Working Paper Series n° 613

September, 2009

"Instrumental Regression in Partially Linear Models"

Jean-Pierre FLORENS,
Jan JOHANNES and
Sébastien VAN BELLEGEM

Institut d'Economie Industrielle (IDEI) – Manufacture des Tabacs
Aile Jean-Jacques Laffont – 21, allée de Brienne – 31000 TOULOUSE – FRANCE
Tél. + 33(0)5 61 12 85 89 – Fax + 33(0)5 61 12 86 37 – **www.idei.fr** – contact@cict.fr

INSTITUT D'ECONOMIE INDUSTRIELLE

INSTRUMENTAL REGRESSION IN PARTIALLY LINEAR MODELS*

Jean-Pierre Florens[1] Jan Johannes[2] Sébastien Van Bellegem[3]

First version: January 24, 2008. This version: September 10, 2009

Abstract

We consider the semiparametric regression $X^t\beta+\phi(Z)$ where β and $\phi(\cdot)$ are unknown slope coefficient vector and function, and where the variables (X, Z) are endogeneous. We propose necessary and sufficient conditions for the identification of the parameters in the presence of instrumental variables. We also focus on the estimation of β. An incorrect parameterization of ϕ may generally lead to an inconsistent estimator of β, whereas even consistent nonparametric estimators for ϕ imply a slow rate of convergence of the estimator of β. An additional complication is that the solution of the equation necessitates the inversion of a compact operator that has to be estimated nonparametrically. In general this inversion is not stable, thus the estimation of β is ill-posed. In this paper, a $\sqrt{n}$-consistent estimator for β is derived under mild assumptions. One of these assumptions is given by the so-called *source condition* that is explicitly interprated in the paper. Finally we show that the estimator achieves the semiparametric efficiency bound, even if the model is heteroscedastic. Monte Carlo simulations demonstrate the reasonable performance of the estimation procedure on finite samples.

Keywords: Partially linear model, semiparametric regression, instrumental variables, endogeneity, ill-posed inverse problem, Tikhonov regularization, root-N consistent estimation, semiparametric efficiency bound

JEL classifications: Primary C14; secondary C30

*We are grateful to R. Lestringant for his help on simulations. We also like to thank X. Chen, A. Chesher, C. De Mol, J. Horowitz, O. Linton, I. Van Keilegom and the participants to research seminars in Yale, Paris I, Queen Mary London, New York University, KULeuven, Texas A&M, Tinbergen Institute Amsterdam, Toulouse III, ENSAI Rennes, CORE, and Dortmund for helpful discussions. This work was supported by the IAP research network nr P5/24 of the Belgian Government (Belgian Science Policy).

[1]Toulouse School of Economics (GREMaQ and IDEI), florens@cict.fr

[2]Université catholique de Louvain, jan.johannes@uclouvain.be

[3]Toulouse School of Economics (GREMaQ and IDEI), svb@tse-fr.eu

1 Introduction

To determine the form of an economic model from a set of explanatory variables, it is often the case that the economist can specify a parametric form for the effect of only a subset of these variables, while he has no precise information about the specification of the other variables. In that context, it is useful to consider a structural model that mixes parametric together with unrestricted components. One meaningful possibility is to assume a partially linear model, that is a semiparametric model of the type

$$Y = X^t\beta + \phi(Z) + U \tag{1.1}$$

where X and Z are multivariate explanatory variables. In this model, the economist imposes his knowledge of a precise parametric specification for the set of the variables X, with unknown parameters β. The effect of the other variables (here denoted Z) on the dependent variable is unknown, or not necessarily linear in Z. The model contains an unrestricted part, $\phi(Z)$, where ϕ is an unknown nonparametric function.

This compromise between an informative but restrictive parametric model and a nonparametric regression model is the foundation of a large number of economic studies. A seminal example was introduced by Engle, Granger, Rice, and Weiss (1986), who analyzed the electricity demand with respect to the average daily temperature T, the monthly price of electricity X_1, income X_2 and a set of eleven monthly dummy variables $X_3, \ldots, X_{13}$. The model used in this study writes the electricity demand as a linear function of variables X_i, plus a smooth function of the monthly temperature, $\phi(T)$. The motivation for introducing a nonparametric part was here the clear nonlinear dependance between the electricity demand and the temperature, because extreme temperatures cause extreme electricity demand.

In many applications, the apparent asymmetry between the effect of the variables X and Z in the partially linear model brings the analyst to include all dummy or categorical variables in the parametric component of the model. Examples include Anglin and Gençay (1996) who estimate hedonic price functions, or Stock (1989) in nonparametric policy analysis to name but a few. This way of doing is not systematic, especially when the economic theory gives a guidance on the effect of some variables. One example is for instance the study of Schmalensee and Stoker (1999), who analyze the log of household gasoline consumption. The nonparametric part of their semiparametric model include the log income and the log age, and the parametric part includes other continuous or discrete variables of interest among which are demographic and geographic variables.

Some economic studies also consider partially linear models as a restricted form of a fully nonparametric model. In that case, semiparametric models such as model (1.1) are considered in order to reduce the dimensionality of the problem, see e.g. Heckman, Ichimura, Smith, and Todd (1998) in the context of training programs evaluation.

The objective of this paper is to study the estimation of the parameter β when the variables X and Z are endogeneous. The underlying idea is that the components of the linear part of the structural model may have some economic significance, and some test of interest can be described in terms of the parameter β only. In that situation, the nonparametric part of the model is of secondary importance for the analysis, but in general influences the inference on β dramatically. Once an accurate estimator of β is found, it has also been suggested to plug it in the partial linear model and to provide inference for ϕ in a second step using standard nonparametric approaches. The above mentioned study of Schmalensee and Stoker (1999) uses that strategy, assuming that the estimator of β in the first step has good statistical properties.

To find an estimator of the parametric part as accurate as possible in spite of the presence of a nonparametric component $\phi(Z)$ in the structural model, and in spite of the possible endogeneity of all variables is then a challenging but important task for economic studies. With that respect, Ai and Chen (2003) provided a significant step forward. Starting from a general model with conditional moment restrictions containing an unknown function, they derive semiparametric efficient estimation of β using the method of sieve.

Our study below is in the continuity of Ai and Chen's results. In the specific model (1.1) we propose new identification results and a new estimation approach under milder assumptions. One desirable property of our estimator is that it is shown to be consistent and efficient even if the function ϕ belongs to an infinite dimensional space such as the space of square integrable functions. Sometimes, no regularity assumption of ϕ is needed in order to recover the consistency and efficiency of the estimation of β, provided that the instruments are strong enough, a situation that we formaly characterize below.

To treat the endogeneity problem, we assume that instrumental variables W are given. These variables are such that the error U in the structural model (1.1) fulfills the mean independence condition $\mathbb{E}(U|W) = 0$. In that context, the paper addresses the following most important issues:

(A) In which circumstance the presence of the nonparametric component has no influence on the *estimate* of β?

(B) Under which conditions does the estimator of β recover the parametric $\sqrt{n}$ rate of convergence and the asymptotic normality?

(C) How to find an efficient estimator of β?

An answer to these questions would help the analyst in his choice of a semiparametric model instead of a fully parametric model. Another way of thinking this problem is to determine the situations or the structural assumptions under which we do not lose too much when we allow the presence of variables in the regression model in an unrestricted form.

In the situation where all variables X and Z are endogeneous, we show below in this paper that issue (A) does not necessarily lead to the trivial condition that there is no nonparametric component in the model (i.e. $\phi \equiv 0$). Among the unobvious situations where the estimator of β does not depend on Z, we point out the particular situation where the instruments W can be separated into two independent components, $W = (W_1, W_2)$, and each component is separately an instrument for the parametric and the nonparametric part. More precisely, Z (resp. X) is independent from W given W_1 (resp. W_2) or, in formal notations,

$$W = (W_1, W_2) \text{ such that } Z \perp\!\!\!\perp W|W_1, X \perp\!\!\!\perp W|W_2 \text{ and } W_1 \perp\!\!\!\perp W_2.$$

Below we characterize all situations where Z has no influence on the estimate of β.

It is also worth mentioning that the simpler situation of estimating β when $\phi \equiv 0$ is a nontrivial subcase of our problem under the nonstandard conditional assumption $\mathbb{E}(U|W) = 0$. Below we solve that case by estimating nonparametrically the conditional expectations $\mathbb{E}(Y|W)$ and $\mathbb{E}(X|W)$. By doing so, we circumvent the discussion about the optimality of the instrument W. That case is developed below in this paper.

In order to address the second issue (B), we admit that the presence of the unrestricted component $\phi(Z)$ might have an influence on the definition of the estimate of β, but we want to find mild, sufficient conditions under which it has no influence on its *rate of convergence*. This question has a long history in the purely exogenous context, see e.g. Robinson (1988) and Andrews (1991). In the result we state below, we give an answer to that question under a set of minimal assumptions. We actually introduce two main types of assumptions:

(i) Smoothness assumptions on the joint density of the observations (Y_i, X_i, Z_i, W_i),

(ii) Strength of the instruments W.

It is worth noticing that the assumptions do not refer directly to the smoothness of the nonparametric function ϕ itself, in contrast to most of the studies written in the exogenous or partially exogenous framework. Instead we work under the above type (ii) assumption for which we introduce a new, objective measure of strength (or weakness) of the instruments. This measure will be called the *source condition*, in reference to formally analogous assumptions used in the numerical analysis literature. However, this notion as it is defined below in this paper, is new in economics (see also Carrasco, Florens, and Renault (2007)) and we give some details and intuitive examples to understand the usefulness of this concept.

The last issue (C) is related to the concept of semiparametric efficiency, see e.g. Stein (1956) and Newey (1990b). That point accurately measures the loss in efficiency when using a partially linear model instead of a fully linear model. The efficiency theorem we prove below covers the general situation where the error U is heteroscedastic. One appealing feature of this result is its formal resemblance with classical theory of optimal GMM estimators in the fully parametric model, which we cover as a particular case.

We close this Introduction by a short comparison between the results below and the economic literature.

Estimation of the parametric part of the partially linear model has been the subject of considerable study in the purely exogenous situation (e.g. Härdle, Liang, and Gao (1990)).

A $\sqrt{n}$-consistent estimator of the parameter has been derived in Robinson (1988) under the quite strong conditions that U is independent from X and Z, and ϕ belongs to some smoothness class. Andrews (1991) extends this result to the case of heteroscedastic errors and Linton (1995) carefully studies the second order properties of the estimator. This model is also the basis of a specification test in Delgado and Stengos (1994). A recent discussion in that context including other semiparametric models can be found in Ichimura and Todd (2007).

If only the parametric variables X are endogeneous, treatment of the endogeneity by instrumental variables is analogous to the idea of GMM estimators, although the presence of the nonparametric exogenous part $\phi(Z)$ leads to substantial technical difficulties. This case is covered e.g. by Chen, Linton, and Van Keilegom (2003) and Ma and Carroll (2006). A very different approach is however needed to treat the endogeneity of the nonparametric variable Z by instrumental variables. As we already mention, Ai and Chen (2003) proposed a sieve estimator for the case where Z is endogeneous, see also Chen and Pouzo (2008). As we shall recall in the next section, the problem becomes *ill-posed* and a stabilization procedure is necessary. That procedure is analogous to the ridge regression for parametric models. In a nonparametric context, this problem was first pointed out by Florens (2003) and further considered in recent economic studies, see e.g. Darolles, Florens, and Renault (2002), Hall and Horowitz (2005), Blundell, Chen, and Kristensen (2007), Chen and Hu (2006), Horowitz and Lee (2007), Carrasco, Florens, and Renault (2007) or Carrasco (2008).

The paper is organized as follows. In the next Section 2, we introduce the formal tools we need to address the above questions about the estimation of parameter β. We also address the questions of existence and identification of solutions β and ϕ. Section 3 first focuses on issue (A) and proposes a first estimator of β. This estimator is actually not optimal since it is not efficient. We then derive a general efficient estimator and give the conditions to answer question (B) about the rate of convergence and the asymptotic normality of the estimator. The "source conditions" are introduced in that section, and explained through meaningful examples. Section 4 addresses the semiparametric efficiency of the estimator. We find also interesting to study the small sample properties of the estimator. Section 5 comments on that aspect of the estimation procedure through a set of Monte Carlo simulations. The results of the estimation shows the very reasonable performance of the estimators for finite sample size. The proof of all results are deferred to a technical appendix.

2 Basic assumption and Identification

The setting of this paper can be summarized through the model

$$Y = \phi(Z) + X^t\beta + U \tag{2.1a}$$

where the random variables $Y \in \mathbb{R}$, $Z \in \mathbb{R}^p$, $X \in \mathbb{R}^k$, and where U is an error term with finite variance such that

$$\mathbb{E}(U|W) = 0 \tag{2.1b}$$

for some instrumental variable $W \in \mathbb{R}^q$.

2.1 Well-posed versus Ill-posed problem

The target function ϕ and parameter β are solution of the functional equation

$$\mathbb{E}(Y|W) = \mathbb{E}(\phi(Z)|W) + \mathbb{E}\left(X^t\beta|W\right). \tag{2.2}$$

Equation (2.2) is an integral equation which can be rewritten as

$$\int y f_{Y|W}(y)dy = \int \phi(z) f_{Z|W}(z)dz + \int x^t \beta f_{X|W}(x)dx$$

where $f_{Y|W}$ denotes the conditional density of Y given W, and similarly for $f_{Z|W}$ and $f_{X|W}$. The estimation of ϕ and β first requires a (nonparametric) estimator of the conditional densities involved in the integral equation. However, once these estimators are defined, it remains a set of intrinsic difficulties in order to solve this equation in (ϕ, β). As noted, for instance, by Newey and Powell (2003) or Florens (2003), one of these problems lies in the noncontinuity of the resulting estimators with respect to joint distribution of the data. This lack of continuity is usually referred as the *ill-posedness* of the problem. In particular it implies that, even if we can find consistent estimators for the conditional densities, it will not lead to a consistent estimator for ϕ or β.

One solution to avoid ill-posedness is to assume that ϕ lies in a compact set of functions, see e.g. Tikhonov, Goncharsky, Stepanov, and Yagola (1995) and, in econometrics, see Newey and Powell (2003), Ai and Chen (2003) or Chen (2007). This assumption automatically eliminates the stability issue in solving the integral equation, and leads to a well-posed problem. Under that assumption, a consistent estimator of β is therefore easier to derive.

The compactness assumption is however difficult to interpret economically and, more importantly, it assumes too much constraint on the set of ϕ.[1] It is possible to deal with the ill-posedness without assuming compactness, and there exists a large literature on techniques to stabilize the inversion of the integral equations such as equation (2.2). In econometric contexts, we refer to Carrasco, Florens, and Renault (2007) for an overview of different methods. The treatment of the fully nonparametric model with this approach can be found in Darolles, Florens, and Renault (2002), Florens (2003) and Hall and Horowitz (2005).

[1] To illustrate this limitation, suppose e.g. that ϕ belongs to a unit ball. The unit ball is compact if and only if it is finite dimensional [e.g. Kress (1999), Theorem 2.10]. In other words, the function ϕ can be expanded as a finite, linear combination of basis functions, which is clearly not a desirable structural assumption for this nonparametric function.

Below we also characterize situations where no regularity assumption on φ is needed in order to derive an efficient estimation for the parameter β.

In this paper, we propose estimators of ϕ and β in the partially linear model (2.1a–2.1b) in the framework of ill-posed inverse problems.

2.2 Operators

As we have seen, a clarification is necessary about the function space involved in the problem. In this paper, we consider L^2 spaces with respect to some specific measure. If this measure is the joint probability measure of the data, then we write $L^2_f(Y)$ or $L^2_f(Z)$ to denote for example functions depending on Y or Z only.

Note that equation (2.2) may be reformulated in different ways (namely by multiplication with functions of W) and leads to different choices of function spaces. One important result of the present paper is to relate this choice to the optimality of the estimator.

Let π and τ be two probability densities. We define

$$T_X : \mathbb{R}^k \to L^2_\tau(\mathbb{R}^q) : \tilde{\beta} \mapsto \mathbb{E}\{X'\tilde{\beta}|W = \cdot\}\frac{f_W(\cdot)}{\tau(\cdot)} \tag{2.3}$$

$$T_Z : L^2_\pi(\mathbb{R}^p) \to L^2_\tau(\mathbb{R}^q) : \tilde{\phi} \mapsto \mathbb{E}\{\tilde{\phi}(Z)|W = \cdot\}\frac{f_W(\cdot)}{\tau(\cdot)} \tag{2.4}$$

where $L^2_\tau(\mathbb{R}^q)$ and $L^2_\pi(\mathbb{R}^p)$ are Hilbert spaces of square integrable functions with respect to the measure τ or π respectively. We can then write (ϕ, β) as the solution of

$$r = T_Z\phi + T_X\beta. \tag{2.5}$$

where $r = \mathbb{E}(Y|W)f_W/\tau$. As we shall prove in this paper, the choice of τ is related to some optimality for the estimation of β.

It is also useful to introduce the corresponding adjoint operators:

$$T_X^\star : L^2_\tau(\mathbb{R}^q) \to \mathbb{R}^k : g \mapsto \mathbb{E}\{Xg(W)\} \tag{2.6}$$

$$T_Z^\star : L^2_\tau(\mathbb{R}^q) \to L^2_\pi(\mathbb{R}^p) : g \mapsto \mathbb{E}\{g(W)|Z = \cdot\}\frac{f_Z(\cdot)}{\pi(\cdot)} \tag{2.7}$$

One interesting point with the introduction of the two functions π and τ is that it allows us to cover different viewpoints taken in the literature. If $\pi = f_Z$ and $\tau = f_W$, then we adopt the setting of Darolles, Florens, and Renault (2002)[2]. If π and τ are $\mathcal{U}[0, 1]$, then we fit to the setting of Hall and Horowitz (2005).

There is however one more fundamental reason to introduce these probability measures in our definition of the operators. The choice of π is related to identification issues, as it is shown in Section 2 below. In particular, we obviously have that ϕ can only be identified on $\operatorname{supp}\pi \cap \operatorname{supp} f_Z$ (the intersection between the supports of f_Z and π). Moreover, the choice

[2]Except Appendix C of Darolles, Florens, and Renault (2002), where a similar generalization is provided in order to model unbounded densities.

of τ will have no influence on the rate of convergence of the proposed estimators, but is related to their asymptotic efficiency, as shown in Section 4.

Throughout the paper, we assume that the operators T_X, T_Z, their dual, and r are well-defined. This point is formalized by the following assumption.

ASSUMPTION 2.1. *With the above notations, we assume that $r \in L^2_\tau(\mathbb{R}^q)$ and that both functions*

$$\mathbb{E}\left(\psi(Z)|W=\cdot\right)\frac{f_W(\cdot)}{\tau(\cdot)} \qquad \text{and} \qquad \mathbb{E}\left(X_i|W=\cdot\right)\frac{f_W(\cdot)}{\tau(\cdot)}$$

belong to $L^2_\tau(\mathbb{R}^q)$ for all $\psi \in L^2_\pi(\mathbb{R}^p)$ and $i = 1,\ldots,k$.

We illustrate this assumption in the next two examples, where we state sufficient conditions such that all quantities are well-defined.

EXAMPLE 2.1. Assumption 2.1 holds true if both $\mathbb{C}\mathrm{ov}(X)$ and $\mathbb{V}\mathrm{ar}(Y)$ are finite and if there exists some positive constants C_1 and C_2 such that $f_W \leqslant C_1 \cdot \tau$ on the support of τ and $f_Z \leqslant C_2 \cdot \pi$ on the support of π. If we set to zero functions outside the support of π and τ, then these conditions imply respectively $L^2_\pi(\mathbb{R}^p) \subseteq L^2_f(Z)$ and $L^2_\tau(\mathbb{R}^q) \subseteq L^2_f(W)$. □

EXAMPLE 2.2. Assumption 2.1 holds true if both $\mathbb{C}\mathrm{ov}(X)$ and $\mathbb{V}\mathrm{ar}(Y)$ are finite and the following Hilbert-Schmidt conditions are fulfilled:

$$\text{(i)} \quad \iint \left(\frac{f_{YW}(y,w)}{f_Y(y)\tau(w)}\right)^2 f_Y(y)\tau(w)dwdy < \infty\,,$$

$$\text{(ii)} \quad \iint \left(\frac{f_{XW}(x,w)}{f_X(x)\tau(w)}\right)^2 f_X(x)\tau(w)dwdx < \infty\,,$$

$$\text{(iii)} \quad \iint \left(\frac{f_{ZW}(z,w)}{\pi(z)\tau(w)}\right)^2 \pi(z)\tau(w)dwdz < \infty\,.$$

In particular, these conditions imply the compactness of the operator $T_Z^\star T_Z$. The Hilbert-Schmidt conditions hold true for instance when all variables are Normal. □

2.3 Existence and identification

We now give conditions for the existence of solutions and for the identification of the parameters from the partial linear model (2.1a–2.1b). Recall that (ϕ,β) are the solution of the equation $r = T_Z\phi + T_X\beta$, where $r = \mathbb{E}(Y|W)f_W/\tau$. A necessary and sufficient condition for the existence of solutions is to assume

$$r \in \mathcal{R}(T_Z) + \mathcal{R}(T_X) = \left\{\psi_Z + \psi_X \text{ such that } \psi_Z \in \mathcal{R}(T_Z) \text{ and } \psi_X \in \mathcal{R}(T_X)\right\}$$

where $\mathcal{R}(T)$ denotes the range of the operator T, i.e. the set of all image elements. [3]

[3]This condition is obviously not always satisfied. When it is possible we can define the parameters of interest (ϕ,β) as

$$(\phi,\beta) = \arg\min\left\{\|r - T_Z\tilde{\phi} - T_X\tilde{\beta}\|_{L^2(W)} \text{ such that } \tilde{\phi} \in L^2_\pi(\mathbb{R}^p) \text{ and } \tilde{\beta} \in \mathbb{R}^k\right\}.$$

The next assumption is a necessary and sufficient condition for the identification of the parameters.

ASSUMPTION 2.2. *The two following conditions hold true:*

(i) The operators T_X and T_Z are injective, i.e. $T_X\beta = 0 \Rightarrow \beta = 0$ and $T_Z\phi = 0 \Rightarrow \phi = 0$,

(ii) $\mathcal{R}(T_X) \cap \mathcal{R}(T_Z) = \{0\}$.

Assumption 2.2 gives conditions on operators. It might be useful or intuitive to translate these conditions on random variables instead. The two following assumptions are together equivalent to Assumption 2.2(i):

(a) The vector Z is strongly identified by W with respect to π, that is

$$\forall h \in L^2_\pi(\mathbb{R}^p) \text{ such that } \frac{f_W}{\tau}\mathbb{E}\{h(Z)|W\} = 0 \ \tau\text{-a.s.} \Longrightarrow h(Z) = 0 \ \tau\text{-a.s.}$$

(b) The matrix

$$\Omega := \mathbb{E}\left\{\mathbb{E}(X|W)\,\frac{f_W(W)}{\tau(W)}\,\mathbb{E}(X^t|W)\right\} \tag{2.8}$$

has full rank.

Condition (a) refers to the concept of strong identification of random variables and corresponds to the notion of complete statistics in the statistical literature (see, e.g., Lehmann and Scheffe (1950)). This condition is weaker than requiring the strong identification of X, Z by W. This weak condition comes from the semiparametric structure of the model. Note also that the matrix Ω of condition (b) is the asymptotic variance of the Generalized Method of Moment estimator for the heteroscedastic model with $\mathbb{V}\mathrm{ar}(U|W)f_W = \tau$ (see Chamberlain (1987)).

Finally observe that, if (Z, X) is jointly strongly identified by W, then the condition (ii) follows if the random variables X and Z are measurable separable[4]. A standard reference on this concept is Chapter 5 of Florens, Mouchart, and Rolin (1990) and a more recent discussion can be found in San Martín, Mouchart, and Rolin (2005).

THEOREM 2.1. *Suppose the model is well-defined (Assumption 2.1). Then Assumption 2.2 is necessary and sufficient for the identification of the function ϕ and the vector β in the model (2.1a)–(2.1b).*

This solution is called *minimal norm solution*, but it can happen that this solution does not exists. Theorem 2.6 of Engl, Hanke, and Neubauer (1996) gives the following necessary and sufficient condition for the existence of the minimal norm solution: the solution r must be such that $r \in \{\mathcal{R}(T_Z)+\mathcal{R}(T_X)\} \oplus \{\mathcal{R}(T_Z)+\mathcal{R}(T_X)\}^\perp$, where $\Omega^\perp$ denotes the orthogonal space to the space Ω. See Chapter 2 of Engl, Hanke, and Neubauer (1996) for details. Note also that, if it exists, the minimal norm solution is not necessarily unique. The general problem of non identifiable nonparametric inverse problems is considered in Johannes (2006), where an estimator of the space of solutions is derived.

[4] X and Z are measurable separable when any function of Z a.s. equal to $X^t\beta$ for a given β is equal to a constant a.s.

Remark 2.1 (Identification of β). As the main object of this paper is the accurate estimation of the parameter β, one can ask what are the necessary and sufficient conditions for the identification of β only. A straightforward adaptation of the above theorem shows that these conditions are given by a relaxed version of Assumption 2.2, where condition (i) is replaced by the injectivity of T_X only. We do not need the injectivity of T_Z, but note that we still need that the single common element between the range of the two operators is zero.

Remark 2.2 (Common variables between X and Z). If the model contains common variables between X and Z, Assumption 2.2 (ii) about the range of the operators T_X and T_Z is no longer fulfilled. That situation requires other identification conditions. One solution is to impose more regularity on ϕ. A sufficient condition for identification is for instance the existence of a measurable set $A \subseteq \mathbb{R}^p$ of nonzero measure, such that $\phi(A) \equiv \partial_k \phi(A) \equiv 0$, where $\partial_k \phi$ denotes the derivative of ϕ with respect to component $k = 1, \dots, p$. This condition restricts the set of ϕ, thus modifies the operator T_Z to another operator $\widetilde{T}_Z$ such that the range condition is fulfilled: $\mathcal{R}(T_X) \cap \mathcal{R}(\widetilde{T}_Z) = \{0\}$. Note that the estimation procedure in that situation is much more complicated and must be adapted to this restriction.

3 Estimation of β

Suppose we observe iid vectors (Y_i, Z_i, X_i, W_i), $i = 1, \dots, n$ from the model (2.1a)–(2.1b) and suppose that the parameters of the model are identified. Recall the definition of the operators in Section 2.2. The normal equations are

$$T_Z^\star r = T_Z^\star T_Z \phi + T_Z^\star T_X \beta \tag{3.1a}$$

$$T_X^\star r = T_X^\star T_Z \phi + T_X^\star T_X \beta. \tag{3.1b}$$

Note that, analogously to the case of the linear regression model, this system projects the problem (2.5) onto the parameter spaces $\mathbb{R}^k$ and $L^2_\pi(\mathbb{R}^p)$ using the adjoint operators.

Before solving this system, we consider the case where the cross terms $T_Z^\star T_X$ and $T_X^\star T_Z$ vanish. In this situation, the estimate of β does not depend on (an estimate of) ϕ.

3.1 Separate estimation of β

The condition that $T_Z^\star T_X$ and $T_X^\star T_Z$ vanish is equivalent to the condition that the range of T_X is orthogonal to the range of T_Z, i.e. $\mathcal{R}(T_X) \perp \mathcal{R}(T_Z)$. This orthogonality condition holds true for instance when we can find two independent sets of instruments for X and Z, i.e. when $W = (W_1, W_2)$ such that $Z \perp\!\!\!\perp W | W_1$, $X \perp\!\!\!\perp W | W_2$ and $W_1 \perp\!\!\!\perp W_2$. However note that we are not limited to this particular case.

When $\mathcal{R}(T_X) \perp \mathcal{R}(T_Z)$ we can study separately the estimation of β and of ϕ, which are given by

$$\phi = (T_Z^\star T_Z)^{-1} T_Z^\star r \tag{3.2a}$$

$$\beta = (T_X^\star T_X)^{-1} T_X^\star r. \tag{3.2b}$$

From the estimation of $T_X^\star T_X$ and $T_X^\star r$, an estimator of β can be defined.

It appears that this estimator is not optimal, in the sense that it is not asymptoticaly efficient. Because we will present an efficient estimator in the next section, we defer the properties of (3.2b) to Appendix A below.

3.2 General estimation approach

In the general case, we consider the system (3.1a–3.1b), where the cross-terms $T_Z^\star T_X$ and $T_X T_Z^\star$ do not vanish. This linear system is equivalent to

$$T_Z^\star (I - P_X) r = T_Z^\star (I - P_X) T_Z \phi \tag{3.3a}$$

$$T_X^\star (I - P_Z) r = T_X^\star (I - P_Z) T_X \beta \tag{3.3b}$$

where $P_X = T_X (T_X^\star T_X)^{-1} T_X^\star$ is the orthogonal projection operator onto the range $\mathcal{R}(T_X)$ of T_X and, similarly, P_Z is the projection onto the closure of the range $\mathcal{R}(T_Z)$. Note that, for all function in the range $\mathcal{R}(T_Z)$, this last projection can be written as $P_Z = T_Z (T_Z^\star T_Z)^{-1} T_Z^\star$.

Below we introduce estimators for the operators involved in this system.

From (3.3a), we see that the estimation of ϕ is again an ill-posed problem since here the inversion of $T_Z^\star (I - P_X) T_Z$ is not stable. We refer to the standard literature on estimation and regularization in nonparametric instrumental regression models, which offer a complete solution to this problem.

The interesting and new fact arises from the equation (3.3b), in which the inversion of $(T_Z^\star T_Z)$ is a source of instability. In consequence, the estimation of β is also ill-posed and a regularized estimate is necessary in order to get a consistent estimator. Ill-posedness however may lead to a very slow rate of convergence of the estimator of β. In the following we give regularity conditions on T_Z, T_X and ϕ such that we get a $\sqrt{n}$-consistent, asymptotically Normal estimate.

In order to define these regularity conditions, we assume that the operator T_Z is compact, which allows to write its singular value decomposition. Namely, there exists a system $\{\varphi_j\}$ of functions of $L^2_\pi(\mathbb{R}^p)$ and a system $\{\psi_j\}$ in $L^2_\tau(\mathbb{R}^q)$ such that

$$T_Z \phi = \sum_{j=1}^{\infty} \lambda_j \langle \phi, \varphi_j \rangle \psi_j \qquad \text{for all } \phi \in L^2_\pi(\mathbb{R}^p)\ , \tag{3.4}$$

where the coefficients λ_j are the strictly positive singular values of T_Z. As the operator T_X is always compact, we also consider a system of eigenvector $\{e_j\}$ in $\mathbb{R}^k$ and a system $\{\tilde{\psi}_j\}$ in $L^2_\tau(\mathbb{R}^q)$ such that

$$T_X \beta = \sum_{j=1}^{k} \mu_j \langle \beta, e_j \rangle \tilde{\psi}_j \qquad \text{for all } \beta \in \mathbb{R}^k\ ,$$

where the coefficients μ_j are the strictly positive singular values of T_X.

Assuming T_Z to be compact allows us to estimate this operator using a kernel smoothing procedure[5]. In the singular value decomposition of T_Z, the ill-posedness comes from the behavior of the singular values λ_j which tend to 0 as j increases. Also, note that the systems of eigenfunctions $\{\varphi_j\}$ and $\{\psi_j\}$ are infinite, while the systems $\{e_j\}$ and $\{\tilde{\psi}_j\}$ contain k elements.

The following assumption presents the regularity conditions for T_Z, T_X and ϕ.

Assumption 3.1 (Source conditions). *There exists $\eta > 0$ and $\nu > 0$ such that*

$$\max_{i=1,\dots,k} \sum_{j=1}^{\infty} \frac{\langle \tilde{\psi}_i, \psi_j \rangle^2}{\lambda_j^{2\eta}} < \infty \ , \tag{3.5}$$

and

$$\sum_{j=1}^{\infty} \frac{\langle \phi, \varphi_j \rangle^2}{\lambda_j^{2\nu}} < \infty \ . \tag{3.6}$$

This assumption quantifies the strength (or weakness) of the instrument W. Since this type of assumption is new in econometrics[6], we will discuss its relevance and some interpretations in the next paragraphs.

The condition (3.5) means that the operator T_X is "adapted" to the operator T_Z, and this adaptation is controlled by the parameter η. If $\mathcal{R}(T_X)$ and $\mathcal{R}(T_Z)$ are orthogonal, then $\eta = \infty$ and this case is discussed in Appendix A below. Then the parameter η may be interpretated as a degree of collinearity between Z and X through the projection onto the instruments W: roughly speaking, the bigger the parameter η, the more orthogonal are the ranges $\mathcal{R}(T_X)$ and $\mathcal{R}(T_Z)$.

In addition to this interpretation, the following examples illuminate Assumption 3.1 in some particular cases.

Example 3.1. Suppose X can be written as $X = m(V)$ for a given function m and a p-dimensional random variable V such that the linear operator

$$T_V : L^2_{f_V}(\mathbb{R}^p) \to L^2_\tau(\mathbb{R}^q) : g \mapsto \mathbb{E}\{g(V)|W = \cdot\} \frac{f_W(\cdot)}{\tau(\cdot)}$$

has a singular value decomposition given by $T_V g = \sum_{j=1}^{\infty} \gamma_j \langle g, \kappa_j \rangle \psi_j$ for all $g \in L^2_{f_V}(\mathbb{R}^p)$. Note that $\{\psi_j\}$ is the singular system of T_Z and T_V. Denote by m_i the i-th component of the vector valued function m and assume that $m_i \in L^2_{f_V}(\mathbb{R}^p)$ for $i = 1, \dots, k$. In that case, by orthonormality of the system $\{\psi_j\}$, condition (3.5) is equivalent to

$$\max_{i=1,\dots,k} \sum_{j=1}^{\infty} \frac{\gamma_j^2}{\lambda_j^{2\eta}} \langle m_i, \kappa_j \rangle^2 < \infty$$

[5] A sufficient condition for the compactness of T_Z is given by the Hilbert Schmidt condition, see Example 2.2.

[6] With one noticeable exception for condition (3.6) that already appears in Darolles, Florens, and Renault (2002).

and a sufficient condition is to check whether $\gamma_j/\lambda_j^\eta \leqslant C$ for some constant C. $\square$

The relevance of this example comes from the fact that the parameters μ_i and λ_i are estimable from the data, and then this assumption is testable. Moreover, these parameters are linked to the correlation between the instruments W and the variables X and Z respectively. The two following examples illustrate this point in some particular cases starting with the Normal model.

Example 3.2 (Normal model). Suppose $(X, Z, W_1, W_2) \sim \mathcal{N}(0, \Sigma)$ and

$$\Sigma = \begin{pmatrix} 1 & 0 & \rho_{X,1} & \rho_{X,2} \\ 0 & 1 & \rho_{Z,1} & 0 \\ \rho_{X,1} & \rho_{Z,1} & 1 & 0 \\ \rho_{X,2} & 0 & 0 & 1 \end{pmatrix}$$

for $0 < |\rho_{X,1}|, |\rho_{X,2}|, |\rho_{Z,1}| \leqslant 1$. Here, note that $Z \perp\!\!\!\perp W_2$ and the case $\rho_{X,1} = 0$ corresponds to the situation where $\mathcal{R}(T_X) \perp \mathcal{R}(T_Z)$. We also take $\pi \in \mathcal{N}(0,1)$ and $\tau \sim \mathcal{N}(0, I_2)$ where I_2 denotes the 2×2 identity matrix. The singular system of T_X reduces to $\{\mu_1, e_1, \tilde{\psi}_1\}$ where $e_1 \equiv 1$ and $\tilde{\psi}_1(w_1, w_2) = (\rho_{X,1}w_1 + \rho_{X,2}w_2)/\mu_1$ with corresponding singular value $\mu_1^2 = \rho_{X,1}^2 + \rho_{X,2}^2$. Moreover, the singular system of T_Z is given by the (univariate) Hermite polynomials H_j in both $L^2_\pi(\mathbb{R})$ and $L^2_\tau(\mathbb{R}^2)$, i.e. $\psi_j(w_1, w_2) = H_j(w_1)$ for all w_1, w_2. The corresponding singular values are $\lambda_j = \rho_{Z,1}^j$. Since $H_1(w_1) = 1$ and $H_2(w_1) = 2w_1$, the orthonormality property of the Hermite polynomials simplifies the regularity condition (3.5) as

$$\sum_{j=1}^{\infty} \frac{\langle \tilde{\psi}_1, \psi_j \rangle^2}{\rho_{Z,1}^{2j\eta}} = \sum_{j=1}^{\infty} \frac{\rho_{X,1}^2}{4\rho_{Z,1}^{2j\eta}} \langle H_2, \psi_j \rangle^2 = \frac{\rho_{X,1}^2}{4\rho_{Z,1}^{4\eta}} ,$$

which is obviously finite for every η. In conclusion, this example always satisfies the source condition for all η. $\square$

Example 3.3. The preceding example can be generalized to the case where the k-dimensional random variable X is not normally distributed. Suppose that $X = m(V)$, where $(V, Z, W_1, W_2) \sim \mathcal{N}(0, \Sigma)$ and

$$\Sigma = \begin{pmatrix} 1 & 0 & \rho_{V,1} & \rho_{V,2} \\ 0 & 1 & \rho_{Z,1} & 0 \\ \rho_{V,1} & \rho_{Z,1} & 1 & 0 \\ \rho_{V,2} & 0 & 0 & 1 \end{pmatrix}$$

for $0 < |\rho_{V,1}|, |\rho_{V,2}|, |\rho_{Z,1}| \leqslant 1$ similarly to Example 3.2. The function m is vector-valued with components in $L^2_{f_V}(\mathbb{R})$ as in Example 3.1. Combining the above Examples 3.1 and 3.2, we see that the source condition is satisfied for all η when m takes a polynomial form. For a general function m, a sufficient condition for (3.5) is to require that $\rho_{V,1}/\rho_{Z,1}^{\eta}$ is bounded by some constant C. The source condition is then directly related to the correlation scheme between the random variables. $\square$

In order to define the estimator, we first recall the definition of the multivariate kernel (see Scott (1992)).

Definition 3.1. *For all $w = (w_1, \dots, w_q) \in \mathbb{R}^q$, K is a multiplicative kernel of order m, i.e. $K(w) = \Pi_{i=1}^q \ell(w_i)$ where ℓ is a univariate, continuous, bounded function such that*

$$\int \ell(u)\, du = 1, \qquad \int u^i \ell(u)\, du = 0$$

for all $i = 1, \dots, m-1$ and there exists two finite constants s_K^m and C_K such that

$$\int u^m \ell(u)\, du = s_K^m, \qquad \int \ell(u)^2\, du = C_K.$$

We now consider nonparametric estimator of the operators and define our estimator of β:

$$\widehat{T}_X \tilde{\beta} = \frac{1}{n} \sum_{i=1}^n X_i^t \tilde{\beta}\, \frac{K_{h_W}(W_i - \cdot)}{\tau(\cdot)} \qquad \text{for all } \tilde{\beta} \in \mathbb{R}^k\ , \tag{3.7}$$

$$\widehat{T}_X^\star \psi = \frac{1}{n} \sum_{i=1}^n X_i \int K_{h_W}(W_i - w)\psi(w)\, dw \qquad \text{for all } \psi \in L_\tau^2(\mathbb{R}^q)\ , \tag{3.8}$$

$$\widehat{T}_Z \tilde{\phi} = \frac{1}{n} \sum_{i=1}^n \frac{K_{h_W}(W_i - \cdot)}{\tau(\cdot)} \int K_{h_Z}(Z_i - z)\tilde{\phi}(z)\, dz \qquad \text{for all } \tilde{\phi} \in L_\pi^2(\mathbb{R}^p)\ , \tag{3.9}$$

$$\widehat{T}_Z^\star \psi = \frac{1}{n} \sum_{i=1}^n \frac{K_{h_Z}(Z_i - \cdot)}{\pi(\cdot)} \int K_{h_W}(W_i - w)\psi(w)\, dw \qquad \text{for all } \psi \in L_\tau^2(\mathbb{R}^q)\ , \tag{3.10}$$

$$\hat{r} = \frac{1}{n} \sum_{i=1}^n Y_i \frac{K_{h_W}(W_i - \cdot)}{\tau(\cdot)}\ , \tag{3.11}$$

for some bandwidth parameters h_W, h_Z that depend on n. It is worth mentioning that these estimators are constructed such that the dual of $\widehat{T}_X$ (resp. $\widehat{T}_Z$) is precisely given by $\widehat{T}_X^\star$ (resp. $\widehat{T}_Z^\star$). This fact is used in the proof of the next theorems. Moreover, with the standard choice for the parameter h, these estimators achieve sufficiently good convergence properties, see Lemma B.3 in the Appendix for details on this convergence. Of course, one could consider other nonparametric estimators and this choice is directly related to the smoothness assumptions we allow on the density f.

Together with sufficient regularity assumptions on the kernel K, $\sqrt{n}$-consistency is achieved if we impose some regularity conditions on the joint density f. The next definition provides the suitable space of regularity for f in order to prove all the results of this paper (see also Definition 2 of Robinson (1988)).

Definition 3.2. *For a given function γ and for $\alpha \geqslant 0$, $s > 0$, the space $\mathfrak{G}_\gamma^{s,\alpha}(\mathbb{R}^\ell)$ is the class of functions $g : \mathbb{R}^\ell \to \mathbb{R}$ satisfying: g is everywhere $(m-1)$-times partially differentiable for $m - 1 < s \leqslant m$; for some $\rho > 0$ and for all x, the inequality*

$$\sup_{y:|y-x|<\rho} \frac{|g(y) - g(x) - Q(y-x)|}{|y-x|^s} \leqslant \psi(x), \tag{3.12}$$

holds true where $Q = 0$ when $m = 1$ and Q is an $(m-1)$th-degree homogeneous polynomial in $y - x$ with coefficients the partial derivatives of g at x of orders 1 through $m-1$ when $m > 1$; ψ is uniformly bounded by a constant when $\alpha = 0$ and the functions g and ψ have finite αth moments with respect to $1/\gamma$ when $\alpha > 0$, i.e. $\int g^\alpha(x)/\gamma(x)\, dx < \infty$ and $\int \psi^\alpha(x)/\gamma(x)\, dx < \infty$. We also write $\mathfrak{G}^{s,\alpha}(\mathbb{R}^l)$ when $\gamma \equiv 1$.

In this definition, (3.12) provides a bound on a Taylor series remainder term.

The parameter β is then estimated by

$$\hat{\beta} = \widehat{M}_\alpha^{-1} \hat{v}_\alpha \tag{3.13}$$

where $\hat{v}_\alpha$ is an estimator of the left hand side of (3.3b) given by

$$\hat{v}_\alpha := \widehat{T}_X^\star \left(I - \widehat{T}_Z(\alpha I + \widehat{T}_Z^\star \widehat{T}_Z)^{-1} \widehat{T}_Z^\star \right) \hat{r}$$

and $\widehat{M}_\alpha$ is an estimator of the RHS given by

$$\widehat{M}_\alpha := \widehat{T}_X^\star \left(I - \widehat{T}_Z(\alpha I + \widehat{T}_Z^\star \widehat{T}_Z)^{-1} \widehat{T}_Z^\star \right) \widehat{T}_X$$

for a positive parameter α that depends on n. We refer to α as the regularization parameter. Note that here we use the Tikhonov regularization method to stabilize the inversion of $T_Z^\star T_Z$. It is of course possible to consider another scheme of regularization (see Carrasco, Florens, and Renault (2007)).

Theorem 3.1. *Consider the nonparametric estimators (3.7–3.11) constructed using a multivariate kernel of order r (Definition 3.1) and for $j = 1, \ldots, k$ suppose (i) the functions $\int x_j^2 f(x, \cdot)dx$ and $\int y^2 f(y, \cdot)dx$ belong to $\mathfrak{G}_\tau^{1,1}(\mathbb{R}^q)$; (ii) the functions $\int x_j f(x, w)dx$ and $\int y f(y, \cdot)dx$ belong to $\mathfrak{G}_\tau^{s,2}(\mathbb{R}^q)$; (iii) the function f_{ZW} belongs to $\mathfrak{G}_{\pi\cdot\tau}^{1,1}(\mathbb{R}^{p+q}) \cap \mathfrak{G}_{\pi\cdot\tau}^{s,2}(\mathbb{R}^{p+q})$. In addition, define $\rho := r \wedge s$ and assume that the bandwidth parameters are such that $h_W = O(n^{-1/(p+q+2\rho)})$ and $h_Z = O(n^{-1/(p+q+2\rho)})$. Suppose that the source condition (Assumption 3.1) is satisfied for some $\nu \geqslant 0$ and $\eta \geqslant 1$. Moreover, if $\eta \geqslant 2$ and $2\rho \geqslant p + q$, we assume*

$$\alpha \cdot n^{\frac{p+q+(2-\nu\wedge 2)\rho}{p+q+2\rho}} = O(1), \quad \alpha^2 \cdot n = O(1)$$

while, if $1 \leqslant \eta < 2$, we assume

$$\alpha^{\eta-2} \cdot n^{\frac{p+q-2\rho}{p+q+2\rho}} = O(1),\ \alpha \cdot n^{\frac{p+q+(2-\nu\wedge 2)\rho}{p+q+2\rho}} = O(1),\ \alpha^2 \cdot n = O(1)\ .$$

Then $\sqrt{n}\|\hat{\beta} - \beta\| = O_p(1)$.

To illustrate this result, we first give some sufficient conditions on the parameter α to get $\sqrt{n}$-consistency. Consider the situation where the source conditions (Assumption 3.1) are fulfilled with $\eta \geqslant 2$ and $2\rho \geqslant p+q$. Then $\alpha = O(n^{-1})$ is a sufficient rate to get the $\sqrt{n}$-consistency. It is interesting to note that α can tend to zero arbitrarily fast (at least faster than $n^{-1+(\nu\wedge 2)\rho/(p+q+2\rho)}$ and no lower bound is necessary for this convergence. This phenomenon is due to the regularity condition imposed on the problem in terms of source condition ($\eta \geqslant 2$). In this situation, a regularization parameter is mandatory in order to have $\sqrt{n}$-consistency, but this parameter can be arbitrarily small. Moreover, note that $\nu = 0$ is also possible. This means that *$\sqrt{n}$-consistency is achieved when no regularity condition on ϕ is assumed.*

The situation differs if $1 \leqslant \eta < 2$, that is if the problem is less regular. In that case the constraint

$$\alpha^{\eta-2} \cdot n^{\frac{p+q-2\rho}{p+q+2\rho}} = O(1)$$

imposes that α cannot converge arbitrarily fast to zero. This implies that, in contrast to the previous case, the rate $O(n^{-1})$ is then no longer valid for all choice of p, q, ρ. Still, the regularity parameter should converge faster than $n^{-1+(\nu\wedge 2)\rho/(p+q+2\rho)}$. In conclusion of this case, $\sqrt{n}$-consistency resulting from the above theorem requires that $\nu > 0$ in some situations. In other words the source condition on ϕ is a sufficient assumption in that situation.

A few more constraints on (α, h_W, h_Z) give the following Central Limit Theorem for $\hat{\beta}$. In particular, we will need some assumptions on the singular value decomposition of the compact operator $T_X^\star(I-P_Z)$. We denote by $\{\mu_j, g_j \in L^2_\tau(\mathbb{R}^q), e_j \in \mathbb{R}^k, j = 1, \ldots, k\}$ of $T_X^\star(I-P_Z)$ this singular value decomposition (similarly to the decomposition (3.5) for instance).

Theorem 3.2. *Consider the nonparametric estimators (3.7–3.11) constructed using a multivariate kernel of order r (Definition 3.1). Suppose assumptions (i) - (iii) of Theorem 3.1 are satisfied. In addition, define $v^2(\cdot) = \mathbb{V}\mathrm{ar}(U|W = \cdot)$ and assume that (iv) the functions $v^2 f_W$ and f_W belong to $\in \mathfrak{G}^{1,1}_\tau(\mathbb{R}^q)$; (v) the eigenfunctions g_j of $T_X^\star(I-P_Z)$ belong to $\mathfrak{G}^{1,0}_\tau(\mathbb{R}^q)$ and (vi) $g_j\sqrt{v^2 \cdot f_W/\tau}$ belong to $L^2_\tau(\mathbb{R}^q)$ for all $j = 1, \ldots, k$. Moreover, define $\rho := r \wedge s$ and suppose that the bandwidth parameters are such that $h_W = O(n^{-1/(p+q+2\rho)})$ and $h_Z = O(n^{-1/(p+q+2\rho)})$. Suppose in addition that the source conditions (Assumption 3.1) are satisfied for some $\nu \geqslant 0$ and $\eta \geqslant 1$. If $\eta \geqslant 2$ and $2\rho \geq p+q$, assume*

$$\alpha \cdot n^{\frac{p+q+(2-\nu\wedge 2)\rho}{p+q+2\rho}} = o(1), \quad \alpha^2 \cdot n = o(1)$$

while, if $1 \leqslant \eta < 2$, assume

$$\alpha^{\eta-2} \cdot n^{\frac{p+q-2\rho}{p+q+2\rho}} = o(1),\ \alpha \cdot n^{\frac{p+q+(2-\nu\wedge 2)\rho}{p+q+2\rho}} = o(1),\ \alpha^2 \cdot n = o(1)\ .$$

Then we have

$$\sqrt{n}(\widehat{\beta}-\beta) \xrightarrow{d} \mathcal{N}\left(0, M^{-1}T_X^\star(I-P_Z)\left[\frac{v^2\cdot f_W}{\tau}(I-P_Z)T_X\right]M^{-1}\right) ,$$

where $M = T_X^\star(I-P_Z)T_X$.

As illustration of the last theorem consider the situation where the source condition (Assumption 3.1) are satisfied with $\eta \geqslant 2$ and $2\rho > p+q$. Then the rate $\alpha = o(n^{-1})$ is sufficient to get the central limit theorem. Again no lower bound is needed for α and the only constraint is that the regularization parameter should be faster than the rate $n^{-1+(\nu\wedge 2)\rho/(p+q+2\rho)}$. Moreover, as in the consistency theorem if $\eta \geqslant 2$ and $2\rho > p+q$ there is no regularity condition on ϕ necessary to obtain the asymptotic normality. The situation differs when $1 \leqslant \eta < 2$. In this less regular problem, α cannot converge arbitrarily fast to zero due to the constraint $\alpha^{\eta-2}\cdot n^{\frac{p+q-2\rho}{p+q+2\rho}} = o(1)$, but has to converge faster than $n^{-1+(\nu\wedge 2)\rho/(p+q+2\rho)}$.

Theorem 3.2 shows explicitly the influence of the function τ on the asymptotic variance of the estimator. If we take for instance τ such that $\mathbb{V}\mathrm{ar}(U|W)f_W(W) = \sigma^2\tau(W)$, then the asymptotic distribution reduces to $\mathcal{N}(0,\sigma^2 M^{-1})$. In the next section we show that this choice for τ gives an estimator that reaches the semiparametric efficiency bound.

4 Semiparametric efficiency of the general estimate

In the following we address the question of the efficiency of our estimator $\hat{\beta}$. Semiparametric efficiency bounds have now a long history and we refer to Newey (1990b) or Bickel, Klaassen, Ritov, and Wellner (1993) for standard references on this concept.

Suppose $\phi = g_\gamma$ is a known function of Z depending on a l-dimensional unknown parameter vector γ and partially differentiable in γ. If $(\hat{\gamma}_{GMM}, \hat{\beta}_{GMM})$ denotes in this parameterized model the optimal GMM estimator of (γ,β) derived from the optimal unconditional moment condition, then it is well known that under regularity conditions the optimal covariance matrix in the limiting normal distribution of $\sqrt{n}[(\hat{\gamma}_{GMM}, \hat{\beta}_{GMM}) - (\gamma,\beta)]$ is

$$\begin{pmatrix} \mathbb{E}\left\{\partial_\gamma g_\gamma(Z)v^{-2}(W)\mathbb{E}(\partial_\gamma g_\gamma(Z)|W)^t\right\} & \mathbb{E}\left\{\partial_\gamma g_\gamma(Z)v^{-2}(W)\mathbb{E}(X|W)^t\right\} \\ \mathbb{E}\left\{Xv^{-2}(W)\mathbb{E}(\partial_\gamma g_\gamma(Z)|W)^t\right\} & \mathbb{E}\left\{Xv^{-2}(W)\mathbb{E}(X|W)^t\right\} \end{pmatrix}^{-1} ,$$

see, e.g., Chamberlain (1987). If we assume $\mathbb{C}\mathrm{ov}(\partial_\gamma g_\gamma(Z)) < \infty$, then the operator

$$T_{g_\gamma(Z)} : \mathbb{R}^l \to L^2_\tau(\mathbb{R}^q) : \theta \mapsto \frac{f_W(W)}{\tau(W)}\,\mathbb{E}(\partial_\gamma g_\gamma(Z)|W)^t\,\theta$$

is well-defined and its adjoint operator is given by

$$T^\star_{g_\gamma(Z)} : L^2_\tau(\mathbb{R}^q) \to \mathbb{R}^l : \psi \mapsto \mathbb{E}\{\partial_\gamma g_\gamma(Z)\psi(W)\} .$$

With these notations, the optimal covariance matrix can be written

$$\begin{pmatrix} T^\star_{g_\gamma(Z)}\left[\frac{\tau}{v^2\cdot f_W}T_{g_\gamma(Z)}\right] & T^\star_{g_\gamma(Z)}\left[\frac{\tau}{v^2\cdot f_W}T_X\right] \\ T^\star_X\left[\frac{\tau}{v^2\cdot f_W}T_{g_\gamma(Z)}\right] & T^\star_X\left[\frac{\tau}{v^2\cdot f_W}T_X\right] \end{pmatrix}^{-1}.$$

By standard matrix calculation we obtain the optimal covariance matrix $M_{g_\gamma(Z)}$ in the limiting normal distribution of $\sqrt{n}(\hat{\beta}_{GMM}-\beta)$ which is given by

$$M^{-1}_{g_\gamma(Z)} = T^\star_X\left[\frac{\tau}{v^2\cdot f_W}T_X\right] - T^\star_X\left[\frac{\tau}{v^2\cdot f_W}T_{g_\gamma(Z)}\right]\cdot\left(T^\star_{g_\gamma(Z)}\left[\frac{\tau}{v^2\cdot f_W}T_{g_\gamma(Z)}\right]\right)^{-1}\cdot T^\star_{g_\gamma(Z)}\left[\frac{\tau}{v^2\cdot f_W}T_X\right].$$

Note that in the heteroscedastic case with τ chosen such that $v^2(\cdot)f_W(\cdot)=\sigma^2\tau(\cdot)$ the optimal covariance matrix is given by

$$\sigma^2 M^{-1}_{g_\gamma(Z)} = T^\star_X T_X - T^\star_X T_{g_\gamma(Z)}\cdot\left(T^\star_{g_\gamma(Z)}T_{g_\gamma(Z)}\right)^{-1}\cdot T^\star_{g_\gamma(Z)}T_X$$

and in the particular homoscedastic case, i.e., $v^2(\cdot)=\sigma^2$, we recover

$$\sigma^2 M^{-1}_{g_\gamma(Z)} = \mathbb{E}\left\{\mathbb{E}(X|W)\mathbb{E}(X|W)^t\right\} - \mathbb{E}\left\{\mathbb{E}(X|W)\mathbb{E}(\partial_\gamma g_\gamma(Z)|W)^t\right\}\cdot \cdot\left(\mathbb{E}\left\{\mathbb{E}(\partial_\gamma g_\gamma(Z)|W)\mathbb{E}(\partial_\gamma g_\gamma(Z)|W)^t\right\}\right)^{-1}\cdot\mathbb{E}\left\{\mathbb{E}(\partial_\gamma g_\gamma(Z)|W)\mathbb{E}(X|W)^t\right\}.$$

We can now state the efficiency result.

THEOREM 4.1. *Consider the nonparametric estimators (3.7–3.11) constructed using a multivariate kernel of order r (Definition 3.1). Suppose assumptions (i) - (vii) of Theorem 3.2 are satisfied and the parameters α, h_Z and h_W are chosen according to Theorem 3.2. If the density τ satisfies $\mathbb{V}\mathrm{ar}(U|W)f_W(W)=\sigma^2\tau(W)$, then the estimator $\hat{\beta}$ achieves the semiparametric efficiency bound, i.e., there exists a parametric model g_γ for ϕ such that $M_{g_\gamma(Z)}=\sigma^2[T^\star_X(I-P_Z)T_X]^{-1}$.*

REMARK 4.1 (Partially endogeneous model). It is interesting to compare this efficiency result with a more general result given in Ai and Chen (2003) that covers a partially endogeneous model. In our setting, assuming only partial endogeneity implies that $T_X\beta = X^t\beta$ and, after some algebra, one can check that the variance of our semiparametric efficient collapses with the one given in Section 6 of Ai and Chen (2003).

5 Finite sample properties

In order to study the finite sample properties of the estimator, we describe below the results of a Monte Carlo simulation. The model considered in the simulation is $Y=X^t\beta+\phi(Z)+U$ with function $\phi(z)=0.25z^2$, parameter $\beta=1$ and $U\sim N(0,1)$. Variables X and Z are here univariate, and are generated as follows: Generate the trivariate Gaussian vector $W = (W_1, W_2, W_3)' \sim N((0,0,0)', Id)$ and consider

$$\begin{cases} Z = W_1 + 2W_2 + W_3 + \eta_Z \ , \\ X = -2W_1 + W_2 - W_3 + \eta_X \ . \end{cases}$$

The random variables η_Z and η_X are generated such that variables X and Z are endogenous:

$$\begin{cases} \eta_Z = U + \varepsilon_Z \ , & \varepsilon_Z \sim N(0,1) \ , \\ \eta_X = -2U + \varepsilon_X \ , & \varepsilon_X \sim N(0,1) \end{cases}$$

where ε_Z and ε_X are independent.

In our Monte Carlo study, data Y_i, X_i, Z_i, W_i are generated from this model and the general estimator (3.13) of β is computed. The choice of the measures π and τ that are used in the estimator is data driven: From each generated sample, we compute the sample mean ($\hat{\mu}_Z$ and $\hat{\mu}_W$) and the sample variance/covariance matrix ($\hat{\sigma}^2_Z$ and $\hat{\Sigma}_W$) from the observations Z_i and W_i. The measure π considered in the estimation is the density function of a Normal random variable with mean $\hat{\mu}_Z$ and variance $\hat{\sigma}^2_Z$. The measure τ is the density of the multivariate Normal random variable with mean $\hat{\mu}_W$ and covariance matrix $1.2 \cdot \hat{\Sigma}_W$.

The kernel used in the nonparametric estimation of all quantities (3.7—3.11) is the Gaussian kernel. In this simulation, the regularization parameter α and the four bandwidths h_Z, h_{W_1}, h_{W_2} and h_{W_3} are fixed. They were determined by the user from a set of preliminary samples generated from the above model. We found that the choice $\alpha = 3 \cdot 10^{-3}$ and $h_Z = h_{W_1} = h_{W_2} = h_{W_3} = 1.95$ gives reasonable results. It is still an open question how to find adaptive data-driven procedures to select these parameters.

	Sample size		
	$n = 100$	$n = 250$	$n = 500$
OLS	0.8190	0.8165	0.8172
	(0.025)	(0.017)	(0.008)
IV	0.9808	0.9939	1.0005
	(0.108)	(0.030)	(0.024)

Table 1: Average and standard error (in parenthesis) of the estimator of β from $M = 1000$ generations of the model explained in text. The true β is 1. OLS is computed from the pseudo-observations $(Y_i - \phi(Z_i), X_i)$ (see text) and IV is the semiparametric instrumental variable estimator.

Table 1 shows the results of the simulation for different sample sizes. From $M = 1000$ samples, the table gives the mean and standard error of the estimators. In order to have a point of comparison, we also compute the OLS estimator of β, ignoring endogeneity, using the pseudo-observations $(Y_i - \phi(Z_i), X_i)$, $i = 1, \ldots, n$, i.e. we subtract the theoretical nonlinear function ϕ from the dependent variable Y_i. The table compares the performance of our estimator with the results of the OLS estimator. The table shows that our procedure offers an appropriate correction for the endogeneity and is reasonably close to the real value of the parameter, even for very small sample sizes ($n = 100$).

It is worth mentioning that, in contrast to the OLS estimator, our semiparametric estimation procedure is computed from the true observations (Y_i, X_i, Z_i, W_i). In other words, it does not consider that the nonparametric function ϕ is known.

Moreover, even if the goal of this study is to investigate the finite sample accuracy for the estimator of the parameter β, we note in passing that an estimator of the nonparametric function $\phi(z)$ is easily obtained. From the normal equation (3.3a), the estimator is given by

$$\hat{\phi} = \left(\alpha I + \hat{T}_Z^\star (I - \hat{P}_X)\hat{T}_Z\right)^{-1} \hat{T}_Z^\star (I - \hat{P}_X)\hat{r} \tag{5.1}$$

where $\hat{P}_X = \hat{T}_X(\hat{T}_X^\star \hat{T}_X)^{-1}\hat{T}_X^\star$. Figure 1 shows this estimator (dashed line) from one single sample size. The true function ϕ is the solid line in the figure, and the gray line is the nonparametric kernel estimator of the regression function without controlling for endogeneity. That figure shows that the fit of the regression function is remarkable, and that the procedure corrects appropriately for the endogeneity of Z.

APPENDIX

A Separate estimation of β

When $\mathcal{R}(T_X) \perp \mathcal{R}(T_Z)$ we can study separately the estimation of β and of ϕ, which are given by (3.2a–3.2b). Even the estimation of β is not a standard problem given our assumption $\mathbb{E}\,(U|W) = 0$ (see (2.1b)). We first need a nonparametric estimator of $T_X^\star T_X$ and $T_X^\star r$. In the following we consider the estimator of $T_X^\star T_X$ given by

$$\hat{M} = \frac{1}{n(n-1)} \sum_{i \neq j} X_i X_j^t \frac{K_h\left(W_i - W_j\right)}{\tau(W_i)}$$

where $K_h(\cdot) = h^{-q}K(\cdot/h)$ for a given bandwidth $h = h(n) > 0$ and a multiplicative kernel K (see Definition 3.1 below). Similarly, an estimator of $T_X^\star r$ is given by

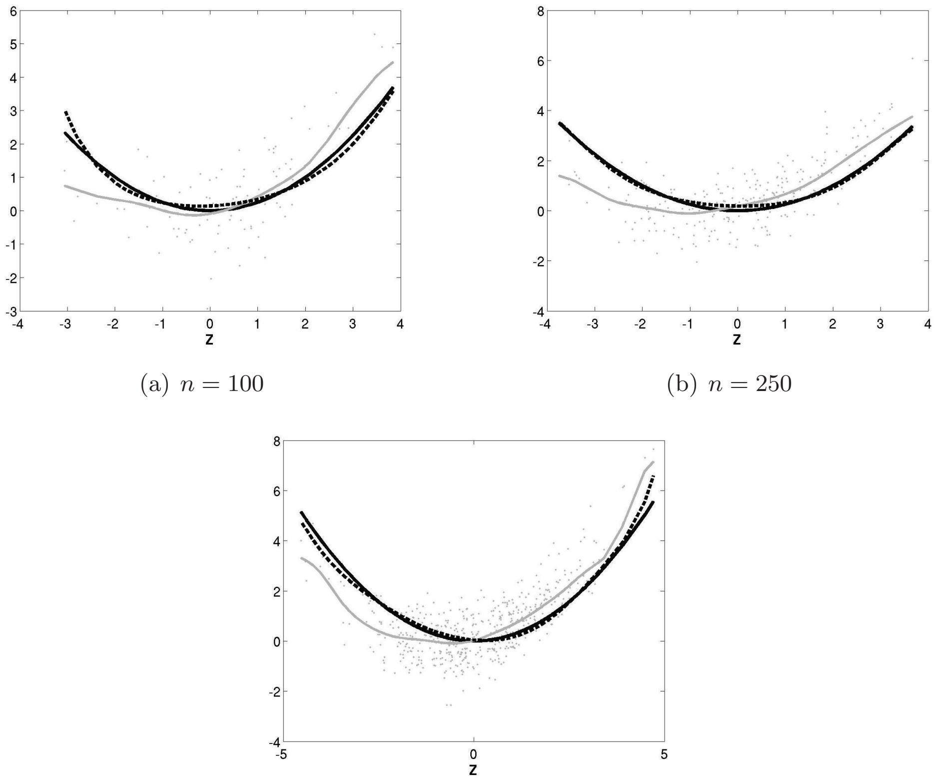
$$\hat{v} = \frac{1}{n(n-1)} \sum_{i \neq j} Y_i X_j \frac{K_h\left(W_i - W_j\right)}{\tau(W_i)}$$

Finally, our estimator of β is

$$\hat{\beta} = \hat{M}^{-1}\hat{v} \tag{A.1}$$

where K is a multivariate kernel (Definition 3.1). In the next results we use a kernel of order 2 to derive the $\sqrt{n}$-consistency of $\hat{\beta}$ and a central limit theorem for $\hat{\beta}$.

Theorem A.1. *Suppose $T_Z^\star T_X = T_X^\star T_Z = 0$ in the system of equations (3.1a–3.1b). If the function $g_1 = \mathbb{E}(\phi(Z)|W)f_W(W)$ belongs to $\mathfrak{G}_\tau^{2,2}(\mathbb{R}^q)$ and each component of the function $g_2 = \mathbb{E}(X|W)$ belongs to $\mathfrak{G}^{2,2}(\mathbb{R}^q)$, then the estimator (A.1) constructed with kernels of order 2 and with a bandwidth $h = O(n^{-1/2})$ is such that $\sqrt{n}\|\hat{\beta} - \beta\| = O_p(1)$.*



(a) $n = 100$

(b) $n = 250$

(c) $n = 500$

Figure 1: The solid line is the true function $\phi(z) = 0.25z^2$. Points represent the data Z_i. The gray line is the nonparametric estimator of ϕ without controlling for endogeneity. The dashed line is the nonparametric instrumental estimator (5.1). Each figure is for one single simulation, with different sample sizes n.

The assumption of the theorem involves the second derivative of f_W as it is usual in the context of kernel density estimation. This type of assumption comes to simplify a second-order expansion in the proof of the result and can be relaxed to milder assumption at the price of a more sophisticated estimation procedures with more technical proofs. This condition then does not appear as a structural restriction on the model.

Theorem A.2. *Under the assumptions of Theorem A.1,*

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}\left(0, (T_X^\star T_X)^{-1} \Lambda (T_X^\star T_X)^{-1}\right)$$

where $\Lambda := \mathbb{V}\mathrm{ar}(X\mathbb{E}(\phi(Z)|W) + (U + \phi(Z))\mathbb{E}(X|W))$ *and we recall that* $T_X^\star T_X$ *is a matrix with entries* $\mathbb{E}(X_s \mathbb{E}(X_t|W))$, $1 \leqslant s, t \leqslant k$.

The asymptotic variance of the theorem is not optimal, in the sense that it does not achieve the semiparametric efficiency bound. It is the consequence of the nuisance term $\phi(Z)$ which cannot be avoided even in the orthogonal situation $\mathcal{R}(T_X) \perp \mathcal{R}(T_Z)$.

The asymptotic variance of the central limit theorem simplifies when the nuisance term disappears, that is when $\phi = 0$. The following result considers this particular situation.

Corollary A.1. *Under the assumptions of Theorem A.1, if $\phi \equiv 0$, then*

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}\left(0, (T_X^\star T_X)^{-1} T_X^\star \left[\frac{v^2 f_W}{\tau} T_X\right] (T_X^\star T_X)^{-1}\right)$$

where $v^2(\cdot) := \mathbb{V}\mathrm{ar}(U|W = \cdot)$.

From this result, we see that if τ is such that $v^2(\cdot) f_W(\cdot) = \sigma^2 \tau(\cdot)$ for some $\sigma^2 > 0$, then the asymptotic covariance simplifies and the central limit theorem becomes

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \Omega^{-1}\right) \ ,$$

where Ω is the matrix $T_X^\star T_X$, see (2.8). In this particular case, the estimator $\hat{\beta}$ is optimal because it is identical to the GMM estimator constructed with optimal instruments in the homoscedastic setting. Indeed, the moment conditions in the homoscedastic model are $\mathbb{E}(Y - X'\beta|W) = 0$. This condition on the conditional moments can be replaced by the following condition on the marginal moments: $\mathbb{E}\{\psi(W)(Y - X'\beta)\} = 0$ for all functions ψ. The optimal GMM estimator corresponds to $\psi(\cdot) = \mathbb{E}(X|W = \cdot)$, in which case the estimator is the solution of

$$\mathbb{E}\left\{\mathbb{E}(X|W)(Y - X'\beta)\right\} = 0$$

which is equivalent to $T_X^\star T_X \beta = T_X^\star r$. This shows that our estimator $\hat{\beta}$ corresponds to the optimal GMM estimator in the homoscedastic model. More details can be found in Newey (1990a).

B Proofs

Proof of Theorem 2.1. Define the operator $T : L^2_\pi(\mathbb{R}^p) \otimes \mathbb{R}^k \to L^2_\tau(\mathbb{R}^q) : (\psi, \gamma) \mapsto T_Z\psi + T_X\gamma$. Note that an equivalent condition for the identification of the parameters (ϕ, β) in the model (2.1a–2.1b) is to assume that T is an injective operator.

First prove the necessary condition and consider a pair (ϕ, β) such that $T(\phi, \beta) = 0$ or equivalently $T_Z\phi = -T_X\beta$. The condition (ii) of Assumption 2.2 implies $T_Z\phi = T_X\beta = 0$ and thus, from condition (i), $\phi = 0$ and $\beta = 0$. Then T is injective.

We now prove the sufficient condition and suppose that T is an injective operator. If T_X or T_Z was not injective, then T would not be injective, this condition (i) of Assumption 2.2 is fulfilled. It reminds to show condition (ii). Suppose this condition does not hold, i.e. there exists a non-null function ψ in $\mathcal{R}(T_Z) \cap \mathcal{R}(T_X)$. This would imply the existence of $\phi_\psi \in L^2_\pi(\mathbb{R}^p) \backslash \{0\}$ and $\beta_\psi \in \mathbb{R}^k \backslash \{0\}$ such that $\phi = T_Z\phi_\psi = T_X\beta_\psi$. Then $T(\psi_\phi, \beta_\psi) = 0$ and, since T is injective, $\psi_\phi = 0$ and $\beta_\psi = 0$, thus we get a contradiction. □

Lemma B.1. *Under the assumptions of Theorem A.1, and if $h \to 0$ as $n \to \infty$,*

$$\mathbb{E}\hat{v} = T_X^\star r + O\left(h^2\right) \ , \tag{B.1}$$

$$\mathbb{E}\|\hat{v}\|^2 = \|T_X^\star r\|^2 + O\left(h^2\right) + O\left(n^{-1}\right) \ , \tag{B.2}$$

$$\mathbb{E}\widehat{M} = T_X^\star T_X + O\left(h^2\right) \ , \tag{B.3}$$

$$\mathbb{E}\|\widehat{M}\|^2 = \|T_X^\star T_X\|^2 + O\left(h^2\right) + O\left(n^{-1}\right) \ . \tag{B.4}$$

Proof. The proof is an application of standard techniques that can be found in the large literature on nonparametric kernel smoothing, see for instance Pagan and Ullah (1999). We only give whole details for the proof of (B.1). Using iterative conditional expectations, we can write

$$\mathbb{E}\hat{v} = \frac{1}{n(n-1)} \sum_{i \neq j} \mathbb{E}\left[Y_i X_j \mathbb{E}\left\{\frac{K_h\left(W_i - W_j\right)}{\tau(W_i)} \Big| Y_i X_j\right\}\right] \ .$$

With $g_1(w) := \int dy\, y f_{WY}(w,y)$ and $g_2(w) := \int dx\, x f_{WX}(w,x)$ (in vector notations),

$$\mathbb{E}\hat{v} = \iint g_1(w_1) g_2(w_2)\, \frac{dw_1 dw_2}{\tau(w_1)} K_h\left(w_1 - w_2\right)\ .$$

We now change variables and define u such that $w_2 = w_1 + uh$. We then write $g_2(w+uh)$ as $g_2(w)$ plus a reminder term. Since $g_2 \in \mathfrak{G}_\tau^{2,2}$ and using that the kernel K integrates to 1, this leads to $\mathbb{E}\hat{v} = T_X^\star r + R$, with[7]

$$\begin{aligned} R &\lesssim \iint g_1(w_1)\left\{Q(uh) + \psi(uh)(uh)^2\right\} K(u)\, \frac{dw_1\, du}{\tau(w_1)} \\ &= \iint g_1(w_1)\psi(uh)(uh)^2 K(u)\, \frac{dw_1\, du}{\tau(w_1)} \end{aligned}$$

where the last equality comes from the fact that $Q(uh)$ is a homogeneous polynomial of order one and that $\int uK(u)du = 0$. By definition of the multivariate kernel, and because g is uniformly bounded, R has rate $O(h^2)$. The proof of the other results is very similar but longer and we skip the details. $\square$

Lemma B.2. *Under the assumptions of Theorem A.1, if $h \to 0$ as $n \to \infty$, then*

$$\sqrt{n}(\hat{v} - \widehat{M}\beta) \xrightarrow{d} \mathcal{N}\left(0, \Lambda\right)$$

where $\Lambda = \mathbb{V}\mathrm{ar}(XT_Z\phi + (U + \phi(Z))T_X)$.

Proof. A straightforward expansion leads to

$$\hat{v} - \widehat{M}\beta = \frac{1}{n(n-1)} \sum_{i \neq j} X_i \left(U_j + \phi(Z_j)\right) \frac{K_h(W_i - W_j)}{\tau(W_i)}\ . \tag{B.5}$$

This U-statistic can be written $e := 2n^{-1}(n-1)^{-1}\sum_{i<j} H(S_i, S_j)$ where $S_i = (W_i, X_i, U_i, Z_i)$ and

$$H(S_i, S_j) = \frac{1}{2}\left\{\frac{X_i}{\tau(W_i)}(U_j + \phi(Z_j)) + \frac{X_j}{\tau(W_j)}(U_i + \phi(Z_i))\right\} K_h\left(W_i - W_j\right)\ .$$

By the asymptotic distribution theory of U-statistics (see Section 5.5 of Serfling (1980)), $\sqrt{n}(e - \mathbb{E}e) \xrightarrow{d} \mathcal{N}(0, 4\zeta)$ where $\zeta = \mathbb{V}\mathrm{ar}_f\, \mathbb{E}_f\left\{H(S_1, S_2)|S_1\right\}$. It remains to compute ζ. With $s_1 = (w_1, x_1, u_1, z_1)$, we define $H(s_1) := \mathbb{E}_f\{H(s_1, S_2)\}$. If $g_1(\tilde{w}) := \iint du\, dz\, (u + \phi(z)) f_{WUZ}(\tilde{w}, u, z)$ and $g_2(\tilde{w}) := \int dx\, x f_{XW}(x, \tilde{w})/\tau(\tilde{w})$, we can write

$$H(s_1) = \frac{x_1}{2\tau(w_1)} \int K_h(w_1 - w) g_1(w)\, dw + \frac{u_1 + \phi(z_1)}{2} \int K_h(w_1 - w) g_2(w)\, dw$$

As in the proof of Lemma B.1, we define v such that $w = w_1 + vh$ and use that $g_1 \in \mathfrak{G}_\tau^{2,2}$ and $g_2 \in \mathfrak{G}^{2,2}$ to write

$$H(s_1) = \frac{x_1}{2\tau(w_1)} g_1(w_1) + \frac{u_1 + \phi(z_1)}{2} g_2(w_1) + R(s_1)$$

with $|R(s_1)| \lesssim h^2 x_1 \psi_1(w_1)/\tau(w_1) + (u_1 + \phi(z_1)) h^2 \psi_2(w_2)$ for some functions ψ_1 and ψ_2 given in Definition 3.2. Using $\mathbb{E}(U|W) = 0$ we can also write

$$H(S) = \frac{1}{2}\frac{f_W(W)}{\tau(W)} X \mathbb{E}(\phi(Z)|W) + \frac{1}{2}\frac{f_W(W)}{\tau(W)}(U + \phi(Z))\mathbb{E}(X|W) + R(S)$$

[7] We write $A \lesssim B$ if there exists a positive constant c such that $A \leqslant cB$.

The leading term of $H(S)$ is $XT_Z\phi + (U + \phi(Z))T_X$ and leads to the result since $\mathbb{V}\mathrm{ar}\, R(S) = o(1)$ as h tends to zero. □

PROOF OF THEOREM A.1. Follows from the proof of Theorem A.2. □

PROOF OF THEOREM A.2. Denote $M := T_X^\star T_X$ and $v := T_X^\star r$ and consider the decomposition

$$\begin{aligned}\hat\beta - \beta &= \widehat{M}^{-1}\hat v - \widehat{M}^{-1}\widehat{M}\beta \\ &= M^{-1}(\hat v - \widehat{M}\beta) + \widehat{M}^{-1}(M - \widehat{M})M^{-1}(\hat v - \widehat{M}\beta)\end{aligned}$$

Using Lemma B.2, the first term of this decomposition leads to the result if we show that the second term is $o_p(n^{-1/2})$. Lemma B.1 with $h = n^{-1/2}$ implies the mean square convergence of $\|\hat M - M\|$. In particular, it holds $\|\hat M - M\| = O_p(n^{-1/2})$. Moreover Lemma B.2 implies that $\|\hat v - \hat M\beta\| = O_p(n^{-1/2})$. Thus the second term is $o_p(n^{-1/2})$, as $\|\widehat{M}^{-1}\|$ is bounded in probability. □

PROOF OF COROLLARY A.1. Conditioning on W, the matrix Λ becomes

$$\Lambda = \mathbb{E}\left[\mathbb{V}\mathrm{ar}\left\{\frac{f_W(W)}{\tau(W)}U\mathbb{E}(X|W)\Big|W\right\}\right] + \mathbb{V}\mathrm{ar}\left[\mathbb{E}\left\{\frac{f_W(W)}{\tau(W)}U\mathbb{E}(X|W)\Big|W\right\}\right]$$

where the second term cancels out using again $\mathbb{E}(U|W) = 0$. An expansion of the first term leads to

$$4\zeta = \mathbb{E}\left\{\left(\frac{f_W(W)}{\tau(W)}\right)^2 \mathbb{E}(X|W)\,\mathbb{V}\mathrm{ar}(U|W)\mathbb{E}(X|W)^t\right\}$$

which gives the announced result. □

LEMMA B.3. *(i) If $\int x_j^2 f(x,w)dx \in \mathfrak{G}_\tau^{1,1}(\mathbb{R}^q)$ and $\int x_j f(x,w)dx \in \mathfrak{G}_\tau^{s,2}(\mathbb{R}^q)$ for each component x_j of x, then*

$$\mathbb{E}\|\hat T_X - T_X\|^2_{L^2_\tau(\mathbb{R}^q)} = O\left((nh_W^q)^{-1} + h_W^{2\rho}\right), \tag{B.6}$$

$$\mathbb{E}\|\hat T_X^\star - T_X^\star\|^2_{\mathbb{R}^k} = O\left((nh_W^q)^{-1} + h_W^{2\rho}\right)\ ; \tag{B.7}$$

(ii) If $f_{ZW} \in \mathfrak{G}^{1,1}_{\pi\cdot\tau}(\mathbb{R}^{p+q})$ and $f_{ZW} \in \mathfrak{G}^{s,2}_{\pi\cdot\tau}(\mathbb{R}^{p+q})$, then

$$\mathbb{E}\|\hat T_Z - T_Z\|^2_{L^2_\tau(\mathbb{R}^q)} = O\left((nh_W^q h_Z^p)^{-1} + (h_Z \vee h_W)^{2\rho}\right), \tag{B.8}$$

$$\mathbb{E}\|\hat T_Z^\star - T_Z^\star\|^2_{L^2_\pi(\mathbb{R}^p)} = O\left((nh_W^q h_Z^p)^{-1} + (h_Z \vee h_W)^{2\rho}\right) \tag{B.9}$$

where $a \vee b = \max(a,b)$;

(iii) If $\int y^2 f(y,w)dx \in \mathfrak{G}_\tau^{1,1}(\mathbb{R}^q)$ and $\int y f(x,w)dx \in \mathfrak{G}_\tau^{s,2}(\mathbb{R}^q)$, then

$$\mathbb{E}\|\hat r - r\|^2_{L^2_\tau(\mathbb{R}^q)} = O\left((nh_W^q)^{-1/2} + h_W^{\rho}\right). \tag{B.10}$$

PROOF. We only give the details for the proof of (B.8). Denote $\hat f_{ZW} = n^{-1}\sum_i K_{h_W}(W_i - w)K_{h_Z}(Z_i - z)$. Using the Cauchy Schwarz inequality,

$$\mathbb{E}\|\hat T_Z - T_Z\|^2_{L^2_\tau(\mathbb{R}^q)} \leqslant \iint \left[\mathbb{V}\mathrm{ar}\{\hat f_{ZW}(z,w)\} + \{\mathbb{E}\hat f_{ZW}(z,w) - f_{ZW}(z,w)\}^2\right]\frac{dz}{\pi(z)}\frac{dw}{\tau(w)}.$$

Then using $f_{ZW} \in \mathfrak{G}^{1,1}_{\pi\cdot\tau}(\mathbb{R}^{p+q})$ the first term is of order $O((nh_W^q h_Z^p)^{-1})$ and with $f_{ZW} \in \mathfrak{G}^{s,2}_{\pi\cdot\tau}(\mathbb{R}^{p+q})$ the second term is bounded by $O((h_W \vee h_q)^{2\rho})$. The proof of the other results is very similar and we skip the details. □

Lemma B.4. *Under Assumption 3.1 and as α tends to zero with $n \to \infty$,*

$$\|T_X^\star(I - P_Z^\alpha)\| = O\left(\alpha^{(\eta\wedge 2)/2}\right) , \tag{B.11}$$

$$\|T_X^\star(I - P_Z^\alpha)P_Z T_X\| = O\left(\alpha^{\eta\wedge 1}\right) , \tag{B.12}$$

$$\|(I - P_Z^\alpha)\, T_Z\phi\| = O\left(\alpha^{1\wedge(\nu+1)/2}\right) . \tag{B.13}$$

Proof. The proof uses the properties $\left\|(T_Z T_Z^\star)^{-\eta/2}T_X\right\| < \infty$ and $\left\|(T_Z^\star T_Z)^{-\nu/2}\phi\right\| < \infty$ which are direct consequences of Assumption 3.1. To show (B.11), we use the decomposition

$$\|T_X^\star(I - P_Z^\alpha)\| \leqslant \left\|(I - P_Z^\alpha)(T_Z T_Z^\star)^{\eta/2}\right\| \cdot \left\|(T_Z T_Z^\star)^{-\eta/2}T_X\right\|$$

where the first factor is $O(\alpha^{(\eta\wedge 2)/2})$ by Theorem 4.3 of Engl, Hanke, and Neubauer (1996) and the second factor is finite. The proof of the other results is similar and we skip the details. □

Proof of Theorem 3.1. Define the operators $\widehat{P}_Z^\alpha := \widehat{T}_Z(\alpha I + \widehat{T}_Z^\star \widehat{T}_Z)^{-1}\widehat{T}_Z^\star$ and $P_Z^\alpha := T_Z(\alpha I + T_Z^\star T_Z)^{-1}T_Z^\star$. The proof is based on the decomposition

$$\hat\beta - \beta = \widehat{M}_\alpha^{-1}\left\{\widehat{T}_X^\star\left(I - \widehat{P}_Z^\alpha\right)\left(\hat r - \widehat{T}_X\beta - \widehat{T}_Z\phi\right) + \widehat{T}_X^\star\left(I - \widehat{P}_Z^\alpha\right)\widehat{T}_Z\phi\right\} . \tag{B.14}$$

Denote $M = T_X^\star(I - P_Z)T_X$. Below we show the three following asymptotic convergences:

$$\begin{aligned}\|\widehat{M}_\alpha^{-1} - M^{-1}\| = O_p\Big(&\left\{1 + \alpha^{\frac{\eta\wedge 2}{2}}\right\}\cdot\left\{(nh_W^q)^{-1/2} + h_W^\rho\right\} \\ &+ \alpha^{\frac{\eta\wedge 2}{2}-1}\cdot\left\{(nh_W^q nh_Z^p)^{-1/2} + (h_W \vee h_Z)^\rho\right\} + \alpha^{\eta\wedge 1}\Big) ,\end{aligned} \tag{B.15}$$

$$\begin{aligned}\|\widehat{T}_X^\star\left(I - \widehat{P}_Z^\alpha\right)\left(\hat r - \widehat{T}_X\beta - \widehat{T}_Z\phi\right)\| = O_p\Big(&\alpha^{\frac{\eta\wedge 2}{2}-1}\cdot\left((nh_W^q h_Z^p)^{-1/2} + (h_W \vee h_Z)^\rho\right)^2 \\ &+ \alpha^{\frac{\eta\wedge 2}{2}}\cdot\left((nh_W^q h_Z^p)^{-1/2} + (h_W \vee h_Z)^\rho\right)\Big)\end{aligned} \tag{B.16}$$

and

$$\|\hat T_X^\star\left(I - \widehat{P}_Z^\alpha\right)\widehat{T}_Z\phi\| = O_p\left(\alpha^{1/2}\cdot\left((nh_W^q h_Z^p)^{-1/2} + (h_W \vee h_Z)^\rho\right)^{\frac{\nu\wedge 2}{2}} + \alpha^{1\wedge\frac{1+\nu}{2}}\right) \tag{B.17}$$

under the assumptions of the theorem. The conditions of the theorem on α, h_W and h_Z ensure that (B.15) has the rate $o_p(1)$ while (B.16) and (B.17) have the rate $O_p(n^{-1/2})$.

Proof of (B.15). First note the inequality

$$\|\widehat{M}_\alpha^{-1} - M^{-1}\| \leqslant \|M^{-1}\| \cdot \|\widehat{M}_\alpha^{-1}\| \cdot \|\hat M_\alpha - M\| .$$

As $\|M^{-1}\|$ is bounded and $\|\widehat{M}_\alpha^{-1}\|$ is bounded in probability we focus on the control of $\|\widehat{M}_\alpha - M\|$:

$$\begin{aligned}\|\widehat{M}_\alpha - M\| \leqslant{}& \|\widehat{T}_X^\star - T_X^\star\|\cdot\left\|(I - \widehat{P}_Z^\alpha)\hat T_X\right\| + \left\|T_X^\star\{(I - \widehat{P}_Z^\alpha) - (I - P_Z^\alpha)\}\hat T_X\right\| \\ &+ \|T_X^\star(I - P_Z^\alpha)\|\cdot\left\|\widehat{T}_X - T_X\right\| + \|T_X^\star\{(I - P_Z^\alpha) - (I - P_Z)\}T_X\| .\end{aligned}$$

Since $(I - \hat P_Z^\alpha)\hat T_X$ is bounded in probability, the first term is controlled by a direct application of Lemma B.3. To bound the second term, we make use of the following relations:

$$(I - \widehat{P}_Z^\alpha) - (I - P_Z^\alpha) = \frac{1}{\alpha}(I - P_Z^\alpha)\left\{\widehat{T}_Z\widehat{T}_Z^\star - T_Z T_Z^\star\right\}\left(I - \widehat{P}_Z^\alpha\right) \tag{B.18}$$

which allows to bound the second term by

$$\frac{1}{\alpha}\,\|T_X^\star(I-P_Z^\alpha)\|\cdot\left\|\widehat{T}_Z\widehat{T}_Z^\star-T_ZT_Z^\star\right\|\cdot\left\|(I-\widehat{P}_Z^\alpha)\widehat{T}_X\right\|=O\left(\alpha^{(\eta\wedge 2)/2-1}\cdot\left((nh_W^qh_Z^p)^{-1/2}+(h_W\vee h_Z)^\rho\right)\right)$$

where the rate comes from Lemma B.3, equation (B.11) of Lemma B.4 above and the relation $\|\widehat{T}_Z\widehat{T}_Z^\star-T_ZT_Z^\star\|=O(\max\{\|\widehat{T}_Z-T_Z\|,\|\widehat{T}_Z^\star-T_Z^\star\|\})$. By similar arguments, the third term is of order $O\left(\alpha^{(\eta\wedge 2)/2}\cdot((nh_W^q)^{-1/2}+h_W^\rho)\right)$. To bound the fourth term we use the identity $(I-P_Z^\alpha)-(I-P_Z)=(I-P_Z^\alpha)\,P_Z$ and, using equation (B.12) of Lemma B.4, find the rate $O_p(\alpha^{\eta\wedge 1})$.

Proof of (B.16). Set $\hat{e}=\hat{r}-\widehat{T}_X\beta-\widehat{T}_Z\phi$. We have $\|\hat{e}\|\leqslant\|\hat{r}-r\|+\|\widehat{T}_X-T_X\|+\|\widehat{T}_Z-T_Z\|$ and hence Lemma B.3 implies that $\|\hat{e}\|$ is of order $O_p((nh_W^qh_Z^p)^{-1/2}+(h_W\vee h_Z)^\rho)$. Consider now the decomposition

$$\widehat{T}_X^\star\left(I-\widehat{P}_Z^\alpha\right)\hat{e}=\left\{\widehat{T}_X^\star\left(I-\widehat{P}_Z^\alpha\right)-T_X^\star\left(I-P_Z^\alpha\right)\right\}\hat{e}+T_X^\star\left(I-P_Z^\alpha\right)\hat{e}\,. \tag{B.19}$$

The norm of first term is bounded by

$$\begin{aligned}\left\|\widehat{T}_X^\star-T_X^\star\right\|\cdot\left\|I-\widehat{P}_Z^\alpha\right\|\cdot\|\hat{e}\|+\left\|T_X^\star\left((I-\widehat{P}_Z^\alpha)-(I-P_Z^\alpha)\right)\right\|\cdot\|\hat{e}\|\\ =O_p\left(\alpha^{\frac{\eta\wedge 2}{2}-1}\cdot\left((nh_W^qh_Z^p)^{-1/2}+(h_W\vee h_Z)^\rho\right)^2\right)\end{aligned}$$

where the rate is derived similarly to the rate of (B.15) and we use, that the first term is negligible wrt. to the second. Analogously the second term of (B.19) is of order $O_p(\alpha^{(\eta\wedge 2)/2}\cdot((nh_W^qh_Z^p)^{-1/2}+(h_W\vee h_Z)^\rho))$.

Proof of (B.17). From Assumption 3.1, in particular (3.6), there exists $g\in L^2_\pi(\mathbb{R}^p)$ such that $\phi=(T_Z^\star T_Z)^{\nu/2}g$ for some $\nu>0$. Then we can write

$$\begin{aligned}\left\|\widehat{T}_X^\star\left(I-\widehat{P}_Z^\alpha\right)\hat{T}_Z\phi\right\|=\left\|\widehat{T}_X^\star\right\|\cdot\left\|\left(I-\widehat{P}_Z^\alpha\right)\hat{T}_Z\right\|\cdot\left\|(T_Z^\star T_Z)^{\nu/2}-(\hat{T}_Z^\star\hat{T}_Z)^{\nu/2}\right\|\cdot\|g\|\\ +\left\|\widehat{T}_X^\star\right\|\cdot\left\|\left(I-\widehat{P}_Z^\alpha\right)\hat{T}_Z(\hat{T}_Z^\star\hat{T}_Z)^{\nu/2}\right\|\cdot\|g\|.\end{aligned}$$

Theorem 4.3 in Engl, Hanke, and Neubauer (1996) leads to $\|(I-\hat{P}_Z^\alpha)\hat{T}_Z\|=O(\alpha^{1/2})$. Moreover, from Section 5.2 of this last reference we get $\|(T_Z^\star T_Z)^{\nu/2}-(\hat{T}_Z^\star\hat{T}_Z)^{\nu/2}\|\leqslant\|T_Z^\star T_Z-\hat{T}_Z^\star\hat{T}_Z\|^{(\nu\wedge 2)/2}$, thus the first term is of order $\alpha^{1/2}((nh_W^qh_Z^p)^{-1/2}+(h_W\vee h_Z)^\rho)^{(\nu\wedge 2)/2}$ from Lemma B.3. Similarly Theorem 4.3 in Engl, Hanke, and Neubauer (1996) gives the rate $\alpha^{1\wedge(1+\nu)/2}$ for the second term. □

Lemma B.5. *Denote $v^2(\cdot)=\mathbb{V}\mathrm{ar}(U^2|W=\cdot)$, $\hat{e}:=\hat{r}-\hat{T}_X\beta-\hat{T}_Z\phi$ and*

$$\hat{e}_U:=\frac{1}{n}\sum_i\frac{U_i}{\tau(\cdot)}K_{h_W}(W_i-\cdot)\,.$$

(i) If $v^2f_W\in\mathfrak{G}_\tau^{1,1}(\mathbb{R}^q)$, then $\mathbb{E}\|\hat{e}_U\|^2_{L^2_\tau(\mathbb{R}^q)}=O((nh_W^q)^{-1})$.

(ii) Let $\{\mu_j,g_j\in L^2_\tau(\mathbb{R}^q),e_j\in\mathbb{R}^k,j=1,\dots,k\}$ be the singular value decomposition of the compact operator $T_X^\star(I-P_Z)$ (see the decomposition (3.4) for instance). If $g_j\in\mathfrak{G}_\tau^{1,0}(\mathbb{R}^q)$ and $g_j\sqrt{v^2\cdot f_W/\tau}\in L^2_\tau(\mathbb{R}^q)$ for all $j=1,\dots,k$, then

$$\sqrt{n}\,(T_X^\star(I-P_Z)e_U)\xrightarrow{d}\mathcal{N}\left(0,T_X^\star(I-P_Z)\left[\frac{v^2\cdot f_W}{\tau}(I-P_Z)T_X\right]\right)\,. \tag{B.20}$$

Proof. We prove the two results separately.

Proof of (i). Using iterative conditional expectation and by definition of v^2 we can write

$$\mathbb{E}\|\hat{e}_U\|^2_{L^2_\tau(\mathbb{R}^q)} = \frac{1}{n}\int \mathbb{E}\Big(v^2(W)K^2_{h_W}(W-w)\Big)\frac{dw}{\tau(w)}.$$

With the standard change of variables, if we denote $g(u) := v^2(u)f_W(u)$,

$$\begin{aligned}\mathbb{E}\|\hat{e}_U\|^2_{L^2_\tau(\mathbb{R}^q)} &= \frac{1}{nh_W^q}\iint K^2(\tilde{w})g(w+h_W\tilde{w})\,d\tilde{w}\,\frac{dw}{\tau(w)}\\ &= \frac{1}{nh_W^q}\Big\{\int g(w)\int K^2(\tilde{w})\,d\tilde{w}\,\frac{dw}{\tau(w)} + R\Big\}\end{aligned}$$

where R is such that $|R| \leq \iint K^2(\tilde{w})|g(w+h_W\tilde{w})-g(w)|\,d\tilde{w}\,\frac{dw}{\tau(w)}$. Using that g belongs to $\mathfrak{G}^{1,1}_\tau(\mathbb{R}^q)$ the first term and $|R|$ are bounded, which proves (i).

Proof of (ii). Using the singular value decomposition of $T^\star_X(I-P_Z)$ we can write

$$\begin{aligned}T^\star_X(I-P_Z)\hat{e}_U &= \frac{1}{n}\sum_i U_i\sum_{j=1}^k \mu_j e_j\int g_j(w)K_h(W_i-w)\,dw\\ &= \frac{1}{n}\sum_i U_i\sum_{j=1}^k \mu_j e_j g_j(W_i) + R\,, \end{aligned}\tag{B.21}$$

where the reminder $R = \frac{1}{n}\sum_i U_i\sum_{j=1}^k \mu_j e_j\int dw\,\{g_j(w)-g_j(W_i)\}K_h(W_i-w)$ has expectation zero and variance

$$\frac{1}{n}\sum_{i,j=1}^k \mu_i\mu_j e_i e_j^t\mathbb{E}\left[\left\{\int dw\;\left(g_j(w)-g_j(W_1)\right)K_h(W_1-w)\right\}^2 v^2(U_1|W_1)\right].$$

Using $g_j \in \mathfrak{G}^{1,0}_\tau(\mathbb{R}^q)$, the reminder R has $\mathbb{V}\mathrm{ar}\,R = O(h_W^2 n^{-1}\,\mathbb{V}\mathrm{ar}(U_1)\cdot\sum_{i=1}^k\mu_i^2)$ and hence is negligible. We derive the asymptotic law by applying a standard central limit theorem on the first term in (B.21) where each summand has a vanishing expectation and a finite variance by assumption $g_j\sqrt{v^2\cdot f_W/\tau} \in L^2_\tau(\mathbb{R}^q)$. It remains to calculate the asymptotic covariance matrix. Using the singular value decomposition of $T^\star_X(I-P_Z)$ we obtain

$$\begin{aligned}\mathbb{C}\mathrm{ov}\left(U_1\sum_{j=1}^k\mu_j e_j g_j(W_1)\right) &= \sum_{i,j=1}^k \mu_i e_i\left\langle g_i, \frac{v^2\cdot f_W}{\tau}g_j\right\rangle_{L^2_\tau}\mu_j e_j^t\\ &= T^\star_X(I-P_Z)\left[\frac{v^2\cdot f_W}{\tau}(I-P_Z)T_X\right],\end{aligned}$$

which proves (ii). □

Proof of Theorem 3.2. Part of this proof is similar to the proof of Theorem 3.1. Here again, we consider the decomposition (B.14). The assumptions of Theorem 3.2 give the rate $o_p(1)$ for (B.15) and the rate $o_p(n^{-1/2})$ for (B.17). The treatment of (B.16) however requires a different decomposition which is considered now.

Denote $\hat{e} = \hat{r} - \widehat{T}_X\beta - \widehat{T}_Z\phi$. We consider the following decomposition of (B.17):

$$\begin{aligned}\widehat{T}^\star_X\left(I-\widehat{P}^\alpha_Z\right)\hat{e} &= \left\{\widehat{T}^\star_X\left(I-\widehat{P}^\alpha_Z\right) - T^\star_X\left(I-P^\alpha_Z\right)\right\}\hat{e} + T^\star_X\left(I-P^\alpha_Z\right)\{\hat{e}-\hat{e}_U\}\\ &\quad + T^\star_X\left\{(I-P^\alpha_Z)-(I-P_Z)\right\}\hat{e}_U + T^\star_X\left(I-P_Z\right)\hat{e}_U\,,\end{aligned}$$

where $\hat{e}_U$ is defined in Lemma B.5. The norm of the first term is controlled as in the proof of Theorem 3.1 and has the rate $o_p(n^{-1/2})$ under the assumptions of the theorem. Using Lemma B.4

and Lemma B.5 above the second term is of order $O_p(\alpha^{(\eta\wedge 2)/2} \cdot ((nh_W^q h_Z^p)^{-1/2} + (h_W \vee h_Z)^\rho))$. To control the third term we use $(I - P_Z^\alpha) - (I - P_Z) = (I - P_Z^\alpha)\, P_Z$ and thus by Lemmas B.4 and B.5, this term is negligible with respect to the second term. With our assumptions on α, h_Z and h_W, the first two terms together are $o_p(n^{-1/2})$. The last term of the decomposition leads to the central limit result by Lemma B.5. □

Proof of Theorem 4.1. In this proof we construct the function g_γ explicitly. If the system $\{\tilde\psi_i \in L^2_\tau(\mathbb{R}^q)\}_{i=1,\dots,k}$ are the eigenfunctions from the spectral decomposition of T_X, then the source condition with $\eta \geqslant 1$ (Assumption 3.1) implies $P_Z\tilde\psi_i \in \mathcal{R}(T_Z)$ for $i = 1, \dots, k$. In other words, there exists for each $i = 1, \dots, k$ a function $\tilde\phi_i \in L^2_\pi(\mathbb{R}^p)$ such that $P_Z\tilde\psi_i = T_Z\tilde\phi_i$. For each $\gamma \in \mathbb{R}^k$ we define $g_\gamma(Z) := \gamma_1\tilde\phi_1 + \cdots + \gamma_k\tilde\phi_k$. Note that g_γ is differentiable w.r.t. γ and is such that

$$T_{g_\gamma(Z)} v = \sum_{i=1}^{k} v_i T_Z \tilde\phi_i$$

for all $v \in \mathbb{R}^k$. The range of the operator $T_{g_\gamma(Z)}$, $\mathcal{R}(T_{g_\gamma(Z)})$, given by the k-dimensional linear subspace $\mathrm{lin}\{T_Z\tilde\phi_i, i = 1, \dots, k\}$ is by definition a subset of $\mathcal{R}(T_Z)$. Hence the projection $P_{g_\gamma(Z)}$ onto $\mathcal{R}(T_{g_\gamma(Z)})$ is the restriction of P_Z onto $\mathcal{R}(T_{g_\gamma(Z)})$ and since $P_Z\tilde\psi_i \in \mathcal{R}(T_{g_\gamma(Z)})$, we also have $P_Z\tilde\psi_i = P_{g_\gamma(Z)}\tilde\psi_i$. This implies $P_Z T_X = P_{g_\gamma(Z)} T_X$ or, equivalently, $M = M_{g_\gamma(Z)}$, and the result is proved. □

References

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

Andrews, D. W. (1991): "Asymptotic normality of series estimators for nonparametric and semiparametric regression models," *Econometrica*, 59, 307–345.

Anglin, P. M., and R. Gençay (1996): "Semiparametric estimation of a hedonic price function," *J. Appl. Econometrics*, 11, 633–648.

Bickel, P. J., C. A. Klaassen, Y. Ritov, and J. A. Wellner (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. The John Hopkins University Press, London.

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75, 1613–1669.

Carrasco, M. (2008): "A regularization approach to the many instruments problem," Working paper, Université de Montréal.

Carrasco, M., J.-P. Florens, and E. Renault (2007): "Linear inverse problems in structural econometrics: Estimation based on spectral decomposition and regularization," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B. Elsevier.

Chamberlain, G. (1987): "Asymptotic efficiency in estimation with condition moment restrictions," *J. Econometrics*, 34, 305–334.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B. Elsevier.

Chen, X., and Y. Hu (2006): "Identification and inference of nonlinear models using two samples wuith arbitrary measurement errors," Discussion paper 1590, Cowles Foundation, Yale University.

Chen, X., O. Linton, and I. Van Keilegom (2003): "Estimation of semiparametric models when the criterion function is not smooth," *Econometrica*, 71, 1591–1608.

Chen, X., and D. Pouzo (2008): "Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals," Discussion paper 1640, Cowles Foundation, Yale University.

Darolles, S., J.-P. Florens, and E. Renault (2002): "Nonparametric Instrumental Regression," Working Paper, Université de Toulouse I, GREMAQ.

Delgado, M. A., and T. Stengos (1994): "Semiparametric Specification Testing of Non-nested Econometric Models," *Rev. Econ. Stud.*, 61, 291–303.

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

Engle, R. F., C. Granger, J. Rice, and A. Weiss (1986): "Semiparametric Estimates of the Relation Between Weather and Electricity Sales," *J. Amer. Statist. Assoc.*, 81, 310–320.

Florens, J.-P. (2003): "Inverse Problems and Structural Econometrics: The Example of Instrumental Variables," in *Advances in Economics and Econometrics: Theory and Applications – Eight World Congress*, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, vol. 36 of *Econometric Society Monographs*. Cambridge University Press.

Florens, J.-P., M. Mouchart, and J.-M. Rolin (1990): *Elements of Bayesian Statistics*. Dekker, New York.

Hall, P., and J. L. Horowitz (2005): "Nonparametric methods for inference in the presence of instrumental variables," *Ann. Statist.*, 33, 2904–2929.

Härdle, W., H. Liang, and J. Gao (1990): *Partially Linear Models*. Springer, Heidelberg.

Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998): "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66.

Horowitz, J., and S. Lee (2007): "Nonparametric instrumental variables estimation of a quantile regression model," *Econometrica*, 75, 1191–1208.

Ichimura, H., and P. Todd (2007): "Implementing nonparametric and semiparametric estimators," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B. Elsevier.

Johannes, J. (2006): "Statistical analysis of non-identifiable inverse problems," Mimeo, Universität Heidelberg.

Kress, R. (1999): *Linear Integral Equations*, no. 82 in Applied Mathematical Sciencs. Springer, New York, second edn.

Lehmann, E., and H. Scheffe (1950): "Completeness Similar Regions and Unbiased Tests Part I," *Sankhyā*, 10, 305–340.

Linton, O. (1995): "Second order approximation in the partially linear regression model," *Econometrica*, 63, 1079–1112.

Ma, Y., and R. J. Carroll (2006): "Locally Efficient Estimators for Semiparametric Models With Measurement Error," *J. Amer. Statist. Assoc.*, 101, 1465–1474.

Newey, W. K. (1990a): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–837.

——— (1990b): "Semiparametric Efficiency Bounds," *J. Appl. Econometrics*, 5, 99–135.

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

Pagan, A., and A. Ullah (1999): *Nonparametric Econometrics*. Cambridge University Press.

Robinson, P. M. (1988): "Root-N-consistent semiparametric regression," *Econometrica*, 56, 931–954.

San Martín, E., M. Mouchart, and J.-M. Rolin (2005): "Ignorable common information, null sets and Basu's first Theorem," *Sankhyā*, 67, 674–698.

Schmalensee, R., and T. M. Stoker (1999): "Household gasoline demand in the united states," *Econometrica*, 67, 645–662.

Scott, D. W. (1992): *Multivariate Density Estimation*. Wiley, New York.

Serfling, R. J. (1980): *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.

Stein, C. (1956): "Efficient nonparametric testing and estimation," in *Proceedings of the third Berheley Symposium on Mathematical STatistics and Probability*.

Stock, J. H. (1989): "Nonparametric Policy Analysis," *J. Amer. Statist. Assoc.*, 84, 567–576.

Tikhonov, A., A. Goncharsky, V. Stepanov, and A. Yagola (1995): *Numerical Methods for the Solution of Ill-Posed Problems*, Mathematics and Its Applications. Springer, New York.