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

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

Statistica Sinica

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

Juan M. Rodriguez-Póo and Alexandra Soberón

Departamento de Economia, Universidad de Cantabria

Abstract:

This study estimates a fixed effects panel data model that adopts a partially linear form: the coefficients of some variables are restricted to be constant, but the coefficients of other variables are assumed to vary depending on some exogenous continuous variables. Moreover, we allow for endogeneity in the structural equation. Conditional moment restrictions are imposed on the first-differences to identify the structural equation. Based on these restrictions, we propose a three-stage estimation procedure, and establish the asymptotic properties of these proposed estimators. Moreover, from the first-differences transformation, we obtain two alternative backfitting estimators to estimate the unknown varying coefficient functions. As a novel contribution, we propose a minimum distance estimator that combines both estimators and, thus, is more efficient and achieves the optimal rate of convergence. The feasibility and benefits of this new procedure are shown by estimating a life-cycle hypothesis panel data model and implementing a Monte Carlo study.

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

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

1. Introduction

Two of the most important issues faced by econometricians when modeling individual choice in demand systems or a market equilibrium are the presence of endogenous variables and individual heterogeneity (see Heckman (2008)). Traditionally, instrumental variable (IV) models have been used to account for endogeneity, whereas heterogeneity is often handled using panel data techniques (e.g., see Arellano (2003)). IV models are also popular when dealing with both issues at the same time in a panel data analysis (e.g., see Hsiao (2003), Chapter 5). In many situations, economic theory does not imply tight functional form specifications for IV models, in which case, it is useful to consider nonparametric and semiparametric extensions. Unfortunately, including such flexible specifications introduces the curse of dimensionality (see Härdle (1990)). Varying coefficient models offer one solution to this problem. These models are clearly motivated by economic theory (see Chamberlain (1992)), encompass many alternative models (e.g., fully nonparametric models and partially linear models); and avoid the so-called ill-posed inverse problem in general nonparametric IV models (see Newey and Powell (2003)).

This study estimates a fixed effects panel data model that adopts a partially linear form. That is, the coefficients of some variables are re-

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

stricted to be constant, but others are assumed to vary depending on some exogenous continuous variables. Moreover, we allow for endogeneity in the structural equation. This structure leads naturally to a semiparametric three-stage estimation procedure based on a transformed (first-order differenced) structural model. In the first stage, endogenous variables are projected onto a set of IVs. In the second stage, constant coefficients are estimated using a profile least squares approach. 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. Following Fan and Zhang (1999), we improve this rate using a one-step backfitting procedure. The resulting estimator is oracle efficient and exhibits an optimal rate of convergence. However, as a result of the first differences transformation we obtain two alternative backfitting estimators for the same unknown function of the varying parameters. To improve the efficiency, we combine the estimators using a minimum distance estimation technique. To the best of our knowledge, this approach and the minimum distance estimation technique applied to this problem are new in the literature.

To avoid the ill-posed inverse problem (see Newey and Powell (2003)), while maintaining some model specification flexibility, we impose condi-

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

tional moment restrictions on the first-differences 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 are available, such as the so-called control function approach of Heckman and Robb (1985), Blundell et al. (2013), Darolles et al. (2004), Gao and Phillips (2013), and Su and Ullah (2008), among others, but these require other 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 coefficients with endogenous covariates; however, they do not consider the panel data case. Several works (see Rodriguez-Poo and Soberon (2017) for a survey) analyze varying-coefficient panel data models, but the resulting estimators are not robust to the presence of endogeneity. Recently, Fève and Florens (2014) consider estimations of nonparametric panel data models using an IV condition. However, their results do not apply straightforwardly to the varying-coefficient model. Finally, IV methods have been proposed in the context of varying-coefficient panel data models with random effects. Cai and Li (2008) propose estimating the unknown functions of interest using the nonparametric generalized method of moments. However, this method does not control for heterogeneity when it is correlated with some explanatory variables, and hence generates asymptot-

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

ically biased estimators when fixed effects are present. The semiparametric partially linear varying-coefficient model encompasses several alternative specifications of interest to econometricians (e.g., partially linear model, fully linear parametric model). Therefore, we propose a Wald-type statistic based on Cai et al. (2017), and the references therein. Furthermore, we provide a technique to compute the confidence bands for the varying coefficients. To show the feasibility and advantages of the proposed procedure, we apply it to extend the life cycle hypothesis (LCH) of Chou et al. (2004) to include panel data.

The remainder of the paper is structured as follows. In Section 2, we set up the econometric model and describe the three-step estimation procedure. In Section 3, we discuss the their asymptotic properties of the model and the procedure. In Section 4, we provide estimators that are more efficient, such as the one-step backfitting and minimum distance estimators. Section 5 develops a Wald-type test for the constant coefficients and the 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 the test statistic. Section 7 applies our methods to estimate the LCH model. Finally, Section 8 concludes the paper. All assumptions and proofs of the main results are relegated to the

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

online Supplementary Material.

2. Model and estimation procedures

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

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

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, respectively, of endogenous random variables, 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. In addition, the structural equation (2.1) includes 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 constant coefficients (i.e., β_1 and β_2) that need to be estimated. Furthermore, denote $\mathcal{L}_{it}$ as an $(\ell \times 1)$ vector of all exogenous variables (i.e., Z, X_2, and U_2) and an M-dimensional vector of other IVs, 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 $\mathcal{L}_{i(t-1)}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

We assume the following moment condition:

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

The above model is general enough to include relevant empirical examples in the economics literature. For example, based on the 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 household features, such as the age of the household head.

Furthermore, the above model allows for two 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 arbitrarily correlated with Z, X, and/or U (i.e., fixed effects). It is well known that, ignoring these sources of endogeneity in a direct estimation of the functions of interest leads to asymptotically biased estimators. The second source of endogeneity can be handled by taking a first-difference transformation in (2.1), obtaining

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

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, and $\Delta U_{it}=(\Delta U_{1it}^{\top},\Delta U_{2it}^{\top})^{\top}$ and $\beta=(\beta_1^{\top},\beta_2^{\top})^{\top}$ are $(k\times 1)$ vectors,

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

with $d = d_1 + d_2$ and $k = k_1 + k_2$. The respective definitions for $X_{i(t-1)}$ and $m(Z_{i(t-1)})$ are similar.

For any given β and $\Delta Y_{it}^* = \Delta Y_{it} - \Delta U_{it}^\top \beta$, Rodriguez-Poo and Soberon (2014) propose estimating 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 $I\!R$, by minimizing the criterion function

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

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 significant because it enables us to overcome the non-negligible asymptotic bias of the differencing nonparametric estimators. If we consider kernels around Z_{it} only, the remainder term in the Taylor approximation is not negligible because the distance between Z_{is} $(s \neq t)$ and z does not vanish asymptotically. This phenomena was noted in Mundra (2005) and Lee and Mukherjee (2014), and solved in Rodriguez-Poo and Soberon (2014, 2015) for a local linear regression.

Unfortunately, although the resulting estimator for γ 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 the expectation on both sides of the structural equation (2.3), conditioning on $\mathcal{L}_{it}$ and $\mathcal{L}_{i(t-1)}$ and using condition (2.2), we can obtain the following:

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

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 criterion function

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

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

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

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

Unfortunately, $\widetilde{m}_{\beta}(z; H_2)$ is an infeasible estimator because the vector of parameters, β, and the nonparametric functions $E(\Delta X_{1it}^{\top}|\mathcal{L}_{it}, \mathcal{L}_{i(t-1)})$

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

and $E(\Delta U_{1it}^{\top}|\mathcal{L}_{it}, \mathcal{L}_{i(t-1)})$ are unknown and need to be estimated. The first stage of our procedure estimates the nonparametric functions. Denote by $\widehat{E}(\Delta X_{1it}^{\top}|\mathcal{L}_{it}, \mathcal{L}_{i(t-1)}; H_1)$ and $\widehat{E}(\Delta U_{1it}^{\top}|\mathcal{L}_{it}, \mathcal{L}_{i(t-1)}; H_1)$ the 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. For example, these might be local linear or constant estimators.

The second stage is to estimate β. Denote $n = N(T-1)$, we propose a conventional profile least squares estimator (see Fan and Huang (2005)),

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

where $\Delta Y = [\Delta Y_{12}, \ldots, \Delta Y_{NT}]^{\top}$, $\Delta\widehat{W}_U = [\Delta\widehat{W}_{U_{12}}, \ldots, \Delta\widehat{W}_{U_{NT}}]^{\top}$, $\Delta\widehat{W}_{U_{it}}$ is the vector $\Delta W_{U_{it}}$ in which the unknown functions have been replaced by consistent estimators, and $\widehat{S}$ is a smoothing matrix; that is,

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

where $\widehat{\Gamma}_{it} = \Delta\widehat{W}_X^{\top}K(Z_{it}, H_2)\Delta\widehat{W}_X$, $\Delta\widehat{W}_X = [\Delta\widehat{W}_{X_{12}}, \ldots, \Delta\widehat{W}_{X_{NT}}]^{\top}$ 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 an $n \times n$ diagonal matrix of the form

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

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

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

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

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, having obtained the estimator for β, and after replacing the unknown quantities with their estimated objects, the resulting three-stage estimator for $m(\cdot)$, at any given value of z, is

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

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

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

3. Statistical properties

In this section, we investigate the 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. Detailed proofs of the following results are given in Sections 3 to 5 of the Supplementary Material.

Theorem 3.1. *Suppose that Assumptions S2.1–S2.10 hold. When* $Ntr(H_2)^2 \to 0$*, because* N *tends to infinity and* T *is fixed, we have*

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

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*

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

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

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

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

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

where

$$
\begin{aligned}
B(z; H_2) &= \mu_2(K) \Big[diag_d(D_f(z) H_2 D_{m_\kappa}(z)) \imath_d f^{-1}_{Z_{it}, Z_{i(t-1)}}(z, z) \\
&\qquad + \frac{1}{2} diag_d \left(tr \left(\mathcal{H}_{m_\kappa}(z) H_2 \right) \right) \imath_d \Big], \\
V(z; H_2) &= 2\sigma^2_\epsilon R^2(K) \mathcal{B}^{-1}_{\Delta W_X \Delta W_X}(z, z).
\end{aligned}
$$

Moreover, $D_{m_\kappa}(z)$ *is the first-order derivative vector of the* κ*th component of* $m(\cdot)$*,* $\mathcal{H}_{m_\kappa}(z)$ *is its Hessian matrix, and* $D_f(z)$ *is the first-order derivative vector of the density function, for* $\kappa = 1, \ldots, d$*. In addition,* $diag_d\left(tr(\mathcal{H}_{m_\kappa}(z) H_2)\right)$ *and* $diag_d(D_f(z) H_2 D_{m_\kappa}(z))$ *denote* $(d \times d)$ *diagonal matrices of elements of* $tr(\mathcal{H}_{m_\kappa}(z) H_2)$ *and* $D_f(z) H_2 D_{m_\kappa}(z)$*, respectively, where* $\imath_d$ *is 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.

Comparing the results of Theorem 3.2 and Corollary S1.2, as expected, we find the best behavior, in terms of bias, of the local linear three-stage estimator against the Naradaya-Watson (NW) version. For other advantages, see Fan and Gijbels (1995). Nevertheless, in this framework with endogenous regressors we need to consider that the local linear estimator requires the use of three different nonparametric estimators as IVs, along

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

with their corresponding bandwidths, whereas the NW version needs only one. Therefore, better performance by the NW estimator in finite samples is expected, as shown in the Monte Carlo experiments. That is why we focus on the NW estimators throughout, although all results can be extended to the local polynomial case.

Furthermore, note that the results from Theorem 3.2 show a bias term that depends asymptotically only on the smoothness of $m(\cdot)$ and $E(\Delta X_{1it}|\mathcal{L}_{it}, \mathcal{L}_{i(t-1)})$. The dependence on $\widehat{\beta}$ and $\widehat{E}(\Delta X_{1it}|\mathcal{L}_{it}, \mathcal{L}_{i(t-1)}; H_1)$ is negligible because $\widehat{\beta}$ is $\sqrt{NT}$-consistent and, under the assumptions established in Section 2 of the Supplementary Material, $\widehat{E}(\Delta X_{1it}|\mathcal{L}_{it}, \mathcal{L}_{i(t-1)}; H_1)$ 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, owing to condition $tr(H_1) = o_p(tr(H_2))$. Note too that Theorem 3.2 includes a variance term with a suboptimal rate of convergence. In this smoothness class, the lower rate of convergence for this type of estimator is $n|H_2|^{1/2}$ (see Härdle (1990) for details). Therefore, in the next section, following Rodriguez-Poo and Soberon (2014) and Rodriguez-Poo and Soberon (2015), we propose a one-step backfitting algorithm that makes the rate of convergence of our estimators optimal.

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

4. Efficient estimators

4.1 One-step backfitting and minimum distance estimators

In this section, we first propose a one-step backfitting algorithm that achieves optimal nonparametric rates of convergence of the estimators for $m(\cdot)$. In addition, as shown later, because of the additive structure of the regression model, the backfitting procedure generates two alternative estimators for $m(\cdot)$. Nevertheless, by combining the two estimators using a minimum distance estimation technique, it is possible to obtain a more efficient estimator for $m\left(\cdot\right)$.

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

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

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 covergence is $\sqrt{n}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

The main reason for applying 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,

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

and

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

Then, the unknown function $m(\cdot)$ in (4.12) can be estimated following the same steps as in (2.6)–(2.9) obtaining (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-definite matrix, then an alternative backfitting estimator for $m(\cdot)$ is

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

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). Substracting $X_{it}^{\top}\widehat{m}_{\widehat{\beta}}(Z_{it}, H_2)$ in both terms of (2.3) and proceeding as above, we obtain (4.13).

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

Therefore, this technique provides two different estimators, $\widehat{m}^{(1)}_{\widehat{\beta}}(z;H_3)$ and $\widehat{m}^{(2)}_{\widehat{\beta}}(z;H_3)$, for the same $m(z)$. We can combine both in an efficient way by minimizing the following criterion function:

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

We propose calculating the estimators $\widehat{m}^{(1)}_{\widehat{\beta}}(z;H_3)$ and $\widehat{m}^{(2)}_{\widehat{\beta}}(z;H_3)$ using subsamples of size N_1T and N_2T respectively, such that $N_1+N_2=N$ and, because $N\to\infty$, $N_j/N\to c_j$, 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 $m(z)$, $\widehat{m}^{(mde)}_{\widehat{\beta}}(z;H_3)$, that minimizes (4.14) is

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

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

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

class of the so-called minimum distance estimators. There are many ways in which to select the weighting matrix, $\mathcal{W}_m(z)$. 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, Hansen (1982) shows that $\mathcal{W}_m^*(z) = V^{(1+2)}(z)$, where $V^{(1+2)}(z)$ stands for the asymptotic variance-covariance matrix of

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

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

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. Because $N \to \infty$ and T is fixed, we have*

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

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

where

$$B(z;H_3) = \mu_2(K)\Big(diag_d\left(D_f(z)H_3D_{m_\kappa}(z)\right)\imath_d f_{Z_{it}}^{-1}(z) + \frac{1}{2}diag_d\left(tr\left(\mathcal{H}_{m_\kappa}(z)H_3\right)\right)\imath_d\Big),$$

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

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

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

The proof of this result is provided in the Supplementary Material.

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

Theorem 4.2. *Suppose that Assumptions S2.1–S2.12 hold. Because* $N\to\infty$ *and* T *is fixed, we have*

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

where

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

The proof of this result is relegated to the Supplementary Material. In Theorem 4.2, the asymptotic bias of the minimum distance estimator is the

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

same as that in Theorem 4.1. Moreover, the asymptotic variance exhibits the optimal rate of convergence for this type of problem. 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 efficient estimators for $m(\cdot)$, while achieving optimality.

5. Inference

The statistical model given in (2.1) and (2.2), includes several models of interest in econometrics and statistics. For example, it is natural to investigate whether certain variables in this component are statistically significant after fitting the model. More generally, one might consider the set of linear hypotheses

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

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:

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

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

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

$$\widehat{\Sigma} = \frac{1}{n}\Delta\widehat{W}_U^\top(I_n-\widehat{S})^\top(I_n-\widetilde{S})\Delta U,$$
$$\widehat{\Sigma}^* = \frac{1}{n}\Delta\widehat{W}_U^\top(I_n-\widehat{S})^\top(I_n-\widetilde{S})\widehat{V}(I_n-\widetilde{S})^\top(I_n-\widehat{S})\Delta\widehat{W}_U,$$

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

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

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

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

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; the proof is provided in the Supplementary Material.

Corollary 5.1. *Suppose that Assumptions S2.1–S2.10 hold. When $Ntr(H_2)^2 \to 0$, because $N \to \infty$ 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.

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

From a nonparametric point of view, we may need to construct a 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 on the local linear version.

6. Monte Carlo experiment

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

$$\begin{aligned} Y_{it} &= X_{1it}m_1(Z_{it}) + U_{1it}\beta_1 + U_{2it}\beta_2 + \mu_i + v_{it}, \qquad (6.20) \\ & i = 1, \ldots, N \quad ; \quad t = 1, \ldots, T, \end{aligned}$$

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, and X_{1it} and U_{1it} are endogenous variables following the reduced-form equations:

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

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

respectively. Here, V_{1it} and V_{2it} are IVs generated independently from a uniform $[0,4]$ distribution, and the noise follows

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

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

To evaluate the performance of the proposed estimators, we set $\rho = 0.7$ and $\sigma_\epsilon^2 = 1$, and conduct simulations in which the number of periods T is equal to four, and we use cross-sections N equal to 100, 200, and 400. For each sample size, we replicate the experiment $1,000$ times. For $K(u)$, we choose the Epanechnikov kernel function $K(u) = 0.75(1-u^2)I(|u| \leq 1)$. To meet the requirement that $Ntr(H_2)^2 \to 0$, 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 we need to use 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 that of the second-stage 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 mean, standard deviation, and root mean squared error (RMSE) of the estimated $1,000$ values for β under different settings. The results show that the performance of 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 increases, all of them converge. However, we have different results for β_1 (i.e., the parameter related to the endogenous variable X_1). As expected, $\widehat{\beta}_{NF}$ presents the best results, but $\widehat{\beta}_F$ and $\widetilde{\beta}_F$ both perform quite well in terms of dealing with the endogeneity problem. Although both methods have similar standard deviations, $\widetilde{\beta}_F$ has a large bias, and $\widehat{\beta}_F$ exhibits the lower RMSE.

In order to verify the aforementioned asymptotic results for the functional coefficient, we compare the finite-sample behavior of different non-

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

Table 1: Mean, standard deviation, and RMSE's of the estimators for β_1 and β_2.

		H_2	$\widehat{\beta}_{E_1}$	$\widehat{\beta}_{NF_1}$	$\widehat{\beta}_{F_1}$	$\widetilde{\beta}_{F_1}$	$\widehat{\beta}_{E_2}$	$\widehat{\beta}_{NF_2}$	$\widehat{\beta}_{F_2}$	$\widetilde{\beta}_{F_2}$
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	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

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

parametric estimators, where $\widehat{\beta}$ is used as a $\sqrt{n}$-consistent estimator of β, and the bandwidth in this stage $H_3 = h_3 I_q$ is chosen using the Silverman's rule-of-thumb; that is, $h_3 = 1.06\widehat{\sigma}_Z N^{-1/5}$, where $\widehat{\sigma}_Z$ 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 $h_2 = 1.75\widehat{\sigma}_Z N^{-1/3}$ and $h_1 = 1.25\widehat{\sigma}_Z N^{-1/3}$. As a measure of accuracy, we use the following RMSE:

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

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

For the sake of comparison, Figure 1 depicts box plots of the $1,000$ RMSE values of the functional coefficient estimators using the proposed estimators. To show the effect of the generated regressors in constructing the feasible estimators, Figure 1(a) collects the results for the NW estimator without the adjustment for endogeneity, and Figure 1(b) collects the results for the nonfeasible NW estimator with ΔW_X instead of $\Delta\widehat{W}_X$. As expected, in Figure 1(b), all RMSE values of the nonfeasible NW estimator converge toward zero. However, this is not true for the NW estimator with endogeneity, presented in Figure 1(a). In addition, to compare the NW and the local linear estimator, Figure 1(c) reflects the results for $\widehat{m}_{\widehat{\beta}}(z;H_2)$, and Figure 1(d) displays the results for the $\widehat{m}_{\widehat{\beta},LL}(z,H_2)$. As the sample size

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

increases, the RMSEs of both estimators shrink to zero. However, as noted in the theoretical results, the NW estimator performs slightly better than the local linear estimator because only one generated regressor is necessary.

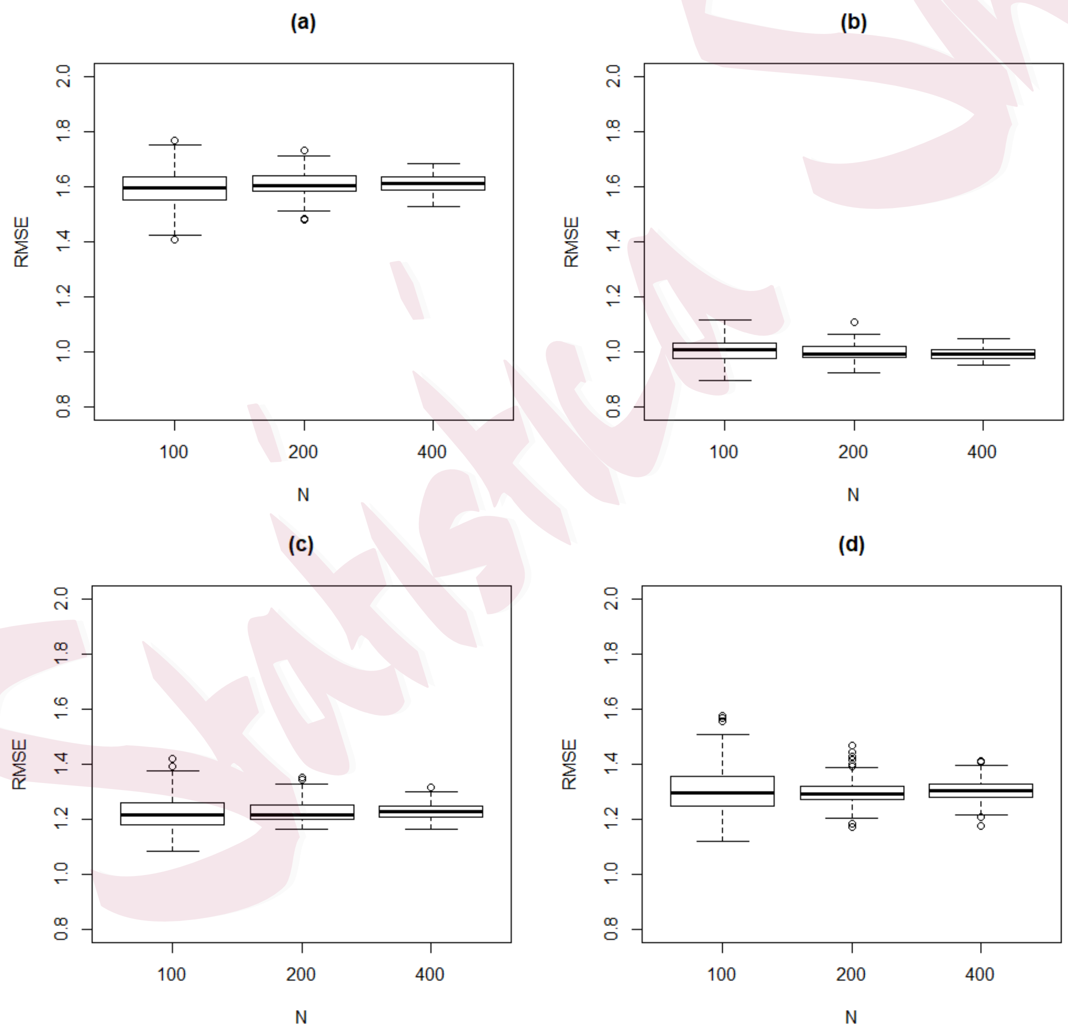
In order to check the efficiency improvement of the backfitting estimators, Figure 2 reports the RMSE of the backfitting estimators relative to that of the NW estimator. Specifically, Figure 2(a) collects the results for $\widehat{m}^{(1)}_{\widehat{\beta}}(z; H_3)$, while Figure 2(b) depicts the results for $\widehat{m}^{(2)}_{\widehat{\beta}}(z; H_3)$. In addition, Figure 3 collects the RMSE of the minimum distance estimator relative to that of the NW estimator. The results show that all figures corroborate the efficiency improvement because their relative RMSEs shrink toward zero. Therefore, the simulation findings confirm the theoretical results.

Finally, we study the size and power performance when testing $H_0 : \beta_1 = -1, \beta_2 = 1$ 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 the sample size $T = 4$ and $N = 200$, and conduct $1,000$ Monte Carlo simulations. 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 the three significance levels.

Figure 4, shows that when $\phi_1 = 0$, the power collapses to the test size.

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

Figure 1: Box plots of the RMSE values of the three-stage nonparametric estimates in 1,000 independent simulations for three sample sizes $N = 100, 200, 400$ and $T = 4$.



Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

Figure 2: Box plots of the nonparametric backfitting estimates in $1,000$ independent simulations for three sample sizes $N = 100, 200, 400$ and $T = 4$.

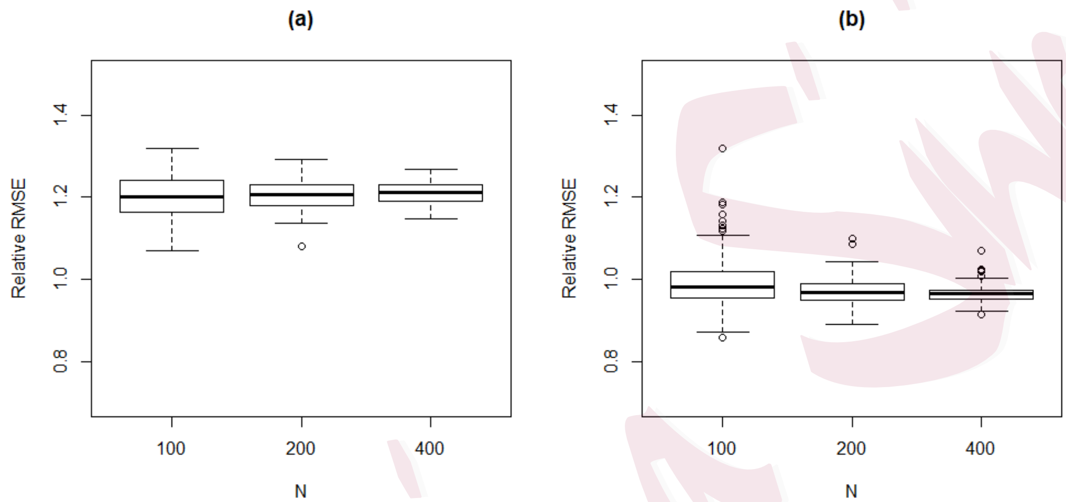
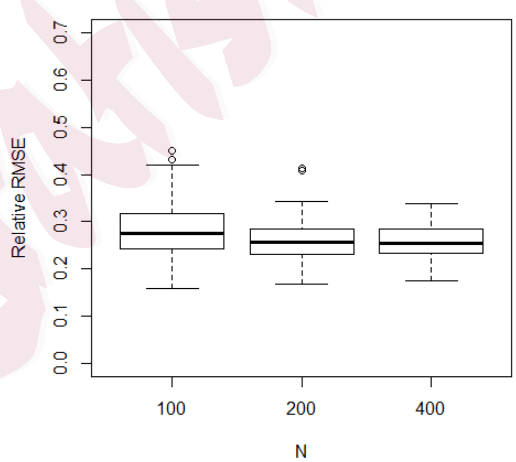
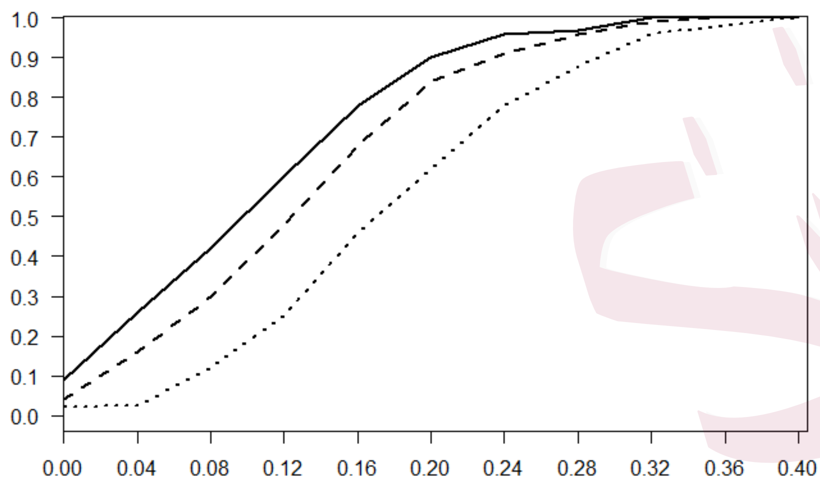


Figure 3: Box plots of the minimum distance estimator in $1,000$ independent simulations for three sample sizes $N = 100, 200, 400$ and $T = 4$.



Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

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



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

More precisely, the simulated sizes of the proposed test are 9%, 3%, and 2%, corresponding to the significance levels 1% (dotted line), 5% (dashed line), and 10% (solid line), respectively. Therefore, the simulated sizes are close to the nominal size, and our test can deliver a correct test size. In contrast, when ϕ_1 deviates from zero, our test is reasonably powerful, because the power curves tend to one quickly. Specifically, the power is over 90% for the significance levels 10% and 5% when $\phi_1 \geq 0.22$.

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

7. Empirical results

To demonstrate the usefulness of the proposed method, in this section, we analyze the effects of unexpected health expenses on household savings. Along with liquidity constraints and habits in consumer preferences, uncertainty about possible economic hardships and household risk aversion are key determinants of a household's consumption/savings decisions; see Friedman (1957). In this situation, precautionary savings can protect individuals against potential income downturns and unforeseen out-of-pocket medical expenses in later life, see Chou et al. (2004) for further discussion.

We propose extending the analysis of Chou et al. (2004) by estimating the following regression model:

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

where i denotes a household, t denotes time, Z_{it} is the age of the household head, X_{1it} is the healthcare expenditure (log), Y_{it} denotes savings, and X_{2it} is permanent income (log). In this sense, household savings is characterized by uncertainty related to future healthcare expenses, $m_1(\cdot)$, and income downturns, $m_2(\cdot)$. Note that household's permanent income is not directly observable. In order to approximate this variable, we follow Chou et al. (2004). Thus, assuming the interest rate is equal to the productivity rate

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

of growth, and 65 years old is the maximum age at which people work, the permanent earnings at age τ_0 is calculated as

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

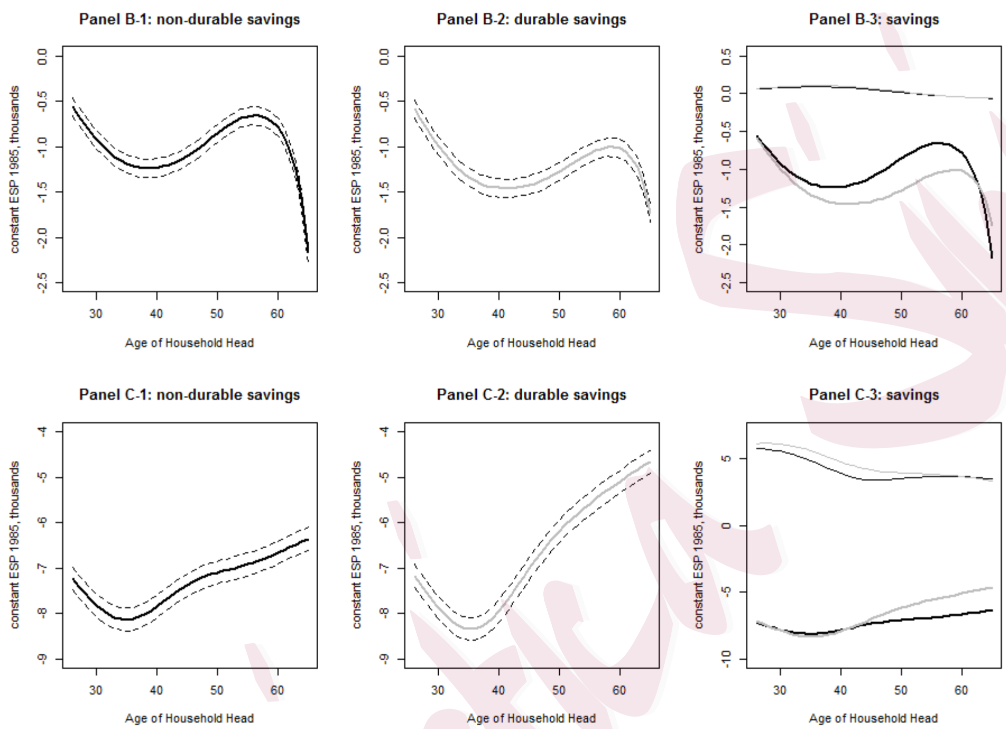
where $f(\tau)$ is the estimated quadratic function of age, Y_{it} is household income, and X_3 is a vector of demographic characteristics.

After choosing the bandwidths in the same way as in the simulation study, we have the estimation results shown in Figure 5. The estimated curves are plotted against the age variable jointly, with the 95% pointwise confidence intervals calculated by adapting the wild bootstrap technique of Härdle et al. (2004) to this context. Figure 5 is divided into two panels, B and C, each of which is split into three graphics. Panel B shows the corresponding elasticity to changes in healthcare expenditure, $\widetilde{m}_1(\cdot)$; Panels C shows the precautionary savings elasticity to changes in household income, $\widetilde{m}_2(\cdot)$. 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.

The results show that when we control for uncertainty about healthcare expenditure (Panel B), younger households (26–33) exhibit a declining savings rate, then a constant rate until age 40, followed by an inverted U-shape.

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

Figure 5: Household savings over the life cycle



Notes: The thick line denotes the estimates for durable savings, the continuous line denotes those nondurable savings, and the dotted line is the 95% pointwise confidence interval.

In addition, if these results are combined with the delay in the wealth accumulation process in Spanish households (age 45 (Spain) vs. age 40 (US)), we realize the negative impact that public health programs have on precautionary savings, confirming the results in Chou et al. (2004). Comparing the elasticities for the different savings results, we find that the consumption

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

of durable goods is particularly sensitive to unexpected changes in income, whereas that of nondurable goods is more sensitive to potential healthcare expenses. This is especially true 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 (gray line) against those obtained without considering endogeneity (black line); see Panel 3 of Figure 5. The results indicate significant differences. When we control for uncertainty about healthcare expenditure, households accumulate assets in the middle of their lives. However, when endogeneity is not taken into account, there is a more or less constant path over the life cycle.

8. Conclusion

This study examined a nonparametric estimation of a varying-coefficient structural panel data model, where the individual heterogeneity is allowed to be correlated with some explanatory variables. This specification is becoming increasingly common in many standard econometric applications, such as studies of household consumption behavior or labor supply analysis. Therefore, we require estimators that keep a reasonable degree of flexibility and are robust to both endogeneity and fixed effects. We attempt to satisfy these requirements using a nonparametric three-stage procedure, where IV

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

techniques are used to deal with the endogeneity, and differencing techniques are used to cope with the fixed effects. Furthermore, to achieve efficiency, a minimum distance estimator is proposed. The feasibility and advantages of the proposed procedure are shown by estimating an LCH panel data model. Simulation results support the empirical findings.

Supplementary Material

In Section S1 of the online Supplementary Material, we present the local linear version of the three-stage estimator in Section 2 for the local constant case. We also give its main asymptotic properties. In Sections S3 to S5, we show the main results claimed in Section 3. Section S2 contains the assumptions needed to prove these results. Section S6 contains the proof of Theorem 4.1, and Section S7 provides that of Theorem 4.2. Both theorems appear in Section 4.2 of this paper.

Acknowledgments

The authors wish to thank the two anonymous referees for their helpful comments and suggestions. The usual disclaimer applies. The authors gratefully acknowledge the 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 Research Project APIE 1/2015-17, "New methods for the empirical analysis of financial markets", of the Santander Financial Institute (SANFI) of the 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". *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 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 paper.

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

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 Surverys 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 Economia. Universidad de Cantabria. Avda. de los Castros s/n. 39005. Santander (Spain)

E-mail: juan.rodriguez@unican.es

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0382

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

Santander (Spain)

E-mail: alexandra.soberon@unican.es