OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 86, 4 (2024) 0305-9049
doi: 10.1111/obes.12609

A Semi-parametric Panel Data Model with Common Factors and Spatial Dependence

ALEXANDRA SOBERON,† ANTONIO MUSOLESI‡ and JUAN M. RODRIGUEZ-POO†

†*Department of Economics, Universidad de Cantabria, Avda. de los Castros s/n Santander, 39011, Spain, e-mail: soberonap@unican.es; rodrigjm@unican.es*
‡*Department of Economics and Business, University of Ferrara, Ferrara, Italy, e-mail: mslntn@unife.it*

Abstract

In the analysis of the Griliches' knowledge capital production function, previous works pointed out the relevance of incorporating slope heterogeneity in the technological parameters, cross-sectional dependence arising simultaneously from common factors and spillovers, and possible nonlinear effects of relevant common observed variables. In order to solve the above problems, in this article we introduce a semi-parametric model in a partially linear form that copes simultaneously with all the previous specification issues. The asymptotic properties of the resulting estimators are obtained and the theoretical findings are further supported for small samples via several Monte Carlo experiments and an empirical application.

I. Introduction

Assessing the effect of common variables (such as technological, institutional, environmental and health factors) on economic activity is of crucial relevance in many empirical cases. These may include, among many others, the study of the impact of real common shocks on productivity or economic growth; the estimation of the effects of changes in oil prices on wages, employment or production activity (Keane and Prasad, 1996; Hamilton, 2003) or the study of the relationship between housing rental prices and labour market conditions (Phillips and Wang, 2021). Furthermore, the impact of these common shocks is likely to vary across different population units and ignoring this fact can lead to misleading inference (see Andrews, 2005).

JEL Classification numbers: C14, C23, D24, C18.
Juan M. Rodriguez-Poo and Alexandra Soberon acknowledge financial support from the I+D+i project Ref. PID2019-105986GB-C22 financed by MCIN/AEI/10.13039/501100011033. In addition, this work is part of the I+D+i project Ref. TED2021-131763A-I00 financed by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR. We also acknowledge Santander Supercomputacion support group at the University of Cantabria who provided access to the supercomputer Altamira Supercomputer at the Institute of Physics of Cantabria (IFCA-CSIC), member of the Spanish Supercomputing Network, for performing simulations. The research was partially carried out while Alexandra Soberon visited the University of Ferrara.

In the panel data literature, the consideration of models with multi-factor error structure (see Pesaran, 2006; Bai, 2009) has enabled researchers to cope in part with the previous challenge because these types of models usually allow for unknown common factors to affect individuals heterogeneously. However, it remains unclear some specification issues related the introduction of observed common covariates in these models. For example, in Eberhardt, Helmers, and Strauss (2013) data on 12 manufacturing industries in 10 countries it is used to estimate the knowledge capital production function of Griliches (1979), that is:

$$Y_{it} = g_i(L_{it}, K_{it}, R_{it}) \exp(\alpha_i + \gamma_i' f_t + \epsilon_{it}),$$

where Y_{it} is value-added, L_{it} and K_{it} are standard labour and capital inputs, and R_{it} is knowledge capital. In addition, the α_i's are individual heterogeneity effects, f_t is a vector of unobserved common factors, γ_i are the corresponding factor loadings and ϵ_{it} is a zero mean error term. Following Griliches (1979), among many others, in Eberhardt *et al.* (2013) it is adopted a Cobb–Douglas technology, $g_i(L_{it}, K_{it}, R_{it}) = L_{it}^{\beta_{1i}} K_{it}^{\beta_{2i}} R_{it}^{\beta_{3i}}$, and taking logs, the following specification is proposed:

$$y_{it} = \alpha_i + \beta_{1i} l_{it} + \beta_{2i} k_{it} + \beta_{3i} r_{it} + \gamma_i' f_t + \epsilon_{it}, \tag{1}$$

where lowercase letters indicate that the variables are in log form.

Despite that the specification proposed in (1) is quite general and incorporates the multifactor error structure, it might ignore some observed common variables such as oil prices shocks, say z_t, that can affect production activity through their effects on production costs or measures of worldwide economic activity (Kilian, 2009). Indeed, while a large body of research has attempted to estimate the effects of oil price shocks on economic activity, the fact that these shocks contribute directly to economic decline remains controversial as the empirical relationship between oil prices and output has been very unstable across previous works. Since the pioneering work of Mork (1989), such instability has been attributed to a misspecification of the linear approximation of the relationship between oil prices and economic activity. Thus, it has been argued that economic activity responds asymmetrically to positive and negative oil prices shocks as rising oil prices should negatively affect aggregate economic activity more than decreasing oil prices should stimulate it (Lescaroux and Mignon, 2008), thus calling for the adoption of flexible specifications (Hamilton, 2003). Indeed, there is generally little prior knowledge – theoretical or empirical – about the shape of the relationship between the common covariates and production. Moreover, erroneously imposing a parametric form for these variables may lead to biased estimates of the technological parameters.

Furthermore, the cross-sectional dependence (CSD) sometimes might not be due only to the presence of latent common factors, but can be a result of the presence of spatial dependence. In Millo (2019) it is provided empirical evidence of spatially correlated residuals even when unobserved common factors are introduced in the knowledge capital production function of Griliches (1979). Spatial processes such as the well-known spatial autoregressive or spatial moving average models are very popular in this approach (See Cliff and Ord, 1972; Arbia, 2006; Lee and Lee and Yu, 2010 among others). However, although both factors and spatial models allow for CSD, the motivations underlying these models differ meaningfully (see, e.g. Ertur and Musolesi, 2017, for a discussion). Indeed, recent works highlight the relevance of jointly modelling both forms of dependence. Nevertheless, while there is a relatively rich literature on linear panel data models with both sources of CSD (see Holly, Pesaran, and Yamagata, 2010; Pesaran and Tosetti, 2011; Bailey, Holly, and Pesaran, 2016; Shi and Lee, 2017, among others), few advances have been made in the study of non-parametric/semi-parametric panel data models in the presence of both sources of CSD. So far, to our knowledge, it does not exist a full analysis of the impact of both sources of dependence on the statistical properties of both the parametric and non-parametric components in a semi-parametric panel data model.

Given the previous considerations, in this article, we propose a semi-parametric heterogeneous panel data model that extends the model in (1) by including an unknown heterogeneous function of the oil prices, $m_i(z_t)$, and different sources of CSD (a factor model combined with a spatial correlation structure). This appears to be suitable for modelling the knowledge capital production function. Hence, we propose the following specification:

$$y_{it} = \alpha_i + \beta_{1i} l_{it} + \beta_{2i} k_{it} + \beta_{3i} r_{it} + m_i(z_t) + \gamma_i' f_t + \epsilon_{it}. \tag{2}$$

A motivation for introducing such a specification is that, despite the well-known limitations of a Cobb–Douglas specification for $g(L_{it}, K_{it}, R_{it})$ (Ivaldi *et al.*, 1996; Ma, Racine, and Ullah, 2020), it remains the cornerstone of a huge literature in empirical economics at all levels of aggregation (Doraszelski and Jaumandreu, 2013; Eberhardt *et al.*, 2013; Ackerberg, Caves, and Frazer, 2015; Antonioli, Gioldasis, and Musolesi, 2021) because it builds on economic theory and generally provides a good approximation of the underlying data in a parsimonious setting. The same cannot be said about the impact of z_t which is much undetermined. As we have mentioned before, we decide to include it non-parametrically in our model. Additionally, note that the heterogeneous effects of these covariates may be the result of unit-specific technological constraints. We emphasize that the model proposed in this article can be applied to many other empirical problems.

In order to obtain estimators of the parameters of interest, we extend the common correlated effects (CCE) technique (see Pesaran, 2006) to consider simultaneously all the above relevant empirical problems and then, we obtain the following outcomes: (i) $\sqrt{T}$-consistent estimators of the slope parameters under heterogeneity and $\sqrt{NT}$-consistent estimators in case of homogeneity. Estimators of the non-parametric components that achieve the optimal rate of convergence are also obtained. ii) The asymptotic properties of the mean group estimator and the pooled estimator of the slope parameters under both homogeneous and heterogeneous parameters are derived. iii) All non-parametric estimators exhibit a rate that is going to depend on the strength of the spatial dependence. In this sense, this result coincides with those obtained in Robinson (2012), Lee and Robinson (2015) and Soberon, Rodriguez-Poo, and Robinson (2022).

The remainder of the article is organized as follows. In section II, we set up the econometric model. In section III, we provide the main theoretical results, i.e. the estimation method and the asymptotic properties. Section IV presents some Monte Carlo simulation results. In section V an empirical application of the knowledge capital production function in Griliches (1979) is conducted by exploiting an annual country-level balanced panel data set covering 24 OECD countries from 1971 to 2014. We conclude in section VI. Finally, all proofs of the mathematical results are relegated to the Appendix B.

II. A semi-parametric panel data model

Let y_{it} be the observation on cross-sectional unit i at time t, and suppose that it is generated according to the following DGP:

$$y_{it} = \alpha_i' d_t + x_{it}' \beta_i + m_i(z_t) + u_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{3}$$

where $d_t = (d_{1t}, d_{2t}, \ldots, d_{r_1 t})'$ is an $r_1 \times 1$ vector of observed common effects, x_{it} is a $p \times 1$ vector of observed explanatory variables, z_t is a $q \times 1$ vector of observed stochastic covariates, and $m_i(\cdot)$ is an unknown smooth function to estimate. In addition, u_{it} is a random error term that follows the following multifactor structure

$$u_{it} = \gamma_i' f_t + \epsilon_{it}, \tag{4}$$

where $f_t = (f_{1t}, f_{2t}, \ldots, f_{r_2 t})'$ is an $r_2 \times 1$ vector of unobserved common factors with associated factor loadings γ_i, and ϵ_{it} is an idiosyncratic error.

In order to allow the unobserved factors f_t to be correlated with the observed data (x_{it}, z_t, d_t), similar to Pesaran (2006), we can model this correlation via the following fairly general semi-parametric model:

$$x_{it} = A_i' d_t + g_i(z_t) + \Gamma_i' f_t + v_{it}, \tag{5}$$

where A_i and Γ_i are $r_1 \times p$ and $r_2 \times p$ factor loading matrices with fixed components, respectively, $g_i(z_t) \equiv (g_{1i}(z_t), \ldots, g_{pi}(z_t))'$ is a $p \times 1$ vector of unknown smooth functions, and v_{it} is a $(p+q) \times 1$ vector of individual-specific components of $(x_{it}', z_t')'$.

The idiosyncratic errors, ϵ_{it}, are assumed to be spatially and temporally correlated and follow an arbitrary form such as

$$\epsilon_{\cdot t} = \Phi^{1/2} \eta_{\cdot t}, \quad \text{for} \quad t = 1, \ldots, T, \tag{6}$$

where $\epsilon_{\cdot t} = (\epsilon_{1t}, \ldots, \epsilon_{Nt})'$ and $\eta_{\cdot t} = (\eta_{1t}, \ldots, \eta_{Nt})'$ are $N \times 1$ vectors, and $\Phi^{1/2}$ is a $N \times N$ matrix.

The model outlined in (3)–(6) is sufficiently general and enables us to deal with several limitations in the existing literature. Firstly, it allows to include common stochastic covariates, z_t, in a nonlinear heterogeneous manner ($m_i(\cdot)$ in equation (3) and $g_i(\cdot)$ in equation (5)). Secondly, we introduce a general non-parametric type of model of spatial correlation through the idiosyncratic error term (see equation (6)) that is very appealing from the empirical point of view. On the one hand, asymptotic normality of the proposed estimators for the parameters of interest can be obtained avoiding parametric descriptions of dependence that may lead to substantial size distortions in tests based on Maximum Likelihood or Quasi-maximum likelihood estimators (see Pesaran and Tosetti, 2011 for further details). On the other hand, it enables us to consider different types of cross sectional dependence (i.e. weak or long-range dependence). More precisely, let ω_{ij} be the ijth element of $E\left(\epsilon_{.t}\epsilon'_{.t}\right)$ for all t. If CSD is limited by the condition

$$\overline{\lim_{N\to\infty}}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\omega_{ij} < \infty, \tag{7}$$

weak dependence is allowed. On the contrary, if condition

$$\lim_{N\to\infty}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\omega_{ij} = 0 \tag{8}$$

is allowed, long-range dependence is permitted (see Lemma 1 for a deeper discussion).

Furthermore, our model is related to other models already considered in the literature. For example, if $m_i = g_i = 0$ and $\Phi = I$ then the model renders to the proposal in Pesaran (2006). In Su and Jin (2012) and Huang (2013) a fully non-parametric model for the x_{it}'s, with common factors, is considered. Unfortunately, they do not allow either common covariates or spatial dependence. Finally, in Lee and Robinson (2015) a fixed effects panel data model is considered, jointly with spatial dependence, but unobserved common factors are not allowed for.

III. Estimation method and asymptotic theory

In this section we estimate β_i and $m_i(\cdot)$ for $i \geq 1$. However, as Heckman (2001) concludes while the representative agent paradigm is shown to lack empirical support, the average person becomes a popular alternative and therefore we provide the corresponding estimators for their weighted averages, i.e., $\overline{\beta} = N^{-1}\sum_{i=1}^{N}\beta_i$ and $\overline{m}_i(\cdot) = N^{-1}\sum_{i=1}^{N} m_i(\cdot)$. We also derive estimators in the case of homogeneous slope parameters. Further their main asymptotic properties are analysed.

Estimation procedure

The following augmented regression model for (3)–(6) is proposed.[1]

$$y_{it} = \beta_i' x_{it} + m_i(z_t) + \delta_i'\lambda_t + e_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{9}$$

where $\lambda_t = (\overline{y}_{At}, \overline{x}_{At}, d_t)$ is a $(1 + p + r_1) \times 1$ vector of observable proxies for f_t, $e_{it} = \epsilon_{it} + o_p(1)$, and δ_i is a nuisance parameter.

In order to obtain estimators for the parameters of interest in (9) we will follow a profile least squares technique (see Fan and Huang, 2005). Note that for any given β_i and δ_i, (9) can be rewritten as

$$y_{it} - x_{it}'\beta_i - \lambda_t'\delta_i = m_i(z_t) + e_{it}, \tag{10}$$

[1]For a more detailed discussion of this equation see Appendix A.

and for z_t in a small neighbourhood of z, one can estimate $m_i(z)$ by minimizing the following weighted local least-squares problem:

$$\sum_{t=1}^{T}[(y_{it} - x_{it}'\beta_i - \lambda_t'\delta_i) - m_i(z) - (z_t - z)'D_{m_i}(z)]^2 K_H(z_t - z), \tag{11}$$

where $K(\cdot)$ is a product kernel function such that for each u it holds that $K_H(u) = |H|^{-1}\prod_{\ell=1}^{q} k(H^{-1}u_\ell)$, for $u = (u_1, \dots, u_q)'$, where $k(\cdot)$ is a univariate kernel function and H is a $q \times q$ bandwidth matrix that is symmetric and positive definite. Let $K_H(z) = \text{diag}\{K_H(z_1 - z), \dots, K_H(z_T - z)\}$ be a $T \times T$ matrix and $Z_z = [Z_{z_1}', \dots, Z_{z_T}']'$ a $T \times (q+1)$ matrix, where $Z_{z_t} = [1, (z_t - z)]$. Assuming that $Z_z'K_H(z)Z_z$ is non-singular, the solution to (11) for $m_i(\cdot)$ is

$$\widehat{m}_i(z, H) = \iota_1'(Z_z'K_H(z)Z_z)^{-1}Z_z'K_H(z)(Y_{i\cdot} - X_{i\cdot}\beta_i - \Lambda\delta_i), \tag{12}$$

where $Y_i \equiv (y_{i1}, \dots, y_{iT})$ is a $T \times 1$ vector, $X_{i\cdot} \equiv (X_{i1}, \dots, X_{iT})'$ and $\Lambda \equiv (\lambda_1, \dots, \lambda_T)'$ are $T \times p$ and $T \times (r_1 + 1 + p)$ matrices, respectively, and ι_1 is a $(1+q) \times 1$ vector having 1 in the first entry and all other entries being 0.

However, this estimator is infeasible since it depends on the unknown terms (β_i, δ_i). To overcome this, we use (12) to obtain a closed-form solution for the parametric estimators. Let $m_i(Z) = (m_i(z_1), \dots, m_i(z_T))'$ be a $T \times 1$ vector of the smooth unknown function, the estimator (12) in vectorial form is

$$\widehat{m}_i(Z, H) = S(Z, H)(Y_{i\cdot} - X_{i\cdot}\beta_i - \Lambda\delta_i), \tag{13}$$

where $S(Z, H)$ is a smoothing matrix that depends only on the observations of z_t, $t = 1, \dots, T$. Writing (10) in vectorial form and substituting (12) in (10), we get

$$\widehat{Y}_{i\cdot} = \widehat{X}_{i\cdot}\beta_i + \widehat{\Lambda}\delta_i + \widehat{e}_{i\cdot}, \tag{14}$$

where $\widehat{e}_{i\cdot}$ is a T-dimensional term such that $\widehat{e}_{it} = e_{it} - (\widehat{m}_i(Z, H) - m_i(Z)) + o_p(1)$. In addition, $\widehat{Y}_{i\cdot} = (I_T - S(Z, H))Y_{i\cdot}$, $\widehat{X}_{i\cdot} = (I_T - S(Z, H))X_{i\cdot}$, and $\widehat{\Lambda} = (I_T - S(Z, H))\Lambda$, where I_T is a $T \times T$ diagonal matrix.

To estimate β_i consistently, we use the idea of partitioned regression and define the projection function $M_{\widehat{\Lambda}} = I_T - \widehat{\Lambda}(\widehat{\Lambda}'\widehat{\Lambda})^{-1}\widehat{\Lambda}'$. Premultiplying both sides of (14) by $M_{\widehat{\Lambda}}$ and applying least squares to the resulting model, we obtain the so-called semi-parametric Common Correlated Effects (SCCE) estimator for β_i,

$$\widehat{\beta}_i = \left(\widehat{X}_{i\cdot}'M_{\widehat{\Lambda}}\widehat{X}_{i\cdot}\right)^{-1}\widehat{X}_{i\cdot}'M_{\widehat{\Lambda}}\widehat{Y}_{i\cdot}. \tag{15}$$

Similarly, a consistent estimator for δ_i is required, so we define the projection matrix $M_{\widehat{X}_i} = I_T - \widehat{X}_{i\cdot}(\widehat{X}_{i\cdot}'\widehat{X}_{i\cdot})^{-1}\widehat{X}_{i\cdot}'$. Then, premultiplying both sides of (14) by $M_{\widehat{X}_i}$ yields

$$\widehat{\delta}_i = \left(\widehat{\Lambda}'M_{\widehat{X}_i}\widehat{\Lambda}\right)^{-1}\widehat{\Lambda}'M_{\widehat{X}_i}\widehat{Y}_{i\cdot}, \tag{16}$$

and using (15) and (16) in (12), the feasible non-parametric estimator for $m_i(\cdot)$ is the so-called Non-parametric CCE estimator,

$$\widehat{m}_i(z,H) = \iota_1'(Z_z'K_H(z)Z_z)^{-1}Z_z'K_H(z)(Y_{i\cdot} - X_{i\cdot}\widehat{\beta}_i - \Lambda\widehat{\delta}_i). \tag{17}$$

Finally, if the parameters of interest are the cross-sectional means of β_i and $m_i(\cdot)$, i.e. $\overline{\beta} = E(\beta_i)$ and $\overline{m}(\cdot) = E[m_i(\cdot)]$, we follow Pesaran and Smith (1995) and propose a SCCE estimator of the mean group $\overline{\beta}$, which is a simple average of the individual SCCE estimators of β_i:

$$\widehat{\beta}_{\text{MG}} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\beta}_i, \tag{18}$$

while the corresponding mean group estimator for $\overline{m}(\cdot)$ is such that

$$\widehat{m}_{\text{MG}}(z,H) = \frac{1}{N}\sum_{i=1}^{N}\widehat{m}_i(z,H). \tag{19}$$

Alternatively, we can generalize the pooled estimator proposed in Pesaran (2006), obtaining the following SCCE pooled estimator for $\overline{\beta}$:

$$\widehat{\beta}_P = \left(\sum_{i=1}^{N}\widehat{X}_{i\cdot}'M_{\widehat{\Lambda}}\widehat{X}_{i\cdot}\right)^{-1}\sum_{i=1}^{N}\widehat{X}_{i\cdot}'M_{\widehat{\Lambda}}\widehat{Y}_{i\cdot}. \tag{20}$$

A non-parametric CCE pooled estimator for $\overline{m}(\cdot)$ is also obtained such that

$$\widehat{m}_P(z,H) = \iota_1'(Z_z'K_H(z)Z_z)^{-1}Z_z'K_H(z)\left[\overline{Y}_{A\cdot} - N^{-1}\sum_{i=1}^{N}X_{i\cdot}\widehat{\beta}_i - N^{-1}\sum_{i=1}^{N}\Lambda\widehat{\delta}_i\right]. \tag{21}$$

Note that a relevant feature of the above estimator is that given the fact that $\overline{m}(\cdot)$ is a non-parametric function of time-varying stochastic regressors, the pooled and mean group non-parametric estimators are the same, i.e. $\widehat{m}_{\text{MG}}(z,H) \equiv \widehat{m}_P(z,H)$.

Finally, from an empirical point of view, it can be also of interest to estimate a restricted submodel of (3) in which homogeneous slopes are assumed, i.e. $\beta_i = \beta$, $\forall i$. For this particular setting, the homogeneous SCCE pooled and mean group estimators ($\widehat{\beta}_{H,\text{MG}}$ and $\widehat{\beta}_{H,P}$, respectively) proposed for β are the same as the corresponding for the heterogeneous case (i.e. $\widehat{\beta}_{H,\text{MG}} = \widehat{\beta}_{\text{MG}}$ and $\widehat{\beta}_{H,P} = \widehat{\beta}_P$), whereas the resulting non-parametric estimator for $\overline{m}(\cdot)$ is of the form

$$\hat{\overline{m}}_{\text{MG}}(z,H) = \iota_1'(Z_z'K_H(z)Z_z)^{-1}Z_z'K_H(z)[\overline{Y}_{A\cdot} - \overline{X}_{A\cdot}\widehat{\beta}_{H,P} - \Lambda\widehat{\delta}_{\text{MG}}], \tag{22}$$

where $\widehat{\delta}_{\text{MG}}$ is the mean group estimator for the mean of δ_i, defined as $\widehat{\delta}_{\text{MG}} = N^{-1}\sum_{i=1}^{N}\widehat{\delta}_i$.

Assumptions and asymptotic theory

For ease of reference, we first state the definition of a strongly mixing sequence. Let $\{\zeta_t\}$ be a strictly stationary process and $\mathcal{F}_s^t$ denotes a σ-algebra of events generated by the random variables $(\zeta_s, \dots, \zeta_t)$ for $s \leq t$. Following Rosenblatt (1956), a process is said to be strongly mixing or α-mixing if

$$\alpha(\tau) = \sup_{s \in \mathcal{N}} \left\{ |P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_{-\infty}^{s}, B \in \mathcal{F}_{s+\tau}^{\infty} \right\} \to 0 \quad \text{as} \quad T \to \infty.$$

Next, we introduce the following notation: $\sigma_{\epsilon_i}^2 \equiv Var(\epsilon_{it})$, $\sigma_\eta^2 \equiv Var(\eta_{it})$, and $\Sigma_{v_i} \equiv Var(v_{it})$, where Σ_{v_i} is a positive definite matrix. Also, $\widetilde{X}_{i\cdot} = X_{i\cdot} - \mathcal{B}_X(z)$, $\widetilde{\Lambda} = \Lambda - \mathcal{B}_\Lambda(z)$, $\widetilde{F} = F - \mathcal{B}_F(z)$, and $\widetilde{D} = D - \mathcal{B}_D(z)$, where $\mathcal{B}_X(z) = E(X_{i\cdot}|z_t = z)$, $\mathcal{B}_\Lambda(z) = E(\Lambda|z_t = z)$, $\mathcal{B}_F(z) = E[F|z_t = z]$, and $\mathcal{B}_D(z) = E[D|z_t = z]$ for $D \equiv (d_1, \dots, d_T)'$ being a $T \times r_1$ matrix.

$$M_{\widetilde{G}} = I_T - \widetilde{G}(\widetilde{G}'\widetilde{G})^{-1}\widetilde{G}', \tag{23}$$

is a $T \times T$ projection matrix where $\widetilde{G} = (\widetilde{D}, \widetilde{F})$ is a $T \times (r_1 + r_2)$ matrix.

We add some additional notation. $\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A})$ are the eigenvalues of the $n \times n$-matrix $\mathbf{A}$. The column norm of A is $\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|$. The row norm of A is $\|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|$. The euclidean norm of A is $\|A\|_2^2 = \text{Tr}\left(A'A\right)$ and $(N, T) \xrightarrow{j} \infty$ denotes N and T tending to infinity jointly but not in particular order.

We now impose the following assumptions that are required to analyse the main asymptotic properties of the proposed estimators. Most of them are inspired by Pesaran (2006) or Pesaran and Tosetti (2011), among others, but they are appropriately modified for this article.

Assumption 1. (Individual-specific errors)

(i) The individual-specific errors η_{it} and v_{jt} are distributed independently for all i, j, and t. Let $\eta_{i\cdot} \equiv (\eta_{i1}, \dots, \eta_{iT})'$ and $v_{i\cdot} \equiv (v_{i1}, \dots, v_{iT})'$, $\eta_{i\cdot}$ and $v_{i\cdot}$ are independently distributed across i with zero means. Also, η_{it} has finite variance, σ_η^2, and finite fourth-order cumulants. Also, v_{it} have covariance matrices, Σ_{v_i}, which are non-singular and satisfy $\sup_i \|\Sigma_{v_i}\| < C < \infty$ and have uniformly bounded fourth-order cumulants. There exists some $\phi > 0$ such that $E|\eta_{it}|^{2(1+\phi)} < \infty$ and $E|v_{it}|^{2(1+\phi)} < \infty$.

(ii) Let $\{(\eta_{it}, v_{it}) : t \geq 1\}$ be independent across i for each fixed t. For each fixed i, the process $\{\eta_{it}, v_{it}\}$ is strictly stationary and α-mixing with the mixing coefficient satisfying $\alpha_i(\tau) = O(\tau^{-\theta})$, where $\tau = |t - s|$, $\theta = (2 + \phi)(1 + \phi)/\phi$, and $\phi > 0$. For some $\phi > 0$, $\sum_{t=1}^T \sum_{s=1}^T [\alpha(|t - s|)]^{2\phi/(2+\phi)} = O(T)$.

Assumption 2. (Common factors and covariates)

(i) $\{(d_t, f_t, z_t) : t \geq 1\}$ are strictly stationary and α-mixing with the mixing coefficients satisfying $\alpha_d(\tau) = O(\tau^{-\theta})$, $\alpha_f(\tau) = O(\tau^{-\theta})$, and $\alpha_z(\tau) = O(\tau^{-\theta})$, respectively.

(ii) (d_t, f_t, z_t) are distributed independently of η_{is} and v_{is} for all i, t, and s.

(iii) $E|f_t - E[f_t|z_t = z]|^{2(1+\phi)} < \infty$ and $E|d_t - E[d_t|z_t = z]|^{2(1+\phi)} < \infty$, for some $\phi > 0$.

Assumption 3. (Spatial weight matrix) $\Phi^{1/2}$ has bounded rows and column norms for all t.

Assumption 4. (Random slope coefficients) The slope coefficients β_i follow the random coefficient model

$$\beta_i = \overline{\beta} + \xi_i, \quad \xi_i \sim i.i.d.(0, \Omega_\xi), \quad \text{for} \quad i = 1, 2, \ldots, N,$$

where $\|\overline{\beta}\| < C$, $\|\Omega_\xi\| < C$, and Ω_ξ is a $p \times p$ symmetric nonnegative definite matrix, for some positive constant $C < \infty$. In addition, the random deviations ξ_i are distributed independently of $(\gamma_j, \Gamma_j, \eta_{jt}, v_{jt}, d_t, z_t, f_t)$ for all i, j, and t.

Assumption 5. (Factor loadings and rank condition)

(i) The unobserved factor loadings (γ_i, Γ_i) are bounded, i.e. $\|\gamma_i\|_2 < C$ and $\|\Gamma_i\|_2 < C$, for all i. Further, it is assumed that the random deviations ξ_i for the slope coefficients are independently distributed of (γ_i, Γ_i).
(ii) Let $\Gamma^* = E(\gamma_i, \Gamma_i) = (\gamma, \Gamma)$, $\text{Rank}(\Gamma^*) = r \le (1 + p)$.

Assumption 6. (Identification of β_i, $\overline{\beta}$, and β) Consider the covariates contained in $\lambda_t = (\overline{y}_{At}, \overline{x}_{At}, d_t)$ and let $M_{\widetilde{G}}$ be defined as in (23). The following conditions hold:

(i) The $p \times p$ matrices $T^{-1}\widetilde{X}_{i\cdot}' M_{\widehat{\Lambda}} \widetilde{X}_{i\cdot}$ and $T^{-1}\widetilde{X}_{i\cdot}' M_{\widetilde{G}} \widetilde{X}_{i\cdot}$ exist and are non-singular for all i. In addition, their corresponding inverse matrices have finite second-order moments for all i.
(ii) The matrix $\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \Sigma_{v_i}$ exists and is non-singular.
(iii) There exists T_0 and N_0 such that for all $T \ge T_0$ and $N \ge N_0$, $(T^{-1}\widetilde{X}_{i\cdot}' M_{\widehat{\Lambda}} \widetilde{X}_{i\cdot})$ and $(T^{-1}\widetilde{X}_{i\cdot}' M_{\widetilde{G}} \widetilde{X}_{i\cdot})$ exist and are nonsingular for all i. In addition, $\sup_i E \left\| \frac{\widetilde{X}_{i\cdot}' M_{\widetilde{G}} \widetilde{X}_{i\cdot}}{T} \right\| < C < \infty$.

Assumption 7. (Density function) The density of z_t satisfies $\rho_{z_t}(z) > 0$ at an interior point $z \in \mathcal{Z}$, where $\mathcal{Z}$ is the support of z_t.

Assumption 8. (Smoothness condition) The functions $m_i(z)$, $\overline{m}(z)$ and $\rho_{z_t}(z)$ are all twice continuously differentiable in the neighbourhood of $z \in \mathcal{Z}$ with bounded derivatives. Furthermore, $\mathcal{B}_X(z)$, $\mathcal{B}_\Lambda(z)$, $\mathcal{B}_D(z)$ and $\mathcal{B}_G(z)$ have continuous second derivatives in the compact support of z_t (i.e., $\mathcal{Z}$).

Assumption 9. (Kernel function) $K(u) = \prod_{l=1}^{q} k(u_l)$ is a product kernel, and the univariate kernel function $k(\cdot)$ is compactly supported and bounded such that $\int k(u)du = 1$, $\int uu'k(u)du = \mu_2(K)I_q$, and $\int k^2(u)du = R(K)$, where $\mu_2(K) \ne 0$ and $R(K) \ne 0$ are scalars and I_q is a $q \times q$ identity matrix. All odd-order moments of k vanish, that is, $\int u_1^{\iota_1}, \ldots, u_q^{\iota_q} k(u)du = 0$, for all non-negative integers $\iota_1, \ldots, \iota_q$ such that their sum is odd.

Assumption 10. (Bandwidth)

(i) Let $c_H = tr\{H^2\} + \{logT/T|H|\}^{1/2}$. The bandwidth matrix H is symmetric and positive definite, where each element of H tends to zero. As $(N,T) \xrightarrow{j} \infty$, $\sqrt{T}c_H^2 \to 0$ and $T|H| \to \infty$.

(ii) $T^{(\theta+1)}|H|^{(2+\phi)/(1+\phi)} \to \infty$, where $\theta = (2+\phi)(1+\phi)/\phi$ and $\phi > 0$.

Assumption 11. (Identification of γ_i) Consider the covariates contained in λ_t and let $M_{\widetilde{G}}$ be defined as in (23). The $\ell \times \ell$ matrices $T^{-1}\widetilde{\Lambda}' M_{\widehat{X}_{i.}}\widetilde{\Lambda}$ and $T^{-1}\widetilde{\Lambda}' M_{\widetilde{G}}\widetilde{\Lambda}$ exist and are non-singular for all i. In addition, their corresponding inverse matrices have finite second-order moments for all i.

Before considering the main asymptotic properties of the proposed estimators, we first establish the following lemma that will be key to determining the impact of the weak or long-range CSD on the rate of convergence of the above estimators.

Lemma 1. Let $\varepsilon_{it} = \left((\epsilon_{it} + v_{it}'\beta_i)\ ,\ v_{it}\right)'$ and $\overline{\varepsilon}_{At} = N^{-1}\sum_{i=1}^{N}\varepsilon_{it}$. If we assume that either $\|\beta_i\| < C$ or that Assumption 4 holds. Under Assumption 1, for each t, we have

- **(a)** $E(\overline{\varepsilon}_{At}) = 0$;
- **(b)** (Weak dependence) $\mathrm{Var}(\overline{\varepsilon}_{At}) = O\left(N^{-1}\right)$, if additionally Assumption 3 holds,
- **(c)** (Long-range dependence) $\mathrm{Var}(\overline{\varepsilon}_{At}) = O(1)$ if each ijth element of $\Phi^{1/2}$ is bounded.

This lemma guarantees that for any process of the form (6), $\overline{\varepsilon}_{At} \xrightarrow{q.m.} 0$ as $N \to \infty$ and the degree of CSD of $\varepsilon_{i.}$ will be bounded by $\nu_N = N^{-2}\iota_N'\Phi\iota_N$, where ι_N is an $N \times 1$ vector of ones. If $\nu_N = O(N^{-1})$ is assumed, we get $\mathrm{Var}(\overline{\varepsilon}_{At}) = O_p(N^{-1})$ which is analogous to the common weak dependence assumption in time series (see Chudik, Pesaran, and Tosetti, 2011). Boundedness of all rows and columns of Φ implies that $\nu_N = O(1)$, so $Var(\overline{\varepsilon}_{At}) = O(1)$, and we allow ''long-range cross-sectional dependence'' (Robinson, 2012). In this article, we use Assumption 3 which implies weak CSD.

Asymptotic properties for the heterogeneous framework

In order to analyse the main asymptotic properties of $\widehat{\beta}_i$, we may replace (3) and (4) in (15) obtaining

$$\widehat{\beta}_i - \beta_i = \left(\widehat{X}_{i.}' M_{\widehat{\Lambda}}\widehat{X}_{i.}\right)^{-1}\widehat{X}_{i.}' M_{\widehat{\Lambda}}(I_T - S)F\gamma_i + \left(\widehat{X}_{i.}' M_{\widehat{\Lambda}}\widehat{X}_{i.}\right)^{-1}\widehat{X}_{i.}' M_{\widehat{\Lambda}}\epsilon_{i.}, + o_p(c_H^2),$$

since $\widehat{X}_{i.}' M_{\widehat{\Lambda}} m_i(Z) = O_p(c_H^2)$ (see the proof of Theorem 1). In the above expression it is shown that $\widehat{\beta}_i$ depends directly on both the unobserved factors and the idiosyncratic error term. However, the following theorem shows that, its asymptotic distribution depends only on the idiosyncratic component.

Theorem 1. Consider the panel data model (3)–(6), and suppose that $\|\beta_i\| \leq C$ and Assumptions 1–3,5 and 6(i), and 7–11 hold. If it is further assumed that $\sqrt{T}/N \to 0$ and $\sqrt{T}c_H^2 \to 0$ as $(N,T) \xrightarrow{j} \infty$,

$$\sqrt{T}(\widehat{\beta}_i - \beta_i) \quad \xrightarrow{d} \quad N\left(0, \Sigma_{v_i}^{-1}\Sigma_{\epsilon_i}\Sigma_{v_i}^{-1}\right),$$

where $\Sigma_{v_i} = \lim_{T\to\infty} T^{-1} E\left[\widetilde{X}'_{i\cdot} M_{\widetilde{G}} \widetilde{X}_{i\cdot}\right]$, $\Sigma_{\epsilon_i} = \lim_{T\to\infty} T^{-1} E\left[\widetilde{X}'_{i\cdot} M_{\widetilde{G}} \Omega_{\epsilon_i} M_{\widetilde{G}} \widetilde{X}_{i\cdot}\right]$ and $\Omega_{\epsilon_i} = E(\epsilon_{i\cdot}\epsilon'_{i\cdot})$ is a $T \times T$ matrix, with $\epsilon_{i\cdot} = (\epsilon_{i1}, \ldots, \epsilon_{iT})'$ being a $T \times 1$ vector.

In Theorem 1 it is shown that the asymptotic distribution of $\widehat{\beta}_i$ will be normal if the rank condition of Assumption 5(ii) holds and if $\sqrt{T}/N \to 0$ as N and T goes to infinity. Furthermore, similar asymptotic results can be obtained whether the rank condition is violated, but that is beyond the scope of this article. See Pesaran and Tosetti (2011) or Kapetanios, Pesaran, and Yamagata (2011) for further details.

For inference reasons, it is very interesting to obtain consistent estimators of the asymptotic variance of $\widehat{\beta}_i$. Hence, using the Newey and West (1987)-type procedure, for example, that variance is given by

$$\widehat{Asy.Var}(\widehat{\beta}_i) = \left(\frac{\widehat{X}'_{i\cdot} M_{\widehat{\Lambda}} \widehat{X}_{i\cdot}}{T}\right)^{-1} \frac{\widehat{X}'_{i\cdot} M_{\widehat{\Lambda}} \widehat{\Omega}_{\epsilon_i} M_{\widehat{\Lambda}} \widehat{X}_{i\cdot}}{T} \left(\frac{\widehat{X}'_{i\cdot} M_{\widehat{\Lambda}} \widehat{X}_{i\cdot}}{T}\right)^{-1},$$

where $\widehat{\Omega}_{\epsilon_i} = T^{-1}\widehat{\epsilon}'_{i\cdot}\widehat{\epsilon}_{i\cdot}$ and $\widehat{\epsilon}_{it} = \widehat{Y}_{it} - \widehat{X}'_{it}\widehat{\beta}_i$. Replacing (3) and (4) in (17) and applying Taylor's expansion of $m_i(z_t)$ around z yields

$$\begin{aligned}\widehat{m}_i(z,H) - m_i(z) \quad &= \frac{1}{2}\iota'_1 (Z'_z K_H(z) Z_z)^{-1} Z'_z K_H(z) Q_{m_i}(z) - \iota'_1 (Z'_z K_H(z) Z_z)^{-1} Z'_z K_H(z) \Lambda(\widehat{\delta}_i - \delta_i) \\ &\quad - \iota'_1 (Z'_z K_H(z) Z_z)^{-1} Z'_z K_H(z) X_{i\cdot}(\widehat{\beta}_i - \beta_i) + \iota'_1 (Z'_z K_H(z) Z_z)^{-1} Z'_z K_H(z) e_{i\cdot} \\ &\quad + o_p(tr\{H^2\}),\end{aligned}$$

where $Q_{m_i}(z) = [(z_1 - z)'\mathcal{H}_{m_i}(z)(z_1 - z), \ldots, (z_T - z)'\mathcal{H}_{m_i}(z)(z_T - z)]'$ is a $T \times 1$ vector and $\mathcal{H}_{m_i}(\cdot)$ is the Hessian matrix of $m_i(\cdot)$. Hence, the difference between $\widehat{m}_i(z,H)$ and its true value is the sum of the above four terms plus a higher-order term that is the remainder of the Taylor expansion. More precisely, the first one is a standard bias term of local linear estimators, which contributes to the asymptotic bias; the second and third terms are due to the approximation error of the fully parametric estimates once f_t is replaced by the observable proxies λ_t, so they also give asymptotic bias. Finally, the fourth term contains the idiosyncratic errors e_{it}, which determine the variance.

Theorem 2. Consider the panel data model presented in (3)–(6) and suppose that Assumptions 1–3,5, and 7–11 hold. If $\sqrt{T|H|}tr\{H^2\} = O(1)$, as $T \to \infty$,

$$\sqrt{T|H|}\left(\widehat{m}_i(z,H) - m_i(z) - \frac{1}{2}\mu_2^q(K) tr\{H^2 \mathcal{H}_{m_i}(z)\}\right) \quad \xrightarrow{d} \quad N\left(0, \frac{\sigma^2_{\epsilon_i} R^q(K)}{\rho_{z_t}(z)}\right).$$

Theorem 2 shows that $\widehat{m}_i(z,H)$ achieves a rate of convergence of $\sqrt{T|H|}$, regardless of the rank condition assumption.

Theorem 3. Consider the panel data model presented in (3)–(6) and suppose that Assumptions 1–6(ii), 7–10 hold. If it is further assumed that $\sqrt{N}c_H^2 \to 0$, as $(N,T) \xrightarrow{j} \infty$ then

$$\sqrt{N}(\widehat{\beta}_{\mathrm{MG}} - \overline{\beta}) \quad \xrightarrow{d} \quad N\left(0, \Omega_\xi\right).$$

Theorem 4. Consider the panel data model presented in (3)-(6) and suppose that Assumptions 1–6(iii), and 7–10 hold. If it is further assumed that $\sqrt{N}c_H^2 \to 0$, as $(N,T) \xrightarrow{j} \infty$ then

$$\sqrt{N}(\widehat{\beta}_P - \overline{\beta}) \quad \xrightarrow{d} \quad N\left(0, \Psi^{*-1}R^{*}\Psi^{*-1}\right),$$

where $\Psi^* = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\Sigma_{v_i}$ and $R^* = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\Sigma_{v_i}\Omega_\xi\Sigma_{v_i}$.

Looking at the above theorems, some results should be pointed out. First, as it can be realized, the asymptotic variance-covariance matrix of both mean group and pooled estimators does not depend on the spatial correlation structure of the model. This is because by Assumption 4, the variability of β_i dominates the other sources of randomness in the model. Second, the rate of convergence of $\widehat{\beta}_{\text{MG}}$ and $\widehat{\beta}_P$ is $\sqrt{N}$, rather than the usual $\sqrt{NT}$. This is due to the rank condition in Assumption 5(ii). Finally, the pooled estimator is more efficient than the mean group one.

Furthermore, for inference reasons a consistent estimator of the asymptotic variance of the mean group estimator can be given by

$$\widehat{Asy.Var}(\widehat{\beta}_{\text{MG}}) = \frac{1}{N(N-1)}\sum_{i=1}^{N}(\widehat{\beta}_i - \widehat{\beta}_{\text{MG}})(\widehat{\beta}_i - \widehat{\beta}_{\text{MG}})', \tag{24}$$

and the corresponding estimator of the asymptotic variance of the pooled estimator is

$$\widehat{\text{Asy.Var}}(\widehat{\beta}_P) = \frac{1}{N}\Psi_{NT}^{-1}R_{NT}\Psi_{NT}^{-1}, \tag{25}$$

where $\Psi_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\widehat{X}_{i\cdot}'M_{\widehat{\Lambda}}\widehat{X}_{i\cdot}$ and

$$R_{NT} = \frac{1}{(N-1)T^2}\sum_{i=1}^{N}(\widehat{X}_{i\cdot}'M_{\widehat{\Lambda}}\widehat{X}_{i\cdot})^{-1}(\widehat{\beta}_i - \widehat{\beta}_{\text{MG}})(\widehat{\beta}_i - \widehat{\beta}_{\text{MG}})'(\widehat{X}_{i\cdot}'M_{\widehat{\Lambda}}\widehat{X}_{i\cdot})^{-1}.$$

A very nice feature of the above non-parametric variance estimators is that they avoid any misspecification problem related to the spatial weights matrix since they do not require prior knowledge of the spatial weights matrix.

Theorem 5. Consider the panel data model presented in (3)-(6) and suppose that Assumptions 1–11 hold. If it is further assumed that $\sqrt{NT|H|}tr\{H^2\} = O(1)$, as $(N,T) \xrightarrow{j} \infty$ then

$$\sqrt{T|H|}\nu_N^{-1/2}\left(\widehat{m}_{\text{MG}}(z,H) - \overline{m}(z) - \frac{1}{2}\mu_2^q(K)tr\{H^2\mathcal{H}_{\overline{m}}(z)\}\right) \quad \xrightarrow{d} \quad N\left(0, \frac{\sigma_\eta^2 R^q(K)}{\rho_{z_t}(z)}\right),$$

where $\mathcal{H}_{\overline{m}}(\cdot)$ is the Hessian matrix of $\overline{m}(\cdot)$.

The non-parametric estimators, $\widehat{m}_i(z,H)$ and $\widehat{m}_{\text{MG}}(z,H)$ exhibit the same rate of convergence, $\sqrt{T|H|}$. Therefore, efficiency gains from pooling observations over the cross-section units are not achieved. However, the reader can observe that for $\widehat{m}_{\text{MG}}(z, H)$ the rate exhibits a new element, ν_N. This term reflects the strengthening of the spatial correlation and depends directly on the particular specification of Φ. Hence, the rate of convergence of this estimator depends on the rate of ν_N, if any, as it was noted in Lemma 1. If weak dependence is assumed as in Assumption 3, the rate of convergence will be $\sqrt{NT|H|}$, whereas we will get $\sqrt{T|H|}$ if strong dependence is allowed for. Furthermore, more efficient estimators could be obtained by taking into account the spatial correlation involved. Note that a similar result is also obtained in Robinson (2012); Lee and Robinson (2015), and Soberon *et al.* (2022) in a different framework.

Finally, the optimal bandwidth is of order $O(T^{-1/(4+q)})$. However, one might choose $O((NT)^{-1/(4+q)})$. If we do so since we are smoothing only in z_t over time series observations, a smaller bandwidth than the optimal will be obtained, especially when N is large. Furthermore, special care is needed when N is large and T is moderate as the optimal rate may lead to a very small bandwidth and the non-parametric estimators can be vulnerable to denominator singularity (see Phillips and Wang, 2021 for further details).

Asymptotic properties for the homogeneous framework

We consider now the main asymptotic properties of the estimators proposed when the slope parameters are homogeneous. The following theorems can be proved following a similar reasoning as for Theorems 3–5, respectively. For the sake of brevity, the specific proofs have been omitted but are available upon request.

Theorem 6. Consider the panel data model presented in (3)-(6) and $\beta_i = \beta$, $\forall i$. Suppose that Assumptions 1–3,5,6(ii), and 7–10 hold. If it is further assumed that $\sqrt{T}/N \to 0$ and $\sqrt{NT}c_H^2 \to 0$, as $(N, T) \xrightarrow{j} \infty$

$$\sqrt{NT}(\widehat{\beta}_{H,\text{MG}} - \beta) \quad \xrightarrow{d} \quad N\left(0, \sigma_\eta^2 \Sigma_{\text{MG}}\right),$$

where $\Sigma_{\text{MG}} = \lim_{N,T\to\infty} \frac{1}{NT}\widetilde{P}'\widetilde{P}$ with $\widetilde{P}' = (\widetilde{W}'_{\cdot 1}\Phi^{1/2}, \ldots, \widetilde{W}'_{\cdot T}\Phi^{1/2})$ and $\widetilde{W}'_{\cdot t} = (\widetilde{w}_{1t}, \ldots, \widetilde{w}_{Nt})$ being $p \times NT$ matrices, whereas $\widetilde{w}_{it}$ is the tth column of $(T^{-1}V'_{i\cdot}M_{\widetilde{G}}V_{i\cdot})^{-1}V'_{i\cdot}M_{\widetilde{G}}$.

Theorem 7. Consider the panel data model presented in (3)-(6) and $\beta_i = \beta$, $\forall i$. Suppose that Assumptions 1–3,5,6(iii), and 7–10 hold. If it is further assumed that $\sqrt{T}/N \to 0$ and $\sqrt{NT}c_H^2 \to 0$, as $(N, T) \xrightarrow{j} \infty$

$$\sqrt{NT}(\widehat{\beta}_{H,P} - \beta) \quad \xrightarrow{d} \quad N\left(0, \sigma_\eta^2 \Psi_H^{*-1} R_H^* \Psi_H^{*-1}\right),$$

where $R_H^* = \lim_{N,T\to\infty} \frac{1}{NT}\widetilde{X}'(I_N \otimes M_{\widetilde{G}})'(\Phi \otimes I_T)(I_N \otimes M_{\widetilde{G}})\widetilde{X}$ and $\Psi_H^* = \lim_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^N V'_{i\cdot}M_{\widetilde{G}}V_{i\cdot}$.

Theorem 8. Consider the panel data model presented in (3)-(6) and and $\beta_i = \beta$, $\forall i$. Suppose that Assumptions 1–3 and 7–11 hold. If it is further assumed that $\sqrt{T|H|}\nu_N^{-1/2}tr\{H^2\} = O(1)$, as N and T tend to infinity

$$\sqrt{T|H|}\nu_N^{-1/2}\left(\hat{\overline{m}}_{\mathrm{MG}}(z,H)-\overline{m}(z)-\frac{1}{2}\mu_2^q(K)tr\{H^2\mathcal{H}_m(z)\}\right)\quad\xrightarrow{d}\quad N\left(0,\frac{\sigma_\eta^2 R^q(K)}{\rho_{z_t}(z)}\right).$$

Looking at the above theorems, several features must be pointed out. As it was noted in Pesaran (2006) for the fully parametric case, the convergence rates of $\widehat{\beta}_{H,\mathrm{MG}}$ and $\widehat{\beta}_{H,P}$ are $\sqrt{NT}$, instead of $\sqrt{N}$ as it was in the heterogeneous case. This is independent of the type of CSD. On its part, the rate of convergence of the non-parametric estimator depends on the rate of ν_N, if any, as it was explained in Theorem 5. Further, unlike in the heterogeneous case, the asymptotic variances of the parametric estimators depend on the particular specification of Φ. However, using results in Ibragimov and Müller (2010) it is possible to show that the robust variance estimators given by (24) and (25) are still valid for the mean group and pooled estimators when the slope parameters are homogeneous (see Pesaran and Tosetti, 2011 for a deeper discussion).

IV. Monte Carlo simulations

To investigate the extent to which the proposed estimators capture the effects of various forms of CSD, we consider two alternative sets of experiments that involve different hypotheses on the data-generating process (DGP). In the first DGP, we consider a semi-parametric model with heterogeneous slope parameters, whereas in the second DGP homogeneous slope parameters are assumed.

For both experiments, we consider the following DGP

$$\begin{aligned} y_{it} &= \alpha_i d_{it} + x_{it}'\beta_i + m_i(z_t) + \gamma_{1i}f_{1t} + \gamma_{2i}f_{2t} + \epsilon_{it}, \\ x_{it} &= A_i' d_t + g_i(z_t) + \Gamma_{1i}' f_{1t} + \Gamma_{2i}' f_{2t} + v_{it}, \end{aligned}$$

for $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$. In the above DGP, there are two individual-specific regressors, $x_{it} = (x_{1it}, x_{2it})'$, two observed common factors (z_t, d_t), two unobserved common factors (f_{1t}, f_{2t}), and three unknown functions $(m_i(z_t)$ and $g_i(z_t) = (g_{1i}(z_t), g_{2i}(z_t)))$. The observed common factor z_t is a random variable generated from a normal distribution with mean 0 and variance 1.

We next specify how to generate the individual-specific errors, unobserved factors, factor loadings, heterogeneous interaction parameters, and other aspects in the DGPs.

a. The factor loadings of the observed common factors are generated as $A_i' \sim \mathrm{IIDN}(0.5\iota_2, 0.5I_2)$, where $\iota_2 = (1,1)'$ and I_2 is a 2×2 identity matrix, and $\alpha_i \sim \mathrm{IIDN}(1,1)$, for $i = 1, \ldots, N$. As Pesaran and Tosetti (2011), α_i and A_i do not change across replications and $d_t = 1$.
b. The unobserved common factors are generated as independent stationary AR(1) processes with zero means and variances 1. More precisely, $s = 1, 2$, $f_{s,t} = 0.5f_{s,(t-1)} + (1 - 0.5^2)^{1/2}\xi_{s,t}$, where $\xi_{s,t} \sim \mathrm{IIDN}(0,1)$ across t, for $t = -49, \ldots, 0, 1, \ldots, T$.
c. The factor loadings $(\gamma_{1i}, \gamma_{2i})$ of the unobserved common factors in the y_{it} equation as generated as $\gamma_{1i} \sim \mathrm{IIDN}(0,1)$ and $\gamma_{2i} \sim \mathrm{IIDN}(0,1)$. Also, for the factor loadings of the unobserved common factors in the x_{it} equation we

consider two different cases for $\Gamma_i = (\Gamma_{1i}, \Gamma_{2i})$ that we denote by A and B, respectively: $\text{vec}(\Gamma_i) = (\Gamma_{11,i}, \Gamma_{12,i}, \Gamma_{21,i}, \Gamma_{22,i})' \sim \text{IIDN}(\Gamma_\tau, I_4)$, $\tau = A, B$. In case A, $\Gamma_A = (1, 0, 0, 1)'$, so the rank condition in Assumption 3.1 is satisfied, whereas in Case B, $\Gamma_B = (1, 1, 0, 0)'$, and the rank condition is not satisfied.

d. The idiosyncratic errors ϵ_{it} of y_{it} are generated according to the following SAR model: $\epsilon_{\cdot t} = \Phi^{1/2}\eta_{\cdot t}$, where $\Phi^{1/2} = (I_N - \theta_0 W_N)^{-1}$, $\eta_{\cdot t}$ is a $N \times 1$ vector generated as independent $N(0, 1)$, and θ_0 is the autoregressive parameter which takes three different values (i.e. 0.3, 0.6, 0.9). Also, W_N is a spatial weight matrix generated from independent $N(0, 1)$ random variables. Specifically, the weights are constructed so W_N is a row-normalized spatial weight based on an exponential distance decay function whose typical element is such as $\varpi_{ij} = exp(-\vartheta_{ij})/\sum_j exp(-\vartheta_{ij})$, where ϑ_{ij} is the distance between units given by the Euclidean distance. On its part, the individual-specific errors of x_{it} are generated independently of each other as stationary AR(1) processes: $v_{s,it} = \rho_{v_{si}} v_{s,i(t-1)} + (1 - \rho^2_{v_{si}})^{1/2}\vartheta_{s,it}$, where $\rho_{v_{si}} \sim \text{IIDU}(0.05, 0.95)$ and $\vartheta_{s,it}$ are i.i.d. $N(0, 1)$ across i and t. For each i, the three processes ϵ_{it}, v_{1it}, and v_{2it} are generated independently of each other.

e. The unknown functions are generated as $m_i(z_t) = \exp(z_t)/(1 + \exp(z_t)) + \varphi_i(0.5z_t - 0.25z_t^2)$, $g_{1i}(z_t) = (1 + \varphi_{1i})(1 + \sin(10z_t))$, and $g_{2i}(z_t) = (1 + \varphi_{2i})\sin(2z_t)$, where $\varphi_i \sim \text{IIDU}(0, 1)$, $\varphi_{1i} \sim \text{IIDU}(0, 0.1^2)$, and $\varphi_{2i} \sim \text{IIDU}(0, 0.1^2)$.

Note that the first 50 observations of v_{1it}, v_{2it}, f_{1t}, and f_{2t} are discarded. Further, two alternative assumptions on the slope coefficients are considered. In particular, heterogeneous slopes are assumed in DGP1 with $\beta_{s,i} = \beta_s + \psi_{s,i}$ where $\beta_s = 1$ and $\psi_{s,i} \sim \text{IIDN}(0, 0.04)$, for $i = 1, 2, \ldots, N$ and $s = 1, 2$ varying across replications, while homogeneous slope parameters are allowed in DGP2 with $\beta_{s,i} = 1$. Each experiment was replicated 1,000 times for $N = 100{,}140, 200$ and T to be either $(25, 50, 75)$. Also, the Epanechnikov kernel $k(u) = 0.75(1 - u^2)\mathbb{1}\{|u| \leq 1\}$ was used and we choose $H = I_q h_0$, where $h_0 = c_0\widehat{\sigma}_z T^{-1/5}$ is the bandwidth term, $\widehat{\sigma}_z$ the sample SD of the smoothing variable $\{z_t\}_{t=1,\ldots,T}$, and $c_0 = 2.34$.

For evaluation of the performance of our estimators, we use the bias, root mean squared errors (RMSE), and coverage rate for the slope parameters, whereas the RMSE is computed for the regression functions. In what follows, we shall focus on β_1, since results for β_2 are very similar and will not be reported. Results for the full rank experiments (case A) and the rank deficient experiments (case B) are summarized in Table 1 and Figure 1. This tables represent the results for the heterogeneous slope case. The corresponding tables and figures for the homogeneous slope setting are relegated to the Appendix C.

Overall, the Monte Carlo results confirm the good performance of the proposed estimators in finite samples. More precisely, the mean group and pooled estimators display very small biases, their RMSEs decline steadily with increases in N and/or T for the different experiments, and their coverage rate oscillates around the standard significance level even in the smallest case. Further, the asymptotic efficiency of the mean group estimators relative to the pooled estimators is confirmed, although the differences between the two estimators are rather slight for relatively large samples. This general conclusion also holds on the rank-deficient case.

Table 1

Small sample properties of the parametric estimators under slope heterogeneity

		Pooled			Mean group			Pooled			Mean group		
θ_0	$N \setminus T$	25	50	75	25	50	75	25	50	75	25	50	75
		Case A: Full rank						Case B: rank deficient					
BIAS ($\times$100)													
0.3	100	0.057	0.040	0.012	0.003	0.009	0.020	0.129	0.035	0.003	0.070	0.022	0.007
	140	0.066	0.066	−0.018	0.080	0.067	−0.009	0.027	0.045	0.013	0.032	0.032	0.010
	200	0.125	0.001	0.042	0.092	0.000	0.023	0.130	0.004	0.054	0.096	0.003	0.032
0.6	100	0.060	0.042	0.010	0.004	0.011	0.018	0.132	0.038	0.001	0.071	0.025	0.005
	140	0.067	0.069	−0.017	0.082	0.069	−0.009	0.029	0.047	0.013	0.034	0.033	0.010
	200	0.127	0.002	0.043	0.094	0.001	0.024	0.081	−0.015	0.035	0.050	−0.007	0.002
0.9	100	0.062	0.044	0.002	0.006	0.014	0.008	0.136	0.041	−0.007	0.072	0.029	−0.005
	140	0.075	0.081	−0.012	0.091	0.075	−0.007	0.043	0.060	0.013	0.045	0.040	0.008
	200	0.129	0.004	0.054	0.100	0.003	0.032	0.083	−0.013	0.040	0.053	−0.005	0.022
RMSE ($\times$100)													
0.3	100	2.771	2.619	1.944	2.670	2.335	1.700	2.588	2.656	1.921	2.535	2.344	1.674
	140	2.108	1.617	1.381	2.069	1.502	1.237	1.896	1.459	1.228	1.900	1.383	1.130
	200	1.858	1.565	1.235	1.788	1.434	1.099	1.860	1.567	1.267	1.790	1.436	1.136
0.6	100	2.770	2.649	2.021	2.667	2.335	1.703	2.587	2.657	1.927	2.533	2.344	1.674
	140	2.111	1.621	1.385	2.070	1.504	1.241	1.900	1.465	1.233	1.903	1.386	1.135
	200	1.858	1.565	1.236	1.788	1.434	1.101	1.679	1.511	1.171	1.670	1.360	1.13045
0.9	100	2.773	2.624	2.131	2.669	2.338	1.865	2.591	2.660	2.103	2.536	2.346	1.837
	140	2.174	1.692	1.470	2.117	1.561	1.321	1.962	1.548	1.339	1.955	1.451	1.231
	200	1.860	1.566	1.267	1.790	1.436	1.136	1.679	1.512	1.211	1.672	1.361	1.088
Coverage rate													
0.3	100	0.955	0.944	0.950	0.955	0.960	0.955	0.952	0.949	0.947	0.952	0.957	0.954
	140	0.947	0.949	0.954	0.950	0.950	0.951	0.953	0.941	0.944	0.957	0.950	0.952
	200	0.951	0.954	0.952	0.950	0.951	0.955	0.950	0.958	0.952	0.950	0.955	0.954
0.6	100	0.951	0.953	0.953	0.960	0.962	0.956	0.950	0.949	0.947	0.954	0.958	0.951
	140	0.948	0.949	0.953	0.950	0.950	0.949	0.955	0.942	0.941	0.957	0.952	0.950
	200	0.949	0.952	0.952	0.948	0.952	0.955	0.951	0.958	0.953	0.950	0.955	0.954
0.9	100	0.953	0.944	0.952	0.954	0.962	0.948	0.952	0.951	0.950	0.955	0.958	0.952
	140	0.950	0.944	0.950	0.945	0.949	0.944	0.953	0.951	0.944	0.958	0.950	0.948
	200	0.949	0.946	0.946	0.946	0.956	0.953	0.953	0.957	0.951	0.954	0.956	0.948

Considering now the results in Figure 1 it can be noted that, as it was expected, the RMSEs of the mean group non-parametric estimators shrink to zero as the sample size increases. Furthermore, all these results corroborate that both parametric and non-parametric estimates are robust to the presence of spatially correlated errors since their RMSEs are steady independently of the θ_0 value.

V. A new look at the knowledge capital production function

Following the discussion already established in section I, we propose to estimate (2) for which we build an annual country-level balanced panel data set covering 24 OECD countries from 1973 to 2014. We use a different dataset than Eberhardt *et al.* (2013) for two main reasons. The first reason is a computational one, as exploiting a balanced data set greatly facilitates the computations, even though the proposed estimators can be

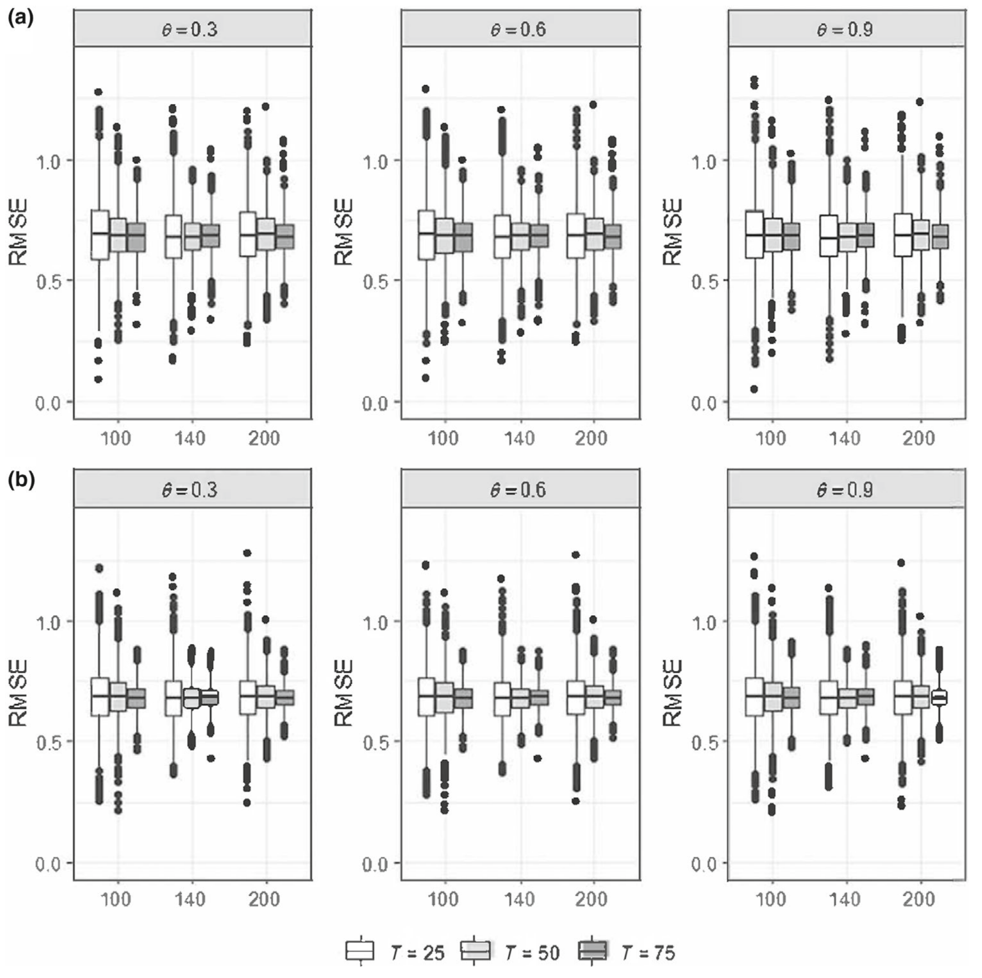

Figure 1. Boxplots of the RMSE values of the non-parametric estimates in 1,000 independent simulations under slope heterogeneity

adapted to the case of unbalanced panels. Second, this is also useful to provide novel and complementary results.

As for the dependent variable (Y_{it}), we use the real GDP at constant prices from Penn World Table version 9.0 (Feenstra, Inklaar, and Timmer, 2015, PWT9). As for the explanatory variables, the capital stock (K_{it}) measured at constant prices is also collected from PWT9. Then, for the labour input (L_{it}), following (Henderson and Parmeter, 2015, p. 142-144) we build a textit‘‘human capital augmented labor’ variable, where employment is also collected from PWT9 while human capital stock is computed as in Ertur and Musolesi (2017). To build an R&D stock variable (R_{it}), we consider gross domestic expenditure on research and development (GERD) flow values collected from the OECD-STATS database. Missing values are filled in a similar way as in Coe, Helpman, and Hoffmaister (2009), and then we calculate the GERD stock using a perpetual inventory method as in Coe and Helpman (1995), assuming the depreciation rate to be 0.05. As for z_t, we introduce oil price shocks.

Table 2
Heterogeneous fully parametric and semi-parametric results.

	Parametric				Semi-parametric
	(i)	(ii)	(iii)	(iv)	(v)
$\ln L_{it}$	0.585***	0.588***	0.599***	0.600***	0.547***
	(0.073)	(0.073)	(0.072)	(0.071)	(0.063)
$\ln K_{it}$	0.308**	0.307**	0.270**	0.267**	0.414***
	(0.098)	(0.010)	(0.096)	(0.092)	(0.123)
$\ln R_{it}$	0.059***	0.060***	0.059***	0.058***	0.056***
	(0.017)	(0.017)	(0.017)	(0.016)	(0.018)
oil		−3.37e−10*	−3.78e−10*	−4.49e−10*	
		(1.38e−10)	(1.58e−10)	(1.89e−10)	
oil^2			5.01e−11*	−1.69e−10*	
			(2.09e−11)	(7.15e−11)	
oil^3				−1.74e−11*	
				(7.36e−12)	
Elasticity of scale	0.952	0.955	0.928	0.925	1.017

Note: The t-values are within brackets.
Significant at ***1%, **5%, and *10% levels.

Previous works have discussed what is an appropriate measure of oil shocks, and in particular, it has been suggested to use an oil shock measure that filters out both price declines and price increases (Hamilton, 1996). Following this reasoning, we similarly construct an oil shock index as Davis and Haltiwanger (2001). Our index equals the log of the ratio of the current crude real oil prices divided by the average of the real prices in the previous 5 years. Note that stationarity of the observed stochastic common covariates is fundamental for valid estimation. In Data S1, we depict some univariate plots, i.e. the oil price index evolution over time and the corresponding autocorrelation function and partial autocorrelation function. These plots suggest that the price oil index is a slightly persistent time series and is consistent with a first-order auto-regressive stationary process. Furthermore, in the Appendix D some formal statistical testing is also conducted. The results support that the oil price index is a stationary time series process.

In Table 2 we consider the heterogeneous slope parameters framework and provide mean group estimates of the elasticities of capital stock, labour, and R&D.[2] In addition, with the aim of assessing the potential misspecification related to the oil price index, different specifications are considered (see columns (ii)−(iv)). In column (ii) we estimate a linear and heterogeneous effect of the oil price, that is, by imposing $m_i(z_t) = \phi_i z_t$ in (2). In column (iii), we estimate the possible nonlinear effect of the oil price by considering a second-order polynomial function, $m_i(z_t) = \phi_{1i} z_t + \phi_{2i} z_t^2$. Also, in column (iv) we consider a third-order polynomial specification of the oil price. Finally, based on Theorem 3 we provide asymptotic standard errors of parameter estimators.

It would be also possible to estimate the technological parameters in (2) under a homogeneous framework. However, the assumptions needed for this estimator to be consistent are not fulfiled by our data (see Theorem 6 for a more detailed discussion) and therefore we prefer not to use them. Moreover, homogeneity in the slope parameters

[2]Pooled estimates of these elasticities are available upon request.

can be questioned also from an economic viewpoint as a theoretical foundation for heterogeneous slope parameters across countries can be found in the ''new growth'' literature, which argues that technology differs across countries. While (Brock and Durlauf, 2001, pp. 8–9) remark that *the assumption of parameter homogeneity seems particularly inappropriate when one is studying complex heterogeneous objects such as countries*, in Durlauf, Kourtellos, and Minkin (2001) it is also suggested that the explanatory power of the Solow growth model is substantially enhanced by allowing for country-specific production functions.

Looking at the results in Table 2, the parametric CCE model (column (*i*)) provides estimated output elasticities with respect to labour, capital, and R&D stock that are equal to 0.585, 0.308, and 0.059, respectively, with an estimated elasticity of scale equals to 0.952. These estimates are similar to those obtained in Eberhardt *et al.* (2013) and Millo (2019). Furthermore, the estimated output elasticity with respect to R&D is consistent with the huge amount of literature surveyed by Hall, Mairesse, and Mohnen (2010). As it can be observed, the oil price variable, both in linear, quadratic, and cubic terms (columns (ii)–(iv)), appears as significant, despite the estimated parameters are implausibly small. This result contradicts a huge amount of empirical literature that has been previously mentioned and it warns us about a possible misspecification error. Indeed, a tacit assumption of the parametric approach is that the curve can be represented in terms of the parametric model. By contrast, non-parametric modelling of a regression relationship does not project the observed data into a Procrustean bed of a fixed parametrization. Then, instead of assuming a parametric functional form for z_t, we opt to estimate it non-parametrically. In order to do so we use our estimator proposed in (19). To compute it, we use the Epanechnikov kernel $k(u) = 0.75(1 - u^2)\mathbb{1}\{|u| \leq 1\}$ and choose $H = I_q h_0$, where $h_0 = c_0 \widehat{\sigma}_z T^{-1/5}$ is the bandwidth term, $\widehat{\sigma}_z$ the sample SD of z_t, and $c_0 = 2.34$. This choice for h_0 fulfils Assumption 10, but we have also performed a sensitivity analysis with different values. Parameter estimates were rather invariant to the bandwidth choice.

Allowing for a non-parametric function $m_i(\cdot)$ may be important to avoid a misspecification bias not only with respect to the estimated effect of the common factors – here, the oil prices index – but also with respect to the estimation of the parameters of interest – here, the technological parameters in a Griliches-type production function. We do find that this bias is empirically sizeable as when estimating the semi-parametric model (column (*v*)) the estimated coefficient of labour decreases to 0.547, while that of capital increases up to 0.414, which represents an increase of about 50% with respect to the specifications where the oil price index enters with a parametric form (columns (*ii*) to (*iv*)). This result is interesting as it is consistent with the reasoning by Romer (1987), who stresses that the true elasticity of output with respect to capital should be greater than the share of capital in total income because of positive externalities associated with investments. Overall, these results now indicate slightly increasing returns to scale, with an estimated elasticity of scale equal to 1.017, which is about 6% to 8% higher with respect to that obtained when the oil price index enters into the model parametrically.

Moreover, and very importantly, as far as the effect of the oil price index is concerned, Figure 2 depicts the estimated functional relationship. The elasticity between output and oil price index is clearly nonlinear: it is about 0.02 for negative values of the explanatory variable, which corresponds to a decrease in oil prices, and then starts to decrease and

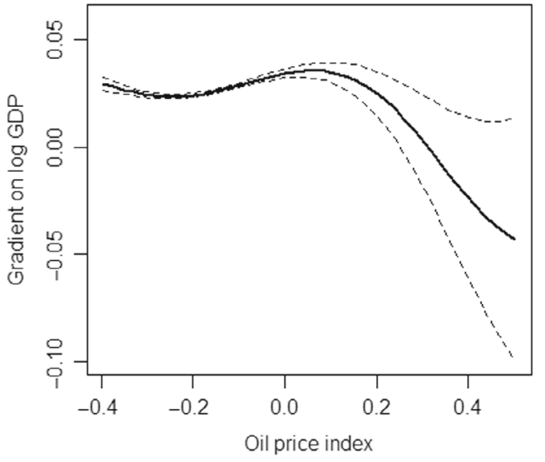

Figure 2. Elasticity of GDP with respect to oil price index

becomes negative for positive values of the oil price index, which correspond to oil price increases, up to −0.05, for the highest positive values of the index. Note that the dotted lines indicate the 95% pointwise confidence interval. This result complements some previous time series evidence and provides additional evidence of an asymmetric effect of oil prices on GDP (Lescaroux and Mignon, 2008; Hamilton, 2003). Although the purpose of this section is simply to illustrate the usefulness of the proposed econometric method, to the best of our knowledge, this article represents the first attempt (i) to estimate the output elasticity with respect to the oil price within a production function framework by exploiting semi-parametric panel data models with several sources of CSD and (ii) to assess the bias associated to the technological parameters when erroneously imposing a parametric specification for the oil price.

VI. Conclusions

In the analysis of the Griliches' knowledge capital production function, at both sectoral and country level, previous works have pointed out the relevance of incorporating slope heterogeneity in the technological parameters and CSD arising from unobserved common factors. However, additional issues to be addressed might be (i) the presence of spatial dependence in the idiosyncratic error term and (ii) possible nonlinear effects of relevant common observed variables, such as the oil price. Accordingly, in this article we have introduced a partially linear panel data model that faces up all previously mentioned drawbacks simultaneouly, namely (i) functional specification; (ii) CSD arising simultaneously from common factors and spatial dependence and (iii) heterogeneous relationships among variables.

Extending the CCE approach to this semi-parametric framework, a $\sqrt{T}$-consistent estimator for the heterogeneous slope parameters has been proposed and an estimator for the non-parametric component has been also obtained. Furthermore, several alternative estimators based on cross-sectional means and pooled data have been also proposed. Their asymptotic properties were obtained under quite standard assumptions in this literature and the theoretical findings were further supported for small samples sizes via several Montecarlo experiments and an empirical illustration. For the latter, we revisit the knowledge capital production function *à la* Griliches. The empirical results have highlighted that modelling the common covariates (here, an oil prices index) with a non-parametric smooth function greatly affects the elasticities estimates. All that has remarked a sizeable empirical bias when estimating a fully parametric model, and ultimately it has enabled us to obtain results that appear to be more reliable, richer, and more consistent with respect to economic theory than those obtained from a fully parametric model.

Final Manuscript Received: July 2023

References

Ackerberg, D. A., Caves, K. and Frazer, G. (2015). 'Identification properties of recent production function estimators', *Econometrica*, Vol. 83, pp. 2411–2451.

Andrews, D. W. K. (2005). 'Cross-section regression with common shocks', *Econometrica*, Vol. 73, pp. 1551–1585.

Antonioli, D., Gioldasis, G. and Musolesi, A. (2021). 'Estimating a non-neutral production function', *Oxford Economic Papers*, Vol. 73, pp. 856–878.

Arbia, G. (2006). *Spatial Econometrics: Statistical Foundations and Applications to Regional Analysis*, Springer-Verlag, Berlin.

Bai, J. (2009). 'Panel data models with interactive fixed effects', *Econometrica*, Vol. 77, pp. 1229–1279.

Bailey, N., Holly, S. and Pesaran, M. H. (2016). 'A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence', *Journal of Applied Econometrics*, Vol. 31, pp. 249–280.

Brock, W. A. and Durlauf, S. N. (2001). 'What have we learned from a decade of empirical research on growth? Growth empirics and reality', *The World Bank Economic Review*, Vol. 15, pp. 229–272.

Chudik, A., Pesaran, M. H. and Tosetti, E. (2011). 'Weak and strong cross-section dependence and estimation of large panels', *The Econometrics Journal*, Vol. 14, pp. C45–C90.

Cliff, A. and Ord, C. (1972). 'Testing for spatial autocorrelation among regression residuals', *Geographical Analysis*, Vol. 4, pp. 267–284.

Coe, D. T. and Helpman, E. (1995). 'International R&D spillovers', *European Economic Review*, Vol. 39, pp. 859–887.

Coe, D. T., Helpman, E. and Hoffmaister, A. W. (2009). 'International R&D spillovers and institutions', *European Economic Review*, Vol. 53, pp. 723–741.

Davis, S. J. and Haltiwanger, J. (2001). 'Sectoral job creation and destruction responses to oil price changes', *Journal of Monetary Economics*, Vol. 48, pp. 465–512.

Doraszelski, U. and Jaumandreu, J. (2013). 'R&D and productivity: estimating endogenous productivity', *Review of Economic Studies*, Vol. 80, pp. 1338–1383.

Durlauf, S. N., Kourtellos, A. and Minkin, A. (2001). 'The local Solow growth model', *European Economic Review*, Vol. 45, pp. 928–940.

Eberhardt, M., Helmers, C. and Strauss, H. (2013). 'Do spillovers matter when estimating private returns to R&D?', *Review of Economics and Statistics*, Vol. 95, pp. 436–448.

Ertur, C. and Musolesi, A. (2017). 'Weak and strong cross-sectional dependence: a panel data analysis of international technology diffusion', *Journal of Applied Econometrics*, Vol. 32, pp. 477–503.

Fan, J. and Huang, T. (2005). 'Profile likelihood inferences on semiparametric varying-coefficient partially linear models', *Bernoulli*, Vol. 11, pp. 1031–1057.

Feenstra, R. C., Inklaar, R. and Timmer, M. P. (2015). 'The next generation of the Penn World Table', *American Economic Review*, Vol. 105, pp. 3150–3182.

Griliches, Z. (1979). 'Issues in assessing the contribution of research and development to productivity growth', *The Bell Journal of Economics*, Vol. 10, pp. 92–116.

Hall, B. H., Mairesse, J. and Mohnen, P. (2010). 'Measuring the returns to R&D', in Hall B. H., and Rosenberg N. (eds), *Handbook of the Economics of Innovation,* Vol. 2. Amsterdampp: North-Holland, 1033–1082.

Hamilton, J. D. (1996). 'This is what happened to the oil price-macroeconomy relationship', *Journal of Monetary Economics*, Vol. 38, pp. 215–220.

Hamilton, J. D. (2003). 'What is an oil shock?', *Journal of Econometrics*, Vol. 113, pp. 363–398.

Heckman, J. J. (2001). 'Micro data heterogeneity and evaluation of public policy: Nobel lecture', *Journal of Political Economy*, Vol. 109, pp. 673–748.

Henderson, D. J. and Parmeter, C. F. (2015). *Applied Nonparametric Econometrics*, Cambridge: Cambridge University Press.

Holly, S., Pesaran, M. and Yamagata, T. (2010). 'A spatio-temporal model of house prices in the USA', *Journal of Econometrics*, Vol. 158, pp. 160–173.

Huang, X. (2013). 'Nonparametric estimation in large panels with cross-sectional dependence', *Econometric Reviews*, Vol. 32, pp. 754–777.

Ibragimov, R. and Müller, U. (2010). 't-statistic based correlation and heterogeneity robust inference', *Journal of Business and Economic Statistics*, Vol. 28, pp. 453–468.

Ivaldi, M., Ladoux, N., Ossard, H. and Simioni, M. (1996). 'Comparing Fourier and translog specifications of multiproduct technology: evidence from an incomplete panel of French farmers', *Journal of Applied Econometrics*, Vol. 11, pp. 649–667.

Kapetanios, G., Pesaran, M. H. and Yamagata, T. (2011). 'Panels with non-stationary multifactor error structures', *Journal of Econometrics*, Vol. 160, pp. 326–348.

Keane, M. P. and Prasad, E. S. (1996). 'The employment and wage effects of oil price changes: a sectoral analysis', *The Review of Economics and Statistics*, Vol. 78, pp. 389–400.

Kilian, L. (2009). 'Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market', *American Economic Review*, Vol. 99, pp. 1053–1069.

Lee, J. and Robinson, P. (2015). 'Panel nonparametric regression with fixed effects', *Journal of Econometrics*, Vol. 188, pp. 346–362.

Lee, L.-F. and Yu, J. (2010). 'Estimation of spatial autoregressive panel data models with fixed effects', *Journal of Econometrics*, Vol. 154, pp. 165–185.

Lescaroux, F. and Mignon, V. (2008). 'On the influence of oil prices on economic activity and other macroeconomic and financial variables', *OPEC Energy Review*, Vol. 32, pp. 343–380.

Ma, S., Racine, J. S. and Ullah, A. (2020). 'Nonparametric estimation of marginal effects in regression-spline random effects models', *Econometric Reviews*, Vol. 39, pp. 792–825.

Millo, G. (2019). 'Private returns to R&D in the presence of spillovers, revisited', *Journal of Applied Econometrics*, Vol. 34, pp. 155–159.

Mork, K. A. (1989). 'Oil and the macroeconomy when prices go up and down: an extension of Hamilton's results', *Journal of Political Economy*, Vol. 97, pp. 740–744.

Newey, W. K. and West, K. D. (1987). 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica*, Vol. 55, pp. 703–708.

Pesaran, M. H. (2006). 'Estimation and inference in large heterogeneous panels with a multifactor error structure', *Econometrica*, Vol. 74, pp. 967–1012.

Pesaran, M. H. and Smith, R. (1995). 'Estimating long-run relationships from dynamic heterogeneous panels', *Journal of Econometrics*, Vol. 68, pp. 79–113.

Pesaran, M. H. and Tosetti, E. (2011). 'Large panels with common factors and spatial correlation', *Journal of Econometrics*, Vol. 161, pp. 182–202.

Phillips, P. and Wang, Y. (2021). 'Functional coefficient panel modeling with communal smoothing covariates', *Journal of Econometrics*, Vol. 227, pp. 371–407.

Robinson, P. M. (2012). 'Nonparametric trending regression with cross-sectional dependence', *Journal of Econometrics*, Vol. 169, pp. 4–14.

Romer, P. M. (1987). 'Crazy explanations for the productivity slowdown', *NBER Macroeconomics Annual*, Vol. 2, pp. 163–202.

Rosenblatt, M. (1956). 'Remarks on some nonparametric estimates of a density function', *The Annals of Mathematical Statistics*, Vol. 27, pp. 832–837.

Shi, W. and Lee, L.-f. (2017). 'Spatial dynamic panel data models with interactive fixed effects', *Journal of Econometrics*, Vol. 197, pp. 323–347.
Soberon, A., Rodriguez-Poo, J. M. and Robinson, P. M. (2022). 'Nonparametric panel data regression with parametric cross-sectional dependence', *The Econometrics Journal*, Vol. 25, pp. 114–133.
Su, J. and Jin, S. (2012). 'Sieve estimation of panel data models with cross section dependence', *Journal of Econometrics*, Vol. 169, pp. 34–47.

Supporting Information

Additional Supporting Information may be found in the online Appendix:

Data S1. Appendices.

Data replication package: the data replication package is available at https://doi.org/10.3886/E188961