

Statistica Sinica Preprint No: SS-2018-0382

Title	Efficient nonparametric three-stage estimation of fixed effects varying coefficient panel data models
Manuscript ID	SS-2018-0382
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202018.0382
Complete List of Authors	Juan M. Rodriguez-Póo and Alexandra Soberón
Corresponding Author	Juan M. Rodriguez-Póo
E-mail	rodrigjm@unican.es
Notice: Accepted version subject to English editing.	

Efficient nonparametric three-stage estimation of fixed effects varying coefficient panel data models

Juan M. Rodriguez-Póo and Alexandra Soberón

Departamento de Economía, Universidad de Cantabria

Abstract:

This paper is concerned with the estimation of a fixed effects panel data model that adopts a partially linear form, in which the coefficients of some variables are restricted to be constant but the coefficients of other variables are assumed to be varying, depending on some exogenous continuous variables. Moreover, we allow for the existence of endogeneity in the structural equation. Conditional moment restrictions on first differences are imposed to identify the structural equation. Based on these restrictions we propose a three stage estimation procedure. The asymptotic properties of these proposed estimators are established. Moreover, as a result of the first differences transformation, to estimate the unknown varying coefficient functions, two alternative backfitting estimators are obtained. As a novelty, we propose a minimum distance estimator that, combining both estimators, is more efficient and achieves the optimal rate of convergence. The feasibility and possible gains of this new procedure are shown by estimating a Life-cycle hypothesis panel data model and a Monte Carlo study is implemented.

Key words and phrases: Panel data; Endogeneity; Fixed effects; Functional-coefficient models; Generalized F-test; Instrumental variables.

1. Introduction

Two of the most important issues that econometricians must face when modeling individual choice in demand systems or market equilibrium are the presence of endogenous variables and individual heterogeneity (see Heckman (2008)). Traditionally, instrumental variable models (IV) have been proposed as the solution to account for endogeneity, whereas heterogeneity has been handled through the use of panel data techniques (see among others Arellano (2003)). Trying to cope with both issues at the same time, instrumental variable models exhibit a long tradition in the panel data analysis literature (see for example Hsiao (2003), Chapter 5). In many situations, economic theory does not imply tight functional form specifications for instrumental variable models so that it is useful to consider nonparametric and semiparametric extensions. Unfortunately, the introduction of this flexible specifications has a cost in terms of curse of dimensionality (see Härdle (1990) for details). One solution to this problem that exhibits a clear motivation from economic theory (see Chamberlain (1992)), encompasses many alternative models (i.e., fully nonparametric models and partially linear models) and avoids the so-called ill-posed inverse problem in general nonparametric instrumental variable models (see Newey and Powell (2003)) are the so-called varying coefficient models.

1. INTRODUCTION

This paper is then concerned with the estimation of a fixed effects panel data model that adopts a partially linear form, in which the coefficients of some variables are restricted to be constant but the coefficients of other variables are assumed to be varying, depending on some exogenous continuous variables. Moreover, we allow for the existence of endogeneity in the structural equation. This structure leads itself naturally to a semiparametric three stage estimation procedure that is based in a transformed (first order differenced) structural model. In the first stage, endogenous variables are projected on a set of instrumental variables, in the second stage constant coefficients are estimated through a profile least squares approach and finally, in the third step nonparametric techniques are used to estimate the varying coefficients. Unfortunately, the estimators obtained in this last stage achieve a rather slow rate of convergence. In order to improve its rate, and following the ideas in Fan and Zhang (1999), a one-step backfitting procedure is developed. It turns out that the resulting estimator is oracle efficient and exhibits an optimal rate of convergence. However, as a result of the first differences transformation, two alternative backfitting estimators for the same unknown function of the varying parameters are obtained. With the aim of improving the efficiency, we combine both estimators through a minimum distance estimation technique and hence, the

1. INTRODUCTION

resulting estimator is more efficient. As far as we know, this approach is completely new and the minimum distance estimation technique applied to this problem has never been used before in the literature.

To avoid the ill-posed inverse problem (see Newey and Powell (2003) for details) but at the same time, to keep some model specification flexibility, conditional moment restrictions on first differences are imposed to identify the structural equation (see Ai and Chen (2003), Hall and Horowitz (2005) and Newey (2013) among others for a similar approach). Other procedures such as the so-called control function approach proposed in Heckman and Robb (1985), Blundell et al. (2013), Darolles et al. (2004), Gao and Phillips (2013), and Su and Ullah (2008) among others are available at the price of assuming other type of identification assumptions. To the best of our knowledge, Cai et al. (2006), Cai and Xiong (2012) and Cai et al. (2017) are the most relevant references for varying coefficient with endogenous covariates, but they completely ignore the panel data case. On its part, several papers (see Rodriguez-Poo and Soberon (2017) for a survey) analyze panel data varying coefficient models, but the resulting estimators are not robust to the presence of endogeneity. Recently, Fève and Florens (2014) consider the estimation of nonparametric panel data models using an instrumental variable condition. However, their results do not apply straightforwardly

1. INTRODUCTION

to the varying coefficient model. Finally, some IV methods have been proposed in the context of varying coefficient panel data models with random effects. In Cai and Li (2008) it is proposed to estimate the unknown functions of interest by using the so-called nonparametric generalized method of moments. However, this method does not control for heterogeneity when it is correlated with some explanatory variables, and hence it renders to asymptotically biased estimators when fixed effects are present. Since the semiparametric partially linear varying coefficient model encompasses several alternative specifications that can be of great interest for econometricians (i.e. partially linear model, fully linear parametric model), based in Cai et al. (2017) and references therein, we propose a Wald-type test statistic. Furthermore, we provide a technique to compute confidence bands for the varying coefficients. To show the feasibility and possible gains of this new procedure, it is applied to extend a Life Cycle Hypothesis (LCH) model as the one proposed in Chou et al. (2004) to a panel data model.

The structure of this paper is as follows. In Section 2 we set up the econometric model and we describe the three-step estimation procedure. In Section 3 we show their asymptotic properties. In Section 4 more efficient estimators such as one-step backfitting and minimum distance estimators are provided. Section 5 develops a Wald-type test for the constant coeffi-

2. MODEL AND ESTIMATION PROCEDURES

icients and a pointwise confidence bands for the functional coefficients. In Section 6 a Monte Carlo study is presented to investigate the finite sample performance of the proposed estimators and test statistic. Section 7 applies our methods to the estimation of a LCH model. Finally, Section 8 concludes the paper. Assumptions and proofs of the main results are relegated to the Supplementary Material.

2. Model and estimation procedures

A partially varying coefficient panel data model assumes the following form:

$$Y_{it} = X_{1it}^\top m_1(Z_{it}) + X_{2it}^\top m_2(Z_{it}) + U_{1it}^\top \beta_1 + U_{2it}^\top \beta_2 + \mu_i + \epsilon_{it},$$
$$i = 1, \dots, N; \quad t = 1, \dots, T, \quad (2.1)$$

where Y_{it} is an observed scalar random variable, X_{1it} and U_{1it} are $(d_1 \times 1)$ and $(k_1 \times 1)$ vectors of endogenous random variables, respectively, X_{2it} and U_{2it} are vectors of exogenous random variables of dimension $(d_2 \times 1)$ and $(k_2 \times 1)$, respectively, ϵ_{it} is the random error, and μ_i denotes the unobserved individual heterogeneity. Also, the structural equation (2.1) includes some unknown functions (i.e., $m_1(\cdot)$ and $m_2(\cdot)$) of a $(q \times 1)$ vector of exogenous continuous random variables, Z_{it} , and some constant coefficients (i.e., β_1 and β_2) that need to be estimated. Furthermore, let us denote \mathcal{L}_{it} as a $(\ell \times 1)$ vector which contains all exogenous variables (i.e., Z , X_2 , and U_2) and a M -

2. MODEL AND ESTIMATION PROCEDURES

dimensional vector of other instrumental variables, where $\ell = q + d_2 + k_2 + M$ and $\ell \geq d_1 + k_1$, which is the identification condition that the number of instruments is larger than the number of endogenous variables. A similar definition is given for $\mathcal{L}_{i(t-1)}$. We assume the following moment condition

$$E(\Delta\epsilon_{it} | \mathcal{L}_{it}, \mathcal{L}_{i(t-1)}) = 0 \quad i = 1, \dots, N; \quad t = 1, \dots, T. \quad (2.2)$$

The above model is general enough to include relevant empirical examples in the economics literature. For example, based on the Life-cycle hypothesis (LCH) theory, Gourinchas and Parker (2002) and Kuan and Chen (2013) show that the elasticity of preventive savings to changes in net wealth and/or medical expenses varies according to certain households' features such as the age of the household head.

Further, the model above allows for two different sources of endogeneity. First, there exists a subset of endogenous explanatory variables (i.e., X_1 and U_1). Second, the heterogeneity term, μ_i , can be also arbitrarily correlated with Z , X and/or U (i.e., fixed effects). It is already well-known that, whether we ignore both sources of endogeneity, direct estimation of the functions of interest tends to estimators asymptotically biased. The second source of endogeneity can be handled by taking a first difference

2. MODEL AND ESTIMATION PROCEDURES

transformation in (2.1) obtaining

$$\begin{aligned} \Delta Y_{it} &= X_{it}^\top m(Z_{it}) - X_{i(t-1)}^\top m(Z_{i(t-1)}) + \Delta U_{it}^\top \beta + \Delta \epsilon_{it}, \quad (2.3) \\ i &= 1, \dots, N; \quad t = 1, \dots, T, \end{aligned}$$

where $X_{it} = (X_{1it}^\top, X_{2it}^\top)^\top$ and $m(Z_{it}) = (m_1(Z_{it})^\top, m_2(Z_{it})^\top)^\top$ are $(d \times 1)$ vectors, while $\Delta U_{it} = (\Delta U_{1it}^\top, \Delta U_{2it}^\top)^\top$ and $\beta = (\beta_1^\top, \beta_2^\top)^\top$ are $(k \times 1)$ vectors, with $d = d_1 + d_2$ and $k = k_1 + k_2$. A similar definition is given for $X_{i(t-1)}$ and $m(Z_{i(t-1)})$.

For any given β , and $\Delta Y_{it}^* = \Delta Y_{it} - \Delta U_{it}^\top \beta$, in Rodriguez-Poo and Soberon (2014) it is proposed to estimate the quantities of interest, $m(\cdot)$, for a given point $z \in \mathcal{A}$ where \mathcal{A} is a compact subset in a nonempty interior of \mathbb{R} , by minimizing the following criterion function

$$\sum_{i=1}^N \sum_{t=2}^T (\Delta Y_{it}^* - \Delta X_{it}^\top \gamma)^2 K_H(Z_{it} - z) K_H(Z_{i(t-1)} - z) \quad (2.4)$$

with respect to γ , where $\gamma = m(z)$. In addition, H is a $q \times q$ symmetric positive definite bandwidth matrix and K is a q -variate such that

$$K_H(u) = \frac{1}{|H|^{1/2}} K(H^{-1/2}u).$$

Note that the kernel weights in (2.4) are related to both Z_{it} and $Z_{i(t-1)}$. This is a significant issue because it enables us to overcome the non-negligible asymptotic bias characteristic of the differencing nonparametric estimators.

2. MODEL AND ESTIMATION PROCEDURES

If we consider kernels only around Z_{it} , the remainder term in the Taylor's approximation will not be negligible since the distance between Z_{is} ($s \neq t$) and z does not vanish asymptotically. This phenomena was already pointed out in Mundra (2005) and Lee and Mukherjee (2014), but it was solved in Rodriguez-Poo and Soberon (2014, 2015) for a local linear regression.

Unfortunately, although the resulting estimator for γ using the latter proposal is robust to fixed effects, it is still subject to the first endogeneity problem (i.e., the endogeneity of X_1 and U_1). Taking expectation on both sides of the structural equation (2.3), conditioning on both \mathcal{L}_{it} and $\mathcal{L}_{i(t-1)}$, and using the condition (2.2) one can obtain the following

$$E[\Delta Y_{it} | \mathcal{L}_{it}, \mathcal{L}_{i(t-1)}] = W_{X_{it}}^\top m(Z_{it}) - W_{X_{i(t-1)}}^\top m(Z_{i(t-1)}) + \Delta W_{U_{it}}^\top \beta, \quad (2.5)$$

where $W_{X_{it}} = (E(X_{1it}^\top | \mathcal{L}_{it}, \mathcal{L}_{i(t-1)}), X_{2it}^\top)^\top$ is a d -dimensional vector and $\Delta W_{U_{it}} = (E(\Delta U_{1it}^\top | \mathcal{L}_{it}, \mathcal{L}_{i(t-1)}), \Delta U_{2it}^\top)^\top$ is a $(k \times 1)$ vector. A similar definition is used for $W_{X_{i(t-1)}}$.

Then, taking into account (2.5) and proceeding as above the coefficient functions $m(\cdot)$ can be estimated by minimizing the following criterion function

$$\sum_{i=1}^N \sum_{t=2}^T (\Delta Y_{it}^* - \Delta W_{X_{it}}^\top \gamma)^2 K_{H_2}(Z_{it} - z) K_{H_2}(Z_{i(t-1)} - z), \quad (2.6)$$

with respect to γ , where $\gamma = m(z)$, where $\Delta W_{X_{it}} = (E(\Delta X_{1it}^\top | \mathcal{L}_{it}, \mathcal{L}_{i(t-1)}), \Delta X_{2it}^\top)^\top$.

2. MODEL AND ESTIMATION PROCEDURES

Assuming that $\sum_{it} K_{H_2}(Z_{it} - z)K_{H_2}(Z_{i(t-1)} - z)\Delta W_{X_{it}}\Delta W_{X_{it}}^\top$ is positive definite, the solution to this problem is

$$\begin{aligned} \tilde{m}_\beta(z; H_2) &= \left(\sum_{i=1}^N \sum_{t=2}^T K_{H_2}(Z_{it} - z)K_{H_2}(Z_{i(t-1)} - z)\Delta W_{X_{it}}\Delta W_{X_{it}}^\top \right)^{-1} \\ &\times \sum_{i=1}^N \sum_{t=2}^T K_{H_2}(Z_{it} - z)K_{H_2}(Z_{i(t-1)} - z)\Delta W_{X_{it}}(\Delta Y_{it} - \Delta W_{U_{it}}^\top \beta). \end{aligned} \quad (2.7)$$

Unfortunately, $\tilde{m}_\beta(z; H_2)$ is an infeasible estimator because the vector of parameters, β , and the nonparametric functions $E(\Delta X_{1it}^\top | \mathcal{L}_{it}, \mathcal{L}_{i(t-1)})$ and $E(\Delta U_{1it}^\top | \mathcal{L}_{it}, \mathcal{L}_{i(t-1)})$ are unknown and need to be estimated. The first stage in our estimation procedure is the estimation of the nonparametric functions. Let us denote by $\hat{E}(\Delta X_{1it}^\top | \mathcal{L}_{it}, \mathcal{L}_{i(t-1)}; H_1)$ and $\hat{E}(\Delta U_{1it}^\top | \mathcal{L}_{it}, \mathcal{L}_{i(t-1)}; H_1)$ nonparametric estimators of $E(\Delta X_{1it}^\top | \mathcal{L}_{it}, \mathcal{L}_{i(t-1)})$ and $E(\Delta U_{1it}^\top | \mathcal{L}_{it}, \mathcal{L}_{i(t-1)})$ respectively, with bandwidth H_1 . They can be for example local linear or constant estimators.

The second stage is to estimate β . We propose a conventional profile least squares estimator (see Fan and Huang (2005))

$$\hat{\beta} = [\Delta \widehat{W}_U^\top (I_n - \widehat{S})^\top (I_n - \widehat{S}) \Delta \widehat{W}_U]^{-1} \Delta \widehat{W}_U^\top (I_n - \widehat{S})^\top (I_n - \widehat{S}) \Delta Y, \quad (2.8)$$

where $\Delta Y = [\Delta Y_{12}, \dots, \Delta Y_{NT}]^\top$, $\Delta \widehat{W}_U = [\Delta \widehat{W}_{U_{12}}, \dots, \Delta \widehat{W}_{U_{NT}}]^\top$, $\Delta \widehat{W}_{U_{it}}$ is the vector $\Delta W_{U_{it}}$ where the unknown functions have been replaced by

2. MODEL AND ESTIMATION PROCEDURES

consistent estimators and finally \widehat{S} is a smoothing matrix, i.e.

$$\widehat{S} = \begin{pmatrix} \widehat{W}_{X_{12}}^\top \widehat{\Gamma}_{12}^{-1} \Delta \widehat{W}_X^\top K(Z_{12}; H_2) - \widehat{W}_{X_{11}}^\top \widehat{\Gamma}_{11}^{-1} \Delta \widehat{W}_X^\top K(Z_{11}; H_2) \\ \vdots \\ \widehat{W}_{X_{NT}}^\top \widehat{\Gamma}_{NT}^{-1} \Delta \widehat{W}_X^\top K(Z_{NT}; H_2) - \widehat{W}_{X_{N(T-1)}}^\top \widehat{\Gamma}_{N(T-1)}^{-1} \Delta \widehat{W}_X^\top K(Z_{N(T-1)}; H_2) \end{pmatrix}$$

with $\widehat{\Gamma}_{it} = \Delta \widehat{W}_X^\top K(Z_{it}, H_2) \Delta \widehat{W}_X$, $\Delta \widehat{W}_X = [\Delta \widehat{W}_{X_{12}}, \dots, \Delta \widehat{W}_{X_{NT}}]^\top$ and $\Delta \widehat{W}_{X_{it}}$ and $\widehat{W}_{X_{it}}$ are the vectors $\Delta W_{X_{it}}$ and $W_{X_{it}}$ respectively, where the unknown functions have been replaced by consistent estimators. Finally $K(z; H_2)$ is a $n \times n$ diagonal matrix of the form

$$K(z; H_2) = \text{diag}(K_{H_2}(Z_{12}-z)K_{H_2}(Z_{11}-z), \dots, K_{H_2}(Z_{NT}-z)K_{H_2}(Z_{N(T-1)}-z)).$$

Note that $\widehat{\beta}$ can be considerably affected by the residuals from the first stage. To overcome it, following Cai et al. (2017) we propose a modified estimator of the form

$$\widetilde{\beta} = [\Delta \widehat{W}_U^\top (I_n - \widehat{S})^\top (I_n - \widetilde{S}) \Delta U]^{-1} \Delta \widehat{W}_U^\top (I_n - \widehat{S})^\top (I_n - \widetilde{S}) \Delta Y, \quad (2.9)$$

where \widetilde{S} is a smoothing matrix of the form

$$\widetilde{S} = \begin{pmatrix} X_{12}^\top \widehat{\Gamma}_{12}^{-1} \Delta \widehat{W}_X^\top K(z_{12}; H_2) - X_{11}^\top \widehat{\Gamma}_{11}^{-1} \Delta \widehat{W}_X^\top K(z_{11}; H_2) \\ \vdots \\ X_{NT}^\top \widehat{\Gamma}_{NT}^{-1} \Delta \widehat{W}_X^\top K(z_{NT}; H_2) - X_{N(T-1)}^\top \widehat{\Gamma}_{N(T-1)}^{-1} \Delta \widehat{W}_X^\top K(z_{N(T-1)}; H_2) \end{pmatrix}$$

Finally, once obtained the estimator for β , replacing the unknown quantities by their estimated objects the resulting three stage estimator for $m(\cdot)$,

3. STATISTICAL PROPERTIES

at any given value of z , is

$$\widehat{m}_{\widetilde{\beta}}(z; H_2) = (\Delta \widehat{W}_X^\top K(z; H_2) \Delta \widehat{W}_X)^{-1} \Delta \widehat{W}_X^\top K(z; H_2) (\Delta Y - \Delta \widehat{W}_U \widetilde{\beta}). \quad (2.10)$$

Note that the criterion function (2.7) stands for the local constant approximation to $m(\cdot)$. A straightforward extension would be to extend our results to the local linear case. In Section 1 of the Supplementary Material we provide all expression of the three stage estimators for this case.

3. Statistical properties

In this section, we investigate some asymptotic properties of the estimators proposed in the previous section. Under some technical assumptions provided in Section 2 of the Supplementary Material, we present their asymptotic behavior. The detailed proofs of the following results are also given in Sections 3 to 5 of the Supplementary Material. From now on, let us denote $n = N(T - 1)$.

Theorem 3.1. *Suppose that Assumptions S2.1-S2.10 hold. When $N \text{tr}(H_2)^2 \rightarrow 0$, as N tends to infinity and T is fixed, we have*

$$\sqrt{n} (\widetilde{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \Sigma^{-1} \Sigma^* \Sigma^{-1}),$$

where $\Sigma = E(\Upsilon_{it} \Upsilon_{it}^\top)$ and $\Sigma^* = 2\sigma_\epsilon^2 E(\Upsilon_{it} \Upsilon_{it}^\top) - \sigma_\epsilon^2 E(\Upsilon_{it} \Upsilon_{i(t+1)}^\top)$ with

$$\Upsilon_{it} = \Delta W_{U_{it}} - \mathcal{B}_{\Delta W_X \Delta W_U}(z, z)^\top \mathcal{B}_{\Delta W_X \Delta W_X}^{-1}(z, z) \Delta W_{X_{it}},$$

3. STATISTICAL PROPERTIES

In addition, the asymptotic normality of the three stage estimator $\widehat{m}_{\widehat{\beta}}(z; H_2)$ can be established as follows

Theorem 3.2. *Suppose that Assumptions S2.1-S2.10. As N tends to infinity and T is fixed, we have*

$$\sqrt{n|H_2|} \left(\widehat{m}_{\widehat{\beta}}(z; H_2) - m(z) - B(z; H_2)(1 + o_p(1)) \right) \xrightarrow{d} \mathcal{N}(0, V(z; H_2)),$$

where

$$B(z; H_2) = \mu_2(K) \left[\text{diag}_d(D_f(z)H_2D_{m_\kappa}(z))\iota_d f_{Z_{it}, Z_{i(t-1)}}^{-1}(z, z) + \frac{1}{2} \text{diag}_d(\text{tr}(\mathcal{H}_{m_\kappa}(z)H_2))\iota_d \right],$$

$$V(z; H_2) = 2\sigma_\epsilon^2 R^2(K) \mathcal{B}_{\Delta W_X \Delta W_X}^{-1}(z, z).$$

Moreover, $D_{m_\kappa}(z)$ is the first order derivative vector of the κ th component of $m(\cdot)$, $\mathcal{H}_{m_\kappa}(z)$ its Hessian matrix, and $D_f(z)$ the first order derivative vector of the density function, for $\kappa = 1, \dots, d$. Also, $\text{diag}_d(\text{tr}(\mathcal{H}_{m_\kappa}(z)H_2))$ and $\text{diag}_d(D_f(z)H_2D_{m_\kappa}(z))$ stands for a $(d \times d)$ diagonal matrix of elements of $\text{tr}(\mathcal{H}_{m_\kappa}(z)H_2)$ and $D_f(z)H_2D_{m_\kappa}(z)$, respectively, being ι_d a $(d \times 1)$ unitary vector.

Under the previous assumptions, the asymptotic normality of the local linear version of the three stage estimators is collected in Corollaries S1.1 and S1.2 that appear in Section 1 of the Supplementary Material.

3. STATISTICAL PROPERTIES

Comparing the results of Theorem 3.2 and Corollary S1.2 it can be noted, as expected, the best behavior, in terms of bias, of the Local Linear three stage estimator against the Nadaraya-Watson version. For other advantages see Fan and Gijbels (1995). Nevertheless, in this framework with endogenous regressors it is also necessary to take into account that the local linear estimator requires the use of three different nonparametric estimators as IV, with their corresponding bandwidths, whereas the Nadaraya-Watson only needs one. Therefore, a better performance of the Nadaraya-Watson estimator in finite samples is expected, as it is going to be shown later in the Monte Carlo experiments. That is the reason why we focus on the Nadaraya-Watson estimators throughout the paper, although all the proposed results could be extended to the local polynomial case.

Furthermore, note that the results from Theorem 3.2 show a bias term that asymptotically depends only on the smoothness of $m(\cdot)$ and $E(\Delta X_{1it}|\mathcal{L}_{it}, \mathcal{L}_{i(t-1)})$. The dependence on $\hat{\beta}$ and $\hat{E}(\Delta X_{1it}|\mathcal{L}_{it}, \mathcal{L}_{i(t-1)}; H_1)$ is negligible because $\hat{\beta}$ is \sqrt{NT} -consistent and, under the assumptions established in Section 2 of the Supplementary Material, $\hat{E}(\Delta X_{1it}|\mathcal{L}_{it}, \mathcal{L}_{i(t-1)}; H_1)$ converges uniformly to $E(\Delta X_{1it}|\mathcal{L}_{it}, \mathcal{L}_{i(t-1)})$. Finally, the dependence on $E(\Delta X_{1it}|\mathcal{L}_{it}, \mathcal{L}_{i(t-1)})$ vanishes because of condition $tr(H_1) = o_p(tr(H_2))$. As it can be also remarked, Theorem 3.2 shows a variance term that exhibits a rate of convergence that

4. MORE EFFICIENT ESTIMATORS

is suboptimal. In this smoothness class the lower rate of convergence for this type of estimators is $n|H_2|^{1/2}$ (see Härdle (1990) for details). For this reason, in the next section, and following Rodriguez-Poo and Soberon (2014) and Rodriguez-Poo and Soberon (2015), we will propose a one-step backfitting algorithm that will make the rate of convergence of our estimators optimal.

4. More efficient estimators

4.1 One-step backfitting and minimum distance estimators

In this section we first propose a one-step backfitting algorithm that will enable us to achieve optimal nonparametric rates of convergence of the estimators for $m(\cdot)$. In addition, as it will be detailed further in this section, because of the additive structure of the regression model, the backfitting procedure generates two alternative estimators for $m(\cdot)$. Nevertheless, by combining both estimators through a minimum distance estimation technique it is possible to obtain a more efficient estimator for $m(\cdot)$.

Applying the well-known one-step backfitting procedure, we propose the following three stage estimator. Assuming that $\sum_{it} K_{H_3}(Z_{it} - z) \widehat{W}_{X_{it}} \widehat{W}_{X_{it}}^\top$

4. MORE EFFICIENT ESTIMATORS

is definite positive,

$$\begin{aligned} \widehat{m}_{\widehat{\beta}}^{(1)}(z; H_3) &= \left(\sum_{i=1}^N \sum_{t=2}^T K_{H_3}(Z_{it} - z) \widehat{W}_{X_{it}} \widehat{W}_{X_{it}}^\top \right)^{-1} \\ &\quad \times \sum_{i=1}^N \sum_{t=2}^T K_{H_3}(Z_{it} - z) \widehat{W}_{X_{it}} (\Delta \widehat{Y}_{1it} - \Delta \widehat{W}_{U_{it}}^\top \widehat{\beta}), \end{aligned} \quad (4.11)$$

where $\Delta \widehat{Y}_{1it} = \Delta Y_{it} + X_{i(t-1)}^\top \widehat{m}_{\widehat{\beta}}(Z_{i(t-1)}, H_2)$, and $\widehat{m}_{\widehat{\beta}}(\cdot, H_2)$ is the estimator defined in (2.10). Note that in this case, we use $\widehat{\beta}$ instead of $\widetilde{\beta}$. In terms of asymptotics the results are the same because in both cases the rate of convergence is \sqrt{n} .

The main idea of the application of the backfitting algorithm here is to sum $X_{i(t-1)}^\top \widehat{m}_{\widehat{\beta}}(Z_{i(t-1)}, H_2)$ in both terms of the first differenced structural equation in (2.3). By doing so, the structural model is transformed into a very simple expression,

$$\Delta \widehat{Y}_{1it} = X_{it}^\top m(Z_{it}) + \Delta \widehat{\epsilon}_{1it} \quad (4.12)$$

and

$$\Delta \widehat{\epsilon}_{1it} = \Delta \epsilon_{it} + \Delta U_{it}^\top \beta + X_{i(t-1)}^\top \left(\widehat{m}_{\widehat{\beta}}(Z_{i(t-1)}; H_2) - m(Z_{i(t-1)}) \right).$$

Then, the unknown function $m(\cdot)$ in (4.12) can be estimated following the same steps as in (2.6)-(2.9) obtaining now (4.11).

Given the additive structure of (2.3), a second estimator for $m(\cdot)$ can be obtained. Assuming that $\sum_{it} K_{H_3}(Z_{i(t-1)} - z) \widehat{W}_{X_{i(t-1)}} \widehat{W}_{X_{i(t-1)}}^\top$ is a positive

4. MORE EFFICIENT ESTIMATORS

definite matrix, then an alternative backfitting estimator for $m(\cdot)$ is

$$\begin{aligned} \widehat{m}_{\widehat{\beta}}^{(2)}(z; H_3) &= \left(\sum_{i=1}^N \sum_{t=2}^T K_{H_3}(Z_{i(t-1)} - z) \widehat{W}_{X_{i(t-1)}} \widehat{W}_{X_{i(t-1)}}^\top \right)^{-1} \\ &\quad \times \sum_{i=1}^N \sum_{t=2}^T K_{H_3}(Z_{i(t-1)} - z) \widehat{W}_{X_{i(t-1)}} (\Delta \widehat{Y}_{2it} - \Delta \widehat{W}_{U_{it}}^\top \widehat{\beta}), \end{aligned} \quad (4.13)$$

where $\widehat{W}_{X_{i(t-1)}} = (\widehat{E}(X_{1i(t-1)}^\top | \mathcal{L}_{it}, \mathcal{L}_{i(t-1)}), X_{2i(t-1)}^\top)^\top$, $\Delta \widehat{Y}_{2it} = X_{it}^\top \widehat{m}_{\widehat{\beta}}(Z_{it}; H_2) - \Delta Y_{it}$ and again $\widehat{m}_{\widehat{\beta}}(\cdot, H_2)$ is the estimator defined in (2.10). Subtracting $X_{it}^\top \widehat{m}_{\widehat{\beta}}(Z_{it}, H_2)$ in both terms of (2.3) and proceeding as above we obtain (4.13).

Therefore, this technique provides two different estimators, $\widehat{m}_{\widehat{\beta}}^{(1)}(z; H_3)$ and $\widehat{m}_{\widehat{\beta}}^{(2)}(z; H_3)$, for the same $m(z)$. A natural idea to combine both in an efficient way would be to obtain such estimator by minimizing the following criterion function

$$\begin{pmatrix} \widehat{m}_{\widehat{\beta}}^{(1)}(z; H_3) - m(z) \\ \widehat{m}_{\widehat{\beta}}^{(2)}(z; H_3) - m(z) \end{pmatrix}^\top \mathcal{W}_m^{-1}(z) \begin{pmatrix} \widehat{m}_{\widehat{\beta}}^{(1)}(z; H_3) - m(z) \\ \widehat{m}_{\widehat{\beta}}^{(2)}(z; H_3) - m(z) \end{pmatrix}. \quad (4.14)$$

We propose to calculate the estimators $\widehat{m}_{\widehat{\beta}}^{(1)}(z; H_3)$ and $\widehat{m}_{\widehat{\beta}}^{(2)}(z; H_3)$ using subsamples of size $N_1 T$ and $N_2 T$ respectively, such that $N_1 + N_2 = N$ and, as $N \rightarrow \infty$, $N_j/N \rightarrow c_j$ for $c_j > 0, j = 1, 2$ and $c_1 + c_2 = 1$. These subsamples need to be chosen randomly across individuals (see Politis et al. (1999) for details). For a given value of z , the value of $m(z)$, $\widehat{m}_{\widehat{\beta}}^{(mde)}(z; H_3)$,

4. MORE EFFICIENT ESTIMATORS

that minimizes (4.14) is

$$\begin{aligned} \widehat{m}_{\widehat{\beta}}^{(mde)}(z; H_3) &= \mathcal{G}_m^{-1}(z) \left\{ \mathcal{J}_{1m}(z) \widehat{m}_{\widehat{\beta}}^{(1)}(z; H_3) \right. \\ &\quad \left. + \mathcal{J}_{2m}(z) \widehat{m}_{\widehat{\beta}}^{(2)}(z; H_3) \right\} \end{aligned} \quad (4.15)$$

where

$$\mathcal{G}_m(z) = \mathcal{W}_m^{11}(z) + 2\mathcal{W}_m^{12}(z) + \mathcal{W}_m^{22}(z), \quad (4.16)$$

$$\mathcal{J}_{1m}(z) = \mathcal{W}_m^{11}(z) + \mathcal{W}_m^{12}(z), \quad (4.17)$$

$$\mathcal{J}_{2m}(z) = \mathcal{W}_m^{22}(z) + \mathcal{W}_m^{12}(z), \quad (4.18)$$

and the $(d \times d)$ matrix $\mathcal{W}_m^{ij}(z)$ is the (i, j) -th component of the block partitioned matrix $\mathcal{W}_m^{-1}(z)$, for $i, j = 1, 2$. This estimator belongs to the class of the so-called minimum distance estimators. In order to select the weighting matrix, $\mathcal{W}_m(z)$ many alternatives are available. We choose the matrix, $\mathcal{W}_m^*(z)$, that minimizes the asymptotic variance-covariance matrix of $\widehat{m}_{\widehat{\beta}}^{(mde)}(z; H_3)$. Indeed, in Hansen (1982) it is shown that $\mathcal{W}_m^*(z) = V^{(1+2)}(z)$, where $V^{(1+2)}(z)$ stands for the asymptotic variance-covariance matrix of

$$\left(\left(\widehat{m}_{\widehat{\beta}}^{(1)}(z_3; H_3) - m(z) \right)^\top, \left(\widehat{m}_{\widehat{\beta}}^{(2)}(z_3; H_3) - m(z) \right)^\top \right)^\top.$$

We now proceed to analyze the asymptotic properties of both the backfitting and the minimum distance estimator.

4. MORE EFFICIENT ESTIMATORS

4.2 Asymptotic properties

The following theorems present the limiting distribution of the backfitting estimators. Note that it achieves the optimal rate of convergence for this smoothness class.

Theorem 4.1. *Suppose that Assumptions S2.1-S2.12 hold. As $N \rightarrow \infty$ and T is fixed, we have*

$$\begin{aligned} \sqrt{N|H_3|^{1/2}} \left(\widehat{m}_{\widehat{\beta}}^{(j)}(z; H_3) - m(z) - B(z; H_3)(1 + o_p(1)) \right) \\ \xrightarrow{d} \mathcal{N} \left(0, V^{(j)}(z; H_3) \right), \end{aligned}$$

where

$$\begin{aligned} B(z; H_3) = \mu_2(K) \left(\text{diag}_d(D_f(z)H_3D_{m_\kappa}(z)) \iota_d f_{Z_{it}}^{-1}(z) \right. \\ \left. + \frac{1}{2} \text{diag}_d(\text{tr}(\mathcal{H}_{m_\kappa}(z)H_3)) \iota_d \right), \end{aligned}$$

for $j = 1$ and $j = 2$,

$$V^{(1)}(z; H_3) = 2\sigma_\epsilon^2 R(K) \mathcal{B}_{W_X W_X}^{-1}(z),$$

$$V^{(2)}(z; H_3) = 2\sigma_\epsilon^2 R(K) \mathcal{B}_{W_{X-1} W_{X-1}}^{-1}(z).$$

The proof of this result is postponed to the Supplementary Material.

We focus now on the asymptotic properties of the minimum distance estimator, $\widehat{m}_{\widehat{\beta}}^{(mde)}(z; H_3)$, getting the following result

Theorem 4.2. *Suppose that Assumptions S2.1-S2.12 hold. As $N \rightarrow \infty$ and T is fixed, we have*

$$\begin{aligned} \sqrt{N|H_3|^{1/2}} \left(\widehat{m}_{\widehat{\beta}}^{(mde)}(z; H_3) - m(z) - B(z; H_3)(1 + o_p(1)) \right) \\ \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_m(z)), \end{aligned}$$

where

$$\mathcal{V}_m(z) = \sigma_\epsilon^2 R(K) \left(\mathcal{B}_{W_X W_X}(z) + \mathcal{B}_{W_{X-1} W_{X-1}}(z) \right)^{-1}. \quad (4.19)$$

The proof of this result is relegated to the Supplementary Material. Finally, focusing on Theorem 4.2 it can be highlighted that the asymptotic bias of the minimum distance estimator is the same as in Theorem 4.1. Moreover, the asymptotic variance exhibits the optimal rate of convergence of this type of problems. Finally, note that it is easy to show that for any vector $b \neq 0$, $b^\top \mathcal{V}_m(z) b \leq b^\top V^{(j)}(z) b$, for $j = 1, 2$. Therefore, it is proved that this technique enables us to obtain more efficient estimators for $m(\cdot)$ and, at the same time, to achieve optimality.

5. Inference

The statistical model that is of our concern, see equations (2.1) and (2.2), is considerably rich and it nests many models of interest in Econometrics and Statistics. For example, it is natural to investigate whether certain variables

5. INFERENCE

in this component are statistically significant after fitting the model. More generally one might consider the set of linear hypothesis

$$H_0 : F\beta = C \quad \text{versus} \quad H_1 : F\beta \neq C,$$

where F is a $(Q \times k)$ full rank matrix with $Q \leq k$, C is a $(Q \times 1)$ vector, and Q is the number of hypotheses on the null. Indeed, using Theorem 3.1, this testing problem can be handled by using the following Wald-type test statistic

$$W_n(F, C) = (F\tilde{\beta} - C)^\top [F\hat{\Sigma}^{-1}\hat{\Sigma}^*\hat{\Sigma}^{-1}F^\top]^{-1}(F\tilde{\beta} - C).$$

Following (2.9) and lemma S3.2 it is easy to show that

$$\begin{aligned} \hat{\Sigma} &= \frac{1}{n} \Delta \widehat{W}_U^\top (I_n - \hat{S})^\top (I_n - \tilde{S}) \Delta U, \\ \hat{\Sigma}^* &= \frac{1}{n} \Delta \widehat{W}_U^\top (I_n - \hat{S})^\top (I_n - \tilde{S}) \hat{V} (I_n - \tilde{S})^\top (I_n - \hat{S}) \Delta \widehat{W}_U, \end{aligned}$$

are consistent estimators of Σ and Σ^* respectively. For \hat{V} we propose

$$\hat{V} = \hat{\sigma}_\epsilon^2 \begin{pmatrix} -1 & 2 & -1 & 0 & \cdots & \cdots \\ 0 & -1 & 2 & -1 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & 0 & -1 & 2 & -1 \end{pmatrix},$$

where $\hat{\sigma}_\epsilon^2 = \frac{1}{NT} \sum_{it} \Delta \hat{\epsilon}_{it}^2$, with

$$\Delta \hat{\epsilon}_{it} = \Delta Y_{it} - \Delta \widehat{W}_{U_{it}}^\top \tilde{\beta} - \widehat{W}_{X_{it}}^\top \hat{m}_{\tilde{\beta}}(Z_{it}; H_2) + \widehat{W}_{X_{i(t-1)}}^\top \hat{m}_{\tilde{\beta}}(Z_{i(t-1)}; H_2).$$

6. MONTE CARLO EXPERIMENT

This is because of Assumption S2.2 and the first difference structure of the model. The level of the test is given by the following result that is proved in the Supplementary material.

Corollary 5.1. *Suppose that Assumptions S2.1-S2.10 hold. When $Ntr(H_2)^2 \rightarrow 0$, as $N \rightarrow \infty$ and T is fixed, under the null hypothesis we have*

$$W_n(F, C) \xrightarrow{d} \chi_Q^2,$$

where χ_Q^2 denotes a chi-square distribution with degrees of freedom Q .

From a nonparametric point of view, we can be interested in the construction of pointwise confidence interval for $m(\cdot)$ for each given point z . In Section 1 of the Supplementary Material we provide confidence bands for all three stage estimators based in the local linear version.

6. Monte Carlo experiment

To assess the finite sample properties of the different estimators and statistical tests proposed in this paper, some Monte Carlo simulations are performed. To this end, we consider the following data generating process (DGP)

$$Y_{it} = X_{1it}m_1(Z_{it}) + U_{1it}\beta_1 + U_{2it}\beta_2 + \mu_i + v_{it}, \quad (6.20)$$

$$i = 1, \dots, N \quad ; \quad t = 1, \dots, T,$$

6. MONTE CARLO EXPERIMENT

where the coefficients $m_1(Z_{it}) = (1.6 + 0.6Z_{it})\exp(-0.4(Z_{it} - 3)^2)$, $\beta_1 = -1$, and $\beta_2 = 1$. The smoothing variable Z_{it} follows a uniform $[2, 6]$ distribution, U_{2it} is exogenous following a $\mathcal{N}(0, 1)$ distribution, whereas X_{1it} and U_{1it} are some endogenous variables following the reduced form equations:

$$X_{1it} = (0.5 + \sin^2(Z_{it}))V_{1it} + \zeta_{1it}, \quad \text{and} \quad U_{1it} = (0.5 + \cos^2(Z_{it}))V_{2it} + \zeta_{2it},$$

where V_{1it} and V_{2it} are instrumental variables independently generated from a uniform $[0, 4]$ distribution and the noises follow

$$\begin{pmatrix} \epsilon_{it} \\ \zeta_{1it} \\ \zeta_{2it} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_\epsilon & \rho\sigma_\epsilon \\ \rho\sigma_\epsilon & \sigma_\epsilon^2 & 0 \\ \rho\sigma_\epsilon & 0 & \sigma_\epsilon^2 \end{pmatrix} \right),$$

where ρ controls the correlation between the residues in the structural equation and in the reduced form equation, and σ_ϵ controls the variation of residues in the reduced form equation. Further, to allow for the presence of heterogeneity in the form of fixed effects, we generate $\mu_i = 0.5\bar{Z}_i + \zeta_i$, where ζ_i is an *i.i.d.* $\mathcal{N}(0, 1)$ random variable and $\bar{Z}_i = T^{-1} \sum_{t=1}^T Z_{it}$.

In order to check the performance of the proposed estimators we set $\rho = 0.7$, $\sigma_\epsilon^2 = 1$, and conduct simulations by considering that the number of periods T equal to 4, while the cross-sections N is 100, 200, and 400. For each sample size, we replicate the experiment 1000 times. For $K(u)$ we choose the Epanechnikov kernel function $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$.

6. MONTE CARLO EXPERIMENT

To meet the requirement that $Ntr(H_2)^2 \rightarrow 0$, we assume $H_2 = h_2 I_q$ and fix the bandwidth for estimating β at three values $h_2 = 1.25N^{-1/3}, 2.5N^{-1/3}$ and $5N^{-1/3}$. Because of the need for undersmoothing in the first stage for asymptotic reasons, required by Assumption S2.8, we set the first stage bandwidth H_1 to be 0.8 times the second stage one H_2 , i.e., $H_1 = 0.8H_2$. For the sake of comparison we analyze the finite sample behavior of the following estimators: $\widehat{\beta}_E$ is the estimator proposed in (2.8) with ΔU instead of $\Delta \widehat{W}_U$ (i.e., when the endogeneity problem has not been solved); $\widehat{\beta}_{NF}$ is the $\widehat{\beta}$ estimator with ΔW_U instead of $\Delta \widehat{W}_U$ (i.e., the nonfeasible estimator); $\widehat{\beta}_F$ is the estimator proposed in (2.8); and $\widetilde{\beta}_F$ is the estimator proposed in (2.9).

In Table 1 we report the means, standard deviation, and root mean squared error (RMSE) of the estimated 1000 values for β under different settings. Analyzing these results, we find that the performance of all these estimators is not sensitive to the choice of bandwidth. All estimators give a similar asymptotically unbiased estimation for β_2 because there is no endogeneity involved in this parameter. As the sample size increase, all of them converge. However, a totally different phenomenon exhibits for β_1 (i.e., the parameter related to the endogenous variable X_1). As it is expected, $\widehat{\beta}_{NF}$ present the best results, but both $\widehat{\beta}_F$ and $\widetilde{\beta}_F$ performs quite

6. MONTE CARLO EXPERIMENT

well when they deal with the endogeneity problem. Although both methods have similar standard deviations, a large bias is resulted from $\tilde{\beta}_F$ and $\hat{\beta}_F$ exhibits the lower RMSE.

In order to verify the asymptotic results analyzed in the previous sections for the functional coefficient, now we compare the finite sample behavior of different nonparametric estimators, where $\hat{\beta}$ is used as a \sqrt{n} -consistent estimator of β and the bandwidth in this stage $H_3 = h_3 I_q$ is chosen by the Silverman's rule-of-thumb, i.e., $h_3 = 1.06\hat{\sigma}_Z N^{-1/5}$, where $\hat{\sigma}_Z$ is the sample standard deviation of Z_{it} . In addition, to meet the requirement that H_1 and H_2 have to be chosen undersmoothed for asymptotic reasons, we set $h_2 = 1.75\hat{\sigma}_Z N^{-1/3}$ and $h_1 = 1.25\hat{\sigma}_Z N^{-1/3}$. As a measure of accuracy, we use the following RMSE

$$RMSE(\hat{m}(z; H)) = \left\{ \frac{1}{R} \sum_{\varphi=1}^R \frac{1}{n} \sum_{i=1}^N \sum_{t=2}^T (\hat{m}(z; H) - m(z))^2 \right\}^{1/2},$$

where φ is the φ -th replication and R is the number of replications.

For the sake of comparison, Figure 1 depicts boxplots of the 1,000 RMSE values of the functional coefficient estimators using different types of estimators proposed previously. To show the effect of the generated regressors in constructing the feasible estimators, Figure 1(a) collects the results for the Nadaraya-Watson (NW) estimator without the adjustment for endogeneity and Figure 1(b) collects the results for the nonfeasible NW

6. MONTE CARLO EXPERIMENT

Table 1: Mean, Standard Deviations and RMSE's of the different estimators for β_1 and β_2 .

		H_2	$\hat{\beta}_{E_1}$	$\hat{\beta}_{NF_1}$	$\hat{\beta}_{F_1}$	$\tilde{\beta}_{F_1}$	$\hat{\beta}_{E_2}$	$\hat{\beta}_{NF_2}$	$\hat{\beta}_{F_2}$	$\tilde{\beta}_{F_2}$
N=100	Mean	$1.25N^{-1/3}$	-0.6113	-0.8045	-0.7150	-0.6472	1.0232	1.0323	1.0271	1.0242
		$2.5N^{-1/3}$	-0.5923	-0.7826	-0.8112	-0.6660	0.9963	1.0018	1.0017	0.9997
		$5.0N^{-1/3}$	-0.5668	-0.7785	-1.0820	-0.7235	0.9837	0.9739	0.9922	0.9843
	SD	$1.25N^{-1/3}$	0.0821	0.2023	0.0971	0.0920	0.0821	0.2024	0.0971	0.0920
		$2.5N^{-1/3}$	0.0890	0.1911	0.1309	0.1308	0.1911	0.1309	0.1308	0.1140
		$5.0N^{-1/3}$	0.0880	0.1879	0.1855	0.1297	0.0880	0.1879	0.1855	0.1297
	RMSE	$1.25N^{-1/3}$	0.3973	0.2814	0.3010	0.3646	0.1466	0.2814	0.1503	0.3642
		$2.5N^{-1/3}$	0.4173	0.2894	0.2297	0.3529	0.1333	0.2894	0.1496	0.3528
		$5.0N^{-1/3}$	0.4420	0.2905	0.2028	0.3054	0.1396	0.2905	0.1617	0.3054
N=200	Mean	$1.25N^{-1/3}$	-0.6049	-0.7829	-0.7539	-0.6560	0.9829	0.9660	0.9797	0.9815
		$2.5N^{-1/3}$	-0.6034	-0.8011	-0.8620	-0.6867	0.9988	1.0010	1.0010	0.9971
		$5.0N^{-1/3}$	-0.5815	-0.7871	-1.0809	-0.7386	0.9968	0.9928	0.9897	0.9866
	SD	$1.25N^{-1/3}$	0.0534	0.1473	0.0676	0.0614	0.0534	0.1473	0.0676	0.0904
		$2.5N^{-1/3}$	0.0572	0.1482	0.0898	0.0898	0.1482	0.0898	0.0898	0.1103
		$5.0N^{-1/3}$	0.0636	0.1425	0.1426	0.0969	0.0636	0.1424	0.1426	0.1144
	RMSE	$1.25N^{-1/3}$	0.3987	0.2623	0.2552	0.3494	0.1058	0.2083	0.1073	0.1110
		$2.5N^{-1/3}$	0.4006	0.2480	0.1646	0.3218	0.1060	0.1958	0.1153	0.1163
		$5.0N^{-1/3}$	0.4234	0.2562	0.1639	0.2787	0.1018	0.1746	0.1207	0.1152
N=400	Mean	$1.25N^{-1/3}$	-0.6203	-0.8089	-0.8242	-0.6901	0.9977	0.9827	0.9943	0.9959
		$2.5N^{-1/3}$	-0.6097	-0.7963	-0.9292	-0.7174	0.9978	0.9969	0.9968	0.9919
		$5.0N^{-1/3}$	-0.5912	-0.7748	-1.1138	-0.7646	0.9960	1.0006	0.9937	0.9973
	SD	$1.25N^{-1/3}$	0.0405	0.1068	0.0591	0.0534	0.0525	0.1151	0.0686	0.0523
		$2.5N^{-1/3}$	0.0350	0.1171	0.0606	0.0491	0.1012	0.0607	0.0820	0.0492
		$5.0N^{-1/3}$	0.0369	0.0881	0.0835	0.0577	0.0753	0.1198	0.0882	0.0577
	RMSE	$1.25N^{-1/3}$	0.3818	0.2189	0.1854	0.3143	0.0585	0.1164	0.0689	0.0698
		$2.5N^{-1/3}$	0.3919	0.2275	0.0932	0.2868	0.0684	0.1179	0.0821	0.0795
		$5.0N^{-1/3}$	0.4104	0.2419	0.1411	0.2423	0.0754	0.1198	0.0884	0.0827

6. MONTE CARLO EXPERIMENT

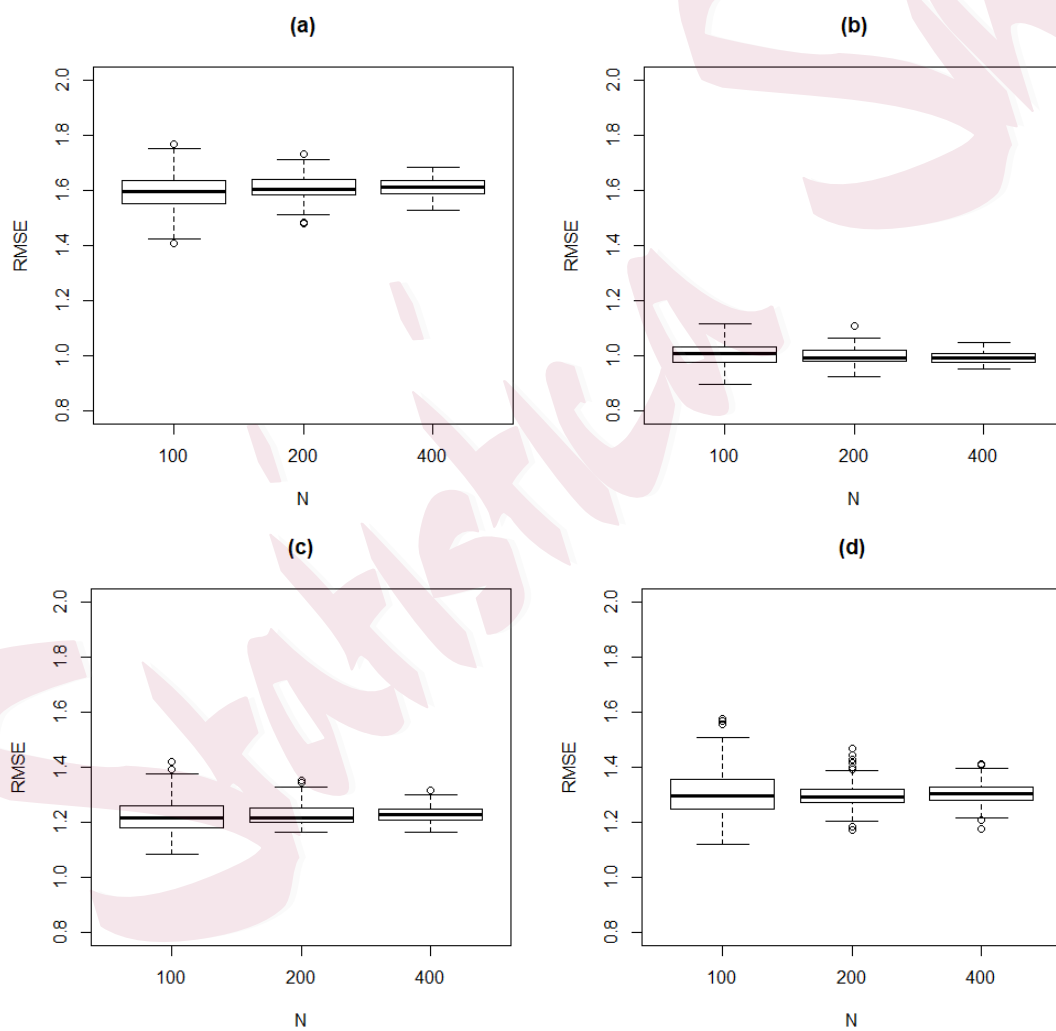
estimator with ΔW_X instead of $\Delta \widehat{W}_X$. As it is expected, in Figure 1(b) all RMSE values of the nonfeasible NW estimator converge toward zero, but this is not true for the NW estimator with endogeneity presented in Figure 1(a). In addition, for the sake of comparison between the NW and the Local linear estimator, Figure 1(c) reflects the results of $\widehat{m}_{\widehat{\beta}}(z; H_2)$ and Figure 1(d) displays the result of the $\widehat{m}_{\widehat{\beta}, LL}(z, H_2)$. As the sample size increases, it can be seen that the RMSE of both estimators shrinks towards zero. However, as it was pointed out in the theoretical results, the NW estimator performs slightly better than the local linear estimator since only one generated regressor is necessary.

In order to check the efficiency improvement of the backfitting estimators, Figure 2 reports the RMSE of the backfitting estimators relative to the NW estimator. Specifically, Figure 2(a) collects the results for $\widehat{m}_{\widehat{\beta}}^{(1)}(z; H_3)$ while Figure 2(b) depicts the results for $\widehat{m}_{\widehat{\beta}}^{(2)}(z; H_3)$. In addition, Figure 3 collects the RMSE of the minimum distance estimator relative to the NW estimator. Looking at these results we find that all figures corroborate the efficiency improvement since their relative RMSE shrink towards zero. Therefore, we can conclude that the findings obtained in the theoretical part are confirmed by the simulation study

Finally, we study the size and power performance of testing $H_0 : \beta_1 =$

6. MONTE CARLO EXPERIMENT

Figure 1: The boxplots of the RMSE values of the three stage nonparametric estimates in 1000 independent simulations for three sample sizes $N = 100, 200, 400$ and $T = 4$.



6. MONTE CARLO EXPERIMENT

Figure 2: The boxplots of the nonparametric backfitting estimates in 1000 independent simulations for three sample sizes $N = 100, 200, 400$ and $T = 4$.

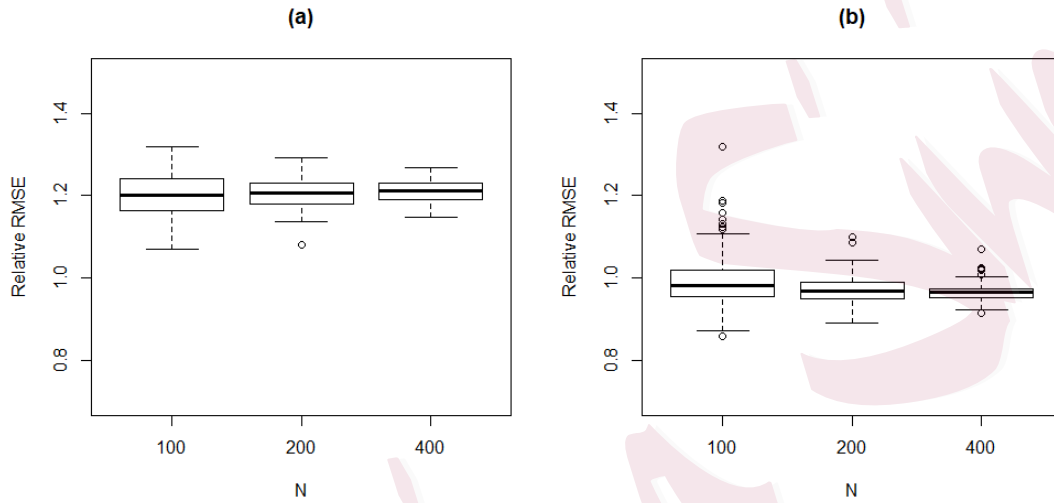
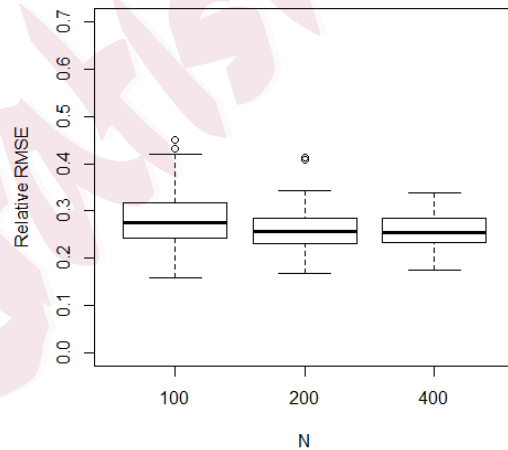


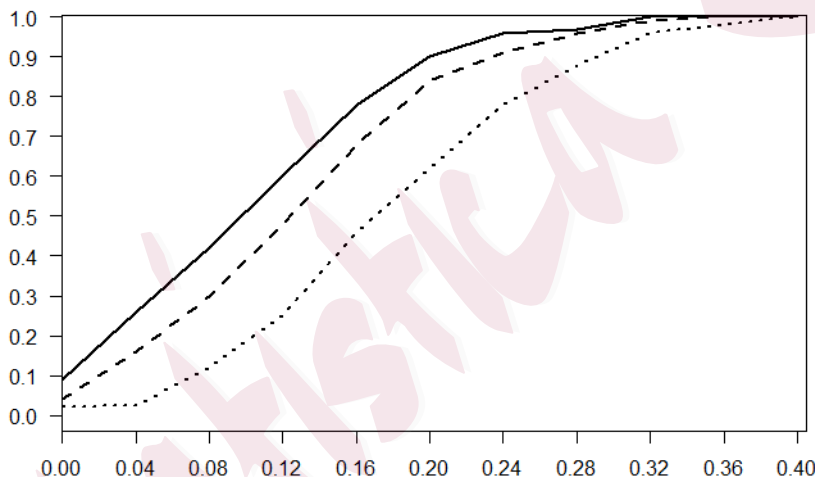
Figure 3: The boxplots of the minimum distance estimator in 1000 independent simulations for three sample sizes $N = 100, 200, 400$ and $T = 4$.



6. MONTE CARLO EXPERIMENT

$-1, \beta_2 = 1$ against the alternative hypothesis $H_1 : \beta_1 = -1 + \phi_1, \beta_2 = 1$, where the power is indexed by ϕ_1 . To this end, we use the Wald test proposed previously for sample size $T = 4$ and $N = 200$ and we conduct 1000 Monte Carlo simulations. Also, the bandwidths used here were $h_2 = 2.5N^{-1/3}$ and $h_1 = 0.8 * h_2$. Figure 4 plots the power curves for three significance levels.

Figure 4: The power curves for sample size $T = 4$ and $N = 200$.



Note: The dotted line is the power curve for 1% significance level, the dashed line and the solid line are for 5% and 10% significance levels, respectively.

From Figure 4, it can be pointed out that when $\phi_1 = 0$, the power collapses to the test size. More precisely, the simulated sizes of the proposed test are 9%, 3% and 2% corresponding to the significance levels 1% (dotted

7. EMPIRICAL RESULTS

line), 5% (dashed line), and 10% (solid line), respectively. Therefore, it is proved that the simulated sizes are close to the nominal size so our test can deliver a correct test size. On the contrary, when ϕ_1 deviates from 0 our test is reasonably powerful since the power curves tend to 1 quickly. Specifically, the power is over 90% for both significance levels 10% and 5% when $\phi_1 \geq 0.22$.

7. Empirical results

With the aim of displaying the usefulness of the proposed method, in this section we analyze the impact of unexpected health expenses on households savings. Along with liquidity constraints and habits in consumer preferences, uncertainty about possible economic hardships and household risk aversion are key determinants of household's consumption/savings decisions; see Friedman (1957). In this situation, precautionary savings appears as a protection tool for the individuals against either potential income downturns or unforeseen out-of-pocket medical expenses in latter stages of life, see Chou et al. (2004) for a further discussion.

To do it, we propose to extend the analysis in Chou et al. (2004) and estimate the following regression model

$$Y_{it} = X_{1it}m_1(Z_{it}) + X_{2it}m_2(Z_{it}) + \mu_i + v_{it}, \quad (7.21)$$

7. EMPIRICAL RESULTS

where i index the household, t the time, Z_{it} is the age of the household head, X_{1it} the health-care expenditures (log), Y_{it} are the savings, and X_{2it} the permanent income (log). In this sense, household savings are characterized by the uncertainty about both future health-care expenses, $m_1(\cdot)$, and income downturns, $m_2(\cdot)$. Note that household permanent income is not directly observable. In order to approximate this variable, we follow the proposal in Chou et al. (2004). Thus, assuming that the interest rate equals to the productivity rate of growth and 65 years old is the maximum age at which people works, the permanent earnings at age τ_0 is calculated as

$$X_2(\tau_0) = X_3^\top \beta + (65 - \tau_0 + 1)^{-1} \sum_{\tau=\tau_0}^{65} f(\tau),$$

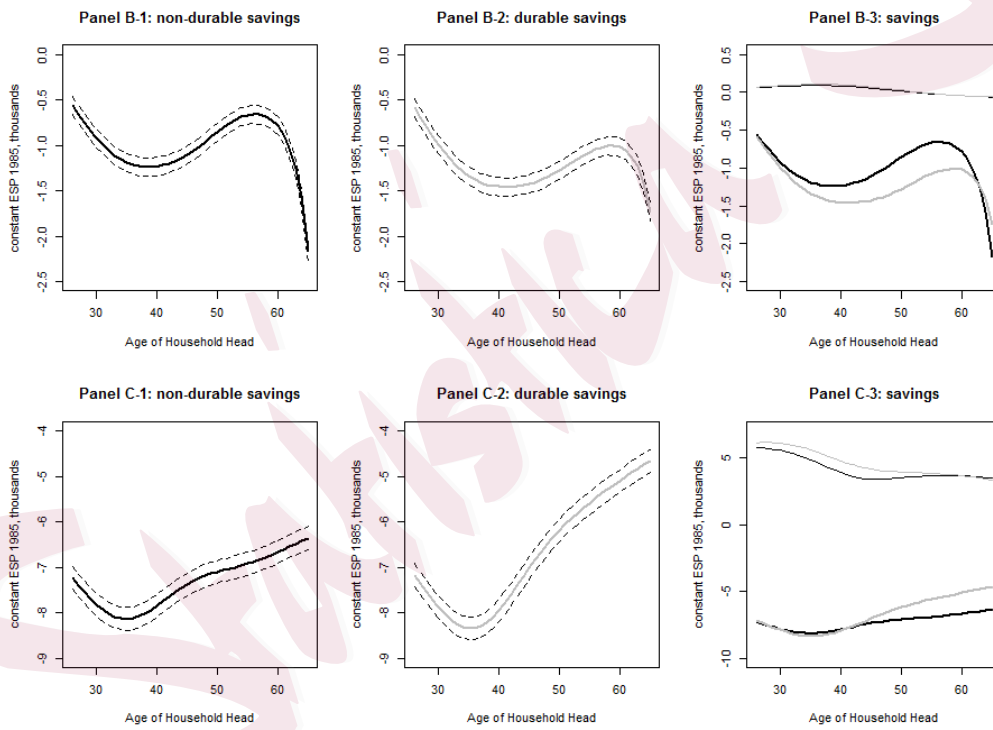
where $f(\tau)$ is the estimated quadratic function of age, Y_{it} the household income and X_3 a vector of demographic characteristics.

Choosing the bandwidths as in the simulation study, estimation results are shown in Figure 5. The estimated curves are plotted against the age variable jointly with 95% pointwise confidence intervals calculated adapting the wild bootstrap technique of Härdle et al. (2004) to this context. Figure 5 is divided into two panels, B and C, which in turn are split up into three graphics. Panels B exhibit the corresponding elasticity to changes in health-care expenditures, i.e., $\tilde{m}_1(\cdot)$, whereas Panels C show the precau-

7. EMPIRICAL RESULTS

tionary savings elasticity to changes in household income; i.e., $\tilde{m}_2(\cdot)$. In addition, Panel B-1 shows the estimated curves when durable goods are not taken into account. Panel B-2 focuses on the second definition of savings, whereas Panel B-3 compares the estimated curves when endogeneity is not considered. This structure is maintained for Panels C. Focusing on

Figure 5: Household savings over the life-cycle



Notes: Thick line denotes the estimates for durable savings, continuous line for non-durable savings while dotted line is the 95% pointwise confidence interval.

the results in Figure 5, it can be noted that when we control for uncertainty

7. EMPIRICAL RESULTS

about health-care expenditures (Panels B) younger households (26-33) exhibit a declining savings rate, following by a constant path till the age of 40, where the hump-shaped appears again. In addition, if these results are combined with the delay in the wealth accumulation process of the Spanish households (note that in the U.S. it begins around 40 age whereas in Spain at 45 age), we realize the negative impact that public health programs have on precautionary savings, confirming the results in Chou et al. (2004). Comparing the behavior of the elasticities for the different savings results, it is obtained that consumption in durable goods reacts more to unexpected changes in income, whereas consumptions in non-durable goods is more sensitive to potential health-care payout. This holds specially for households over 45 years old.

Finally, in order to evaluate the empirical relevance of the endogeneity problem we compare the results of our technique (line grey) against those obtained without considering endogeneity (line black); see Panels 3 of Figure 5. By looking at these results, there are some significant differences. When we control for uncertainty about health-care expenditures, households accumulate assets in the middle of their life, whereas when endogeneity is not taken into account there is a more or less constant path over the life cycle.

8. Conclusion

This paper is concerned with the nonparametric estimation of a structural panel data varying coefficient model, where the individual heterogeneity is allowed to be correlated with some explanatory variables. This specification is rather frequent nowadays in many standard econometric applications such as the study of household consumption behavior or labor supply analysis. Therefore, it is of interest to have available estimators that, at the same time, keep a reasonable degree of flexibility and are robust to both endogeneity and fixed effects. Trying to satisfy these requirements, in this paper a nonparametric three-stage procedure is developed where IV techniques are used to deal with endogeneity, and differencing techniques are used to cope with fixed effects. Furthermore, to achieve efficiency, a minimum distance estimator is proposed. The feasibility and possible gains of this new procedure are shown by estimating a LCH panel data model and simulation results support the empirical findings.

Acknowledgements

The authors wish to thank two anonymous referees for their very helpful comments and suggestions. The usual disclaimer applies. The authors gratefully acknowledge financial support from the Programa Estatal de Fomento

de la Investigación Científica y Técnica de Excelencia/Spanish Ministry of Economy and Competitiveness. Ref. ECO2016-76203-C2-1-P. In addition, this work is part of the Research Project APIE 1/2015-17: "New methods for the empirical analysis of financial markets" of the Santander Financial Institute (SANFI) of UCEIF Foundation resolved by the University of Cantabria and funded with sponsorship from Banco Santander.

References

- Ai, C. and X. Chen (2003). "Efficient estimation of models with conditional moment restrictions containing unknown functions". *Econometrica* 71, 1795–1843.
- Arellano, M. (2003). *Panel Data Econometrics*. Oxford University Press.
- Blundell, R., D. Kristensen, and R. Matzkin (2013). "Control functions and simultaneous equations methods". *The American Economic Review* 103, 563–569.
- Cai, Z., M. Das, H. Xiong, and X. Wu (2006). "Functional-coefficient instrumental variables models". *Journal of Econometrics* 133, 207–241.
- Cai, Z., Y. Fang, M. Lin, and J. Su (2017). "Inferences for a partially varying coefficient model with endogenous regressors". *Journal of Business & Economic Statistics* 35, 1–13.
- Cai, Z. and Q. Li (2008). "Nonparametric estimation of varying coefficient dynamic panel data models". *Econometric Theory* 24, 1321–1342.
- Cai, Z. and H. Xiong (2012). "Partially varying coefficient instrumental variables models".

REFERENCES

- Statistica Neerlandica* 66, 85–110.
- Chamberlain, G. (1992). “Efficiency bounds for semiparametric regression”. *Econometrica* 60, 567–596.
- Chou, S. Y., J. Liu, and C. Huang (2004). “Health insurance and savings over the life cycle: a semiparametric smooth coefficient estimation”. *Journal of Applied Econometrics* 19, 295–322.
- Darolles, S., Y. Fan, and J. P. Florens (2004). “Nonparametric instrumental variables”. *Econometrica* 79, 1541–1565.
- Fan, J. and I. Gijbels (1995). *Local Polynomial Modelling and its Applications*. Chapman & Hall.
- Fan, J. and T. Huang (2005). “Profile likelihood inferences on semiparametric varying-coefficient partially linear models”. *Bernoulli* 11, 1031–1057.
- Fan, J. and W. Zhang (1999). “Statistical estimation in varying coefficient models”. *Annals of Statistics* 27, 1491–1518.
- Fève, F. and J. Florens (2014). “Non parametric analysis of panel data models with endogenous variables”. *Journal of Econometrics* 181, 151–164.
- Friedman, M. (1957). *A Theory of the Consumption Function*. Princeton: Princeton University Press.
- Gao, Y. and P. C. B. Phillips (2013). “Semiparametric estimation in triangular system equations

REFERENCES

- with nonstationarity". *Journal of Econometrics* 176, 59–79.
- Gourinchas, P. and J. Parker (2002). "Consumption over the life cycle". *Econometrica* 70, 47–89.
- Hall, P. and J. L. Horowitz (2005). "Nonparametric methods for inference in the presence of instrumental variables". *Annals of Statistics* 33, 2904–2929.
- Hansen, L. (1982). "Large sample properties of the Generalized Method of Moments estimators". *Econometrica* 50, 1029–1054.
- Härdle, W. (1990). *Applied nonparametric regression*. Cambridge University Press.
- Härdle, W., S. Huet, E. Mammen, and S. Sperlich (2004). "Bootstrap inference in semiparametric generalized additive models". *Econometric Theory* 20, 265–300.
- Heckman, J. J. (2008). "Econometric causality". *International Statistical Review* 76, 1–27.
- Heckman, J. J. and R. Robb (1985). "Alternative methods for evaluating the impact of interventions". In J. J. Heckman and B. S. Singer (Eds.), *Longitudinal Analysis of Labor Market Data*, Econometric Society Monographs, pp. 156–246. Cambridge University Press.
- Hsiao, C. (2003). *Analysis of Panel Data* (2nd ed.). Econometric Society Monographs.
- Kuan, C. and C. Chen (2013). "Effects of national health insurance on precautionary saving: new evidence from Taiwan". *Empirical Economics* 44, 942–943.
- Lee, Y. and D. Mukherjee (2014). "New nonparametric estimation of the marginal effects in fixed-effects panel models: an application on the environmental Kuznets curve". Working

REFERENCES

paper.

Mundra, K. (2005). “Nonparametric slope estimation for fixed-effect panel data”. Working paper, Department of Economics, San Diego State University.

Newey, W. (2013). “Nonparametric instrumental variable estimation”. *American Economic Review: Papers & Proceedings* 103, 550–556.

Newey, W. and J. Powell (2003). “Instrumental variable estimation of nonparametric models”. *Econometrica* 71, 1565–1578.

Politis, D., J. P. Romano, and M. Wolf (1999). *Subsampling*. Springer Series in Statistics. Springer, New York, NY.

Rodriguez-Poo, J. M. and A. Soberon (2014). “Direct semi-parametric estimation of fixed effects panel data varying coefficient models”. *Econometrics Journal* 17, 107–138.

Rodriguez-Poo, J. M. and A. Soberon (2015). “Nonparametric estimation of fixed effects panel data varying coefficient models”. *Journal of Multivariate Analysis* 133, 95–122.

Rodriguez-Poo, J. M. and A. Soberon (2017). “Nonparametric and semiparametric panel data models: recent developments”. *Journal of Economic Surveys* 31, 923–960.

Su, L. and A. Ullah (2008). “Local polynomial estimation of nonparametric simultaneous equation models”. *Journal of Econometrics* 144, 193–218.

Departamento de Economía. Universidad de Cantabria. Avda. de los Castros s/n. 39005. Santander (Spain)

REFERENCES

E-mail: juan.rodriguez@unican.es

Departamento de Economía. Universidad de Cantabria. Avda. de los Castros s/n. 39005.

Santander (Spain)

E-mail: alexandra.soberon@unican.es

Statistica Sinica