$k(x, z) = \frac{\tau}{2} \exp\left(-\frac{|\tau(z-x)|}{\sqrt{2}}\right) \sin\left(|\tau(z-x)| \frac{2}{\sqrt{2}} + \frac{\pi}{4}\right); \quad (5)$$

see p. 324 in [Berlinet and Thomas-Agnan \(2004\)](#). In  $\mathcal{G} = W_2^2$  the point-evaluation derivative functional  $Lg = \partial g(x_0)/\partial x$  is continuous, with a Riesz representer

$$\ell(x) = \frac{\partial k(x, z)}{\partial z} \Big|_{z=x_0}, \quad (6)$$

where  $k$  is given in (5). Hence, the results on the average partial effect in Example 1 can be extended to heterogeneous partial effects  $\partial g(x_0)/\partial x$  using these tools. In particular, the derivative  $\partial g(x_0)/\partial x$  is point identified if  $\|P_{\mathcal{N}(T)}\ell(x)\|_{\mathcal{G}} = 0$ , with  $\ell$  given in (6). ■

**Example 5 (Discrete Variables, cont.):** The NPIVR in the discrete case is indeed parametric, and hence, many of the complications of the nonparametric case disappear in the discrete case, e.g. the ill-posedness. Yet, this is an important case given its practical relevance, and it also serves to illustrate how the quantities involved in our partial identification result can be estimated from a sample.

In the context of the application in [Angrist and Evans \(1998\)](#) and [Freyberger and Horowitz \(2013\)](#), let  $T$  be the  $2 \times 4$  matrix whose  $(i, j)$  element is  $P(X = x_j | Z = z_i)$ , where  $x_j \in \{2, 3, 4, 5\}$  and  $z_i \in \{0, 1\}$ , for  $i = 1, 2$  and  $j = 1, 2, 3, 4$ ; let  $r$  be the  $2 \times 1$  vector obtained from stacking  $\{E(Y|Z = z_i)\}_{i=1,2}$  and define  $g$  and  $c$  as before. Equation (1) is then equivalent to  $Tg = r$ . In this example  $g$  is not identified, as this would require  $\text{rank}(T) = 4$ . A linear functional  $Lg = c'g$  is identified if and only if  $c$  is generated by the rows of  $T$ , denoted here by  $t_1$  and  $t_2$ , respectively. This identification condition can be tested. Assume the instrument is strong, so that  $t_1$  and  $t_2$  are linearly independent. Then, the dimension of

the space of linear functionals that are identified is  $\text{rank}(T) = 2$ . We will investigate in a practical situation the identification of different linear functionals, such as, for example,  $L_{23}g = g(3) - g(2)$  and  $L_{P23}g = P(X = 3)g(3) - P(X = 2)g(2)$ , corresponding to  $c = (-1, 1, 0, 0)'$  and  $c = (-P(X = 2), P(X = 3), 0, 0)'$ , respectively. Notice that in the latter  $c$  is unknown, but it can be estimated from the data. Applying our results to the general example and using Theorem 2, p. 65 in [Luenberger \(1997\)](#) to compute  $g_0$ , we obtain

$$c'g_0 - \rho \leq c'g \leq c'g_0 + \rho,$$

where  $g_0 = T'(TT')^{-1}r$ , (this is consistent with  $T^\dagger = T'(TT')^{-1} = (T'T)^\dagger T'$ ), and

$$\rho = (B - |g_0|) |Ac|,$$

where  $A$  is a  $2 \times 4$  matrix with orthonormal rows, which are in turn orthogonal to  $t_1$  and  $t_2$ . This matrix is easily obtained from statistical packages.

We now apply these results to a subset of the sample used in [Angrist and Evans \(1998\)](#). The data consist of 150,618 women who are 21-35 years old, have 2-5 children, and whose oldest child is between 8 and 12 years old. The sample data is a subsample of the 1980 Census Public Use Micro Samples (PUMS), and it has been used in [Freyberger and Horowitz \(2013\)](#) to empirically illustrate the role of shape restrictions in the identification of linear functionals  $Lg$ .

We proceed first by estimating  $g_0$  as follows. Let  $M$  denote the  $2 \times 4$  matrix with  $(i, j)$  element  $m_{ij} = P(X = x_j, Z = z_i)$ , and let  $s$  denote the  $2 \times 1$  vector with  $i^{\text{th}}$  element  $s_i = E(Y|Z = z_i)P(Z = z_i)$ . Then, we estimate  $g_0$  from  $Mg_0 = s$ , replacing  $M$  and  $s$  by the sample analogues in the expression  $g_0 = M'(MM')^{-1}s$ . That is, we obtain  $\hat{g}_0 = \hat{M}'(\hat{M}\hat{M}')^{-1}\hat{s}$ , where the  $(i, j)$  element of  $\hat{M}$  is given by  $\hat{m}_{ij} = n^{-1}(\sum_{k=1}^n I(X_k = x_j)I(Z_k = z_i))$  and the  $i^{\text{th}}$  element of  $\hat{s}$  is given by  $n^{-1}\sum_{k=1}^n Y_k I(Z_k = z_i)$ . In the Appendix, we consider a general and convenient method of proof to show that

$$\sqrt{n}(\hat{g}_0 - g_0) \rightarrow_d N(0, V),$$

where  $V$  is a positive definite matrix whose expression is given in the Appendix. From this result, we obtain asymptotic 95% confidence intervals for  $g_0$  as well as for  $c'g_0$ .

Table 2 reports the results of inferences based on the midpoint  $g_0$  for several linear functionals of  $g_0$ . We provide the estimates  $\hat{g}_0$ , the functionals  $L_{kj}\hat{g}_0 = \hat{g}_0(j) - \hat{g}_0(k)$  and their probability-weighted versions  $L_{Pkj}\hat{g}_0 = \hat{\pi}_j\hat{g}_0(j) - \hat{\pi}_k\hat{g}_0(k)$ , for selected values of  $k$  and  $j$ , and where  $\hat{\pi}_j$  is the sample estimate of  $P(X = j)$ , with  $j, k = 2, 3, 4, 5$ . We also report their corresponding 95% confidence intervals based on the asymptotic theory above. Inference based on  $g_0$  suggests a highly nonlinear behaviour of weeks worked as a function of children. Having one more children after 2 leads to an expected reduction of 3 weeks of work, but going from 3 to 4 leads to an estimated decrease of 15 weeks. These results contrast with those obtained from the standard linear IV model, which yield a constant negative effect of 5 weeks, see [Angrist and Evans \(1998\)](#). All the estimates based on  $\hat{g}_0$  are quite precise.

Table 3 reports the results of the set estimates. Point estimates for the radius  $\rho$  are provided in this table, together with its factors and the estimated bounds for the linear functionals discussed above. The norm  $|\hat{g}_0|$  equals 30.379, so the first factor of  $\rho$ , which is common to all functionals, is estimated at

73 ( $B = 104$ ). Hence, norm constraints seem to provide little information for the global identification of  $g_0$  in this application. However, as we will see, despite the large values of the first factor, there are still functionals that are well identified and for which norm constraints become quite relevant. The overall results suggest that the point-evaluation functionals and the linear functionals  $L_{kj}$  are not well identified. Only for  $g_0(2)$  and  $L_{23}$  the norm constraints provide some partial information, with the former functional being particularly well identified. In many cases, the obtained bounds fall outside the natural range of  $g$ , so in these cases we report the intersection of our bounds with the pointwise bounds  $0 \leq g \leq 52$ . See also [Blundell, Chen and Kristensen \(2007\)](#) for a similar approach. Incorporating prior information, such as the shape restrictions in [Freyberger and Horowitz \(2013\)](#), seems to be the only available alternative to shrink the identified sets for the functionals  $L_{kj}$ . Using monotonicity and concavity constraints, [Freyberger and Horowitz \(2013\)](#) found substantial reductions in the size of the identified set for  $L_{23}$ . Monotonicity seems a plausible assumption in the 80’s, with relatively low labor participation for women and low rates of non-parental child care. Nevertheless, our wide identified sets suggest that testing for this prior information might be hard, if not impossible.

In stark contrast to previous functionals, probability-weighted functionals lead to much smaller identified sets. This is particularly the case for  $L_{P25}$ , with an estimated radius of 1.36. These functionals contain important structural interpretation. For instance,  $L_{P25}$  measures the relative (i.e. per-capita over the whole population) aggregated difference of weeks worked for women with 5 children and those with 2 children. Our results suggest an estimate of approximately -14 weeks for this effect. Accounting for the relative sizes of the populations lead to more informative inference in this application.

In summary, this application highlights one of the main points of our identification result; some functionals are relatively “well identified”, whereas others are not. This information can be assessed from the data, by relatively simple tools when observations are discrete. Norm bounds are useful, since without them there will be strictly no information on the linear functionals considered (cf. [Theorem 3.1](#) with  $B = \infty$ ).

## 5 Conclusions and future research

In this paper, we have characterized the identified set of linear continuous functionals of the NPIVR under norm constraints. We have determined the main factors driving the length of the identified set. A major role is played by the “identifiable component”  $g_0$  of the structural function. A main implication of our result is that inference based on  $g_0$  possesses certain robustness properties in our context. The length of the identified set depends inversely on the norm of  $g_0$  relative to  $B$ , but also on the maximum absolute bias arising from omitting normalized “unidentifiable components” in the computation of the linear functional. The example to BLA shows how the formula of the length extends the popular omitted variables formula of classical linear regression. Our theory is quite general. When applied to RKHS, it leads to new results for the point-evaluation functionals, including a new identification condition for  $g(x_0)$  at a fixed  $x_0$ . These results can be easily extended to derivatives.

We have illustrated the main ideas with an application to the effect of children on women’s weeks of work, using a subsample of [Angrist and Evans \(1998\)](#). This is an example where  $g$  is known to



be unidentified. It is also an example where asymptotic results can be developed using relatively standard tools. We have shown how different functionals lead to different lengths in the corresponding identified sets. A companion paper will investigate estimation for continuous variables, as well as other aspects of set inference. The required tools are non-standard in the continuous case, given the infinite dimensionality of the problem and the lack of identification. In particular, as shown by [Severini and Tripathi \(2006, 2012\)](#), under some conditions, estimation of  $Lg_0$  cannot be obtained at a  $\sqrt{n}$ -rate.

Our results are geometric in nature and, as such, are also applicable to other settings where linear inverse problems arise, and where the interest is in linear continuous functionals. For instance, the distribution of random coefficients in many random coefficients econometric models satisfies a linear inverse problem similar to that in (1), but with a known operator  $T$ ; see [Ichimura and Thompson \(1998\)](#), [Hoderlein, Klemela, and Mammen \(2010\)](#), [Gautier and Kitamura \(2013\)](#) and [Hoderlein, Nesheim and Simoni \(2012\)](#), among others. Our results can be applied to characterize, for instance, the identified set for counterfactual effects which are linear functionals of the random coefficients' distribution.

Another interesting extension of our identification results replaces the norm constraints with general convex restrictions. That is, the constraint  $\|g\|_{\mathcal{G}} \leq B$  could be replaced by the restriction  $g \in C$ , where  $C$  is a general convex, closed and bounded set of  $\mathcal{G}$ , such that  $C \cap \mathcal{G}_0 \neq \emptyset$ . This setting is very general and includes shape restrictions, such as monotonicity and concavity, consideration of pseudo-norms  $\|\cdot\|_{\mathcal{G}}$  rather than norms, and point-wise bounds  $a \leq g(x) \leq b$  for all  $x \in \mathcal{S}_X$ , among many others. However, for this general case much of the simple geometrical symmetry of our case is lost, and an identification analysis seems to require a case-by-case study. See, for instance, [Hoderlein, Nesheim and Simoni \(2012\)](#) for the case of spaces of densities. There are natural extensions of  $g_0$  in this more general setting, see [Engl, Hanke and Neubauer \(1996\)](#), p. 140. However, the extended version of  $g_0$  in general loses much of its dominant role in the partial identified case, and its interest seems to be confined to the point identified case. It can be shown that a general sufficient condition for identification of  $Lg$  in this more general context is  $P_{\mathcal{N}(T) \cap \overline{\text{span}}(C-g_0)} \ell = 0$ , where  $\overline{\text{span}}$  denotes the closed linear span. The proof of this follows the arguments given above for the norm bounded case. Whether or not this result can be improved, and when the point identification condition fails, how the expression of the identified set looks like is beyond the scope of this paper, and will be investigated in future research.



## 6 Appendix

### 6.1 Figures and Tables

Figure 1: Geometrical illustration of partial identification

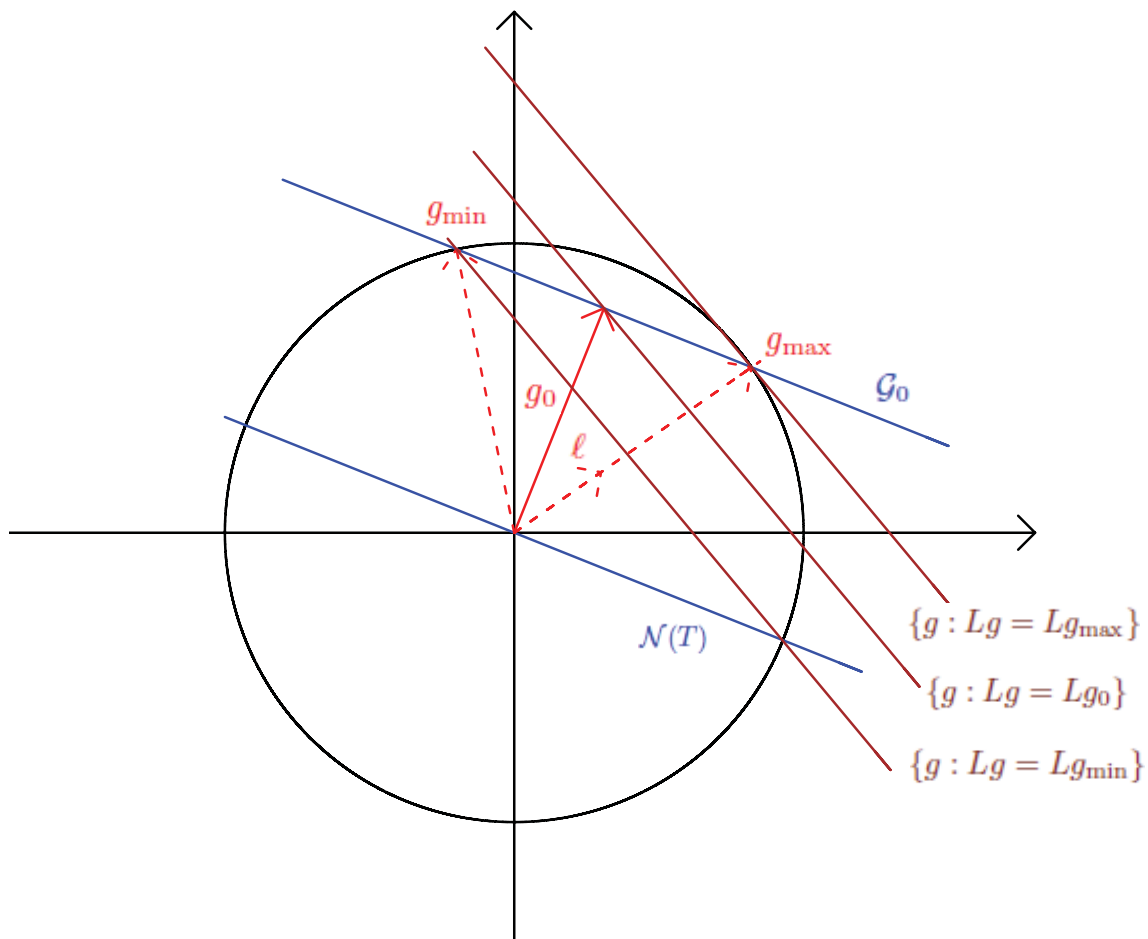


Table 1: Different formulations of  $g_0$  in the literature

Formulations	References and Comments
$g_0 = \min_{g \in \mathcal{G}_0} \ g\ _{\mathcal{G}}$	See, e.g., <a href="#">Engl, Hanke and Neubauer (1996)</a> .
$g_0 = T^\dagger r = (T^*T)^\dagger T^*r$	See, e.g., <a href="#">Engl, Hanke and Neubauer (1996)</a> .
$g_0 = \sum_i^\infty \frac{1}{\lambda_i} \langle r, v_i \rangle u_i$	For $T$ compact. $\{u\}_i^\infty$ is an orthonormal basis of $\mathcal{N}(T)^\perp$ , $\{v\}_i^\infty$ is an orthonormal basis of $\mathcal{N}(T^*)^\perp$ and $\{\lambda\}_i^\infty$ is a sequence of singular values for $T$ . See, e.g., <a href="#">Engl, Hanke and Neubauer (1996)</a> .
$g_0 = \lim_\lambda g_\lambda$ $g_\lambda := \arg \min_{g \in \mathcal{G}} \ Tg - r\ _{\mathcal{G}} + \lambda \ g\ _{\mathcal{G}}$	See, e.g., <a href="#">Engl, Hanke and Neubauer (1996)</a> .
$g_0 = \lim_\lambda g_\lambda$ $g_\lambda := \arg \min_{g \in \mathcal{G}} \ Tg - r\ _{\mathcal{G}} + \lambda \ Lg\ _{\mathcal{G}}$	$L$ could be a differential operator. See, for instance, <a href="#">Engl, Hanke and Neubauer (1996)</a> , <a href="#">Wahba (1990)</a> , <a href="#">Blundell, Chen and Kristensen (2007)</a> and <a href="#">Chen and Pouzo (2012)</a> .
$g_0 = P_{\mathcal{N}(T)^\perp} g$	See, e.g., <a href="#">Severini and Tripathi (2006, 2012)</a> , <a href="#">Santos (2011)</a> .
$g_0(x) = \sum_{i=1}^\infty \sum_{j=1}^i \beta_{ij} r(z_j) \bar{\psi}_i(x)$	<a href="#">Du and Cui (2008)</a> .
$g_0(x) = \langle T^*r, T^*Tk(\cdot, x) \rangle_{\mathcal{H}_{k'}}$	$k(\cdot, \cdot)$ is reproducing kernel for $\mathcal{G}$ and $\mathcal{H}_{k'}$ is a RKHS with kernel $k'$ . See <a href="#">Saitoh (2007)</a> for $k'$ and other details.

Table 2: Estimates of  $g_0$  and  $Lg_0$  and their 95% Confidence interval.

$Lg_0$	Estimate	95% CI	
$g_0(2)$	22.749	22.746	22.751
$g_0(3)$	19.633	19.572	19.693
$g_0(4)$	4.391	4.108	4.674
$g_0(5)$	0.788	0.667	0.909
$L_{23}g_0$	-3.115	-3.173	-3.058
$L_{34}g_0$	-15.241	-15.584	-14.898
$L_{45}g_0$	-3.603	-3.906	-3.300
$L_{25}g_0$	-21.960	-22.082	-21.839
$LP_{23}g_0$	-8.353	-8.369	-8.337
$LP_{34}g_0$	-5.497	-5.535	-5.460
$LP_{45}g_0$	-0.291	-0.310	-0.271
$LP_{25}g_0$	-14.142	-14.145	-14.140

Table 3: Estimation of  $Lg$ 

$Lg$	$B - \ g_0\ $	$\ P_N \ell\ $	$\rho$	$[Lg_{\min}, Lg_{\max}]$	
$g(2)$	73.620	0.0113	0.835	21.913	23.584
$g(3)$	73.620	0.210	15.473	4.160	35.106
$g(4)$	73.620	0.978	72.011	-52.000	52.000
$g(5)$	73.620	0.999	73.581	-52.000	52.000
$L_{23g}$	73.620	0.199	14.719	-17.834	11.603
$L_{34g}$	73.620	1.186	87.328	-52.000	52.000
$L_{45g}$	73.620	1.403	103.309	-52.000	52.000
$L_{25g}$	73.620	1.005	74.045	-52.000	52.000
$L_{P23g}$	73.620	0.055	4.105	-12.458	-4.248
$L_{P34g}$	73.620	0.128	9.495	-14.993	3.997
$L_{P45g}$	73.620	0.068	5.060	-5.352	4.769
$L_{P25g}$	73.620	0.018	1.360	-15.503	-12.782

## 6.2 Asymptotics for the discrete case

To obtain the asymptotic distribution of  $\hat{g}_0 = \hat{M}'(\hat{M}\hat{M}')^{-1}\hat{s}$ , where the  $(i, j)$  element of the  $d(Z) \times d(X)$  matrix  $\hat{M}$  is given by  $\hat{m}_{ij} = n^{-1}(\sum_{k=1}^n I(X_k = x_j)I(Z_k = z_i))$  and the  $i^{\text{th}}$  element of the  $d(Z) \times 1$  vector  $\hat{s}$  is given by  $n^{-1}\sum_{k=1}^n Y_k I(Z_k = z_i)$ , we note that  $\hat{s} = \hat{M}\hat{g}_0$ , and then

$$\hat{P}\hat{g}_0 = 0,$$

where  $\hat{P} = I - \hat{M}'(\hat{M}\hat{M}')^{-1}\hat{M}$  is a projection matrix. Similarly,  $Pg_0 = 0$ , with  $P = I - M'(MM')^{-1}M$ . Then, by simple algebra and the standard central limit theorem (CLT),

$$0 = P(\hat{g}_0 - g_0) + (\hat{P} - P)g_0 + o_P(n^{-1/2}),$$

and hence, using that  $P$  is symmetric, idempotent and satisfies  $P^\dagger P = P$ ,

$$\begin{aligned} \sqrt{n}(\hat{g}_0 - g_0) &= -P\sqrt{n}(\hat{P} - P)g_0 + o_P(1) \\ &= -(g_0' \otimes P)\sqrt{n}\text{vec}(\hat{P} - P) + o_p(1). \end{aligned}$$

We then use the delta method to obtain the asymptotic distribution of  $\sqrt{n}\text{vec}(\hat{P} - P)$ . Write  $\text{vec}(\hat{P}) = \phi(\text{vec}(\hat{M}'))$  and similarly  $\text{vec}(P) = \phi(\text{vec}(M'))$ . We need to compute the Jacobian  $\Delta(M)$  of  $\phi$  at  $\text{vec}(M')$ . By Exercise 13.24 in [Abadir and Magnus \(2005\)](#),

$$\frac{\partial \text{vec}(P)}{(\partial \text{vec}(M'))'} = -(I + K_c)(M'(MM')^{-1} \otimes P) =: \Delta(M),$$

where  $K_c$  is the  $d(X)^2 \times d(X)^2$  commutation matrix for an  $d(X) \times d(X)$  matrix  $A$  such that  $K_c \text{vec}A = \text{vec}(A')$ . Therefore, by the delta method,

$$\sqrt{n}(\hat{g}_0 - g_0) = (g_0' \otimes P)(I + K_c)(M'(MM')^{-1} \otimes P)\sqrt{n}(\text{vec}(\hat{M}') - \text{vec}(M')).$$

It is easy to show by a standard CLT argument that,

$$\sqrt{n}(\text{vec}(\hat{M}') - \text{vec}(M')) \xrightarrow{d} N(0, \Omega),$$

where  $\Omega$  is given by

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega'_{12} & \Omega_{22} \end{pmatrix},$$

with

$$[\Omega_{11}]_{(i,j)} = 1(i = j)m_{1i} - m_{1i}m_{1j},$$

$$[\Omega_{12}]_{(i,j)} = -m_{1i}m_{2j},$$

$$[\Omega_{22}]_{(i,j)} = 1(i = j)m_{2i} - m_{2i}m_{2j},$$

and where recall  $m_{ij} := P(X = x_j, Z = z_i)$ . Let define

$$J := (g'_0 \otimes P)(I + K_c)(M'(MM')^{-1} \otimes P).$$

Then  $\sqrt{n}(\hat{g}_0 - g_0) \xrightarrow{d} N(0, V)$  where  $V := J\Omega J'$ . For a nonrandom  $c$ , clearly  $\sqrt{n}c'(\hat{g}_0 - g_0) \xrightarrow{d} N(0, c'Vc)$ . For a random  $\hat{c} = (-\hat{P}(X = 2), \hat{P}(X = 3), 0, 0)$ , we have  $\sqrt{n}\hat{c}'(\hat{g}_0 - g_0) = \sqrt{n}c'(\hat{g}_0 - g_0) + o_p(1) \xrightarrow{d} N(0, c'Vc)$ .

## References

- ABADIR, K. M., AND J. R. MAGNUS (2005): *Matrix Algebra*. New York: Cambridge University Press.
- ADAMS, R. A. (1975): *Sobolev Spaces*. New York: Academic Press.
- AI, C., AND X. CHEN (2003): “Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions,” *Econometrica*, 71(6), 1795–1843.
- ANGRIST, J. D., AND A. B. KRUEGER (1991): “Does Compulsory School Attendance Affect Schooling and Earnings?, ” *The Quarterly Journal of Economics*, 106(4), 979-1014.
- ANGRIST, J. D., K. GRADDY, AND G. W. IMBENS (2000): “The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish,” *The Review of Economic Studies*, 67.3: 499-527.
- ANGRIST, J. D., AND W. N. EVANS (1998): “Children and Their Parents’ Labor Supply: Evidence from Exogenous Variation in Family Size,” *American Economic Review*, 88, 450-477.
- BERLINET, A., AND C. THOMAS-AGNAN (2004): *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. 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.

- BRONARS, S. G., AND J. GROGGER (1994): “The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment,” *American Economic Review*, vol. 84(5), 1141-56.
- BROWN, B. W., AND W. K. NEWEY (1998): “Efficient Semiparametric Estimation of Expectations,” *Econometrica*, 66, 453–464.
- CARD, D. (1995): “The Wage Curve: A Review,” *Journal of Economic Literature*, vol. 33(2), 285-299.
- CARRASCO, M., J. P. FLORENS, AND E. RENAULT (2006): “Linear Inverse Problem in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization,” in *Handbook of Econometrics*, vol. 6, ed. by J. J. Heckman and E. E. Leamer. Amsterdam: North-Holland, 5633–5751.
- CHEN, X., AND D. POUZO (2009): “Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals,” *Journal of Econometrics*, 152, 46-60.
- CHEN, X., AND D. POUZO (2012): “Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Generalized Residuals,” *Econometrica*, 80(1), 277–321.
- CHEN, X., AND M. REISS (2011): “On Rate Optimality for Ill-Posed Inverse Problems in Econometrics,” *Econometric Theory*, 27, 497–521.
- DAROLLES, S., Y. FAN, J. P. FLORENS, AND E. RENAULT (2011): “Nonparametric Instrumental Regression,” *Econometrica*, 79(5), 1541–1565.
- DU, H., AND M. CUI (2008): “Approximate Solution of the Fredholm Integral Equation of the First Kind in a Reproducing Kernel Hilbert Space,” *Applied Mathematics Letters*, 21, 617–623.
- ENGL, H. W., M. HANKE, AND A. NUEBAUER (1996): *Regularization of Inverse Problems*. Dordrecht: Kluwer Academic Publishers.
- FLORENS, J.-P., J. JOHANNES, AND S. VAN BELLEGEM (2011): “Identification and Estimation by Penalization in Nonparametric Instrumental Regression,” *Econometric Theory*, 27, 472–496.
- FREYBERGER, J., AND J. HOROWITZ (2013): “Identification and Shape Restrictions in Nonparametric Instrumental Variables Estimation,” Cemmap working paper.
- GAGLIARDINI, P., AND O. SCAILLET (2012): “Nonparametric Instrumental Variable Estimation of Structural Quantile Effects,” *Econometrica*, 80: 1533–1562.
- GAUTIER, E., AND Y. KITAMURA (2013): “Nonparametric Estimation in Random Coefficients Binary Choice Models,” *Econometrica*, 81, 581-607.
- HALL, P., AND J. HOROWITZ (2005): “Nonparametric Methods for Inference in the Presence of Instrumental Variables,” *Annals of Statistics*, 33, 2904–2929.

- HODERLEIN, S., J. KLEMELA, AND E. MAMMEN (2010): “Analyzing the Random Coefficient Model Nonparametrically,” *Econometric Theory*, 26, 804-837.
- HODERLEIN, S., L. NESHEIM, AND A. SIMONI (2012): “Semiparametric Estimation of Random Coefficients in Structural Economic Models,” Working paper.
- HOROWITZ, J. (2011): “Applied Nonparametric Instrumental Variables Estimation,” *Econometrica*, 79(2), 347–394.
- ICHIMURA, H., AND T. S. THOMPSON (1998): “Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution,” *Journal of Econometrics*, 86, 269-295.
- KURDILA, A. J., AND M. ZABRANKIN (2005): *Convex Functional Analysis*. Switzerland: Birkhäuser Verlag.
- LOCHNER, L., AND E. MORETTI (2012): “The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports,” *American Economic Review*, vol. 94(1), 155-189.
- LUENBERGER, D. G. (1997): *Optimization by Vector Space Methods*. New York: John Wiley & Sons. 1969 edition.
- MANSKI, C. (2003): *Partial Identification of Probability Distributions*. New York: Springer-Verlag.
- MANSKI, C. (2007): *Identification for Prediction and Decision*. Cambridge: Harvard University Press.
- NEWHEY, W. K. (2013): “Nonparametric Instrumental Variables Estimation,” *American Economic Review Papers and Proceedings*, 103, 550-556.
- NEWHEY, W. K., AND D. MCFADDEN (1994): “Large Sample Estimation and Hypothesis Testing,” in *Handbook of Econometrics*, vol. 4, ed. by R. F. Engle and D. McFadden. Amsterdam: North-Holland, 2111–2245.
- NEWHEY, W. K., AND J. POWELL (2003): “Instrumental Variables Estimation for Nonparametric Models,” *Econometrica*, 71, 1565–1578.
- PARZEN, E. (1967): *Time series analysis papers*. California: Holden-Day.
- POWELL, J. L., J. H. STOCK, AND M. STOKER (1998): “Semiparametric Estimation of Index Coefficients,” *Econometrica*, 1403-1430.
- SAITOH, S.(2007) “Applications of Tikhonov Regularization to Inverse Problems Using Reproducing Kernels ,” *Journal of Physics: Conference Series*, vol. 73, no. 1.
- SANTOS, A. (2011): “Instrumental Variable Methods for Recovering Continuous Linear Functionals,” *Journal of Econometrics*, 161(2), 129–146.
- SANTOS, A. (2012): “Inference in Nonparametric Instrumental Variables with Partial Identification,” *Econometrica*, 80(1), 213–275.

- 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 (2012): “Efficiency Bounds for Estimating Linear Functionals of Nonparametric Regression Models with Endogenous Regressors,” *Journal of Econometrics*, 170(2), 491–498.
- SONG, K. (2013): “Point Decisions for Interval-Identified Parameters, ” forthcoming in *Econometric Theory*.
- STOCK, J. H. (1989): “Nonparametric Policy Analysis,” *Journal of the American Statistical Association*, 84(406), 567–575.
- TAMER, E. (2010): “Partial Identification in Econometrics,” *Annual Review of Economics*, Vol. 2: 167-195.
- WAHBA, G. (1990): *Spline Models for Observational Data*. Philadelphia: SIAM.