DEPARTMENT OF ECONOMICS

WORKING PAPER SERIES

2016-02

McMASTER UNIVERSITY

Department of Economics
Kenneth Taylor Hall 426
1280 Main Street West
Hamilton, Ontario, Canada
L8S 4M4

<http://www.mcmaster.ca/economics/>

SEMIPARAMETRIC VARYING COEFFICIENT MODELS WITH ENDOGENOUS COVARIATES

S. CENTORRINO AND J. S. RACINE

ABSTRACT. Though parametric methods are popular in applied settings, practitioners often require nonparametric alternatives. However, fully nonparametric methods are known to suffer from the curse-of-dimensionality, which limits their practical application. Semiparametric methods occupy a middle ground, have the desirable feature that they are both flexible, and provide an attractive alternative to fully nonparametric methods, while attenuating the curse-of-dimensionality. Traditional semiparametric methods, such as the popular ‘varying coefficient’ specification, do not account for endogenous covariates, which restricts their application. In this paper we consider the estimation of semiparametric varying coefficient models when the functional coefficients may contain (continuous) endogenous covariates thereby extending the reach of this flexible and powerful class of models.

1. INTRODUCTION

Instrumental regression techniques are necessary for valid estimation and inference when a model’s disturbances (denoted by U) are correlated with its covariates (denoted by Z). Letting Y denote the model’s outcome, the relationship of interest is given by $Y = \varphi(Z) + U$, where $\varphi(\cdot)$ is the function of interest. Standard regression techniques typically presume that the disturbances are uncorrelated with the predictors, i.e., that $E(U|Z) = 0$, hence, the study of $\varphi(Z)$ is equivalent to the study of the conditional expectation function $E(Y|Z)$, which leads to classical regression techniques such as ordinary least squares.

However, when $E(U|Z) \neq 0$, which can occur surprisingly often in applied settings, the relationship of interest no longer coincides with the conditional expectation function and, therefore, requires an alternative and, often, more complex treatment. While parametric instrumental regression techniques have been widely studied and are available to the empirical researcher (see e.g. Wooldridge 2010), sometimes practitioners instead find themselves in need of nonparametric alternatives. Recent advances in nonparametric estimation with endogenous regressors include Darolles,

Date: March 15, 2016.

Racine would like to gratefully acknowledge support from the Social Sciences and Humanities Research Council of Canada (SSHRC:www.sshrc.ca), and the Shared Hierarchical Academic Research Computing Network (SHARC-NET:www.sharcnet.ca).

Fan, Florens and Renault (2011) and Horowitz (2011), to name but a few. A common unifying element in this literature is the need to confront and solve the so-called ‘ill-posed inverse problem’ – see the recent authoritative survey by Horowitz (2014) and the references therein. While Darolles et al. (2011) use a locally-weighted kernel approach and Horowitz (2011) uses a sieve-based approach, both approaches, being fully nonparametric, suffer from the well-known ‘curse-of-dimensionality’ problem which limits the application of these methods to settings involving a small number of predictors. Practitioners, on the other hand, frequently face situations where they are dealing with a non-trivial number of predictors. As an alternative, semiparametric approaches are often adopted as they attenuate the curse-of-dimensionality while providing far more flexibility than their fully parametric counterparts.

In this paper, we consider one such semiparametric approach known as the ‘smooth coefficient’ model (see, among others Hastie and Tibshirani 1993, Fan and Zhang 1999, Li, Huang, Li and Fu 2002), and presume that $\varphi(Z, X) = X'\beta(Z)$, where Z is correlated with U but X is uncorrelated with U . This model is particularly appealing because other popular semiparametric and nonparametric approaches can be cast as special cases, including the popular partially linear model (Robinson 1988, Ai and Chen 2003, Florens, Johannes and Van Bellegem 2012), among others. We shall consider a sieve approach and will adopt Legendre polynomials for its implementation. See the related work of Chen (2007), Chen and Pouzo (2015) and Chen and Christensen (2015a), to name but a few. Our approach will also be seen to be practically appealing since unknown conditional moments will be approximated with sieves, which can themselves be computed via least squares and, as such, are much less computationally demanding than their locally weighted counterparts, which ought to make them particularly attractive to the practitioner.

A well-established literature exists that allows for the endogeneity of X in a varying or partially linear model using instrumental variable methods (Cai, Das, Xiong and Wu 2006, Cai and Xiong 2012). However, to the best of our knowledge, ours is the first paper that tackles potential endogeneity in Z in an instrumental variable varying coefficient setting.¹ We anticipate that our method would appeal to the practitioner, who encounters such a situation, as it allows the marginal effects for the exogenous regressors to be a smooth and heterogeneous function of the endogenous covariates.

¹There exists a recent paper of Delgado, Ozabaci, Sun and Kumbhakar (2015), in which the authors discuss the estimation of such a model using a control function approach.

By way of illustration, consider estimating returns to education (Mincer 1974, Card 1999). Denote as Y , the logarithm of wage, as Z the level of education, and, finally, as X the race of the individual. We could consider the following linear parametric specification, which includes an interaction term between X and Z , i.e.,

$$Y = \gamma_0 + Z\gamma_1 + X\gamma_2 + XZ\gamma_3 + U.$$

Much of the literature to date has focused on the possibility that returns to education are heterogeneous in race,

$$Y = \underbrace{(\gamma_0 + X\gamma_2)}_{\beta_0(X)} + Z \underbrace{(\gamma_1 + X\gamma_3)}_{\beta_1(X)} + U,$$

so that, for instance, we can assess whether the wage gap between individuals with high and low education differs among races (Li, Ouyang and Racine 2013). The potential endogeneity of Z is typically modelled using the standard parametric two stage least squares framework. That is, Z is replaced with the predicted values from a first step regression of Z on X and some instruments, W (Cai et al. 2006). These models might, however, be too rigid for some settings and practitioners might find themselves in need of a more flexible alternative.

In such cases, one might entertain a slightly different and more flexible semiparametric formulation. Although there is overwhelming evidence that wage gaps vary with race, it seems plausible to allow for the marginal effect to vary with regards to the level of education, i.e.,

$$Y = \underbrace{(\gamma_0 + Z\gamma_1)}_{\beta_0(Z)} + X \underbrace{(\gamma_2 + Z\gamma_3)}_{\beta_1(Z)} + U,$$

where one may expect the magnitude of the coefficient $\beta_1(Z)$ to differ across education levels, i.e., one might expect that the wage gap differs with respect to the level of education.

We allow for just these types of effects in our proposed approach, but naturally our approach is much more flexible than the simple linear structure used for this illustration, and we focus on a general, smooth, nonparametric specification for the functions $\beta_0(Z)$ and $\beta_1(Z)$ that delivers flexibility while attenuating the aforementioned curse.

The rest of the paper proceeds as follows. Section 2 describes the statistical model and some of the underlying assumptions. Sections 3 and 4 discuss the identification conditions required for our

semiparametric estimation framework based on sieve approximations. We conclude the paper with a Monte Carlo simulation that assesses the finite-sample performance of our approach in Section 5, along with an empirical application, which highlights the practical appeal of our proposed approach.

2. MODEL

In what follows, we adopt the flexible and popular smooth coefficient specification, i.e.,

$$Y = \beta_0(Z) + X_1\beta_1(Z) + \cdots + X_d\beta_d(Z) + U,$$

where Y is a continuous response variable, Z is a p -dimensional vector of endogenous covariates, and X is a $d+1$ -dimensional vector of exogenous predictors that enter the model linearly and that include a constant term. The components of the $d+1$ -dimensional coefficient vector, β , are unknown functions of the endogenous vector, Z . This framework allows us to consider, in a parsimonious manner, settings, where we allow for a broad range of interactions between the exogenous regressors and the endogenous covariates.

Economizing on notation, this model can be expressed by stacking the exogenous predictors and the functional coefficients into two vectors of dimension $d+1$, as follows:

$$X = [1, X_1, X_2, \dots, X_d]', \quad \beta(Z) = [\beta_0(Z), \beta_1(Z), \beta_2(Z), \dots, \beta_d(Z)]',$$

so that the model becomes:

$$(1) \quad Y = X'\beta(Z) + U.$$

Adopting the terminology that is used in this literature, we refer to the X 's as the 'linear' variables (or predictors), and to the Z 's as the 'smoothing' variables (or covariates) (Fan and Zhang 2008, Park, Mammen, Lee and Lee 2015).

Our primary interest lies in inference that exploits the nonparametric estimator of $\beta(z)$, which no longer coincides with the marginal effects associated with conditional mean of Y , since we consider cases in which $E(XU|z) \neq 0$. That is,

$$E(XY|z) = E(XX'|z)\beta(z) + E(XU|z) \neq E(XX'|z)\beta(z).$$

In order to be able to identify and estimate the vector of coefficients $\beta(Z)$, we assume that one observes a q -dimensional vector of exogenous instruments, W , with $q \geq p$, such that:

$$(2) \quad E(XU|w) = 0.$$

Thus, the model outlined in (1) nests the standard nonparametric instrumental variable regression, when X is equal to a vector of ones (see Newey and Powell 2003, Hall and Horowitz 2005, Darolles et al. 2011, Horowitz 2011, Chen and Pouzo 2012), and the partially linear model with endogenous covariates, when the interaction coefficients $\{\beta_1, \dots, \beta_d\}$ are constant functions (see Ai and Chen 2003, Florens et al. 2012).

The random vector (Y, X, Z, W) is characterized by its cumulative distribution function (CDF) F . We denote by $\mathbb{L}^2$ the space of square integrable function with respect to this CDF F . We assume the following.

Assumption 2.1.

- (i) *The marginal distribution of the smoothing variables, Z , and the instruments, W , is absolutely continuous w.r.t. the Lebesgue measure, and bounded away from 0 and ∞ on its support.*
- (ii) *The marginal distributions of the linear regressors X are either absolutely continuous w.r.t. the Lebesgue measure or they are discrete measures with finite support.*
- (iii) *The dependent variable Y belongs to $\mathbb{L}^2$. That is, $E(Y^2) < \infty$.*

Assumption 2.2. *The functional coefficients $\{\beta_0, \dots, \beta_d\}$ belong to $\mathbb{L}^2(Z)$, and they are Hölder continuous of some order $s > 0$.*

Assumption 2.3. *The conditional variance of U given the instrument W is uniformly bounded away from 0 and ∞ on the support of W .*

Assumption 2.1 (i) requires the joint distribution of the endogenous covariates and the instruments to be continuous, and is a standard condition invoked in the literature on nonparametric instrumental regressions. This assumption implies that the singular values of the conditional expectation operator converge to zero at some rate, and is needed for solving the ill-posed inverse

problem that arises in this setting. Assumption 2.1 (ii) loosens restrictions on the linear predictors thereby allowing them to be either continuous or discrete in nature. We do not place any restrictions on the number of continuous linear predictors, and furthermore, X could potentially be purely categorical if appropriate. Assumption 2.1 (iii) imposes a regularity condition on the dependent variable Y . In particular, Y is square integrable w.r.t. to the joint distribution F .

The last two assumptions, 2.2 and 2.3, impose smoothness conditions on the varying coefficients and require the variance of the error term to be bounded, both of which are standard assumptions in the literature (see, e.g., Newey 1997).

3. IDENTIFICATION

Without loss of generality, we take the support of the vector (Z, W) to be the $[0, 1]$ hypercube of dimension $p+q$ (this is for the proofs that follow and in no way restricts the support in applications to satisfy such bounds).

As mentioned above, when we take the linear predictors X to be equal to a vector of ones, our model becomes the standard nonparametric instrumental variable regression:

$$(3) \quad Y = \beta_0(Z) + U.$$

Given a valid vector of instruments W , for which $E(U|w) = 0$, we can recover β_0 as the solution to

$$E(Y|w) = E(\beta_0(Z)|w),$$

(see Newey and Powell 2003, Darolles et al. 2011, Horowitz 2011, Chen and Pouzo 2012, among others). The function $\beta_0(z)$ in (3) is then identified as long as the following assumption holds.

Assumption 3.1 (Completeness Condition). *For all functions $\phi \in \mathbb{L}^2(Z)$, we have that*

$$E(\phi(Z)|w) \stackrel{a.s.}{=} 0 \Rightarrow \phi(Z) \stackrel{a.s.}{=} 0.$$

This *completeness condition* is generically satisfied for pairs (Z, W) , which are continuously distributed, and when the dimension of the instrumental vector W is at least as large as the dimension of the endogenous covariate vector Z , which is an assumption that we maintain in our approach (Andrews 2011, D'Haultfoeuille 2011). Unfortunately, it is not possible to construct a

test for this condition that controls size uniformly over the class of complete distributions (Canay, Santos and Shaikh 2013). However, Freyberger (2015) provides the first positive result about the testability of this condition and demonstrates that local deviations from completeness still ensure the existence of a consistent estimator for $\beta_0(Z)$ in (3).

We now consider the identification of the varying coefficient model for a general vector X . First, notice that, using the parametric specification of the model with respect to X , we have:

$$E(XU|W) = 0,$$

which implies that $\beta(z)$ needs to satisfy the following moment restriction:

$$E[X(Y - X'\beta(Z))|W] = 0.$$

We, therefore, need to generalize the completeness condition in (3.1), in order to take into account the varying coefficient structure of our model. We first make the following assumption.

Assumption 3.2. *The second moment matrix of the vector X conditional on (Z, W) , denoted as $E(XX'|z, w)$, is a symmetric positive definite matrix for all $z, w \in [0, 1]^{p+q}$.*

This assumption is analogous to a standard condition found in the literature, namely, that the matrix $E(XX')$ is of full rank in the parametric linear regression model. As a matter of fact, when Z is a constant, the varying coefficient model resorts to a standard linear regression model, so in this instance Assumption 3.2 is a standard condition required for the identification of β in the least-squares framework (i.e., $E(XX')$ must be of full rank). Equivalently, when Z is treated as exogenous, $Z = W$, and Assumption 3.2 requires that the smallest eigenvalue of $E(XX'|z)$ be bounded away from zero (Xue and Yang 2006, Lee, Mammen and Park 2012).

Denote by 0_{d+1} a vector of zeros of dimension $d + 1$. We can state the following result.

Proposition 3.1. *Under the completeness condition in Assumption 3.1 and Assumption 3.2, we have that:*

$$E(XX'\beta(Z)|W) \stackrel{a.s.}{=} 0_{d+1} \Rightarrow \beta_l(Z) \stackrel{a.s.}{=} 0, \forall l = 0, \dots, d.$$

Proof. Notice that:

$$E(XX'\beta(Z)|W) = 0_{d+1} \Rightarrow (E(XX'\beta(Z)|W))' E(XX'\beta(Z)|W) = 0.$$

By Assumption 3.2, $E(XX'|z, w)$ is positive definite for all (z, w) . Therefore, there exists a strictly positive constant ρ , such that:

$$E(XX'|z, w) \succeq \rho I_{d+1},$$

where $\succeq$ denotes a partial ordering for positive semidefinite matrices, and I_{d+1} is the identity matrix of dimension $d + 1$. Hence,

$$(E(XX'\beta(Z)|W))' E(XX'\beta(Z)|W) \geq \rho^2 (E(\beta(Z)|W))' E(\beta(Z)|W),$$

which implies:

$$\rho^2 (E(\beta(Z)|W))' E(\beta(Z)|W) \leq 0.$$

This can hold, however, if and only if $E(\beta_l(Z)|W) = 0$, for every $l = 0, \dots, d$. The result of the proposition follows from the completeness condition in Assumption 3.1. $\square$

The result of this proposition shows that the varying coefficient model is a very straightforward extension of existing results and does not require any specific additional conditions on the smooth coefficients. In particular, it is interesting to compare our identification conditions to those of Ai and Chen (2003) in the special case of a partially linear model,

$$Y = \beta_0(Z) + X'_{-1}\gamma + U,$$

where X_{-1} is the vector of linear regressors without the constant component. Ai and Chen (2003) require the distribution of Z , given both X_{-1} and W , to be complete, $E(\beta_0(Z)|X_{-1}, W)$ not to belong to the linear span of the columns of X_{-1} , and $E(X_{-1}X'_{-1})$ to be full rank. Compared to their identification strategy we, therefore, make a more restrictive assumption on the second moment matrix of the linear regressors, but we ease the completeness condition for the nonparametric identification of the smooth coefficients.

In spite of the parametric and additive structure in X , the usual ill-posed inverse problem common to this literature arises. Define the following space of joint functions of X and Z ,

$$\mathcal{E}_{XZ} = \{\varphi : \mathbb{R}^{p+d} \rightarrow \mathbb{R}, \varphi \in \mathbb{L}^2(X, Z), \varphi = XX'\beta(Z)\}.$$

Our conditional expectation operator is

$$T : \mathcal{E}_{XZ} \rightarrow \mathbb{L}^2(W),$$

which is taken to map square integrable functions of the pair (X, Z) restricted to have a varying coefficient form, into square integrable functions of the instrument W . Further define $r = E(XY|W)$. Thus,

$$(4) \quad TXX'\beta = r$$

defines a system of Fredholm integral equations of the first kind, and therefore, β is a solution to an ill-posed inverse problem (Kress 1999, Carrasco, Florens and Renault 2007).

In order to solve the integral equation in (4), having defined the operator T we need to find the adjoint operator of T denoted as T^* . Generically, we could define T^* as a mapping of functions of W into joint functions of (X, Z) . That is,

$$T^* : \mathbb{L}^2(W) \rightarrow \mathbb{L}^2(X, Z).$$

However, this definition does not use the varying coefficient structure we have imposed on our model, in which X enters in a known, linear parametric form. Similar to the above, define

$$\mathcal{E}_{XW} = \{\omega : \mathbb{R}^{q+d} \rightarrow \mathbb{R}, \omega \in \mathbb{L}^2(X, W), \omega = XX'\pi(W)\},$$

the space of square integrable functions of the pair (X, W) , which assumes a varying coefficient structure in the instrument W .

Denote as $f_{ZW}(\cdot)$ the joint probability density function (PDF) of the pair (Z, W) . We denote the marginal and conditional PDFs in a similar way. Take a function $\omega \in \mathbb{L}^2(W)$ and write:

$$\begin{aligned}\langle Txx'\beta, \omega \rangle_{\mathbb{L}^2(W)} &= \int \int (\mathbb{E}[XX'|z, w] \beta(z))' \omega(w) f_{ZW}(z, w) dz dw \\ &= \int \beta(z)' \int \mathbb{E}[XX'|z, w] \omega(w) f_W(w|z) dw f_Z(z) dz = \langle \beta, T^*xx'\omega \rangle_{\mathbb{L}^2(Z)}.\end{aligned}$$

Therefore, the adjoint operator T^* can be defined in a manner similar to the definition of T , i.e., as

$$T^* : \mathcal{E}_{XW} \rightarrow \mathbb{L}^2(Z).$$

This definition is intuitive. As X enters the conditional expectation function in a known parametric form, we can effectively embed this information about the structure of the model into the definition of our operator. Therefore, the adjoint operator is simply projecting into the space of functions of Z , which are the object of interest in our framework.

Regularization methods can be used to remove the discontinuity in the identifying mapping in order to render estimation possible. For nonparametric instrumental regressions, we can use different regularization methods (e.g., Tikhonov (1943), Landweber (1951) – Fridman (1956), or Petrov (1940) – Galerkin (1915), also known in econometrics as sieve regularization). We adopt here a sieve regularization approach in the spirit of Newey and Powell (2003), Horowitz (2011), and Chen and Pouzo (2012), while alternative semi- and nonparametric approaches include the use of control functions (Newey, Powell and Vella 1999) and locally weighted kernel approaches (Darolles et al. 2011).

The main reason for adopting the sieve regularization approach (also known as Petrov-Galerkin regularization) is that its implementation is quite straightforward in this context. In particular, it is a fairly straightforward exercise to impose the additive structure in X that characterizes the varying coefficient model.

4. ESTIMATION

In what follows, we presume that one observes an i.i.d. realization of the random variables (Y, X, Z, W) , that we denote by (y_i, x_i, z_i, w_i) , for $i = 1, \dots, n$.

The main moment restriction that generates the estimating equation is given by:

$$(5) \quad E(XY|W) = E(XX'\beta(Z)|W).$$

This model can, therefore, be considered as a special case of the general framework presented in Chen and Pouzo (2012, 2015). Take the following families of basis functions $\pi_K(\cdot) = \{\pi_{1,K}(\cdot), \dots, \pi_{K,K}(\cdot)\}$, and $\psi_J(\cdot) = \{\psi_{1,J}(\cdot), \dots, \psi_{J,J}(\cdot)\}$ of dimension $K > 0$ and $J > 0$, respectively.

We presume that the linear span of $\psi_J(\cdot)$ is dense in $\mathbb{L}^2(Z)$, so that we can approximate the smooth coefficients as follows:

$$\beta_l^J(z) = \sum_{j=1}^J c_{jl} \psi_{j,J}(z), \quad \forall l = 0, \dots, d.$$

The $J(d+1)$ -vector of coefficients $c = \{c_{jl}, j = 1, \dots, J; l = 0, \dots, d\}$ satisfies the following system of moment restrictions:

$$E[(X \otimes \pi_K(W))(Y - (X \otimes \psi_J(Z))'c)] = 0,$$

where $\otimes$ denotes the Kronecker product, such that $u \otimes v = \text{vec}(uv')$, and the $\text{vec}(\cdot)$ operator stacks a matrix into a column vector. Finally, denote as:

$$\begin{aligned} \mathbf{\Pi} &= [x_1 \otimes \pi_K(w_1) \dots x_n \otimes \pi_K(w_n)]', \\ \mathbf{\Psi}_X &= [x_1 \otimes \psi_J(z_1) \dots x_n \otimes \psi_J(z_n)]'. \end{aligned}$$

Then, the estimator of a is given by:

$$\hat{c} = \left[\mathbf{\Psi}_X' \mathbf{\Pi} (\mathbf{\Pi}' \mathbf{\Pi})^{-} \mathbf{\Pi}' \mathbf{\Psi}_X \right]^{-} \mathbf{\Psi}_X' \mathbf{\Pi} (\mathbf{\Pi}' \mathbf{\Pi})^{-} \mathbf{\Pi}' y,$$

where $(\cdot)^{-}$ denotes the generalized inverse of a matrix, and

$$\hat{\beta}^J(z) = \mathbf{\Psi}_J^{\oplus}(z)' \hat{c},$$

where, abusing notation slightly,

$$\Psi_J^\oplus(z) = \begin{bmatrix} \psi_J(z) & & 0_{J \times d} \\ 0_{J \times 1} & \psi_J(z) & 0_{J \times (d-1)} \\ & 0_{J \times 2} & \ddots & 0_{J \times (d-2)} \\ & \vdots & & \ddots & \vdots \\ & & 0_{J \times d} & & \psi_J(z) \end{bmatrix}$$

is the direct sum of the vector $\psi_J(z)$ repeated $d+1$ times to form a $J(d+1) \times (d+1)$ matrix.

The parameters K and J act as smoothing and regularization parameters, respectively, with $K \geq J$, to ensure identification of the model (Blundell, Chen and Kristensen 2007, Newey 2013, Centorrino, Fève and Florens 2015, Chen and Christensen 2015a).

In particular, the regularization parameter J *compactifies* the parameter space and renders estimation possible in this context. Define as:

$$\mathcal{B}_{Js} = \left\{ \beta_l = \sum_{j=1}^J c_{jl} \psi_J, \|\beta_l\| = 1 \right\},$$

the space of Hölder continuous functions of order s that belong to the sieve space of dimension J . We define the following *sieve measure of ill-posedness* (Blundell et al. 2007):

$$\tau_{lJ} = \sup_{\beta_l \in \mathcal{B}_{Js}} [\|T\beta_l\|]^{-1},$$

such that, $\tau_{lJ} \rightarrow \infty$, as $J \rightarrow \infty$. When τ_{lJ} diverges at a polynomial rate, we say that the inverse problem is *mildly ill-posed*, while it is *severely ill-posed* if τ_{lJ} diverges at an exponential rate. We further denote as:

$$\tilde{\tau}_J = \tau_{0J} \vee \tau_{1J} \vee \cdots \vee \tau_{dJ},$$

the fastest diverging measure of ill-posedness for the smooth coefficients. Define:

$$G_\pi = E [\pi_K(w_i) \pi_K(w_i)'],$$

$$G_\psi = E [\psi_J(z_i) \psi_J(z_i)'],$$

$$S = E [\pi_K(w_i) \psi_J(z_i)'].$$

The following assumptions are standard in the sieve literature (Newey 1997, Huang 2003, Chen 2007, Belloni, Chernozhukov, Chetverikov and Kato 2015, Chen and Christensen 2015a, 2015b).

Assumption 4.1. *The vectors $\pi_K(\cdot) = \{\pi_{1,K}(\cdot), \dots, \pi_{K,K}(\cdot)\}$, and $\psi_J(\cdot) = \{\psi_{1,J}(\cdot), \dots, \psi_{J,J}(\cdot)\}$ form two families of basis functions in $\mathbb{L}^2$, such that:*

- (i) $\pi_K(\cdot)$ and $\psi_J(\cdot)$ are Hölder continuous of order equal to or greater than s ;
- (ii) $\pi_K(\cdot)$ and $\psi_J(\cdot)$ satisfy:

$$\sup_{\xi \in [0,1]} |G_\pi^{-1/2} \pi_K(\xi)| \leq \zeta_{\pi K}, \text{ and } \sup_{\xi \in [0,1]} |G_\psi^{-1/2} \psi_J(\xi)| \leq \zeta_{\psi J},$$

where $\zeta_{\pi K}$ and $\zeta_{\psi J}$ are two sequences of constants, with $\zeta = \zeta_{\pi K} \vee \zeta_{\psi J}$;

- (iii) The smallest eigenvalues of G_π and G_ψ are bounded away from zero, for all $J, K > 0$;
- (iv) There exists a vector of coefficients c_J , such that, for all $\phi \in \mathbb{L}^2(Z)$:

$$\|\phi(\xi) - \psi_J(\xi)c_J\|^2 = O_P(J^{-2s/p});$$

- (v) The dimensions K and J of the basis function satisfy:

$$\tilde{\tau}_J \zeta^2 / \sqrt{n} = O(1), \quad \tilde{\tau}_J \zeta \sqrt{\log J/n} = o(1), \text{ and } K = O(J);$$

- (vi) There exists a constant, such that, for all $\phi \in \mathbb{L}^2(Z)$:

$$\tilde{\tau}_J \|T(\phi(\xi) - \psi_J(\xi)c_J)\|^2 \leq \text{const} \|\phi(\xi) - \psi_J(\xi)c_J\|^2,$$

- (vii) Denote as Π_K the orthogonal projection onto the space generated by the basis functions π_K .

Then

$$\|(\Pi_K T - T)\phi\|^2 = o(\tilde{\tau}_J^{-2}).$$

Part (i) of Assumption 4.1 places some regularity conditions on the basis functions. Part (ii) gives an upper bound on the sup-norm of the basis function. Part (iii) is a further regularity condition on the second moment matrices of the basis functions, which are taken to be invertible for all $J, K > 0$. Part (iv) gives the order of approximation of the unknown function by sieves, which is increasing in the smoothness of the function, and decreasing in the dimension of the endogenous covariates. This is satisfied by most linear sieves, including polynomial, Fourier, spline, and Cohen-Deubeshies-Vial

wavelet series (Chen 2007). Part (v) restricts the rate of divergence of $\tilde{\tau}_j$ and of the dimension of the additive basis with respect to the sample size, and bounds the order of the instrumental basis, K , in such a way that the latter is diverging at most as fast as the truncation parameter J . Part (vi) includes the so-called $\mathbb{L}^2$ -stability condition, which is common in this literature (Blundell et al. 2007, Chen and Pouzo 2012). Finally, part (vii) constitutes an approximation property of the instrumental basis (Horowitz 2014, Chen and Christensen 2015a). The latter is equal to 0 whenever ψ_J and π_K form a Riesz, or an eigenfunction, basis for the conditional expectation operator.

We, thus, obtain the following result.

Theorem 4.1. *Under Assumption 4.1, we have that:*

$$\|\hat{\beta}^J - \beta\|^2 = O\left(\tilde{\tau}_J^2 \frac{J}{n} + J^{-2s/p}\right).$$

The $\mathbb{L}^2$ rate of convergence obtained here coincides with the rates obtained in Chen and Reiss (2011), Horowitz (2011), and Chen and Christensen (2015a). The only difference in our result is in the term $\tilde{\tau}_J^2$ that appears in the variance component. This term arises because we do not select a different regularization parameter J for each component function, as would be optimal in order to account for potential differences in smoothness.

Hence, strictly speaking, we do not achieve the minimax rate of convergence for the estimation of all functional coefficients. The result of this theorem can be improved by selecting as many values of J as there are functional coefficients to be estimated. The implementation and properties of such an optimal approach go beyond the scope of the present paper and are deferred for future research.

5. MONTE-CARLO SIMULATIONS

We generate samples of size $n = 1000$ from the following scheme. We first sample the exogenous covariates from a joint normal distribution:

$$\begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.2 & -0.3 \\ 0.2 & 1 & 0.4 \\ -0.3 & 0.4 & 1 \end{bmatrix}\right),$$

such that $W = \zeta_1$, $X_1 = \Phi(\zeta_2)$, and $X_2 = B(1, \Phi(\zeta_3))$, where Φ denotes the CDF of a standard normal distribution, and $B(1, p)$ is the Bernoulli distribution with probability of success equal to p . We use three separate specifications for the smooth coefficients,

$$\beta_0(Z) = 2 \exp(-Z),$$

$$\beta_1(Z) = 2Z^2,$$

$$\beta_2(Z) = \sin(\pi Z),$$

with the coefficients β_0 and β_2 being infinitely smooth, and β_1 of finite smoothness.

We then draw two random terms, $V \sim \mathcal{N}(0, 0.01)$, and $U \sim \mathcal{N}(0, (1/3)^2)$.

The endogenous variables are sampled according to two distinct schemes.

Scheme 1. We use only the continuous exogenous variable (W, X_1) to obtain

$$Z = [1 + \exp(2W + 2.5X_1 - 5U + V)]^{-1},$$

$$Y = \beta_0(Z) + X_1\beta_1(Z) + U.$$

Scheme 2. We use both the continuous and discrete exogenous components, to produce

$$Z = [1 + \exp(2W + 2.5X_1 - X_2 - 5U + V)]^{-1},$$

$$Y = \beta_0(Z) + X_1\beta_1(Z) + X_2\beta_2(Z) + U.$$

We generate 1000 sample paths for each scheme. We use Legendre polynomials of order J for the approximation of the unknown functional coefficients. The parameter J is chosen adaptively by applying the data-driven procedure suggested in Horowitz (2014). That is, we select the value of J , which minimizes the following criterion function,

$$\begin{aligned} \hat{J} = \arg \min_{1 \leq J \leq J_n} & \frac{2}{3n^2} \log(n) \sum_{i=1}^n \left\{ \left[y_i - \tilde{\beta}_0^{J_n}(z_i) - x_{1i}\tilde{\beta}_1^{J_n}(z_i) - x_{2i}\tilde{\beta}_2^{J_n}(z_i) \right]^2 \right. \\ & \left. \times \sum_{j=1}^J \left[\left(\hat{T}\hat{T}^* \right)^{-1} \hat{T}\psi_j(z) \right] (w_i)^2 \right\} - \|\hat{\beta}_0^J(z_i) + x_{1i}\hat{\beta}_1^J(z_i) + x_{2i}\hat{\beta}_2^J(z_i)\|^2, \end{aligned}$$

where,

$$\left(\hat{T}\hat{T}^*\right)^{-1}\hat{T} = \Pi \left[\Pi' \Psi_X (\Psi_X' \Psi_X)^{-} \Psi_X' \Pi \right]^{-} \Pi',$$

and J_n satisfies,

$$J_n = \arg \min_{J=1,2,\dots} \left\{ \tilde{\tau}_J^2 J^{3.5}/n \quad : \quad \tilde{\tau}_J^2 J^{3.5}/N - 1 \geq 0 \right\},$$

i.e., J_n is the smallest integer such that $\tilde{\tau}_J^2 J^{3.5}/N \geq 1$. An estimator of $\tilde{\tau}_J^2$ can be obtained by computing the smallest eigenvalues of the matrix

$$\Psi_X' \Pi (\Pi' \Pi)^{-} \Pi' \Psi_X.$$

Finally, $\{\hat{\beta}_l^{J_n}, l = 0, 1, 2\}$ are the estimators of the functional coefficients obtained using $J = J_n$. Other possible selection rules include the model averaging criterion of Liu and Tao (2014) and the sup-norm adaptive procedure suggested by Chen and Christensen (2015a).

The smoothing parameter K is chosen to be proportional to J . That is, $K = c_K J$, where $c_K \geq 1$. Selecting c_K greater than 1 has been shown to reduce the variance of the final estimator, so that we pick $c_K = 2$ (Blundell et al. 2007, Centorrino et al. 2015). We have not conducted sensitivity analysis regarding the particular choice of K , as this goes beyond the scope of the present paper. Interested readers are referred to the work of Blundell et al. (2007) and Chen and Christensen (2015a) for more extensive results on this particular issue.

Table 1 reports some summary statistics for the order of Legendre polynomials selected as explained above.

	Mean	St.Dev	Min	Max
Scheme 1	4.11	0.70	1.00	5.00
Scheme 2	4.18	0.51	2.00	5.00

TABLE 1. Order of Legendre Polynomials for each scheme

The results from the simulated data generating processes are reported in figures 1 and 2. The dashed light grey line represents the true function. The continuous black line denotes the mean of our sieve based nonparametric instrumental variable estimator of the functional coefficients. The dashed black lines give its 95% simulated confidence intervals. Finally, the dotted-dashed grey line depicts the mean of the estimator that overlooks the endogeneity of Z .

The Monte-Carlo simulations seem to suggest that the estimator behaves well in finite samples. The importance of accounting for endogeneity in this class of models can be appreciated by looking at the discrepancy between the grey line and the true regression function. We also briefly remark that, in line with our assumptions, the estimator does not suffer from the presence of discrete predictors.

FIGURE 1. Results from Simulation Scheme 1

6. EMPIRICAL APPLICATION: RETURNS TO EDUCATION

An extensive body of economic literature has examined the degree to which returns to education factor into the well-documented and growing disparity between the more and the less educated workers at the microeconomic level. Returns to education also serve as a widely used proxy for the effect of human capital on economic growth at the macroeconomic level. It is widely accepted that there exists a substantial degree of heterogeneity across individuals with respect to the marginal effect of an additional year of education on earnings. Plausible explanations include heterogeneity in the marginal return to schooling, heterogeneity in the cost associated with an additional year of education (Card 2001), and so forth.

FIGURE 2. Results from Simulation Scheme 2

Denote as Y , the natural logarithm of the hourly wage, as Z , the level of education achieved by an individual (in years), and as X , other variables that may include parental education and various worker-specific characteristics.

Perhaps the most general way to model heterogeneous responses is to consider the following non-separable model,

$$Y = \psi(X, Z, U),$$

where Z is allowed to interact nonlinearly both with X and with the unobservable component U , which may represent, for instance, unobservables such as individual ability (Heckman and Vytlacil 2005).

Although such an approach might place minimal structure on the underlying process, interpretation presents many challenges. One popular simplification is to instead presume that the error term U is separable (i.e., additive), which implies that heterogeneity in returns to education arises from the observable components only. Given that there may exist proxies for unobservables, this is not a particularly restrictive simplification. The additively separable specification is given by

$$Y = \psi_1(X, Z) + U,$$

where ψ_1 is an unknown function. One recent development involves modelling the function ψ_1 so that it is linear in the level of education, Z , but nonlinear with respect to the X variables, so that,

$$\psi_1(X, Z) = Z'\beta(X),$$

(see e.g. Su, Murtazashvili and Ullah 2013, Su, Chen and Ullah 2014). This modeling strategy is inspired by structural models of educational choice, in which the individual's utility and education production functions are both presumed to be linear in Z . For this strategy, the functional coefficients are such that the heterogeneous response of each individual is a functional of observable characteristics.

This modelling strategy could be extended to include higher order functions of Z (Card 1999, Card 2001), which might be required because, in a model of optimal educational choice, the marginal returns and costs of schooling may not remain constant with respect to Z .

Our approach, on the other hand, explicitly allows for heterogeneity and nonlinearity with respect to Z by expressing the model as

$$Y = X'\beta(Z) + U.$$

Notice that this model still allows for heterogeneous returns to education, as the derivative of Y with respect to Z remains a function of X . However, the linearity assumption with respect to X permits separate consideration of the interaction between the level of education and each of the individual’s characteristics. For instance, as mentioned in the introduction, one could explore the dynamics of the wage gap across racial groups as a function of their educational choices.

It is widely appreciated that, in this context, the level of education, Z , is endogenous. The potential endogeneity of Z arises from an omitted variable bias, as some of the unobservable characteristics of the individual, such as ability, may be correlated with school choice and thereby bias the regression coefficient in a standard regression setting. Similarly, it has long been recognized that schooling decisions may be measured with error, and that this measurement error may also lead to endogeneity viz Z .

By way of illustration, we shall apply our approach to a dataset used in Card (1993). This data consists of a sample of 3010 individuals drawn from the National Longitudinal Survey of Young Men (NLSYM). The dependent variable is the logarithm of the hourly wage earned in 1976, Z is the level of education, measured in years, in 1976, and X includes labor market experience, a black/non-black dummy, a dummy for 4 year college proximity, a regional dummy, and a dummy for urban areas.

Card (1993) considers several specifications, where additional control variables are included/omitted, with college proximity instrumenting for education. Card advocates using this dummy variable as an instrument for educational choice, arguing that young people living close to a college exhibit higher levels of education and earnings, after controlling for regional differences and differences in parental education levels.

This observation suggests that living close to a college should have a larger impact on the educational attainment of children from poorer families. In order to corroborate this intuition, Card (1993) considers a specification, in which college proximity is included among the exogenous regressors X , and he uses the interaction between college proximity and parental background to instrument for education.

We replicate this specification using as instruments, W , which is the interaction of parents’ education and the college proximity. We remove all observations for which parents’ education is

missing, and we end up with a sample of 2200 individuals. Summary statistics are reported in Table 2.

	Mean	St.Dev	Min	Max
Log Hourly Wage	6.27	0.44	4.61	7.79
Education (years)	13.38	2.66	1.00	18.00
Mother's education	10.34	3.12	0.00	18.00
Father's education	9.84	3.43	0.00	18.00
Experience	8.67	4.07	0.00	23.00
Black dummy	0.21	0.41	0.00	1.00
College proximity	0.69	0.46	0.00	1.00
South dummy	0.40	0.49	0.00	1.00
Urban dummy	0.72	0.45	0.00	1.00

TABLE 2. Summary statistics

Notice that neither the endogenous variable Z , nor the instruments W , are measured as purely continuous random variables, since they both exhibit a finite number of discrete support points. However, as long as the number of support points of W is greater than the number of support points of Z , we know that the functional coefficients in our model, $\beta(Z)$, are identified (Newey 2013). Moreover, strictly speaking, experience should also be treated as an endogenous covariate as it is functionally related to education. However, for the purposes of illustration, we overlook this issue in our reduced form specification and beg the reader's forgiveness – our aim is to provide a concrete illustration of our method and we do not intend this to be the last word on wage gaps.

As in all semi- and nonparametric approaches, we need to select smoothing parameters. We elect to use the data-adaptive method outlined in Horowitz (2014), which delivers a cubic approximation for the functional coefficients (i.e., $\hat{J} = 3$). We also construct nonparametric bootstrap 90% coverage confidence bounds via pairwise resampling from the joint distribution of (Y, X, Z, W) (Horowitz and Lee 2012). Results are reported in Figure 3.

In each figure, the dashed light grey line denotes the classic linear parametric two stage least square (TSLS) estimator. The dashed-dotted grey line denotes the estimator from a partially linear specification, in which all functional coefficient except the intercept are constant functions of Z . Finally, the black line depicts the estimated functional coefficient from our proposed varying coefficient specification. Results from the TSLS estimator are also reported in Table 3.

	Intercept	Exper	Exper Sq	Black	College Prox	South	Urban	Education
Coefficient	4.26	0.10	-0.24	-0.15	0.00	-0.11	0.15	0.10
Std.Error	0.23	0.01	0.04	0.03	0.02	0.02	0.02	0.01

TABLE 3. Results from TSLS estimation. Dependent variable: log of hourly wage.

The first thing to notice is that the nonparametric confidence bounds appear to be somewhat large. This is, of course, to be expected and is a known feature of instrumental variable estimators in these settings, which tend to exhibit more variability than other estimators that ignore endogeneity (Card 2001). Moreover, the amount of sample information available at very low levels of education, say 0 – 2 years, is quite limited and this is clearly reflected in the nonparametric confidence bounds.

Next, note that the estimator of the intercept does not differ substantially across the three specifications, although it appears that the varying coefficient specification, if anything, suggests a lower return to education for the constant term, relative to both the standard TSLS specification, and, in particular, relative to the partially linear model.

The other coefficients exhibit the expected signs and magnitudes. Furthermore, note that the return on experience is increasing with respect to years of education, while the marginal effect of experience squared (the second derivative of log wage w.r.t. experience) exhibits a sharper decline for more educated workers. The coefficient for race indicates that the wage gap narrows as the level of education increases.

Our results are consistent with those of Card (1993). In particular, we find that the coefficient for college proximity that we do not report here has no explanatory power. This is the case for the TSLS model, and it appears to hold also for our proposed semiparametric estimator. Also, the regional dummy coefficient remains negative and the urban dummy coefficient is positive, although they seem to have both negligible effects on wages at the highest levels of educational attainment.

7. CONCLUSION

We propose a novel semiparametric varying coefficient model, for which the functional coefficients depend on potentially endogenous covariates, Z . We show that the coefficients can be identified under standard assumptions invoked in the literature on ill-posed problems, and we devise a simple sieve based estimator that we hope will be appealing to practitioners. Not only is the approach extremely flexible nesting other popular semiparametric (and even fully nonparametric) estimators,

(a) Constant

(b) Experience

(c) Experience Sq

(d) Black

(e) South

(f) Urban

FIGURE 3. Estimation of the functional coefficients. The black line is the varying coefficient estimator (dashed are confidence intervals), the dashed-dotted grey line is the value of the coefficient from a partially linear specification, and the dashed light grey line is the TSLS estimator. 23

but it also can attenuate the curse-of-dimensionality that limits the practical application of existing fully nonparametric approaches. Finally, we illustrate the method by reexamining returns to education in a sample extracted from the NLSYM.

There are a number of extensions to the framework that one might consider. One such extension involves ‘adaptation’, that is, each one of the varying coefficients may possess differing intrinsic degrees of smoothness, and the rate of convergence we have provided might be sub-optimal in such cases. Following Fan and Zhang (1999), we could pursue a two-step estimation procedure that could overcome this potential limitation.

Another important issue is related to model selection. One possibility, overlooked in the present paper, is that Z may also contain exogenous components, while X may, in fact, contain endogenous components as well. Investigating solutions for such cases would hold much practical appeal, especially for the estimation of demand functions, where price and income are both potentially endogenous (Hausman and Newey 1995, Yatchew and No 2001, Vanhems 2010, Blundell, Horowitz and Parey 2012). Moreover, another case would include the possibility that some of the functional coefficients may, in fact, be constants, and one might devise a more flexible partially linear specification, which would take such cases into account.

APPENDIX

A. Proofs.

A.1. *Preliminary results.* We start by defining the following objects,

$$\begin{aligned} G_{\pi_X} &= E \left[(x_i \otimes \pi_K(w_i)) (x_i \otimes \pi_K(w_i))' \right] = E \left[x_i x_i' \otimes \pi_K(w_i) \pi_K(w_i)' \right], \\ G_{\psi_X} &= E \left[(x_i \otimes \psi_J(z_i)) \text{vec}(x_i \otimes \psi_J(z_i))' \right] = E \left[x_i x_i' \otimes \psi_J(z_i) \psi_J(z_i)' \right], \\ S_X &= E \left[(x_i \otimes \pi_K(w_i)) \text{vec}(x_i \otimes \psi_J(z_i))' \right] = E \left[x_i x_i' \otimes \pi_K(w_i) \psi_J(z_i)' \right], \end{aligned}$$

by using the properties of Kronecker products so that we can rewrite our estimator of β as

$$\hat{\beta}^J = \Psi_J^\oplus(z)' \left(\hat{S}_X' \hat{G}_{\pi_X}^{-1} \hat{S}_X \right)^{-1} \hat{S}_X' \hat{G}_{\pi_X}^{-1} y.$$

Further define:

$$\tilde{\beta}^J = \Psi_J^\oplus(z)' \left(\hat{S}_X' \hat{G}_{\pi_X}^{-1} \hat{S}_X \right)^{-1} \hat{S}_X' \hat{G}_{\pi_X}^{-1} \mathbf{\Pi}' \mathbf{X}' \mathbf{B},$$

with,

$$\mathbf{B} = \begin{bmatrix} \beta_0(x_1) & \dots & \beta_0(x_n) \\ \vdots & \dots & \vdots \\ \beta_d(x_1) & \dots & \beta_d(x_n) \end{bmatrix}, \quad \mathbf{X} = [x_1, \dots, x_n]',$$

the $(d+1) \times n$ matrices of coefficients and regressors, and

$$\beta^J = \Psi_J^\oplus(z)' c.$$

We will adopt much of the notation in Chen and Christensen (2015a, CC henceforth), and will exploit some of their results about the $\mathbb{L}^2$ convergence of sieve estimators for nonparametric instrumental regressions.

For a matrix A , we denote as $v_{\min}(A)$ its smallest singular value, and as $\lambda_{\min}(A'A)$ and $\lambda_{\max}(A'A)$, respectively, the smallest and the largest eigenvalues of $A'A$. Also, we denote as $A_l^- = (A'A)^- A'$, the left pseudo-inverse of A .

Notice that,

$$(6) \quad \|A_l^-\|^2 = \lambda_{\max} \left(A_l^- (A_l^-)' \right) = \lambda_{\max} \left((A'A)^{-1} \right) = 1/\lambda_{\min} (A'A),$$

(see also Lemma E.5 in CC). Further define:

$$v_{JK} = v_{\min} \left(G_{\pi}^{-1/2} S G_{\psi}^{-1/2} \right).$$

Proposition A.1. *Under the completeness condition in Assumption 3.1, and for all $\beta \in \mathcal{B}_{sJ}$, $v_{JK} = O(\tilde{\tau}_J^{-1})$.*

Proof. See CC, p. 4 of the Online Appendix. □

Lemma A.2. *Under the conditions and the result of Proposition A.1, and Assumption 3.2, we have that:*

$$\tilde{v}_{JK} = v_{\min} \left(G_{\pi_X}^{-1/2} S_X G_{\psi_X}^{-1/2} \right) = O(\tilde{\tau}_J^{-1}).$$

Proof. First of all, notice that:

$$\rho_{\min} I_{d+1} \preceq E(x_i x_i' | z_i, w_i) \preceq \rho_{\max} I_{d+1},$$

by the positive definiteness in Assumption 3.2. Denote as $G_{\pi}^{\oplus} = I_{d+1} \otimes G_{\pi}$, the block diagonal matrix obtained by the Kronecker product of the identity matrix of dimension $d + 1$ and G_{π} . Equivalently define $G_{\psi}^{\oplus}$ and $S^{\oplus}$. Now we have that:

$$\begin{aligned} \left\| \left(G_{\psi_X}^{-1/2} S_X' G_{\pi_X}^{-1} S_X G_{\psi_X}^{-1/2} \right)^{-1} \right\|^2 &\leq \left(\frac{\rho_{\max}}{\rho_{\min}} \right)^2 \left\| \left(G_{\psi}^{\oplus -1/2} S^{\oplus'} G_{\pi}^{\oplus -1} S^{\oplus} G_{\psi}^{\oplus -1/2} \right)^{-1} \right\|^2 \\ &= O(1) \left(1 / \lambda_{\min} \left(G_{\psi}^{-1/2} S' G_{\pi}^{-1} S G_{\psi}^{-1/2} \right) \right) = v_{JK}^{-2} = O(\tilde{\tau}_J^2) \end{aligned}$$

by the properties of the eigenvalues of block diagonal matrices, and the result of Proposition A.1. □

We also need the following result on the convergence of the orthonormalized versions of the projection matrices. This result is given as an assumption in this paper. The proof can be found in CC.

Assumption A.3. Define the orthonormalized versions of the projection matrices:

$$\begin{aligned}\hat{G}_{\pi_X}^o &= G_{\pi_X}^{-1/2} \hat{G}_{\pi_X} G_{\pi_X}^{-1/2}, \\ \hat{G}_{\psi_X}^o &= G_{\psi_X}^{-1/2} \hat{G}_{\psi_X} G_{\psi_X}^{-1/2}, \\ \hat{S}_X^o &= G_{\pi_X}^{-1/2} \hat{S}_X G_{\psi_X}^{-1/2},\end{aligned}$$

where $G_{\pi_X}^o = I_{K \times (d+1)}$ and $G_{\psi_X}^o = I_{J \times (d+1)}$. Under the conditions of Assumption 4.1 (v), we have that:

$$\begin{aligned}\|\hat{G}_{\pi_X}^o - G_{\pi_X}^o\|^2 &= O(\zeta_{\pi_X}^2(\log K)/n), \\ \|\hat{G}_{\psi_X}^o - G_{\psi_X}^o\|^2 &= O(\zeta_{\psi_X}^2(\log J)/n), \\ \|\hat{S}_X^o - S_X^o\|^2 &= O(\zeta^2(\log J)/n),\end{aligned}$$

as $J, K \rightarrow \infty$.

A.2. *Proof of Theorem 4.1.* We can write:

$$\|\hat{\beta}^J - \beta\|^2 \leq \|\hat{\beta}^J - \tilde{\beta}^J\|^2 + \|\tilde{\beta}^J - \beta^J\|^2 + \|\beta^J - \beta\|^2 = \|R_1\|^2 + \|R_2\|^2 + \|R_3\|^2.$$

The last term can be easily bounded using Assumption 4.1. Then we have,

$$\|R_3\|^2 = O\left(J^{-2s/p}\right).$$

We can now consider the term:

$$\begin{aligned}\|R_2\|^2 &= \|\Psi_J^\oplus(z)' \left(\hat{S}_X' \hat{G}_{\pi_X}^{-1} \hat{S}_X \right)^{-1} \hat{S}_X' \hat{G}_{\pi_X}^{-1} \Pi' (\mathbf{X}' \mathbf{B} - \Psi_X c) / n\|^2 \\ &\leq \|\Psi_J^\oplus(z)' \left[\left(\hat{G}_{\pi_X}^{-1/2} \hat{S}_X \right)_l^- \hat{G}_{\pi_X}^{-1/2} G_{\pi_X}^{1/2} - \left(G_{\pi_X}^{-1/2} S_X \right)_l^- \right] G_{\pi_X}^{-1/2} \Pi' (\mathbf{X}' \mathbf{B} - \Psi_X c) / n\|^2 \\ &\quad + \|\Psi_J^\oplus(z)' \left(G_{\pi_X}^{-1/2} S_X \right)_l^- G_{\pi_X}^{-1/2} \Pi' (\mathbf{X}' \mathbf{B} - \Psi_X c) / n\|^2 \\ &= \|R_{21}\|^2 + \|R_{22}\|^2\end{aligned}$$

We first notice that:

$$\begin{aligned}\Psi_X c &= [x_1 \otimes \psi_J(z_1) \dots x_n \otimes \psi_J(z_n)]' c = \mathbf{X}' [\Psi_J^\oplus(z_1)' c, \dots, \Psi_J^\oplus(z_n)' c] \\ &= \mathbf{X}' [\beta^J(z_1), \dots, \beta^J(z_n)] = \mathbf{X}' \mathbf{B}^J.\end{aligned}$$

We then bound the term in R_{22} :

$$\begin{aligned}\|R_{22}\|^2 &= \|\Psi_J^\oplus(z)' \left(G_{\pi_X}^{-1/2} S_X\right)_l^- G_{\pi_X}^{-1/2} \mathbf{\Pi}' \mathbf{X}' (\mathbf{B} - \mathbf{B}^J) / n\|^2 \\ &\leq O(1) \left\| \left(G_{\pi_X}^{-1/2} S_X G_{\psi_X}^{-1/2}\right)_l^- \right\|^2 \|G_{\pi_X}^{-1/2} \mathbf{\Pi}' \mathbf{X}' (\mathbf{B} - \mathbf{B}^J) / n\|^2 \\ &\leq O(\tilde{\tau}_J^2) \left\| G_{\pi_X}^{-1/2} \left\{ \mathbf{\Pi}' \mathbf{X}' (\mathbf{B} - \mathbf{B}^J) / n - E[(x_i \otimes \pi_K(w_i)) x_i' (\beta(z_i) - \beta^J(z_i))] \right\} \right\|^2 \\ &\quad + \|E \left[G_{\pi_X}^{-1/2} (x_i \otimes \pi_K(w_i)) x_i' (\beta(z_i) - \beta^J(z_i)) \right]\|^2 \\ &= \left(O \left(\frac{\tilde{\tau}_J^2 \zeta_{\pi_K}^2}{n} \right) + o(1) \right) \|R_3\|^2 = O(J^{-2s/p}),\end{aligned}$$

where the second line comes from the boundedness of the $\mathbb{L}^2$ norm of the sieve basis – the third line comes from the result of lemmas A.2 and B.1, and finally, the last line comes from Assumption 4.1 (v).

We can bound the remaining terms in a similar fashion:

$$\begin{aligned}\|R_{21}\|^2 &\leq O(1) \left\| \left(\hat{G}_{\pi_X}^{-1/2} \hat{S}_X\right)_l^- \hat{G}_{\pi_X}^{-1/2} G_{\pi_X}^{1/2} - \left(G_{\pi_X}^{-1/2} S_X\right)_l^- \right\|^2 \|G_{\pi_X}^{-1/2} \mathbf{\Pi}' (\mathbf{X}' \mathbf{B} - \Psi_X c) / n\|^2 \\ &= O(\tilde{\tau}_J^2) \|G_{\pi_X}^{-1/2} \mathbf{\Pi}' (\mathbf{X}' \mathbf{B} - \Psi_X c) / n\|^2 \\ &= \left(O \left(\frac{\tilde{\tau}_J^2 \zeta_{\pi_K}^2}{n} \right) + o(1) \right) \|R_3\|^2 = O(J^{-2s/p}),\end{aligned}$$

where the second line follows from Assumption A.3 and Lemma B.2, while the last step follows the same lines as above.

We are now left with the term R_1 . We decompose it as before,

$$\begin{aligned}
\|R_1\|^2 &= \|\Psi_J^\oplus(z)' \left(\hat{S}'_X \hat{G}_{\pi_X}^{-1} \hat{S}_X \right)^{-1} \hat{S}'_X \hat{G}_{\pi_X}^{-1} \mathbf{\Pi}' u/n\|^2 \\
&\leq \|\Psi_J^\oplus(z)' \left[\left(\hat{G}_{\pi_X}^{-1/2} \hat{S}_X \right)_l^- \hat{G}_{\pi_X}^{-1/2} G_{\pi_X}^{1/2} - \left(G_{\pi_X}^{-1/2} S_X \right)_l^- \right] G_{\pi_X}^{-1/2} \mathbf{\Pi}' u/n\|^2 \\
&\quad + \|\Psi_J^\oplus(z)' \left(G_{\pi_X}^{-1/2} S_X \right)_l^- G_{\pi_X}^{-1/2} \mathbf{\Pi}' u/n\|^2 \\
&\leq \|\Psi_J^\oplus(z)' \left[\left(\hat{G}_{\pi_X}^{-1/2} \hat{S}_X \right)_l^- \hat{G}_{\pi_X}^{-1/2} G_{\pi_X}^{1/2} - \left(G_{\pi_X}^{-1/2} S_X \right)_l^- \right]\|^2 \|G_{\pi_X}^{-1/2} \mathbf{\Pi}' u/n\|^2 \\
&\quad + \|\Psi_J^\oplus(z)' \left(G_{\pi_X}^{-1/2} S_X \right)_l^- \|^2 \|G_{\pi_X}^{-1/2} \mathbf{\Pi}' u/n\|^2.
\end{aligned}$$

Using Assumption 2.3, we have that:

$$\|G_{\pi_X}^{-1/2} \mathbf{\Pi}' u/n\|^2 = O(K/n),$$

(see, Newey 1997). Following the same line of proof as above, it is straightforward to obtain

$$\|R_1\|^2 = O(\tilde{\tau}_J^2 J/n).$$

The result of the theorem follows.

B. Supplementary Lemmas.

Lemma B.1. *Let $\beta^J(z) = \Psi_J^\oplus(z)'c$, for a deterministic coefficient $c \in \mathbb{R}^{J(d+1)}$, and define $\mathbf{B}^J = [\beta^J(z_1), \dots, \beta^J(z_n)]$. Then, we obtain that:*

$$\|G_{\pi_X}^{-1/2} \{ \mathbf{\Pi}' \mathbf{X}' (\mathbf{B} - \mathbf{B}^J) /n - E[(x_i \otimes \pi_K(w_i)) x'_i (\beta(z_i) - \beta^J(z_i))] \} \| = O\left(\frac{\zeta_{\pi_K}}{\sqrt{n}} \|\beta - \beta^J\|\right).$$

Proof. The proof follows along the lines of Lemma E.9 in Chen and Christensen (2015a). Denote as $\tilde{x}_i \otimes \tilde{\pi}_K(w_i)$, the orthonormalized version of $x_i \otimes \pi_K(w_i)$, i.e. $G_{\pi_X}^{-1/2}(x_i \otimes \pi_K(w_i))$. Then, we have that,

$$G_{\pi_X}^{-1/2} \mathbf{\Pi}' \mathbf{X}' (\mathbf{B} - \mathbf{B}^J) /n = \frac{1}{n} \sum_{i=1}^n (\tilde{x}_i \otimes \tilde{\pi}_K(w_i)) x'_i (\beta(z_i) - \beta^J(z_i)),$$

and, therefore,

$$E \left[G_{\pi_X}^{-1/2} \mathbf{\Pi}' \mathbf{X}' (\mathbf{B} - \mathbf{B}^J) /n \right] = E \left[(\tilde{x}_i \otimes \tilde{\pi}_K(w_i)) x'_i (\beta(z_i) - \beta^J(z_i)) \right],$$

because of the i.i.d. assumption. Finally,

$$\begin{aligned}
& E \left[\| G_{\pi_X}^{-1/2} \mathbf{\Pi}' \mathbf{X}' (\mathbf{B} - \mathbf{B}^J) / n - E \left[(\tilde{x}_i \otimes \tilde{\pi}_K) x_i' (\beta(z_i) - \beta^J(z_i)) \right] \|^2 \right] \\
& \leq \frac{1}{n} E \left[\sum_{k=1}^K (\beta(z_i) - \beta^J(z_i))' x_i (\tilde{x}_i' \tilde{x}_i \otimes \tilde{\pi}_{K,k}^2(w_i)) x_i' (\beta(z_i) - \beta^J(z_i)) \right] \\
& \leq \rho_{max} \frac{\zeta_{\pi_K}^2}{n} \|\beta - \beta^J\|^2
\end{aligned}$$

The result of the Lemma follows from Chebyshev's inequality. $\square$

Lemma B.2. *Let*

$$\left\| \left(\hat{G}_{\pi_X}^{-1/2} \hat{S}_X \right)_l^- \hat{G}_{\pi_X}^{-1/2} G_{\pi_X}^{1/2} - \left(G_{\pi_X}^{-1/2} S_X \right)_l^- \right\| = O(\tilde{v}_{JK}^{-2})$$

Proof. We first rewrite each term using the orthonormalized versions of the projection matrices:

$$\left(\hat{G}_{\pi_X}^{-1/2} \hat{S}_X \right)_l^- \hat{G}_{\pi_X}^{-1/2} G_{\pi_X}^{1/2} - \left(G_{\pi_X}^{-1/2} S_X \right)_l^- = \hat{G}_{\psi_X}^{-1/2} \left[\left(\hat{G}_{\pi_X}^{o-1/2} \hat{S}_X^o \right)_l^- \hat{G}_{\pi_X}^{o-1/2} - (S_X^o)_l^- \right].$$

Then we have,

$$\begin{aligned}
\left\| \left(\hat{G}_{\pi_X}^{o-1/2} \hat{S}_X^o \right)_l^- \hat{G}_{\pi_X}^{o-1/2} - (S_X^o)_l^- \right\|^2 & \leq \left\| \left(\hat{G}_{\pi_X}^{o-1/2} \hat{S}_X^o \right)_l^- - (S_X^o)_l^- \right\|^2 \|\hat{G}_{\pi_X}^{o-1/2}\|^2 \\
& + \|\hat{G}_{\pi_X}^{o-1/2} - I_{K \times (d+1)}\|^2 \|(S_X^o)_l^-\|^2.
\end{aligned}$$

The last term can be easily bound by using Assumption A.3, the spectral norm of the left pseudo-inverse in (6) and the result in Lemma A.2. First notice that

$$\|\hat{G}_{\pi_X}^{o-1/2} - I_{K \times (d+1)}\| \leq \|\hat{G}_{\pi_X}^{o-1/2}\| \|\hat{G}_{\pi_X}^{o1/2} - I_{K \times (d+1)}\| \leq \frac{\lambda_{min}(\hat{G}_{\pi_X}^{o1/2})}{1 + \sqrt{\lambda_{min}(\hat{G}_{\pi_X}^o)}} \|\hat{G}_{\pi_X}^o - I_{K \times (d+1)}\|,$$

where the last inequality follows from Lemma B.3. Since, $\|\hat{G}_{\pi_X}^o - I_{K \times (d+1)}\| = o(1)$, by Assumption A.3, it has to be that $\lambda_{min}(\hat{G}_{\pi_X}^o) = O(1)$, and $\lambda_{min}(\hat{G}_{\pi_X}^{o1/2}) = O(1)$. Hence,

$$\|\hat{G}_{\pi_X}^{o-1/2} - I_{K \times (d+1)}\|^2 \|(S_X^o)_l^-\|^2 = O(\tilde{v}_{JK}^{-4} \zeta_{\pi_K}^2 (\log K)/n) = O(\tilde{v}_{JK}^{-2}).$$

For the first term, we again use the results in Assumption A.3 and the result above to claim that $\|\hat{G}_{\pi_X}^{o-1/2}\|^2 = 1 + o(1)$. Notice that, by this statement and the conditions in Assumption 4.1(v), we

have

$$\mathbb{P} \left(\|\hat{G}_{\pi_X}^{o-1/2} \hat{S}_X^o - S_X^o\| \geq \frac{1}{2} v_{JK} \right) = o(1).$$

We can thus apply the result in Lemma B.4. Finally, by virtue of Assumption 4.1(v),

$$\begin{aligned} \left\| \left(\hat{G}_{\pi_X}^{o-1/2} \hat{S}_X^o \right)_l^- - (S_X^o)_l^- \right\|^2 &\leq O(\tilde{v}_{JK}^{-4}) \left(\|\hat{G}_{\pi_X}^{o-1/2} - I_K\|^2 \|\hat{S}_X^o\|^2 + \|\hat{S}_X^o - S_X^o\|^2 \right) \\ &= O(\tilde{v}_{JK}^{-4} \zeta^2 (\log K)/n) = O(\tilde{v}_{JK}^{-2}). \end{aligned}$$

□

Lemma B.3 (Schmitt (1992). See also CC, Lemma E.3 p.40). *Let $A, B \in \mathbb{R}^{n_r \times n_r}$ be positive definite. Then*

$$\|B^{1/2} - A^{1/2}\| \leq \frac{1}{\sqrt{\lambda_{\min}(A)} + \sqrt{\lambda_{\min}(B)}} \|A - B\|.$$

Lemma B.4. *Let $A, B \in \mathbb{R}^{n_r \times n_c}$, with $n_r \leq n_c$ and full row rank n_r , such that $\|A - B\| \leq 0.5 v_{\min}(A)$. Then*

$$\|B_l^- - A_l^-\| \leq \frac{2(1 + \sqrt{5})}{v_{\min}^2(A)} \|A - B\|.$$

Proof. See CC Lemma E.4 p. 40.

□

REFERENCES

- Ai, C. and Chen, X. (2003), ‘Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions’, *Econometrica* **71**(6), 1795–1843.
- Andrews, D. W. K. (2011), ‘Examples of L^2 -Complete and Boundedly-Complete Distributions’, *Cowles Foundation Discussion Paper* **1801**.
- Belloni, A., Chernozhukov, V., Chetverikov, D. and Kato, K. (2015), ‘Some New Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results’, *Journal of Econometrics* **186**(2), 345–366.
- Blundell, R., Chen, X. and Kristensen, D. (2007), ‘Semi-Nonparametric IV Estimation of Shape Invariant Engel Curves’, *Econometrica* **75**, 1613–1669.
- Blundell, R., Horowitz, J. L. and Parey, M. (2012), ‘Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation’, *Quantitative Economics* **3**(1), 29–51.
- Cai, Z., Das, M., Xiong, H. and Wu, X. (2006), ‘Functional Coefficient Instrumental Variables Models’, *Journal of Econometrics* **133**(1), 207 – 241.
- Cai, Z. and Xiong, H. (2012), ‘Partially varying coefficient instrumental variables models’, *Statistica Neerlandica* **66**(2), 85–110.
- Canay, I. A., Santos, A. and Shaikh, A. M. (2013), ‘On the Testability of Identification in Some Nonparametric Models with Endogeneity’, *Econometrica* **81**(6), 2535–2559.
- Card, D. (1993), Using Geographic Variation in College Proximity to Estimate the Return to Schooling, in L. Christofides, E. Grant and R. Swidinsky, eds, ‘Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp’, Toronto: University of Toronto Press.
- Card, D. (1999), The Causal Effect of Education on Earnings, Vol. 3, Part A of *Handbook of Labor Economics*, Elsevier, pp. 1801 – 1863.
- Card, D. (2001), ‘Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems’, *Econometrica* **69**(5), 1127–1160.
- Carrasco, M., Florens, J.-P. and Renault, E. (2007), Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization, in J. J. Heckman and E. E. Leamer, eds, ‘Handbook of Econometrics’, North Holland, pp. 733–751.
- Centorrino, S., Fève, F. and Florens, J.-P. (2015), ‘Additive Nonparametric Instrumental Regressions: a Guide to Implementation’, *Journal of Econometric Methods* **In Press**.
- Chen, X. (2007), Large Sample Sieve Estimation of Semi-Nonparametric Models, in J. J. Heckman and E. E. Leamer, eds, ‘Handbook of Econometrics’, Vol. 6, Part B, Elsevier, pp. 5549 – 5632.
- Chen, X. and Christensen, T. M. (2015a), ‘Optimal Uniform Convergence Rates and Adaptive Estimation of Non-parametric Instrumental Variables Models’, *Mimeo - NYU*.

- Chen, X. and Christensen, T. M. (2015*b*), ‘Optimal Uniform Convergence Rates and Asymptotic Normality for Series Estimators under Weak Dependence and Weak Conditions’, *Journal of Econometrics* **188**(2), 447–465.
- Chen, X. and Pouzo, D. (2012), ‘Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals’, *Econometrica* **80**, 277–321.
- Chen, X. and Pouzo, D. (2015), ‘Sieve Wald and QLR Inferences on Semi/Nonparametric Conditional Moment Models’, *Econometrica* **83**(3), 1013–1079.
- Chen, X. and Reiss, M. (2011), ‘On Rate Optimality For Ill-Posed Inverse Problems In Econometrics’, *Econometric Theory* **27**(Special Issue 03), 497–521.
- Darolles, S., Fan, Y., Florens, J.-P. and Renault, E. (2011), ‘Nonparametric Instrumental Regression’, *Econometrica* **79**(5), 1541–1565.
- Delgado, M. S., Ozabaci, D., Sun, Y. and Kumbhakar, S. C. (2015), ‘Smooth Coefficient Models with Endogenous Environmental Variables’, *Mimeo - Purdue University*.
- D’Haultfoeuille, X. (2011), ‘On the Completeness Condition in Nonparametric Instrumental Problems’, *Econometric Theory* **27**(Special Issue 03), 460–471.
- Fan, J. and Zhang, W. (1999), ‘Statistical Estimation in Varying Coefficient Models’, *Ann. Statist.* **27**(5), 1491–1518.
- Fan, J. and Zhang, W. (2008), ‘Statistical Methods with Varying Coefficient Models’, *Statistics and Its Interface* **1**, 179–195.
- Florens, J.-P., Johannes, J. and Van Bellegem, S. (2012), ‘Instrumental Regressions in Partially Linear Models’, *The Econometrics Journal* **15**(2), 304–324.
- Freyberger, J. (2015), ‘On Completeness and Consistency in Nonparametric Instrumental Variable Models’, *Mimeo - University of Wisconsin Madison*.
- Fridman, V. M. (1956), ‘A Method of Successive Approximations for Fredholm Integral Equations of the First Kind’, *Uspekhi, Math. Nauk.* **11**, 233–334.
- Galerkin, B. G. (1915), ‘Rods and plates. Series Occurring in Various Questions Concerning the Elastic Equilibrium of Rods and Plates’, *Vestnik Inzhenerov* **19**, 897–908.
- Hall, P. and Horowitz, J. L. (2005), ‘Nonparametric Methods for Inference in the Presence of Instrumental Variables’, *Annals of Statistics* **33**(6), 2904–2929.
- Hastie, T. and Tibshirani, R. (1993), ‘Varying-Coefficient Models’, *Journal of the Royal Statistical Society. Series B (Methodological)* **55**(4), 757–796.
- Hausman, J. A. and Newey, W. K. (1995), ‘Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss’, *Econometrica* **63**(6), 1445–1476.
- Heckman, J. J. and Vytlacil, E. (2005), ‘Structural Equations, Treatment Effects, and Econometric Policy Evaluation’, *Econometrica* **73**(3), 669–738.
- Horowitz, J. L. (2011), ‘Applied Nonparametric Instrumental Variables Estimation’, *Econometrica* **79**, 347–394.

- Horowitz, J. L. (2014), ‘Ill-Posed Inverse Problems in Economics’, *Annual Review of Economics* **6**(1), 21–51.
- Horowitz, J. L. and Lee, S. (2012), ‘Uniform Confidence Bands for Functions Estimated Nonparametrically with Instrumental Variables’, *Journal of Econometrics* **168**(2), 175 – 188.
- Huang, J. Z. (2003), ‘Local asymptotics for polynomial spline regression’, *Ann. Statist.* **31**(5), 1600–1635.
- Kress, R. (1999), *Linear integral equations*, Applied mathematical sciences, Springer-Verlag.
- Landweber, L. (1951), ‘An iterative formula for Fredholm integral equations of the first kind’, *American Journal of Mathematics* **73**, 615–624.
- Lee, Y. K., Mammen, E. and Park, B. U. (2012), ‘Flexible generalized varying coefficient regression models’, *Ann. Statist.* **40**(3), 1906–1933.
- Li, Q., Huang, C. J., Li, D. and Fu, T.-T. (2002), ‘Semiparametric Smooth Coefficient Models’, *Journal of Business & Economic Statistics* **20**(3), 412–422.
- Li, Q., Ouyang, D. and Racine, J. S. (2013), ‘Categorical Semiparametric Varying-Coefficient Models’, *Journal of Applied Econometrics* **28**(4), 551–579.
- Liu, C.-A. and Tao, J. (2014), ‘Model Selection and Model Averaging in Nonparametric Instrumental Variables Models’, *Working Paper*.
- Mincer, J. (1974), *Schooling, Experience, and Earnings*, Human Behavior and Social Institutions, National Bureau of Economic Research.
- Newey, W. K. (1997), ‘Convergence Rates and Asymptotic Normality for Series Estimators’, *Journal of Econometrics* **79**(1), 147 – 168.
- Newey, W. K. (2013), ‘Nonparametric Instrumental Variables Estimation’, *American Economic Review* **103**(3), 550–56.
- Newey, W. K. and Powell, J. L. (2003), ‘Instrumental Variables Estimation of Nonparametric Models’, *Econometrica* **71**(5), 1565–1578.
- Newey, W. K., Powell, J. L. and Vella, F. (1999), ‘Nonparametric estimation of triangular simultaneous equations models’, *Econometrica* **67**, 565–603.
- Park, B. U., Mammen, E., Lee, Y. K. and Lee, E. R. (2015), ‘Varying Coefficient Regression Models: A Review and New Developments’, *International Statistical Review* **83**(1), 36–64.
- Petrov, G. I. (1940), ‘Application of Galerkin’s method to the problem of stability of flow of a viscous fluid’, *Journal of Applied Mathematics and Mechanics* **4**, 3–12.
- Robinson, P. M. (1988), ‘Root-n consistent semiparametric regression’, *Econometrica* **56**, 931–954.
- Schmitt, B. A. (1992), ‘Perturbation Bounds for Matrix Square Roots and Pythagorean Sums’, *Linear Algebra and its Applications* **174**, 215 – 227.

- Su, L., Chen, Y. and Ullah, A. (2014), Functional Coefficient Estimation with both Categorical and Continuous Data, *in* J. Racine, L. Su and A. Ullah, eds, ‘The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics’, Oxford University Press, pp. 129–148.
- Su, L., Murtazashvili, I. and Ullah, A. (2013), ‘Local Linear GMM Estimation of Functional Coefficient IV Models With an Application to Estimating the Rate of Return to Schooling’, *Journal of Business & Economic Statistics* **31**(2), 184–207.
- Tikhonov, A. N. (1943), ‘On the stability of inverse problems’, *Doklady Akademii Nauk SSSR* **39**(5), 195–198.
- Vanhems, A. (2010), ‘Non-parametric estimation of exact consumer surplus with endogeneity in price’, *Econometrics Journal* **13**(3), S80–S98.
- Wooldridge, J. (2010), *Econometric Analysis of Cross Section and Panel Data*, second edn, MIT Press.
- Xue, L. and Yang, L. (2006), ‘Additive Coefficient Modeling via Polynomial Splines’, *Statistica Sinica* **16**(4), 1423–1446.
- Yatchew, A. and No, J. A. (2001), ‘Household Gasoline Demand in Canada’, *Econometrica* **69**(6), 1697–1709.

(S. Centorrino) DEPARTMENT OF ECONOMICS, SUNY STONY BROOK, LONG ISLAND, NEW YORK, USA

(J. S. Racine) DEPARTMENT OF ECONOMICS, GRADUATE PROGRAM IN STATISTICS, DEPARTMENT OF MATHEMATICS AND STATISTICS, MCMASTER UNIVERSITY, HAMILTON, ONTARIO, CANADA