CORE DISCUSSION PAPER

2006/25

INSTRUMENTAL REGRESSION IN PARTIALLY LINEAR MODELS

Jean-Pierre FLORENS[1], Jan JOHANNES[2] and Sébastien VAN BELLEGEM[3]

December 2005

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 parametrization of ϕ generally leads to an inconsistent estimator of β, whereas consistent nonparametric estimators for β have a slow rate of convergence. An additional complication is that the solution of the equation necessitates the inversion of a compact operator which can 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* which we explicit and interpret in the paper. Finally we show that the estimator achieves the semiparametric efficiency bound, even if the model is heteroskedastic.

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

[1]Université de Toulouse I, GREMAQ and IDEI, Place Anatole France, F-31042 Toulouse Cedex 9, France, Email: florens@cict.fr.

[2]Universität Heidelberg, Institut für Angewandte Mathematik, Im Neuenheimer Feld, 294, D-69120 Heidelberg, Germany, johannes@statlab.uni-heidelberg.de.

[3]Université catholique de Louvain, Institut de statistique and CORE, Voie du Roman Pays, 20, B-1348 Louvain-la-Neuve, Belgium, vanbellegem@stat.ucl.ac.be.

We like to thank X. Chen, C. De Mol, J. Horowitz for helpful discussions. This work was supported by the IAP research network nr P5/24 of the Belgian Government (Belgian Science Policy). Financial support from GREMAQ and IDEI is gratefully acknowledged.

This text presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

1 Introduction

The instrumental variable regression is characterized by a relation

$$Y = \varphi(Z) + U \tag{1.1a}$$

together with a mean independence condition

$$\mathbb{E}(U|W) = 0 \tag{1.1b}$$

where W are the instruments, Y and Z are endogeneous variables and φ is a nonparametric function that defines the relationship of interest.

One difficult question is to give regularity conditions in order to determine the speed of convergence of the estimator of ϕ. This speed of convergence is typically related to the smoothness of ϕ and the dependence between W and Z. The latest is described by the sequence of singular values on the conditional expectation operator[1] and on the regularity of φ characterized by the possibility to represent $\varphi(Z)$ as a conditional expectation. These conditions are difficult to interpret in the underlined economic model. Moreover, from an i.i.d. sample of n data, the speeds of convergence are typically slower than $\sqrt{n}$ in particular cases[2]. The usual curse of dimensionality also applies in that model, in particular through the dependence scheme between Z and W: The bigger is the dimension of Z, the weaker is this dependence.

In standard regression models, a classical method to reduce the dimensionality of the problem is to impose more restrictions on the object of interest. One possibility is to assume that φ has an additive structure, that is we can find two sets of endogeneous variables (Z, X) such that $\varphi(Z, X) = \phi(Z) + \psi(X)$ for some functions ϕ and ψ. More specifically, this paper analyses the situation where ψ takes a linear structure. Then the model considered in this paper is

$$Y = \phi(Z) + X^t\beta + U \tag{1.2a}$$

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

$$\mathbb{E}\left(U|W\right) = 0 \tag{1.2b}$$

for some instrumental variable $W \in \mathbb{R}^q$. In the following, we consider the nonparametric estimation of ϕ and the estimation of the parameter β from an i.i.d. sample of the vector (Y, Z, X, W).

An elementary application of the model (1.2a) arises when $\phi = 0$. This case includes the parametric analysis of instrumental variable regression which has been extensively studied

[1] i.e. the λ_j such that there exists φ_j satisfying $\mathbb{E}\{\mathbb{E}(\varphi_j(Z)|W)|Z\} = \lambda_j^2\varphi_j$.

[2] This is for instance the case when (Z, W) is jointly normally distributed and the function φ is poorly approximated by polynomials.

under the assumption $\mathbb{C}\text{ov}(U, W) = 0$ (see standard textbooks in econometrics, e.g. Chapter 3 of Hayashi (2000)). The weaker condition (1.2b) has been less considered and treats the joint distribution of (Z, W) nonparametrically.

The model (1.2a–1.2b) can be viewed as an extension of the so-called partially linear model defined with $\mathbb{E}(U|X, Z) = 0$. This situation has been the subject of considerable study (see, e.g. Härdle, Liang, and Gao (1990)). In that context, $\sqrt{n}$-consistent estimator of the parameter β is derived in Robinson (1988) under the strong conditions that U is independent from X and Z, and ϕ belongs to some smoothness class.

Model (1.2a) with partial endogeneity has also been studied in some papers. For instance, Chen, Linton, and Van Keilegom (2003) consider the situation where the variable X only is endogeneous, while Ai and Chen (2003) study the case where the variable Z is endogeneous.

In the following, we consider the case where both X and Z are endogeneous. There is however another deeper contrast between our approach and the aforementioned papers. This difference lies in the underlying assumption specified for the space of parameters. This question is connected to the notion of *well-* or *ill-posedness* that we discuss now.

1.1 Well-posed versus Ill-posed problem

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

$$\mathbb{E}\left(Y|W\right) = \mathbb{E}\left(\phi(Z)|W\right) + \mathbb{E}\left(X^t\beta|W\right) \tag{1.3}$$

which links $\beta \in \mathbb{R}^k$ with functions assumed to be elements of Hilbert spaces. In this paper, all Hilbert spaces are L^2 spaces with respect to some specific measure. If this measure is the joint probability density f of the data generating process, then we write $L^2_f(Y)$ or $L^2_f(Z)$ to denote for example functions depending on Y or Z only.

Equation (1.3) is an integral equation which can be rewritten as

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

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 require 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 for (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. 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-posed problems is to assume that ϕ lies in a compact set of functions, see e.g. Tikhonov, Goncharsky, Stepanov, and Yagola (1995). This assumption automatically eliminates ill-posedness of the problem and leads to a well-posed problem[3]. This type of assumption is used, e.g., in Newey and Powell (2003), Ai and Chen (2003) or Chen (2006) to circumvent the inherent instability of the problem. The compactness assumption is however an extremely strong assumption which, in addition, is difficult to test.

It is however possible to deal with the ill-posedness, and a large literature on techniques exists to stabilize the inversion of the integral equations such as equation (1.3). In econometric contexts, we refer to Carrasco, Florens, and Renault (2006) for an overview of different methods. The treatment of the fully nonparametric model (1.1a–1.1b) with this approach can be found in Darolles, Florens, and Renault (2002), Florens (2003) and Hall and Horowitz (2005). See also Blundell and Horowitz (2004) for an application to a test of exogeneity.

In this paper, we propose estimators of ϕ and β in the partially linear model (1.2a–1.2b) in the framework of ill-posed inverse problems. It is first helpful to give a definition of the functional operators involved in (1.3).

1.2 Operators

Note that equation (1.3) 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

$$T_X : \mathbb{R}^k \to L^2_\tau(\mathbb{R}^q) : \tilde{\beta} \mapsto \mathbb{E}\{X'\tilde{\beta}|W = \cdot\}\frac{f_W(\cdot)}{\tau(\cdot)} \tag{1.4}$$

$$T_Z : L^2_\pi(\mathbb{R}^p) \to L^2_\tau(\mathbb{R}^q) : \tilde{\phi} \mapsto \mathbb{E}\{\tilde{\phi}(Z)|W = \cdot\}\frac{f_W(\cdot)}{\tau(\cdot)} \tag{1.5}$$

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

$$r = T_Z\phi + T_X\beta. \tag{1.6}$$

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:

$$T^\star_X : L^2_\tau(\mathbb{R}^q) \to \mathbb{R}^k : g \mapsto \mathbb{E}\{Xg(W)\} \tag{1.7}$$

$$T^\star_Z : L^2_\tau(\mathbb{R}^q) \to L^2_\pi(\mathbb{R}^p) : g \mapsto \mathbb{E}\{g(W)|Z = \cdot\}\frac{f_Z(\cdot)}{\pi(\cdot)} \tag{1.8}$$

[3] i.e. the estimators of ϕ or β depend continuously on the estimators of the conditional densities in the integral equation.

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)[4]. 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 of τ will have no influence on the rate of convergence of the proposed estimators, but is related to their asymptotic efficiency, as showed in Section 3.

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 1.1. *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 1.1. Assumption 1.1 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 1.2. Assumption 1.1 holds true if both $\mathbb{C}\mathrm{ov}(X)$ and $\mathbb{V}\mathrm{ar}(Y)$ are finite and the following Hilbert-Schmidt conditions are fullfiled:

$$\text{(i)} \int dy \int dw \left(\frac{f_{YW}(y,w)}{f_Y(y)\tau(w)}\right)^2 f_Y(y)\tau(w) < \infty \,,$$

$$\text{(ii)} \int dx \int dw \left(\frac{f_{XW}(x,w)}{f_X(x)\tau(w)}\right)^2 f_X(x)\tau(w) < \infty \,,$$

$$\text{(iii)} \int dz \int dw \left(\frac{f_{ZW}(z,w)}{\pi(z)\tau(w)}\right)^2 \pi(z)\tau(w) < \infty \,.$$

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

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

1.3 Objectives of the paper

Below we show that the estimation of ϕ in the partially linear model (1.2a–1.2b) is very similar to the estimation of φ in the fully nonparametric model (1.1a–1.1b) using estimators and results of Darolles, Florens, and Renault (2002) or Carrasco, Florens, and Renault (2006). Estimation of β however leads to a set of new important issues among which is the question whether the parametric speed of convergence can be recovered.

The first question we address in Section 2 is the identification of the parameters in the partially linear model. Then we address the question of finding a consistent estimator of β and the paper shows that it is possible to construct an estimator of the parameters of the linear part that exhibits $\sqrt{n}$ consistency. Efficiency of the estimator is discussed next, including the situation where the error term is conditionaly heteroskedastic. All proofs are written in an appendix.

It is worth mentionning that the results are derived under mild and realistic assumptions. One of these assumptions is given by the so-called *source condition* which measure how ill-posed is the problem at hand. We give this condition explicitly in Section 2 and propose some econometric interpretations on simple models.

2 Existence and Identification

In this section we give conditions for the existence of solutions and for the identification of the parameters from the partial linear model model (1.2a–1.2b). 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

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

where $\mathcal{R}(T)$ denotes the range of the operator T. However, this condition is obviously not always satisfied, therefore we define the parameters of interest (ϕ, β) as

$$(\phi, \beta) = \arg\min \big\{\|r - T_Z\tilde{\phi} - T_X\tilde{\beta}\|_{L^2(W)} \text{ such that } \tilde{\phi} \in L^2_\pi(\mathbb{R}^p) \text{ and } \tilde{\beta} \in \mathbb{R}^k\big\}. \quad (2.1)$$

This solution is called *minimal norm solution*, and is not necessary unique[5].

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

[5]The general problem of non identifiable nonparametric inverse problems is considered in Johannes (2005), where an estimator of the space of solutions is derived.

Assumption 2.1. *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) $\mathcal{R}(T_X) \cap \mathcal{R}(T_Z) = \{0\}$.

Theorem 2.1. *Suppose the model is well-defined (Assumption 1.1). Then Assumption 2.1 is necessary and sufficient for the identification of the function ϕ and the vector β in the model (1.2a)–(1.2b).*

Assumption 2.1 gives conditions on operators, but can be interpreted as conditions on random variables. The two following assumptions are together equivalent to Assumption 2.1(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

$$\Omega := \mathbb{E}\left\{\mathbb{E}(X|W)\,\frac{f_W(W)}{\tau(W)}\,\mathbb{E}(X^t|W)\right\} \tag{2.2}$$

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 to require the strong identification of X, Z by W (as used, e.g., in Darolles, Florens, and Renault (2002) or Hall and Horowitz (2005)). Note also that the matrix Ω of condition (b) is the asymptotic variance of the Generalized Method of Moment estimator for the heteroskesdastic model with $\mathbb{V}\text{ar}(U|W)f_W = \tau$ (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[6]. 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 (2006).

3 Estimation

Suppose we observe iid vectors (Y_i, Z_i, X_i, W_i), $i = 1, \dots, n$ from the model (1.2a)–(1.2b) and suppose that the parameters of the model are identified. Recall the definition of the operators in Section 1.2. The minimal norm solution (2.1) is such that

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

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

[6] 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.

Note that, analogously to the case of the linear regression model, this system projects the problem (1.6) onto the parameter spaces $\mathbb{R}^k$ and $L^2_\pi(\mathbb{R}^p)$ using the adjoint operators.

To define our estimators, we shall consider two situations, depending on the behavior of the cross terms $T_Z^\star T_X$ and $T_X^\star T_Z$ in the linear system (3.1a)–(3.1b). First, we consider the situation where $T_Z^\star T_X = T_X^\star T_Z = 0$. An equivalent condition is that the range of T_X is orthogonal to the range of T_Z, i.e. $\mathcal{R}(T_X) \perp \mathcal{R}(T_Z)$. In that situation, we can separate the estimation of β from the estimation of ϕ. The general situation where $\mathcal{R}(T_X)$ is not orthogonal to $\mathcal{R}(T_Z)$ is discussed next.

3.1 Estimation when $\mathcal{R}(T_X) \perp \mathcal{R}(T_Z)$ or $\phi = 0$

We first consider the situation where $\mathcal{R}(T_X) \perp \mathcal{R}(T_Z)$, and discuss the case $\phi = 0$ at the end of this section. The orthogonality condition $\mathcal{R}(T_X) \perp \mathcal{R}(T_Z)$ 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

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

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

The estimation of ϕ is an ill-posed problem because the inversion of $T_Z^\star T_Z$ is not stable. This situation has been extensively studied and we refer to Darolles, Florens, and Renault (2002), Florens (2003) or Hall and Horowitz (2005) for the estimation of this quantity via regularization methods.

However, the estimation of β is not standard given our assumption $\mathbb{E}(U|W) = 0$ (see (1.2b)). We first introduce 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

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

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

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

Finally, our estimator of β is

$$\hat{\beta} = \hat{M}^{-1} \hat{v} \tag{3.3}$$

The definition of the multivariate K involved in the estimators is given below (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 r, i.e. $K(w) = \Pi_{i=1}^q \ell(w_i)$ where ℓ is a univariate, continuous, bounded, positive function such that*

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

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

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

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*

$$\sup_{y:|y-x|<\rho} \frac{|g(y) - g(x) - Q(y-x)|}{|y-x|^s} \leqslant \psi(x), \tag{3.4}$$

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 wrt $1/\gamma$ when $\alpha > 0$, i.e. $\int dx\ g^\alpha(x)/\gamma(x) < \infty$ and $\int dx\ \psi^\alpha(x)/\gamma(x) < \infty$. We also write $\mathfrak{G}^{s,\alpha}(\mathbb{R}^l)$ when $\gamma \equiv 1$.

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 3.1. *Suppose $T_Z^\star T_X = T_X^\star T_Z = 0$ in the system of equations (3.1a–3.1b). If the function $g_1 = \mathbb{E}(U + \phi(Z)|W) f_W(W)$ belongs to $\mathfrak{G}_\tau^{2,2}(\mathbb{R}^q)$ and each component of the function $g_2 = T_X$ belongs to $\mathfrak{G}^{2,2}(\mathbb{R}^q)$, then the estimator (3.3) constructed with kernels of order 2 and with a bandwidth $h = O(n^{-1/2})$ is such that $\sqrt{n}\|\hat{\beta} - \beta\| = O_p(1)$.*

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 3.2. *Under the assumptions of Theorem 3.1,*

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

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

The asymptotic variance of the theorem is not optimal, in the sense that it does not achieves the semiparametric efficiency bound. It is the consequence of the nuisance term $\phi(Z)$ which cannot be avoided even in the orthogonal situation $\mathcal{R}(T_X) \perp \mathcal{R}(T_Z)$. This phenomenon actually appears in even simpler situations, for instance when ϕ takes a parametric form. This setting is considered in the following remark.

Remark 3.1. When ϕ has a parametric form, the model becomes $Y = Z^t\gamma + X^t\beta + U$ for some random variables $Y \in \mathbb{R}$, $Z \in \mathbb{R}^p$, $X \in \mathbb{R}^k$, and $U \in \mathbb{R}$ such that $\mathbb{E}\,(WU) = 0$ for some instrumental variable $W \in \mathbb{R}^q$. The parameters γ and β are the solutions of the moment equation $\mathbb{E}(WY) = \mathbb{E}(WZ^t)\gamma + \mathbb{E}(WX^t)\beta$. Using a positive definite $q \times q$ weight matrix V, the parameters are equivalently characterized as minimizer of the quadratic form $\mathbb{E}[W(Y - Z^t\gamma - X^t\beta)]^t V^{-1}\mathbb{E}[W(Y - Z^t\gamma - X^t\beta)]$. If we observe an i.i.d. sample of the random vector (Y, Z, X, W), we can replace the expectations in the quadratic form by their empirical counterparts. It is well known that the moment estimator of γ, β obtained as the minimizers of the empirical counterpart of the quadratic form is consistent and asymptotically normal, but only efficient if we use the specific weigth matrix $V = \mathbb{C}\mathrm{ov}(WU)$. Note, that the weight matrix V plays the same role as the density τ used in our definition of the above operators (see equation (1.4)). Moreover, the parameters γ, β minimize the quadratic form if and only if they solve the following system of linear equations:

$$\begin{aligned}
\mathbb{E}(ZW^t)V^{-1}\mathbb{E}(WY) &= \mathbb{E}(ZW^t)V^{-1}\mathbb{E}(WZ^t)\gamma + \underbrace{\mathbb{E}(ZW^t)V^{-1}\mathbb{E}(WX^t)\beta}_{(i)}, \\
\mathbb{E}(XW^t)V^{-1}\mathbb{E}(WY) &= \underbrace{\mathbb{E}(XW^t)V^{-1}\mathbb{E}(WZ^t)\gamma}_{(ii)} + \mathbb{E}(XW^t)V^{-1}\mathbb{E}(WX^t)\beta.
\end{aligned} \tag{3.5}$$

As in the general nonparametric case that we have treated above, in this system of equations we have projected the moment condition onto the parameter space and the projectors are the adjoint operators, here given by the transposed matrices. The orthogonality condition corresponds now to the case where (i) and (ii) in (3.5) vanish. This situation arises for instance when we can separate the instruments into two sets of variables, $W = (W_1, W_2)$, such that (a) W_1 and W_2 are uncorrelated, (b) W_1 and X are uncorrelated, and (c) W_2 and Z, are uncorrelated. In the orthogonal case, the estimator $\hat{\beta}$ of $\beta = M^{-1}v$ with $M = \mathbb{E}(XW^t)V^{-1}\mathbb{E}(WX^t)$ and $v = \mathbb{E}(XW^t)V^{-1}\mathbb{E}(WY)$ is then given by replacing the theoretical expectations by their empirical counterparts. The asymptotic variance of this estimator is again $M^{-1}\Lambda^\circ M^{-1}$, with

$$\Lambda^\circ := \mathbb{E}(XW^t)V^{-1}\,\mathbb{C}\mathrm{ov}\left(W(U + Z^t\gamma)\right)V^{-1}\mathbb{E}(WX^t).$$

Similarly to what happens in Theorem 3.2, the asymptotic variance of the estimator is affected by the parameter α, which prevents to reach the efficiency bound. Note that this phenomenon also appears in the classical parametric regression without instrumental variables. □

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 3.1. *Under the assumptions of Theorem 3.1, 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 (2.2). In this particular case, the estimator $\hat{\beta}$ is optimal because it is identical to the GMM estimator constructed with optimal instruments in the homoskedastic setting. Indeed, the moment conditions in the homoskedastic 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 homoskedastic model. More details can be found in Newey (1990a).

The next section gives an estimator in the general setting, that is for general function ϕ, and with no orthogonality condition. The proposed estimator achieves the semiparametric efficiency bound.

3.2 Estimation in the general case

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

$$T_Z^\star (I - P_X) r = T_Z^\star (I - P_X) T_Z \phi \tag{3.6a}$$

$$T_X^\star (I - P_Z) r = T_X^\star (I - P_Z) T_X \beta \tag{3.6b}$$

where $P_X = T_X (T_X^\star T_X)^{-1} T_X^\star$ is the orthogonal projection operator onto the closure of the range $\mathcal{R}(T_X)$ of T_X and, similarly, $P_Z = T_Z (T_Z^\star T_Z)^{-1} T_Z^\star$ is the projection onto the closure of the range $\mathcal{R}(T_Z)$.

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

From (3.6a), 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 (3.6b), 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

$$T_Z\phi = \sum_{j=1}^{\infty} \lambda_j \langle \phi, \varphi_j \rangle \psi_j \qquad \text{for all } \phi \in L^2_\pi(\mathbb{R}^p) \tag{3.7}$$

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[7]. A sufficient condition for compactness is to assume T_Z to be a Hilbert-Schmidt operator, see Example 1.2 above. 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*

$$\max_{i=1,\ldots,k} \sum_{j=1}^{\infty} \frac{\langle \tilde{\psi}_i, \psi_j \rangle^2}{\lambda_j^{2\eta}} < \infty \ , \tag{3.8}$$

and

$$\sum_{j=1}^{\infty} \frac{\langle \phi, \varphi_j \rangle^2}{\lambda_j^{2\nu}} < \infty \ . \tag{3.9}$$

[7] A sufficient condition for the compactness of T_Z is given by the Hilbert Schmidt condition, see Example 1.2.

Since this type of assumption is new in econometrics[8], we will discuss its relevance and some interpretations in the next paragraphs.

The condition (3.8) means that the operator T_X is "adapted" to the operator T_Z, and this adaptation is controled by the parameter η. An equivalent condition is to impose that each function $\tilde{\psi}_i \in \mathcal{R}((T_Z T_Z^\star)^{\eta/2}) \otimes \mathrm{Ker}(T_Z^\star)$, where $\mathrm{Ker}(T_Z^\star)$ stands for the null space of $T_Z^\star$. This last condition is often called the *source condition* in the numerical literature on ill-posed inverse problems (see e.g. Engl, Hanke, and Neubauer (2000)). For specific (λ_j, ψ_j), there is a characterization of the source condition in terms of the differentiability of the $\tilde{\psi}_i$'s, see Johannes and Vanhems (2005). The second condition (3.9) can also be viewed as a source condition for ϕ, i.e. $\phi \in \mathcal{R}((T_Z^\star T_Z)^{\nu/2})$.

If $\mathcal{R}(T_X)$ and $\mathcal{R}(T_Z)$ are orthogonal, then $\eta = \infty$ and this case has been discussed above. Then the parameter η may be interpretated as a degree of colinearity 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 $i = 1, \ldots, k$. In that case, by orthonormality of the system $\{\psi_j\}$, condition (3.8) is equivalent to

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

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

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.

[8] With one noticeable exception for condition (3.9) that already appears in Darolles, Florens, and Renault (2002).

Example 3.2 (Normal model). Suppose X, Z are univariate Normal and $W = (W_1, W_2)$ is bivariate standard Normal with $(X, Z, W_1, W_2) \sim \mathcal{N}(0, \Sigma)$ 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 $\mathcal{R}(T_X) \perp \mathcal{R}(T_Z)$ which has been treated above. 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 (3.8) 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. □

Example 3.3. The preceeding 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 (3.8) 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, and this property can be easily tested. □

We now consider nonparametric estimator of the operators and define our estimator of β. Let K be a multivariate kernel (Definition 3.1) and set

$$\widehat{T}_X\tilde{\beta} = \frac{1}{n}\sum_{i=1}^{n} X_i^t\tilde{\beta}\,\frac{K_{h_W}(W_i-\cdot)}{\tau(\cdot)} \qquad \text{for all } \tilde{\beta}\in\mathbb{R}^k\ , \tag{3.10}$$

$$\widehat{T}_X^\star\psi = \frac{1}{n}\sum_{i=1}^{n} X_i\int dw\; K_{h_W}(W_i-w)\psi(w) \qquad \text{for all } \psi\in L^2_\tau(\mathbb{R}^q)\ , \tag{3.11}$$

$$\widehat{T}_Z\tilde{\phi} = \frac{1}{n}\sum_{i=1}^{n} \frac{K_{h_W}(W_i-\cdot)}{\tau(\cdot)}\int dz\; K_{h_Z}(Z_i-z)\tilde{\phi}(z) \qquad \text{for all } \tilde{\phi}\in L^2_\pi(\mathbb{R}^p)\ , \tag{3.12}$$

$$\widehat{T}_Z^\star\psi = \frac{1}{n}\sum_{i=1}^{n} \frac{K_{h_Z}(Z_i-\cdot)}{\pi(\cdot)}\int dw\; K_{h_W}(W_i-w)\psi(w) \qquad \text{for all } \psi\in L^2_\tau(\mathbb{R}^q)\ , \tag{3.13}$$

$$\hat{r} = \frac{1}{n}\sum_{i=1}^{n} Y_i\frac{K_{h_W}(W_i-\cdot)}{\tau(\cdot)}\ , \tag{3.14}$$

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 A.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.

The paramater β is then estimated by

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

where $\hat{v}_\alpha$ is an estimator of the left hand side of (3.6b) 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 used the Tikhonov regularization method to stabilize the inversion of $T_Z^\star T_Z$. It is of course possible to consider here other scheme of regularization, such as the Landweber-Fridman iterative regularization for instance (see Carrasco, Florens, and Renault (2006)).

Theorem 3.3. *Consider the nonparametric estimators (3.10–3.14) constructed using a multivariate kernel of order r (Definition 3.1) and for $j=1,\dots,k$ 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 fulfiled 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 also achieved when no source 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

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

impose that α cannot converge arbitrarily fastly 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 (3.8) for instance).

Theorem 3.4. *Consider the nonparametric estimators (3.10–3.14) constructed using a multivariate kernel of order r (Definition 3.1). Suppose assumptions (i) - (iii) of Theorem 3.3 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* $\mathfrak{G}_\tau^{1,0}(\mathbb{R}^q)$ *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 fastly 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 3.4 shows explicitely the influence of the function τ on the asymptotic variance of the estimator. If we take for instance τ such that $\mathbb{V}\text{ar}(U|W)f_W(W) = \sigma^2\tau(W)$, 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.

3.3 Efficiency of $\hat{\beta}$

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 parametrized 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

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

see, e.g., Chamberlain (1987). If we assume $\mathbb{C}\text{ov}(\partial_\gamma g_\gamma(Z)) < \infty$, 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 $\sqrt{n}(\hat\beta_{GMM} - \beta)$ wich 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 choosen 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 3.5. *Consider the nonparametric estimators (3.10–3.14) constructed using a multivariate kernel of order r (Definition 3.1). Suppose assumptions (i) - (vii) of Theorem 3.4 are satisfied and the parameters α, h_Z and h_W are choosen according to Theorem 3.4. If the density τ satisfies $\mathbb{V}\text{ar}(U|W)f_W(W) = \sigma^2\tau(W)$, then the estimator $\hat\beta$ achieves the semiparametric efficiency bound, i.e., there exists a parametric model g_γ for ϕ such that $M_{g_\gamma(Z)} = \sigma^2[T^\star_X(I - P_Z)T_X]^{-1}$.*

4 Final comments

One important conclusion from the above results is the usefulness of defining appropriate function spaces in order to define optimal estimators. This definition depends on the two probability densities π and τ introduced in (1.4) and (1.5). By doing so, we include and extend in particular the setting of Darolles, Florens, and Renault (2002) or Hall and Horowitz (2005) and allow, for instance, that the random variables are Normally distributed. In particular we show that the choice of the density τ has no influence on the rate of convergence of the estimator, but is related to the asymptotic efficiency. With that respect, the density τ plays the same role as the weight matrix of GMM estimators. As in two steps GMM estimation, a two steps procedure would be a natural extension of our approach, in which an estimator of the optimal density τ would be used.

The paper also defines an appropriate assumption on ϕ given by the source condition that relates the behavior of ϕ and the conditional expectation operator T_X with the conditional expectation operator T_Z (see Assumption 3.1). When the operator T_X is sufficiently regular with respect to T_Z ($\eta \geqslant 2$), the source condition assumption on ϕ is not necessary to get a $\sqrt{n}$-consistent estimator of β. If T_X is less regular, then the source condition on ϕ is a sufficient assumption for the $\sqrt{n}$-consistent estimation of β. It is worth mentioning that no regularity on ϕ is assumed in terms of smoothness. The only regularity on ϕ required for the consistency is the source condition in some situations when T_X is not regular enough.

The source condition has a simple interpretation in some models, including the Normal model. It may be viewed as a measure of the dependence between the endogeneous variables (X, Z) and the instruments W. It assumes in particular the compactness of the operator T_Z. If this operator is not compact[9], then the source condition can be replaced by an assumption of the type $\tilde{\psi}_i \in \mathcal{R}((T_Z T_Z^\star)^{\eta/2}) \otimes \mathrm{Ker}(T_Z^\star)$, where $\tilde{\psi}_i$ are the eigenfunctions of the (always compact) operator T_X. If T_Z is not a compact operator but we have an estimator that converges with an appropriate rate[10], the regularization procedure derived in this paper would still lead to a $\sqrt{n}$-consistent estimator of β.

A Appendix: 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 (1.2a–1.2b) 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 2.1 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 2.1 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 β_ψ, thus we get a contradiction. □

[9] Non compactness of T_Z appears for instance when there is at least one common variable between the endogeneous variables Z and the instruments W.

[10] This appropriate rate is given in Lemma A.3 below.

Lemma A.1. *Under the assumptions of Theorem 3.1, and if $h \to 0$ as $n \to \infty$,*

$$\begin{aligned} \mathbb{E}\hat{v} &= T_X^\star r + O\left(h^2\right) \ , && \text{(A.1)}\\ \mathbb{E}\|\hat{v}\|^2 &= \|T_X^\star r\|^2 + O\left(h^2\right) + O\left(n^{-1}\right) \ , && \text{(A.2)}\\ \mathbb{E}\widehat{M} &= T_X^\star T_X + O\left(h^2\right) \ , && \text{(A.3)}\\ \mathbb{E}\|\widehat{M}\|^2 &= \|T_X^\star T_X\|^2 + O\left(h^2\right) + O\left(n^{-1}\right) \ . && \text{(A.4)}\end{aligned}$$

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 (A.1). Using iterative conditional expectations, we can write

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

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),

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

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[11]

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

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. □

Lemma A.2. *Under the assumptions of Theorem 3.1, 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 $\Lambda = \mathbb{V}\mathrm{ar}(X T_Z \phi + (U + \phi(Z)) T_X)$.

Proof. A straightforward expansion leads to

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

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

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

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

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 $\zeta = \mathbb{V}\mathrm{ar}_f\, \mathbb{E}_f \{H(S_1, S_2)|S_1\}$. 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

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

As in the proof of Lemma A.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

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

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

Proof of Theorem 3.1. Follows from the proof of Theorem 3.2. □

Proof of Theorem 3.2. 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 A.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 A.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 A.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 3.1. Conditionning 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 A.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*

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

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

(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*

$$\mathbb{E}\|\hat{T}_Z - T_Z\|^2_{L^2_\tau(\mathbb{R}^q)} = O\left((nh^q_W h^p_Z)^{-1} + (h_Z \vee h_W)^{2\rho}\right), \tag{A.8}$$

$$\mathbb{E}\|\hat{T}^\star_Z - T^\star_Z\|^2_{L^2_\pi(\mathbb{R}^p)} = O\left((nh^q_W h^p_Z)^{-1} + (h_Z \vee h_W)^{2\rho}\right) \tag{A.9}$$

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

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

$$\mathbb{E}\|\hat{r} - r\|^2_{L^2_\tau(\mathbb{R}^q)} = O\left((nh^q_W)^{-1/2} + h^\rho_W\right). \tag{A.10}$$

Proof. We only give the details for the proof of (A.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,

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

Then using $f_{ZW} \in \mathfrak{G}^{1,1}_{\pi\cdot\tau}(\mathbb{R}^{p+q})$ the first term is of order $O((nh^q_W h^p_Z)^{-1})$ 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 A.4. *Under Assumption 3.1 and as* α *tends to zero with* $n \to \infty$*,*

$$\|T^\star_X(I - P^\alpha_Z)\| = O\left(\alpha^{(\eta\wedge 2)/2}\right), \tag{A.11}$$

$$\|T^\star_X(I - P^\alpha_Z)P_Z T_X\| = O\left(\alpha^{\eta\wedge 1}\right), \tag{A.12}$$

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

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

$$\|T^\star_X(I - P^\alpha_Z)\| \leqslant \left\|(I - P^\alpha_Z)(T_Z T^\star_Z)^{\eta/2}\right\| \cdot \left\|(T_Z T^\star_Z)^{-\eta/2}T_X\right\|$$

where the first factor is $O(\alpha^{(\eta\wedge 2)/2})$ by Theorem 4.3 of Engl, Hanke, and Neubauer (2000) and the second factor is finite. The proof of the other results is similar and we skip the details. □

Proof of Theorem 3.3. Define the operators $\widehat{P}^\alpha_Z := \widehat{T}_Z(\alpha I + \widehat{T}^\star_Z\widehat{T}_Z)^{-1}\widehat{T}^\star_Z$ and $P^\alpha_Z := T_Z(\alpha I + T^\star_Z T_Z)^{-1}T^\star_Z$. The proof is based on the decomposition

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

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

$$\|\widehat{M}^{-1}_\alpha - M^{-1}\| = O_p\Big(\left\{1 + \alpha^{\frac{\eta\wedge 2}{2}}\right\}\cdot\left\{(nh^q_W)^{-1/2} + h^\rho_W\right\} + \alpha^{\frac{\eta\wedge 2}{2}-1}\cdot\left\{(nh^q_W nh^p_Z)^{-1/2} + (h_W \vee h_Z)^\rho\right\} + \alpha^{\eta\wedge 1}\Big), \tag{A.15}$$

$$\|\widehat{T}^\star_X\left(I - \widehat{P}^\alpha_Z\right)\left(\hat{r} - \widehat{T}_X\beta - \widehat{T}_Z\phi\right)\| = O_p\Big(\alpha^{\frac{\eta\wedge 2}{2}-1}\cdot\left((nh^q_W h^p_Z)^{-1/2} + (h_W \vee h_Z)^\rho\right)^2 + \alpha^{\frac{\eta\wedge 2}{2}}\cdot\left((nh^q_W h^p_Z)^{-1/2} + (h_W \vee h_Z)^\rho\right)\Big) \tag{A.16}$$

and

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

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

Proof of (A.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 controled by a direct application of Lemma A.3. To bound the second term, we make use of the following relations:

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

which allows to bound the second term by

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

where the rate comes from Lemma A.3, equation (A.11) of Lemma A.4 above and the relation $\|\widehat{T}_Z\widehat{T}_Z^\star - T_Z T_Z^\star\| = O(\max\{\|\widehat{T}_Z - T_Z\|, \|\widehat{T}_Z^\star - T_Z^\star\|\})$. 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 (A.12) of Lemma A.4, find the rate $O_p(\alpha^{\eta\wedge 1})$.

Proof of (A.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 A.3 implies that $\|\hat{e}\|$ is of order $O_p((nh_W^q h_Z^p)^{-1/2} + (h_W \vee h_Z)^\rho)$. Consider now the decomposition

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

The norm of first term is bounded by

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

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

Proof of (A.17). From Assumption 3.1, in particular (3.9), there exists $g \in L_\pi^2(\mathbb{R}^p)$ 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 (2000) 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 $\alpha^{1/2}((nh_W^q h_Z^p)^{-1/2} + (h_W \vee h_Z)^\rho)^{(\nu\wedge 2)/2}$ from Lemma A.3. Similarly Theorem 4.3 in Engl, Hanke, and Neubauer (2000) gives the rate $\alpha^{1\wedge(1+\nu)/2}$ for the second term. □

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

(ii) Let $\{\mu_j, g_j \in L^2_\tau(\mathbb{R}^q), e_j \in \mathbb{R}^k, j = 1,\ldots,k\}$ be the singular value decomposition of the compact operator $T_X^\star(I - P_Z)$ (see the decomposition (3.7) 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 $j = 1,\ldots,k$, then

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

Proof. We prove the two results separately.

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

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

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

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

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

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

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

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

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

Using $g_j \in \mathfrak{G}_\tau^{1,0}(\mathbb{R}^q)$, 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 (A.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

$$\mathbb{C}\text{ov}\left(U_1\sum_{j=1}^k\mu_je_jg_j(W_1)\right)=\sum_{i,j=1}^k\mu_ie_i\left\langle g_i,\frac{v^2\cdot f_W}{\tau}g_j\right\rangle_{L^2_\tau}\mu_je_j^t$$
$$=T^\star_X(I-P_Z)\left[\frac{v^2\cdot f_W}{\tau}(I-P_Z)T_X\right],$$

which proves (ii). □

Proof of Theorem 3.4. Part of this proof is similar to the proof of Theorem 3.3. Here again, we consider the decomposition (A.14). The assumptions of Theorem 3.4 give the rate $o_p(1)$ for (A.15) and the rate $o_p(n^{-1/2})$ for (A.17). The treatment of (A.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 (A.17):

$$\widehat T^\star_X\left(I-\widehat P^\alpha_Z\right)\hat e=\left\{\widehat T^\star_X\left(I-\widehat P^\alpha_Z\right)-T^\star_X\left(I-P^\alpha_Z\right)\right\}\hat e+T^\star_X\left(I-P^\alpha_Z\right)\left\{\hat e-\hat e_U\right\}$$
$$+T^\star_X\left\{(I-P^\alpha_Z)-(I-P_Z)\right\}\hat e_U+T^\star_X\left(I-P_Z\right)\hat e_U\;,$$

where $\hat e_U$ is defined in Lemma A.5. The norm of the first term is controled as in the proof of Theorem 3.3 and has the rate $o_p(n^{-1/2})$ under the assumptions of the theorem. Using Lemma A.4 and Lemma A.5 above the second term is of order $O_p(\alpha^{(\eta\wedge 2)/2}\cdot((nh^q_Wh^p_Z)^{-1/2}+(h_W\vee h_Z)^\rho))$. To control the third term we use $(I-P^\alpha_Z)-(I-P_Z)=(I-P^\alpha_Z)\,P_Z$ and thus by Lemmas A.4 and A.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 A.5. □

Proof of Theorem 3.5. In this proof we construct the function g_γ explicitely. If the system $\{\tilde\psi_i\in L^2_\tau(\mathbb{R}^q)\}_{i=1,\ldots,k}$ 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 $i=1,\ldots,k$. In other words, there exists for each $i=1,\ldots,k$ 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

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

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 $\text{lin}\{T_Z\tilde\phi_i,i=1,\ldots,k\}$ 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 $P_ZT_X=P_{g_\gamma(Z)}T_X$ 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.

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., and J. Horowitz (2004): “A Nonparametric Test of Exogeneity,” Working Paper CWP15/04, University College London, CEMMAP.

Carrasco, M., J.-P. Florens, and E. Renault (2006): “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. 6. North Holland, to appear.

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

Chen, X. (2006): “Large Sample Sieve Estimation of Semi-Nonparametric Models,” in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6. North Holland, to appear.

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

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

Engl, H. W., M. Hanke, and A. Neubauer (2000): *Regularization of Inverse Problems*. Kluwer Academic, Dordrecht.

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, to appear.

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

Hayashi, F. (2000): *Econometrics*. Princeton University Press, Princeton.

Johannes, J. (2005): “Statistical analysis of non-identifiable inverse problems,” Mimeo, Universität Heidelberg.

Johannes, J., and A. Vanhems (2005): “Regularity spaces in inverse problems,” Mimeo, Universität Heidelberg.

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

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 (2006): “Ignorable common information, null sets and Basu’s first Theorem,” *Sankhyā*, forthcoming.

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

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

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.