Journal Pre-proof

Efficient estimation of a partially linear panel data model with cross-sectional dependence

Alexandra Soberon, Massimiliano Mazzanti, Antonio Musolesi, Juan M. Rodriguez-Poo

PII: S0047-259X(24)00100-3
DOI: https://doi.org/10.1016/j.jmva.2024.105393
Reference: YJMVA 105393

To appear in: *Journal of Multivariate Analysis*

Received date : 22 April 2024
Revised date : 16 November 2024
Accepted date : 18 November 2024

Please cite this article as: A. Soberon, M. Mazzanti, A. Musolesi et al., Efficient estimation of a partially linear panel data model with cross-sectional dependence, *Journal of Multivariate Analysis* (2024), doi: https://doi.org/10.1016/j.jmva.2024.105393.

Efficient estimation of a partially linear panel data model with cross-sectional dependence

Alexandra Soberon[a,*], Massimiliano Mazzanti[b], Antonio Musolesi[b], Juan M. Rodriguez-Poo[a]

[a]*Department of Economics, Universidad de Cantabria, Spain*
[b]*Department of Economics and Business, University of Ferrara, Italy*

Abstract

This paper considers efficiency improvements in a partially linear panel data model that accounts for possible nonlinear effects of common covariates and allows for cross-sectional dependence arising simultaneously from unobserved common factors and spatial dependence. A generalized least squares-type estimator is proposed by taking into account this dependence structure. Also, possible gains in terms of the rate of convergence are studied. A Monte Carlo study is carried out to investigate the proposed estimators' finite sample performance. Further, an empirical application is conducted to assess the impact of the carbon price linked to the European Union Emission Trading System on carbon dioxide emissions.

Keywords: Climate policy effects, Cross-sectional dependence, European Union Emissions Trading System Partially linear models, Semiparametric efficient estimators.
2020 MSC: Primary 62G05, Secondary 62P12

1. Introduction

Nonparametric and semiparametric panel data models traditionally assume independence across individuals. However, economic agents (regions, states, or countries, among others) are typically interdependent due to externalities, spillovers, or common shocks. Therefore, ignoring this type of dependence, typically known as cross-sectional dependence (CSD), may be inappropriate as standard estimation procedures can lead to inefficient and even inconsistent estimators, as shown in [26, 43] and the references therein. Recently, the question of how to characterize CSD has received considerable attention, emerging two prominent (non-exclusive) strands in the literature.

On the one hand, the multifactor error approach states that the correlation structure can be characterized by the presence of a finite number of unobserved common factors that affect all individuals with different intensities. A prominent approach within this strand of the literature is the common correlated effect estimator (CCE) introduced by [39], which has also been extended to a fully nonparametric regression setting by [27, 50], for example.

On the other hand, the spatial econometric approach assumes that the correlation structure can be modeled through a pre-specified spatial weight matrix that may depend on either the geographic locations of the cross-sectional units or more general economic variables. In this case, the question of efficiency improvements using the correlation structure emerges naturally, independently of whether the dependence is allowed in either time or cross-sectional dimension (or both). For example, [33, 45] assume an unknown structure of the CSD and shown that a simple Nadaraya-Watson estimator is dominated by a Generalized Least Squares (GLS)-type one in efficiency terms, under some conditions on the rate at which the cross-sectional dimension, N, is allowed to growth with the time series length, T.

While these two approaches are developed separately, several empirical problems have led researchers to pay more attention to the development of consistent estimation procedures in the presence of both types of dependence. In a fully parametric setting with both heterogeneous and homogeneous slope parameters, [42] considers both types of CSD. However, these types of results are scarce in the nonparametric literature. Recently, [48] extends the CCE

*Corresponding author. Email address: alexandra.soberon@unican.es

approach to consistently estimate the parameters of interest in a partially linear panel data model. This model treats the common observed variables in a nonparametric way and accounts for both the presence of unknown common factors and spatial dependence. Nevertheless, different results in terms of efficiency are obtained for the heterogeneous and homogeneous cases. When the slope parameters are heterogeneous, it is shown that the asymptotic variance does not depend on the spatial correlation structure of the model. Unfortunately, this is not the case in the homogeneous slope parameters case. Under this latter setting, the asymptotic variance of the resulting estimator depends on the particular specification of the correlation structure, so it is possible to obtain more efficient estimators.

The above results suggest the possibility of achieving efficiency gains in such semiparametric specifications. This is precisely the focus of our paper: proposing a GLS-type estimation technique that delivers more efficient estimators for a partially linear model. This model features homogeneous slope parameters, heterogeneous smooth functions for the common observed variables, and accounts for both types of CSD. According to our theoretical results, efficiency gains will not only affect the inference analysis, but can also lead to a substantially different estimate of the shape of the unknown curve and the parameter estimates.

This noteworthy result is further corroborated by a simulation exercise and an empirical illustrative example where a macroeconomic panel dataset is exploited to assess the impact of the European Union Emissions Trading System (EU ETS) on CO_2 emissions. To do so, we propose a partially linear Environmental Kuznets Curve (EKC) specification (see [6, 36], among others) where the key policy variable, which is invariant across units, is the price of carbon linked to the European market of allowances. Since there is a high degree of uncertainty surrounding the shape (and sign) of this policy effect and there are no ex-ante theoretical or empirical reasons to impose a specific parametric relation between CO_2 emissions and the price of polluting, a nonparametric relationship between these two variables is adopted. Furthermore, following the postulates in the EKC literature, the rest of the covariates are specified through a fully parametric model. Additionally, unobservable common factors, which could affect individual countries in a heterogeneous manner [35], as well as spatial dependence [46] are also introduced simultaneously.

The rest of the paper is organized as follows. Section 2 introduces the model and the estimation method. Section 3 refers to the efficient estimation techniques. Section 4 presents some Monte Carlo simulations to analyze the finite sample performance of the proposed estimators, while Section 5 applies that methodology to evaluate the effect of the EU ETS on CO_2 emissions. Section 6 concludes the paper. All mathematical proofs are relegated to the Appendix.

2. Econometric model and estimation procedures

Let y_{it} be the response variable for the cross-sectional unit i at time t. We consider the following partially linear panel data model with both unobserved common factors and spatially correlated errors,

$$y_{it} = \alpha_i^\top d_t + x_{it}^\top \beta + m_i(z_t) + \gamma_i^\top f_t + \epsilon_{it}, \quad i \in \{1, \ldots, N\}, \quad t \in \{1, \ldots, T\}, \tag{1}$$

where x_{it} is a $p \times 1$ vector of individual-specific explanatory variables and $d_t = (d_{1t}, \ldots, d_{nt})^\top$ is a $n \times 1$ vector that could contain deterministic terms or commonly observed variables that enter linearly in the model. We allow for the presence of observed continuous common stochastic covariates (common policy effects), $z_t \in \boldsymbol{R}^q$, that enter through a nonparametric heterogeneous function (i.e., $m_i(\cdot)$) in the model. Further, f_t is a $r \times 1$ vector of unobserved common factors that are allowed to simultaneously affect all cross-section units, albeit with different intensities measured with the factor loadings, γ_i, and ϵ_{it} is the idiosyncratic error term. In addition, β and $m_i(\cdot)$ are unknown objects that need to be estimated. Through the paper, we assume that, if an intercept term exists, it is included in d_t. If this is the case, in order to identify $m_i(\cdot)$, we need to impose the following condition: $E[m_i(z_t)] = 0$ for each i.

In general, as it is noted in [39, 50], among others, the unobserved common factors, f_t, are allowed to be correlated with the observed data (x_{it}, z_t, d_t). The covariates x_{it} are control variables that are determined in the system and z_t can be political variables that are common among the cross-sectional units [2] and are not determined in the system. These variables can be technological, institutional, environmental, or health factors. The x_{it}'s variables are determined in the system according to the following fairly general specification:

$$x_{it} = A_i^\top d_t + g_i(z_t) + \Gamma_i^\top f_t + v_{it}, \tag{2}$$

where A_i and Γ_i are $n \times p$ and $r \times p$ factor loadings matrices with fixed components, respectively, v_{it} is a $p \times 1$ vector of individual-specific components of x_{it}, and $g_i(z_t)$ is a $p \times 1$ vector of unknown smooth functions. The spatial dependence is introduced by assuming that the idiosyncratic error term, ϵ_{it}, is conditionally correlated and heteroscedastic.

The econometric model introduced in (1) and (2) can be motivated from various perspectives. One rationale for such a specification is that in several circumstances such as wage, cost, or production functions, parametric specifications for the main explanatory variables are well established and build on economic theory. However, there is generally a high degree of uncertainty surrounding the way in which common observed variables may affect the cross-sectional units. For example, allowing for flexible forms may be suitable when focusing on the effect of real common shocks on productivity or economic growth or when estimating the effects of oil prices on wages, employment, or production activity [24, 30]. In our empirical application, we will consider an extended EKC model where y_{it} stands for CO_2 emissions per capita, the vector x_{it} contains standard explanatory variables in the EKC (such as GDP per capita, its square, and research and development activities), and the common observed variable z_t is the carbon price linked to the European market of allowances that is allowed to have a nonlinear heterogeneous effect among countries [2].

To estimate β and $m_i(\cdot)$, [48] make an extension of the CCE approach of [39] from fully parametric to partially linear models and proposes to approximate the unobserved factors, f_t, by a suitable proxy that does not depend on an initial estimate of β and $m_i(\cdot)$. In particular, in its Online Appendix A it is shown that f_t can be approximated by the cross-sectional averages of the observed variables (y_{it}, x_{it}) assuming: i) rank$(\Gamma^*) = r \le (1+p)$ for sufficiently large N, where $\Gamma^* = E(\gamma_i, \Gamma_i) = (\gamma, \Gamma)$; ii) $N^{-1}\sum_{i=1}^N v_{it} \overset{q.m.}{\to} 0$ and $N^{-1}\sum_{i=1}^N \epsilon_{it} \overset{q.m.}{\to} 0$ for each t; and iii) $N^{-1}\sum_{i=1}^N g_i(z_t)$ and $N^{-1}\sum_{i=1}^N m_i(z_t)$ are twice-continuously differentiable in the neighborhood of $z \in$int$(\mathcal{Z})$, where $\mathcal{Z}$ is the support of z_t. Following this approach and let $\overline{y}_{At} = N^{-1}\sum_{i=1}^N y_{it}$ and $\overline{x}_{At} = N^{-1}\sum_{i=1}^N x_{it}$, we propose to approximate f_t by some linear function of $\lambda_t = (\overline{y}_{At}, \overline{x}_{At}, d_t)$ that is a $\ell \times 1$ vector of observable proxies for f_t with $\ell = (1+p+n)$ plus a term $o_p(1)$. Hence, the following augmented regression model is considered

$$y_{it} = x_{it}^\top \beta + m_i(z_t) + \delta_i^\top \lambda_t + e_{it}, \quad i \in \{1, \dots, N\}, \quad t \in \{1, \dots, T\}, \tag{3}$$

where $e_{it} = \epsilon_{it} + o_p(1)$ is the error term and δ_i is a $\ell \times 1$ vector of nuisance parameters.

Let Z_z be a $T \times (1+q)$ matrix whose t-th element is $Z_{z_t} = [1, (z_t - z)^\top]$ for z being a fixed point and $K_{H_1}(z)$ be a $T \times T$ diagonal matrix as $K_{H_1}(z) = diag\{K_{H_1}(z_1 - z), \dots, K_{H_1}(z_T - z)\}$, where H_1 is a $q \times q$ symmetric and positive definite matrix and $K(\cdot)$ is a nonnegative product kernel function such that, for each u, $K_{H_1}(u) = |H_1|^{-1}\prod_{j=1}^q k(H_1^{-1}u_j)$ where $u = (u_1, \dots, u_q)^\top$ and $k(\cdot)$ is a univariate kernel function. Assuming that $Z_z^\top K_{H_1}(z) Z_z$ is invertible, in [48] is shown that, for $i \in \{1, \dots, N\}$, the following nonparametric estimator can be proposed for $m_i(\cdot)$,

$$\widetilde{m}_i(z, H_1) = \iota_1^\top \left\{Z_z^\top K_{H_1}(z) Z_z\right\}^{-1} Z_z^\top K_{H_1}(z)\left(Y_{i\cdot} - X_{i\cdot}\beta - \Lambda\delta_i\right), \tag{4}$$

where $Y_{i\cdot} \equiv (y_{i1}, \dots, y_{iT})^\top$ is a $T \times 1$ vector, $X_{i\cdot} \equiv (x_{i1}, \dots, x_{iT})^\top$ and $\Lambda \equiv (\lambda_1, \dots, \lambda_T)^\top$ are $T \times p$ and $T \times \ell$ matrices, respectively, and ι_1 is a $(1+q) \times 1$ vector having 1 in the first entry and 0 in all other entries.

Nevertheless, $\widetilde{m}_i(z, H_1)$ is an infeasible estimator since it depends on the unknown parameters (δ_i, β). To overcome it, we propose to rewrite (4) in matrix notation by denoting $\widetilde{m}_i(Z, H_1) = S[Y_{i\cdot} - X_{i\cdot}\beta - \Lambda\delta_i]$, where S is a $T \times T$ smoothing matrix that only depends on the values of z_t and whose definition is apparent from (4). Hence, plugging the resulting expression in (3) we get the following regression model

$$\widehat{Y}_{i\cdot} = \widehat{X}_{i\cdot}\beta + \widehat{\Lambda}\delta_i + \widehat{e}_{0,i\cdot}, \tag{5}$$

where $\widehat{Y} = (I_T - S)Y_{i\cdot}$, $\widehat{X}_{i\cdot} = (I_T - S)X_{i\cdot}$, and $\widehat{\Lambda} = (I_T - S)\Lambda$. Also, $\widehat{e}_{0,i\cdot} = (\widehat{e}_{0,i1}, \dots, \widehat{e}_{0,iT})^\top$ is a $T \times 1$ vector whose tth element is $\widehat{e}_{0,it} = e_{it} + O_p\left\{tr(H_1^2)\right\}$. Then, by assuming that $\sum_{i=1}^N \widehat{X}_{i\cdot}^\top M_{\widehat{\Lambda}}\widehat{X}_{i\cdot}$ and $\widehat{\Lambda}^\top M_{\widehat{X}_i}\widehat{\Lambda}$ are invertible matrices, where $M_{\widehat{\Lambda}} = I_T - \widehat{\Lambda}\left(\widehat{\Lambda}^\top\widehat{\Lambda}\right)^{-1}\widehat{\Lambda}^\top$ and $M_{\widehat{X}_i} = I_T - \widehat{X}_{i\cdot}\left(\widehat{X}_{i\cdot}^\top\widehat{X}_{i\cdot}\right)^{-1}\widehat{X}_{i\cdot}^\top$ are $T \times T$ projection matrices, and following the procedure in [48], we obtain

$$\widehat{\beta} = \left(\sum_{i=1}^N \widehat{X}_{i\cdot}^\top M_{\widehat{\Lambda}}\widehat{X}_{i\cdot}\right)^{-1}\sum_{i=1}^N \widehat{X}_{i\cdot}^\top M_{\widehat{\Lambda}}\widehat{Y}_{i\cdot}, \tag{6}$$

$$\widehat{m}_i(z, H_1) = \iota_1^\top \left\{Z_z^\top K_{H_1}(z) Z_z\right\}^{-1} Z_z^\top K_{H_1}(z)\left(Y_{i\cdot} - X_{i\cdot}\widehat{\beta} - \Lambda\widehat{\delta}_i\right), \tag{7}$$

$$\widehat{\overline{m}}(z; H_1) = \iota_1^\top \left\{Z_z^\top K_{H_1}(z) Z_z\right\}^{-1} Z_z^\top K_{H_1}(z)\left(\overline{Y}_A - \overline{X}_A\widehat{\beta} - \Lambda\widehat{\overline{\delta}}\right), \tag{8}$$

where $\widehat{\overline{\delta}} = N^{-1}\sum_{i=1}^{N}\widehat{\delta}_i$ for $\widehat{\delta}_i = \left(\widehat{\Lambda}^{\top}M_{\widehat{X}_i}\widehat{\Lambda}\right)^{-1}\widehat{\Lambda}^{\top}M_{\widehat{X}_i}\widehat{Y}_{i\cdot}$. Also, $\overline{X}_A \equiv \left(\overline{x}_{A1}^{\top},\ldots,\overline{x}_{AT}^{\top}\right)^{\top}$ is a $T\times p$ matrix and $\overline{Y}_A \equiv (\overline{y}_{A1},\ldots,\overline{y}_{AT})^{\top}$ is a $T\times 1$ vector. If d_t contains a constant term, we need to impose some identification conditions to obtain unique estimators for both $m_i(\cdot)$ and $\overline{m}(\cdot) = E[m_i(\cdot)]$. More precisely, to identify $m_i(\cdot)$ we impose the condition $E[m_i(z_t)] = 0$, so the proposed estimator for $m_i(z_t)$ is $\widehat{m}_i^{*}(z_t;H_1) = \widehat{m}_i(z_t;H_1) - E[\widehat{m}_i(z_t;H_1)]$, whereas to identify $\overline{m}(\cdot)$ we impose $\sum_{i=1}^{N}\alpha_i = 0$ and the proposed estimator for $\overline{m}(z_t)$ is (8).

The asymptotic properties of the estimators derived in (6)–(8) have been already obtained in [48] in a different context where the spatial dependence follows a rather restricted form. However, a more general form for the correlation structure is allowed in this paper by assuming that the conditional variance-covariance matrix of the idiosyncratic error term is an unknown smooth function of the common observable variable, z_t. Hence, to obtain the limiting behavior of these estimators under this more general setting, we need to introduce the following definitions and assumptions.

Let $\rho_{z_t}(z)$ be the pdf of z_t, $\mathcal{Z}$ be the support of z_t, and $D \equiv (d_1,\ldots,d_T)^{\top}$ be a $T\times N$ matrix. Denote $\widetilde{X}_{i\cdot} = X_{i\cdot} - \mathcal{B}_X(z)$, $\widetilde{\Lambda} = \Lambda - \mathcal{B}_{\Lambda}(z)$, $\widetilde{F} = F - \mathcal{B}_F(z)$, $\widetilde{D} = D - \mathcal{B}_D(z)$, and $\widetilde{X}^{(\varpi)} = X^{(\varpi)} - \mathcal{B}_{X^{(\varpi)}}(z)$, where $\mathcal{B}_X(z) = E(X_{i\cdot}|z_t = z)\rho_{z_t}(z)$, $\mathcal{B}_{\Lambda}(z) = E(\Lambda|z_t = z)\rho_{z_t}(z)$, $\mathcal{B}_F(z) = E[F|z_t = z]\rho_{z_t}(z)$, $\mathcal{B}_D(z) = E[D|z_t = z]\rho_{z_t}(z)$, and $\mathcal{B}_{X^{(\varpi)}}(z) = E[X^{(\varpi)}|z_t = z]\rho_{z_t}(z)$. We define $M_{\widetilde{G}} = I_T - \widetilde{G}(\widetilde{G}^{\top}\widetilde{G})^{-1}\widetilde{G}^{\top}$ as a $T\times T$ projection matrix, where $\widetilde{G} = (\widetilde{D},\widetilde{F})$ is a $T\times(N+r)$ matrix.

Assumption 1. For $t\in\{1,\ldots,T\}$ and $i\in\{1,\ldots,N\}$, $E(\epsilon_{it}|z_t = z) = 0$. Further, for $t = s$, $E\left(\epsilon_{\cdot t}\epsilon_{\cdot t}^{\top}\middle|z_t\right) = \Omega_N(z_t)$ is a $N\times N$ matrix, and for $t\neq s$, $E\left(\epsilon_{\cdot t}\epsilon_{\cdot s}^{\top}\middle|z_t,z_s\right) = 0$. Let $\Omega(Z) = diag_{t\in\{1,\ldots,T\}}\{\Omega_N(z_t)\}$ and $\Omega_N(z_t) = \left\{\omega_{ij}(z_t)\right\}_{i,j\in\{1,\ldots,N\}}$. The functions $\omega_{ij}(z)$ have uniformly bounded derivative of second order at z, where $z\in\mathrm{int}(\mathcal{Z})$, and $\Omega(Z)$ is nonsingular.

Assumption 2. The $(n+r+q)\times 1$ vector of common components $(d_t^{\top}, f_t^{\top}, z_t^{\top})^{\top}$ is covariance stationary with absolute summable autocovariances, distributed independently of the individual-specific errors, ϵ_{it} and v_{it}, for all i and t.

Assumption 3. The individual-specific errors ϵ_{it} and $v_{jt'}$ are distributed independently for all i, j, t and t', and for each i, v_{it} follows a linear stationary process with absolute summable autocovariances given by $v_{it} = \sum_{\tau=0}^{\infty}S_{i\tau}\vartheta_{i,t-\tau}$, where, for each i, ϑ_{it} is a $p\times 1$ vector of serially uncorrelated random variables with mean zero, I_p variance matrix, and finite fourth-order cumulants. For each i, the coefficient matrices $S_{i\tau}$ satisfy the condition $E\left(v_{it}v_{it}^{\top}\right) = \sum_{\tau=0}^{\infty}S_{i\tau}S_{i\tau}^{\top} = \Sigma_{v_i} \leq C < \infty$, where Σ_{v_i} is a $p\times p$ positive definite matrix such that $\sup_i\|\Sigma_{v_i}\|_2 < \infty$ and C is some positive constant.

Assumption 4. The unobserved factor loadings (γ_i,Γ_i) are bounded, i.e., $\|\gamma_i\|_2 < C$ and $\|\Gamma_i\|_2 < C$, for all i.

Assumption 5. Let $\Gamma^{*} = E(\gamma_i,\Gamma_i) = (\gamma,\Gamma)$, $\mathrm{rank}(\Gamma^{*}) = r \leq (p+1)$.

Assumption 6. The following $p\times p$ matrices $(NT)^{-1}\sum_{i=1}^{N}\widetilde{X}_{i\cdot}^{\top}M_{\widetilde{\Lambda}}\widetilde{X}_{i\cdot}$ and $(NT)^{-1}\sum_{i=1}^{N}\widetilde{X}_{i\cdot}^{\top}M_{\widetilde{G}}\widetilde{X}_{i\cdot}$ exist and are nonsingular. They also have finite second-order moments.

Assumption 7. The probability density function of z_t, $\rho_{z_t}(\cdot)$, is continuous and bounded away from zero. Also, $\rho_{z_t}(\cdot)$, $g_i(\cdot)$, $m_i(\cdot)$, and $\overline{m}(\cdot)$ have bounded derivatives of order two in a neighborhood of $z\in\mathrm{int}(\mathcal{Z})$.

Assumption 8. All second-order derivatives of $E(\lambda_t|z_t)$, $E(\overline{x}_{At}|z_t)$, and $E(\overline{y}_{At}|z_t)$ are bounded and uniformly continuous at z, where $z\in\mathrm{int}(\mathcal{Z})$.

Assumption 9. $K(u) = \prod_{j=1}^{q}k(u_j)$ is a product kernel, and the univariate kernel function $k(\cdot)$ is compactly supported and bounded such that $\int k(u)du = 1$, $\int u^2k(u)du = \mu_2(K)$, and $\int k^2(u)du = R(K)$, where $\mu_2(K)\neq 0$ and $R(K)\neq 0$ are scalars. 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. Let $c_{H_1} = tr(H_1^2) + (\ln T/T|H_1|)^{1/2}$. The bandwidth matrix H_1 is symmetric and positive definite, where each element of H_1 tends to zero. As $(N,T)\to\infty$, $\sqrt{N}c_{H_1}^2\to 0$, $\sqrt{NT}c_{H_1}^2\to 0$, $NT|H_1|\to\infty$, and $T|H_1|\to\infty$.

Assumption 11. For some $\varsigma > 0$, $E[|\epsilon_{it}|^{(2+\varsigma)}|z_t = z]$ exists and is bounded.

For the sake of generality, $\omega_{ij}(z)$ is considered in Assumption 1 as an unknown smooth function that needs to be estimated. Assumptions 2-5 are rather common conditions concerning the individual-specific errors of x_{it}, common factors, and rank condition (see [39, 42] for further details). Assumption 6 is required to identify β. In addition, Assumptions 7-8 are standard smoothness and boundedness conditions on the density function and moment functionals.

Assumptions 9-10 are kernel and bandwidth conditions quite common in the local linear literature, and Assumption 11 is required for the Lyapunov condition. Note that the kernel function having a compact support in Assumption 9 is imposed for the sake of brevity and can be removed at the cost of lengthy proofs. Specifically, this assumption implies that the product kernel satisfies $\int vv^\top K(v)dv = \mu_2(K)I_q$ and $\int K^2(v)dv = R^q(K)$, where I_q is a $q \times q$ identity matrix, and the Gaussian kernel is allowed.

Theorem 1. *Suppose that Assumptions 1-10 hold, $\widehat{\beta}$ and $\widehat{\delta}_i$ are consistent estimators for β and δ_i, respectively. If it is further assumed that $\sqrt{T}c^2_{H_1} \to 0$ and $\sqrt{T}/N \to 0$, as $(N, T) \to \infty$,*

$$\sqrt{NT}(\widehat{\beta} - \beta) \quad \xrightarrow{d} \quad N\left(0, Q^{-1}\Psi Q^{-1}\right),$$

where $\Psi = \lim_{N,T\to\infty}(NT)^{-1}E\left\{\widetilde{X}^\top \left(I_N \otimes M_{\widetilde{G}}\right)^\top \Omega(Z)\left(I_N \otimes M_{\widetilde{G}}\right)\widetilde{X}\right\}$ and $Q = \lim_{N,T\to\infty}(NT)^{-1}\sum_{i=1}^N E\left(\widetilde{X}_{i\cdot}^\top M_{\widetilde{G}}\widetilde{X}_{i\cdot}\right)$ are $p \times p$ matrices, and $\widetilde{X}$ is a $NT \times p$ matrix.

The proof of this theorem is done in the Appendix. Theorem 1 shows that $\widehat{\beta}$ is a root-NT consistent estimator of β in the presence of unobserved common factors. Nevertheless, the asymptotic variance depends on the particular specification of $\Omega(Z)$. Therefore, an alternative estimator with better asymptotic properties in terms of variance-reduction can be obtained by considering this correlation structure.

Theorem 2. *Suppose that Assumptions 1-11 hold and that $\sqrt{T|H_1|}tr(H_1^2) = O(1)$, as $(N, T) \to \infty$, then*

$$\sqrt{T|H_1|}\left[\widehat{m}_i(z; H_1) - m_i(z) - \frac{1}{2}\mu_2^q(K)tr\left\{H_1^2\mathcal{H}_{m_i}(z)\right\}\right] \quad \xrightarrow{d} \quad N\left(0, \frac{\omega_{ii}(z)R^q(K)}{\rho_{z_t}(z)}\right),$$

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

The proof of Theorem 2 follows directly from the proof of Theorem 2.1 in [47]. In Theorem 2 it is shown that $\widehat{m}_i(\cdot; H_1)$ is asymptotically normal and exhibits a rate of convergence of order $\sqrt{T|H_1|}$, regardless of the rank condition (see Assumption 5) holds. Nevertheless, $\widehat{m}_i(z; H_1)$ completely ignores the information that characterizes the idiosyncratic error term (see Assumption 1).

Theorem 3. *Suppose that Assumptions 1-11 hold and that $\sqrt{T|H_1|\nu_N^{-1}(z)}tr(H_1^2) = O(1)$, as $(N, T) \to \infty$, then*

$$\sqrt{T|H_1|\nu_N^{-1}(z)}\left[\widehat{\overline{m}}(z, H_1) - \overline{m}(z) - \frac{1}{2}\mu_2^q(K)tr\left\{H_1^2\mathcal{H}_{\overline{m}}(z)\right\}\right] \quad \xrightarrow{d} \quad N\left(0, \frac{R^q(K)}{\rho_{z_t}(z)}\right),$$

where $\nu_N(z) = N^{-2}\iota_N^\top E(\epsilon_{\cdot t}\epsilon_{\cdot t}^\top|z_t = z)\iota_N$ is a scalar term and $\mathcal{H}_{\overline{m}}(\cdot)$ is the Hessian matrix of $\overline{m}(\cdot)$.

In Theorem 3 is shown that $\widehat{\overline{m}}(\cdot; H_1)$ is asymptotically normal, but the rate of convergence is different concerning the one obtained for $\widehat{m}_i(\cdot; H_1)$. It shows a new element, $\nu_N(z)$, which reflects the strengthening of the spatial correlation and heteroscedasticity and depends directly on the particular specification of $\Omega(Z)$. Then, more efficient estimators could be obtained by considering the information in $\Omega(Z)$. Furthermore, unlike the parametric estimator, the rate of convergence of this nonparametric estimator depends on the rate of increase of $\nu_N(z)$, if any. Therefore, under weak spatial dependence, $\nu_N(z) = O(N^{-1})$, the rate of convergence is of order $(NT|H_1|)^{-1/2}$, whereas it is $(T|H_1|)^{-1/2}$ under strong spatial dependence, i.e., $\nu_N(z) = O(1)$. Note that the proof of this theorem is done following a similar proof scheme as the corresponding for Theorem 2.2 in [49] in a different context, and it is therefore omitted.

3. Efficient estimation techniques

In this section, we propose alternative GLS-type estimators that are more efficient than those presented in (6) and (8). We will show that the efficiency gains not only affect the inference analysis, but can also lead to a different estimate of the shape of the unknown curve and the parameter estimates. Specifically, efficiency improvements are achieved by considering the potential CSD incorporated in Assumption 1.

Let $V_{\widehat{\Lambda}}$ be a $T \times (T - \ell)$ orthonormal eigenvector matrix of $M_{\widehat{\Lambda}}$ that corresponds to the eigenvalues of one. We propose to premultiply both sides of the regression model (5) by $V_{\widehat{\Lambda}}^{\top}$. By stacking the resulting observations over NT we can denote $\widehat{Y}$ and $\widehat{e_0}$ as $NT \times 1$ vectors whose ith elements are $\widehat{Y_{i\cdot}}$ and $\widehat{e}_{0,i\cdot}$, respectively, and $\widehat{X}$ is a $NT \times p$ matrix whose ith elements is $\widehat{X}_{i\cdot}$. Hence, the resulting regression model is of the form

$$\widehat{Y}^{*} = \widehat{X}^{*}\beta + e^{*}, \tag{9}$$

where $\widehat{Y}^{*} = \left(I_N \otimes V_{\widehat{\Lambda}}\right)^{\top} \widehat{Y}$ and $e^{*} = \left(I_N \otimes V_{\widehat{\Lambda}}\right)^{\top} \widehat{0,e}$ are $NT \times 1$ vectors, $\otimes$ is the Kronecker product, and $\widehat{X}^{*} = \left(I_N \otimes V_{\widehat{\Lambda}}\right)^{\top} \widehat{X}$ is a $NT \times p$ matrix. Using the fact that $\Omega_{e^{*}} \equiv \mathrm{Var}(e^{*}|Z) = \left(I_N \otimes V_{\widehat{\Lambda}}^{\top}\right) \Omega(Z) \left(I_N \otimes V_{\widehat{\Lambda}}\right)$ and $V_{\widehat{\Lambda}} V_{\widehat{\Lambda}}^{\top} = M_{\widehat{\Lambda}}$, we propose to premultiply (9) by $\Omega_{e^{*}}^{-1/2}$ to obtain

$$\Omega_{e^{*}}^{-1/2}\widehat{Y}^{*} = \Omega_{e^{*}}^{-1/2}\widehat{X}^{*}\beta + \Omega_{e^{*}}^{-1/2}e^{*},$$

and the resulting Generalized Least Squares (GLS) estimator for β is

$$\widehat{\beta}_{GLS} = \left\{\widehat{X}^{\top}\left(I_N \otimes M_{\widehat{\Lambda}}\right)\Omega^{-1}(Z)\left(I_N \otimes M_{\widehat{\Lambda}}\right)\widehat{X}\right\}^{-1}\widehat{X}^{\top}\left(I_N \otimes M_{\widehat{\Lambda}}\right)\Omega^{-1}(Z)\left(I_N \otimes M_{\widehat{\Lambda}}\right)\widehat{Y}. \tag{10}$$

However, this GLS estimator is again infeasible since it depends on $\Omega(Z)$ that is generally unknown. To overcome it, we propose the following estimator for $\widehat{\Omega}(Z) = diag_{t\in\{1,\dots,T\}}\left\{\widehat{\Omega}_N(z_t)\right\}$ with

$$\widehat{\Omega}_N(z) = \frac{\sum_{t=1}^{T} K^{*}_{H_2}(z_t - z)\widehat{e}_{\cdot t}\widehat{e}_{\cdot t}^{\top}}{\sum_{t=1}^{T} K^{*}_{H_2}(z_t - z)}, \tag{11}$$

where $K^{*}(\cdot)$ is a nonnegative kernel function as the defined in (4), H_2 is a $q \times q$ symmetric and positive definite matrix, and $\widehat{e}_{\cdot t} \equiv (\widehat{e}_{1t}, \dots, \widehat{e}_{Nt})^{\top}$ is a $N \times 1$ vector of residuals defined as $\widehat{e}_{it} = y_{it} - x_{it}^{\top}\widehat{\beta} - \widehat{m}_i(z_t; H_1) - \widehat{\delta}_i^{\top}\lambda_t$. Note that H_2 satisfies different conditions from H_1 and will thus be chosen differently. Therefore, replacing $\Omega(Z)$ by $\widehat{\Omega}(Z)$ in (10) we get the Feasible Generalized Least Square Estimator,

$$\widehat{\beta}_{FGLS} = \left\{\widehat{X}^{\top}\left(I_N \otimes M_{\widehat{\Lambda}}\right)\widehat{\Omega}^{-1}(Z)\left(I_N \otimes M_{\widehat{\Lambda}}\right)\widehat{X}\right\}^{-1}\widehat{X}^{\top}\left(I_N \otimes M_{\widehat{\Lambda}}\right)\widehat{\Omega}^{-1}(Z)\left(I_N \otimes M_{\widehat{\Lambda}}\right)\widehat{Y}. \tag{12}$$

Focusing now on the nonparametric estimator for $\overline{m}(\cdot)$, we rewrite the model to estimate in matrix form obtaining

$$Y_{\cdot t} - X_{\cdot t}\beta - \Delta\lambda_t = \iota_N \overline{m}(z_t) + U_{\cdot t}, \tag{13}$$

where $Y_{\cdot t} \equiv (y_{1t}, \dots, y_{Nt})^{\top}$ and $U_{\cdot t} \equiv (u_{1t}, \dots, u_{Nt})^{\top}$ are $N \times 1$ vectors, for $u_{it} = \epsilon_{it} + [m_i(z_t) - \overline{m}(z_t)] + o_p(1)$, whereas $X_{\cdot t}$ and Δ are $N \times p$ and $N \times \ell$ matrices, respectively. Following [33] and [49], among others, and by imposing the identification condition $\varpi^{\top}\iota_N = 1$ to identify $\overline{m}(\cdot)$, we premultiply (13) by a given $N \times 1$ weight vector ϖ obtaining

$$\varpi^{\top}(Y_{\cdot t} - X_{\cdot t}\beta - \Delta\lambda_t) = \overline{m}(z_t) + \varpi^{\top}U_{\cdot t}. \tag{14}$$

Note that if d_t contains a constant term, we follow [49] and impose the following identification conditions $\varpi^{\top}\iota_N = 1$ and $\varpi^{\top}\alpha = 0$, where $\alpha = (\alpha_1, \dots, \alpha_N)^{\top}$. To estimate this regression model we choose ϖ to minimize $\mathrm{Var}(\varpi^{\top}U_{\cdot t}|z_t) = \varpi^{\top}\Phi_N(z_t)\varpi$, subject to $\varpi^{\top}\iota_N = 1$, where $\Phi_N(z_t) = E(u_{\cdot t}u_{\cdot t}^{\top}|z_t) = \{\varphi_{ij}(z_t)\}_{i,j\in\{1,\dots,N\}}$. By solving this optimization problem we obtain

$$\varpi^{*}(z) = \left\{\iota_N^{\top}\Phi_N^{-1}(z)\iota_N\right\}^{-1}\Phi_N^{-1}(z)\iota_N. \tag{15}$$

Replacing (15) into (14) and following a similar procedure as in the previous section, the following GLS weighted local-least squares estimator for $\overline{m}(\cdot)$ is proposed

$$\widehat{\overline{m}}_{GLS}(z; H_1, \varpi) = \iota_1^{\top}\left\{Z_z^{\top}K_{H_1}(z)Z_z\right\}^{-1}Z_z^{\top}K_{H_1}(z)\widetilde{Y}\varpi, \tag{16}$$

where $\widetilde{Y}$ is a $T \times N$ matrix whose it-th element is such as $\widetilde{y}_{it} = y_{it} - x_{it}^{\top}\beta - \lambda_t^{\top}\delta_i$. Finally, using the definition of u_{it} and applying Assumption 1 we obtain $\mathrm{Var}(u_{it}|z) \equiv \varphi_{ij}(z) = \omega_{ij}(z) - \{m_i(z) - \overline{m}(z)\}^2 + o_p(1)$, for $i, j \in \{1, \dots, N\}$.

Since β, δ_i, and ϖ are unknown elements this estimator is unfeasible. However, following a similar procedure as in (11), with $\widehat{u}_{it} = y_{it} - x_{it}^\top\widehat{\beta} - \widehat{\delta}_i^\top\lambda_t - \widehat{\overline{m}}(z_t; H_1)$ instead of $\widehat{e}_{it}$, it is possible to obtain a consistent estimator for $\Phi_N(z)$, i.e., $\widehat{\Phi}_N(z)$. Therefore, the resulting FGLS weighted local-least squares estimator for $\overline{m}(\cdot)$ is

$$\widehat{\overline{m}}_{FGLS}(z; H_1, \widehat{\varpi}) = \iota_1^\top \left\{Z_z^\top K_{H_1}(z)Z_z\right\}^{-1} Z_z^\top K_{H_1}(z)\widehat{\widetilde{Y}}\widehat{\varpi}, \tag{17}$$

where $\widehat{\widetilde{Y}}$ is a $T \times N$ matrix whose it-th element is $\widehat{\widetilde{y}}_{it} = y_{it} - x_{it}^\top\widehat{\beta} - \lambda_t^\top\widehat{\delta}_i$ and

$$\widehat{\varpi} = \left\{\iota_N^\top\widehat{\Phi}_N^{-1}(z)\iota_N\right\}^{-1}\widehat{\Phi}_N^{-1}(z)\iota_N.$$

To obtain the asymptotic properties of the GLS estimators in (10) and (16), the following additional conditions are required.

Assumption 12. $(NT)^{-1}E\left\{\widetilde{X}^\top\left(I_N \otimes M_{\widehat{\Lambda}}\right)^\top\Omega^{-1}(Z)\left(I_N \otimes M_{\widehat{\Lambda}}\right)\widetilde{X}\right\}$ and $(NT)^{-1}E\left\{\widetilde{X}^\top\left(I_N \otimes M_{\widetilde{G}}\right)^\top\Omega^{-1}(Z)\left(I_N \otimes M_{\widetilde{G}}\right)\widetilde{X}\right\}$ are $p \times p$ matrices that exist and are non-singular. They also have finite second-order moments.

Assumption 13. $K^*(u) = \prod_{j=1}^q k^*(u_j)$ is a product kernel where the univariate kernel function $k^*(\cdot)$ is even and uniformly bounded with bounded support. Moreover, $k^*(\cdot)$ is integrable on the bounded support.

Assumption 14. The bandwidth matrix H_2 is symmetric and positive definite, where each element of H_2 tends to zero. As $(N, T) \to \infty$, $T|H_1| \to \infty$, $T|H_1|^2 = o(|H_2|)$, and $N^3/T|H_1| + Ntr(H_2^2)/tr(H_1^2) \to 0$.

Assumption 15. $\widehat{\rho}(\cdot)$ and $\widehat{\mathcal{H}}_{\overline{m}}(\cdot)$ are consistent estimators of $\rho(\cdot)$ and $\mathcal{H}_{\overline{m}}(\cdot)$, respectively, where $\mathcal{H}_m(\cdot) = \partial m(\cdot)/\partial z\partial z^\top$, and it holds

$$\begin{aligned}
\rho_z(z) - \widehat{\rho}(z) &= O_p\left(\|\Omega_N(z)\|^{-1}\left\|\widehat{\Omega}_N(z) - \Omega_N(z)\right\|\right), \\
\{\mathcal{H}_{\overline{m}}(z)\}^2 - \left\{\widehat{\mathcal{H}}_{\overline{m}}(z)\right\}^2 &= O_p\left(\|\Omega_N(z)\|^{-1}\left\|\widehat{\Omega}_N(z) - \Omega_N(z)\right\|\right).
\end{aligned}$$

Assumptions 1 and 13 together help to ensure that the bias of each element of the estimators of $\Omega(z)$ are $O_p\left\{tr(H_2^2)\right\}$. Assumption 14 shows the relationship between H_1, H_2, N, and T. They are necessary to show the consistency of these efficient estimators. Assumption 15 is required to establish the asymptotic theory of the efficient estimators without involving too much technicality and simplify the proofs.

Assumption 16. Let $X^\varpi \equiv (X_{1\cdot}, \ldots, X_{N\cdot})\varpi$ be a $T \times d$ matrix, the matrices $T^{-1}Z_z^\top K_{H_1}(z)X^\varpi$ and $T^{-1}Z_z^\top H_{H_1}(z)\Lambda$ exist.

Assumption 17. For some $\varsigma > 0$, $E[|u_{it}|^{(2+\varsigma)}|z_t = z]$ exists and is bounded.

Assumption 18. As $N \to \infty$, $\left\|\Phi_N^{-1}(z)\right\| + N\left\{\iota_N^\top\Phi_N^{-1}(z)\iota_N\right\}^{-2}\iota_N^\top\Phi_N^{-2}(z)\iota_N = O_p(1)$.

Furthermore, to obtain an efficient estimator of the unknown function, Assumption 1 imposes the smoothness of the covariance function. Assumption 17 is necessary to check the Lyapunov condition for the CLT. Finally, Assumption 18 was discussed in detail in [45] where it was found that a sufficient (but not necessary) condition for the second term on the left-hand side to be bounded is that the largest eigenvalue of $\Phi_N(z)$ is bounded.

Theorem 4. *Suppose that Assumptions 1-5, 7-10, and 12 hold, the GLS estimator of β is consistent. If it is further assumed that $\sqrt{T}/N \to 0$ as $(N, T) \to \infty$,*

$$\sqrt{NT}\left(\widehat{\beta}_{GLS} - \beta\right) \xrightarrow{d} N\left(0, Q_\varpi^{-1}\right),$$

where $Q_\varpi = \lim_{N,T\to\infty}(NT)^{-1}E\left\{\widetilde{X}^\top\left(I_N \otimes M_{\widetilde{G}}\right)^\top\Omega^{-1}(Z)\left(I_N \otimes M_{\widetilde{G}}\right)\widetilde{X}\right\}$.

Theorem 5. *Suppose that Assumptions 1-5, 7-10, 13-14, and 16-17 hold and denote* $\nu_N^{(\varpi)}(z) = \left\{\iota_N^\top \Phi_N^{-1}(z)\iota_N\right\}^{-1}$. *If it is further assumed* $\sqrt{T|H_1|\left\{\nu_N^{(\varpi)}(z)\right\}^{-1}} tr(H_1^2) = O(1)$ *as* $(N,T)\to\infty$,

$$\sqrt{T|H_1|\left\{\nu_N^{(\varpi)}(z)\right\}^{-1}}\left[\widehat{\widehat{\overline{m}}}_{GLS}(z,H_1,\varpi) - \overline{m}(z) - \frac{\mu_2^q(K)}{2} tr\left\{H_1^2\mathcal{H}_{\overline{m}}(z)\right\}\right] \quad \xrightarrow{d} \quad N\left(0, \frac{R^q(K)}{\rho_{z_t}(z)}\right).$$

Theorem 4 shows that there exists an efficiency gain in $\widehat{\beta}_{GLS}$ with respect to $\widehat{\beta}$. On its part, for N and T sufficiently large, in Theorem 5 is proved that the distribution of $\widehat{\widehat{\overline{m}}}_{GLS}(z,H_1,\varpi)$ will be asymptotically normal if N and T are of the same order of magnitude (i.e., if $T/N\to\kappa$, where κ is a positive finite constant) and the rate of convergence will depend on the rate of increase, if any, of $\nu_N^{(\varpi)}(z)$. Further, the efficiency improvement of this new estimation procedure is corroborated if $\nu_N^{(\varpi)}(z) < \nu_N(z)$ (see [33, 45] for further details).

To finish the asymptotic analysis of the proposed estimators we must show that both parametric and nonparametric FGLS estimators (see (12) and (17)) are asymptotically equivalent to their GLS counterparts (see (10) and (16)).

Theorem 6. *Suppose that Assumptions 1-5, 7-10, and 14 hold. If it is further assumed that* $T/N\to 0$ *as* $(N,T)\to\infty$,

$$\widehat{\beta}_{FGLS} - \widehat{\beta}_{GLS} = o_P\left(\frac{1}{\sqrt{NT}}\right).$$

Theorem 7. *Suppose that Assumptions 1-5, 7-10, and 13-18 hold. As* $(N,T)\to\infty$,

$$\widehat{\widehat{\overline{m}}}_{FGLS}(z,H_1,\varpi) - \widehat{\widehat{\overline{m}}}(z,H_1,\varpi) = o_p\left(\frac{\left\{\nu_N^{(\varpi)}(z)\right\}^{-1/2}}{\sqrt{NT|H_1|}} + tr(H_1^2)\right).$$

Theorems 6-7 show the asymptotic equivalence between the GLS and FGLS estimators. Note that they are crucial results to prove that $\widehat{\beta}_{FGLS}$ and $\widehat{\widehat{\overline{m}}}_{FGLS}(z,H_1,\varpi)$ have the same limiting distribution as $\widehat{\beta}_{GLS}$ and $\widehat{\widehat{\overline{m}}}_{GLS}(z,H_1,\varpi)$, respectively. Furthermore, to prove Theorem 7 we need to assume $\iota_N^\top\Phi_N^{-1}(z)\iota_N \geq N/\|\Phi_N(z)\|$, where $\|\Phi_N(z)\|$ denotes the square root of the largest eigenvalue of $\Phi_N(z)^\top\Phi_N(z)$. Therefore, we can conclude that the variance rate of $\widehat{\widehat{\overline{m}}}(z,H_1,\varpi)$ is $(NT|H_1|)^{-1}$, whether $\|\Phi_N(z)\|$ remains bounded.

4. Monte Carlo simulation

To analyze the finite sample performance of the proposed estimators, in the following we report the results of several simulation studies to compare the behavior of the three proposed estimators for $\overline{m}(\cdot)$, namely $\widehat{\widehat{\overline{m}}}(\cdot;H_1)$ (initial estimator), $\widehat{\widehat{\overline{m}}}_{GLS}(\cdot;H_1)$ (infeasible improved estimator), and $\widehat{\widehat{\overline{m}}}_{FGLS}(\cdot;H_1)$ (feasible improved estimator). Taking as benchmark [42], for all experiments we consider the following DGP based on Eq. (1)-(2):

$$\begin{aligned} y_{it} &= \alpha_i d_{1it} + x_{it}^\top\beta + m_i(z_t) + \gamma_{1i}f_{1t} + \gamma_{2i}f_{2t} + \epsilon_{it}, \\ x_{lit} &= a_{l1i}d_{1t} + a_{l2i}d_{2t} + g_{li}(z_t) + \gamma_{l1i}f_{1t} + \gamma_{l3i}f_{3t} + v_{lit}, \end{aligned}$$

for $i\in\{1,\ldots,N\}$, $t\in\{1,\ldots,T\}$, $\ell=1,2$. Hence, there are two individual-specific regressors (i.e., $x_{it}=(x_{1it},x_{2it})^\top$), three observed common factors (z_t,d_{1t},d_{2t}), and three unobserved common factors (f_{1t},f_{2t},f_{3t}). For $\rho\in\{0,0.2,0.5,0.8\}$ and $t\in\{-49,\ldots,0,1,\ldots,T\}$, the observed common factors are generated as stationary AR(1) process:

$$\begin{aligned} &d_{1t}=1, \quad d_{2t}=\rho_{2(t-1)}+u_{dt}, \quad u_{dt}\sim IIDN(0,1-\rho^2), \quad d_{2,-50}=0, \\ &z_t=\rho z_{(t-1)}+u_t, \quad u_t\sim IIDN(0.5,1/16), \quad z_{-50}=0, \end{aligned}$$

and the unobserved common factors and individual-specific errors of x_{it} are generated as stationary AR(1) process:

$$\begin{aligned} &f_{lt}=\rho f_{l(t-1)}+u_{f_l,t}, \quad u_{f_l,t}\sim IIDN(0,1-\rho^2), \\ &v_{lit}=\rho_{v_l i}v_{li(t-1)}+v_{lit}, \quad v_{ilt}\sim IIDN\left(0,1-\rho_{v_{l_i}}^2\right), \quad \rho_{v_{l_i}}\sim IIDU(0.05,0.95), \end{aligned}$$

where $\ell \in \{1,2,3\}$, $f_{l,-50} = 0$, and $v_{li,-50} = 0$. Furthermore, the factor loadings of the observed common effects do not change across replications and are generated as follows: $(\alpha_1, \ldots, \alpha_{(N-1)}) \sim IIDU(0,1)$ and $(a_{11i}, a_{21i}, a_{12i}, a_{22i}) \sim IIDN(0.5\iota_4, 0.5I_4)$, where $\iota_4 = (1,1,1,1)^\top$, I_4 is the 4×4 identity matrix, and $\alpha_N = -\sum_{i=1}^{N-1} \mu_i$. Note that the first 50 observations of z_t,v_{1it}, v_{2it}, f_{1t}, f_{2t}, and f_{3t} are discarded.

The loading coefficients of unobserved common factors in the y_{it} and x_{it} are generated as follows:

$$\Gamma_i^\top = \begin{pmatrix} \gamma_{11i} & 0 & \gamma_{13i} \\ \gamma_{21i} & 0 & \gamma_{23i} \end{pmatrix} \sim IIDN \begin{pmatrix} N(0.5,0.5) & 0 & N(0,0.5) \\ N(0,0.5) & 0 & N(0.5,0.5) \end{pmatrix},$$

where $\gamma_{i1} \sim IIDN(1,0.2)$ and $\gamma_{i2} \sim IIDN(1,0.2)$, so the rank condition is satisfied. Further, the heterogeneous unknown functions are generated such as $m_i(z_t) = (1/(1+z_t^2)) + \upsilon_i$, $g_{1i}(z_t) = (1+z_t^2) + \upsilon_{1i}$, and $g_{2i}(z_t) = sin(2z_t) + \upsilon_{2i}$, where $(\upsilon_i, \upsilon_{1i}, \upsilon_{2i}) \sim IIDN(0,1)$ are fixed across simulations. The idiosyncratic errors ϵ_{it} of y_{it} are generated according to $\epsilon_{it} = b_i(z_t)\eta_t + \sqrt{0.5}\epsilon_{0it}$, where $b_i(z_t) = b_i z_t$ with b_i being generated as independent $N(0,10)$ variates, kept fixed across replications, whereas (η_t, ϵ_{0it}) are generated as independent Gaussian AR(1), with innovations having unit variance and using the ρ's value as the autoregressive coefficient.

The points at which the functions are estimated and the second stage bandwidth choice, are in line with those of [33, 48]. In other words, the one-dimensional regressor was generated to have mean 0.5 and variance 1/16, so the bulk of observations lie in the interval [0, 1]. In addition, we set the first stage bandwidth $H_1 = h_1 I_q$ to be 1.2 times the second stage one $H_2 = h_2 I_q$ (i.e., $h_1 = 1.2h_2$) because of the need for oversmoothing in the first stage, and three bandwidth values are used (i.e., $h_2 \in \{0.1, 0.3, 0.5\}$). Note that even though ($h_1 = 1.2h_2$) does not imply oversmoothing asymptotically, in finite sample applications it effectively oversmooths.

In each DGP, we consider the pairs $(N,T) \in \{(125,75), (150,100), (175,125)\}$, and use the Epanechnikov kernel functions. For the evaluation of the performance of the slope estimators we use the averaged bias and standard deviation (Sd) for the slope parameters. We use 1,000 replications and report the results in Table 1 for β_1. The results for β_2 are quite similar and will not be reported for brevity, but they are available upon request.

Table 1 shows the very good performance of the parametric estimators. All of them display very small bias and their average standard deviation decline steadily with increases in N or T and decreases of ρ, and they do not seem to be very sensitive to the bandwidth selection. Furthermore, the estimator's efficiency is improved by taking the spatial dependence of the error term and the heterogeneity into account and $\widehat{\beta}_F$ presents the best results.

The performance of the nonparametric estimators proposed in this paper is assessed via the square-root of the averaged squared errors of the different estimators (RASE). For 1000 simulations and, under different strengths of cross-sectional dependence (i.e., ρ), Fig. 1 shows the boxplots of the RASE. To show the efficiency gain of the proposed estimators and facilitate the comparison between them we compute the Mean Squared Error (MSE) for the regression functions at three evaluation points (i.e., $z_1 = 0.25$, $z_2 = 0.5$, and $z_3 = 0.75$). Table 2 reports the relative Monte Carlo MSE of $\widehat{\widehat{m}}_{GLS}(\cdot; H_1)$ and $\widehat{\widehat{m}}_{FGLS}(\cdot; H_1)$ to $\widehat{\widehat{m}}(\cdot; H_1)$.

Analyzing the results for the nonparametric estimators summarized in Fig. 1 it can be pointed out that the nonparametric procedure proposed in this paper is robust to the bandwidth selection. Another important finding is that an increase in the sample size leads to a decrease in the RASEs of the nonparametric estimators. In addition, as was expected from the results in Section 3, the GLS and FGLS nonparametric estimators that take into account the information contained in the error term (i.e., the spatial dependence and unobserved common factors) give smaller RASEs than the initial estimator in all the cases considered, although the GLS estimator performs the best.

Finally, we analyze the results in Table 2 for the relative MSEs under different sample sizes, different values of ρ, and fixed points (i.e., $z \in \{0.25, 0.5, 0.75\}$). For a fixed point, increases in the sample size lead to decreases in the relative MSEs that shrink to zero and the figures when $\rho = 0$ and $\rho = 0.8$ are very close. Also, the estimators are not very sensitive to the bandwidth choice since there is a slight reduction of the relative MSEs when the bandwidth increases. Furthermore, when the sample size, ρ, and z are fixed, all the ratios are lower to 1 which means that the improved estimators are more efficient than the initial estimator. Therefore, it is corroborated that taking the unobserved common factors and spatial dependence into account improves the estimator's efficiency.

Table 1: Simulation results for the Bias (x100) and Sd of the parametric estimators for β_1 that are computed as $Bias = n_0^{-1}\sum_{\varphi=1}^{1000}(\widehat{\beta}_\varphi - \beta)$ and $Sd = \left\{n_0^{-1}\sum_{\varphi=1}^{1000}(\widehat{\beta}_\varphi - \widehat{\overline{\beta}})^2\right\}^{1/2}$, where $\widehat{\overline{\beta}} = n_0^{-1}\sum_{\varphi=1}^{1000}\widehat{\beta}_\varphi$ and $n_0 = 1,000$. These expressions are used for the three parametric estimators considered in this study: $\widehat{\beta}$ (initial estimator), $\widehat{\beta}_G$ (infeasible improved estimator), and $\widehat{\beta}_F$ (feasible improved estimator).

			N=125,T=75		N=150,T=100		N=175,T=125	
ρ	h_2		Bias	Sd	Bias	Sd	Bias	Sd
0	0.1	$\widehat{\beta}$	0.295	0.092	0.178	0.080	-0.301	0.060
		$\widehat{\beta}_G$	0.070	0.008	0.022	0.006	0.019	0.005
		$\widehat{\beta}_F$	0.291	0.090	0.176	0.079	-0.298	0.059
	0.3	$\widehat{\beta}$	0.187	0.092	0.159	0.078	-0.329	0.060
		$\widehat{\beta}_G$	0.034	0.007	0.023	0.005	0.011	0.004
		$\widehat{\beta}_F$	0.186	0.090	0.155	0.077	-0.322	0.059
	0.5	$\widehat{\beta}$	0.164	0.092	0.219	0.078	-0.314	0.060
		$\widehat{\beta}_G$	0.032	0.006	0.021	0.005	0.011	0.004
		$\widehat{\beta}_F$	0.164	0.089	0.213	0.076	-0.305	0.058
0.2	0.1	$\widehat{\beta}$	-0.213	0.100	-0.290	0.079	-0.089	0.062
		$\widehat{\beta}_G$	0.012	0.008	0.007	0.007	-0.003	0.005
		$\widehat{\beta}_F$	-0.209	0.098	-0.287	0.078	-0.087	0.062
	0.3	$\widehat{\beta}$	-0.178	0.096	-0.276	0.078	-0.029	0.062
		$\widehat{\beta}_G$	-0.002	0.007	0.016	0.006	0.005	0.005
		$\widehat{\beta}_F$	-0.176	0.093	-0.271	0.077	-0.028	0.061
	0.5	$\widehat{\beta}$	-0.201	0.095	-0.306	0.078	-0.033	0.062
		$\widehat{\beta}_G$	-0.005	0.007	0.015	0.006	0.007	0.005
		$\widehat{\beta}_F$	-0.195	0.092	-0.305	0.076	-0.036	0.061
0.5	0.1	$\widehat{\beta}$	-0.165	0.130	-0.213	0.104	0.251	0.087
		$\widehat{\beta}_G$	-0.016	0.010	0.028	0.008	-0.044	0.006
		$\widehat{\beta}_F$	-0.165	0.128	-0.208	0.102	0.249	0.086
	0.3	$\widehat{\beta}$	-0.277	0.123	-0.236	0.104	0.309	0.085
		$\widehat{\beta}_G$	-0.034	0.009	0.029	0.007	-0.034	0.006
		$\widehat{\beta}_F$	-0.273	0.120	-0.229	0.102	0.303	0.084
	0.5	$\widehat{\beta}$	-0.422	0.122	-0.204	0.104	0.288	0.086
		$\widehat{\beta}_G$	-0.035	0.009	0.031	0.007	-0.034	0.005
		$\widehat{\beta}_F$	-0.410	0.118	-0.194	0.100	0.280	0.083
0.8	0.1	$\widehat{\beta}$	1.114	0.231	-0.134	0.220	0.252	0.087
		$\widehat{\beta}_G$	0.040	0.016	-0.058	0.013	-0.044	0.006
		$\widehat{\beta}_F$	1.076	0.225	-0.135	0.216	0.249	0.086
	0.3	$\widehat{\beta}$	1.222	0.229	-0.041	0.213	0.309	0.085
		$\widehat{\beta}_G$	0.024	0.013	-0.047	0.011	-0.034	0.006
		$\widehat{\beta}_F$	1.169	0.219	-0.046	0.205	0.303	0.084
	0.5	$\widehat{\beta}$	1.231	0.229	-0.107	0.213	0.288	0.086
		$\widehat{\beta}_G$	0.014	0.013	-0.052	0.010	-0.034	0.005
		$\widehat{\beta}_F$	1.157	0.216	-0.110	0.201	0.280	0.083

5. Climate policy analysis: assessing the effects of EU ETS on CO_2 emissions

5.1. Brief Overview of EU ETS studies

Over the past few decades, growing concern about environmental degradation has prompted major economies to take action against global warming and climate change trough agreements like the Kyoto Protocol and the 2015 Paris Agreement. Reducing greenhouse gas emissions is a key goal for signatory countries with the EU ETS as the cornerstone of the European Union's strategy to decarbonize the economy and mitigate climate change [9].

The literature exploring the effects of carbon pricing on economic and environmental performance has evolved over decades, with a primary focus on green taxation [1]. The other pillar of carbon pricing is the EU ETS, which

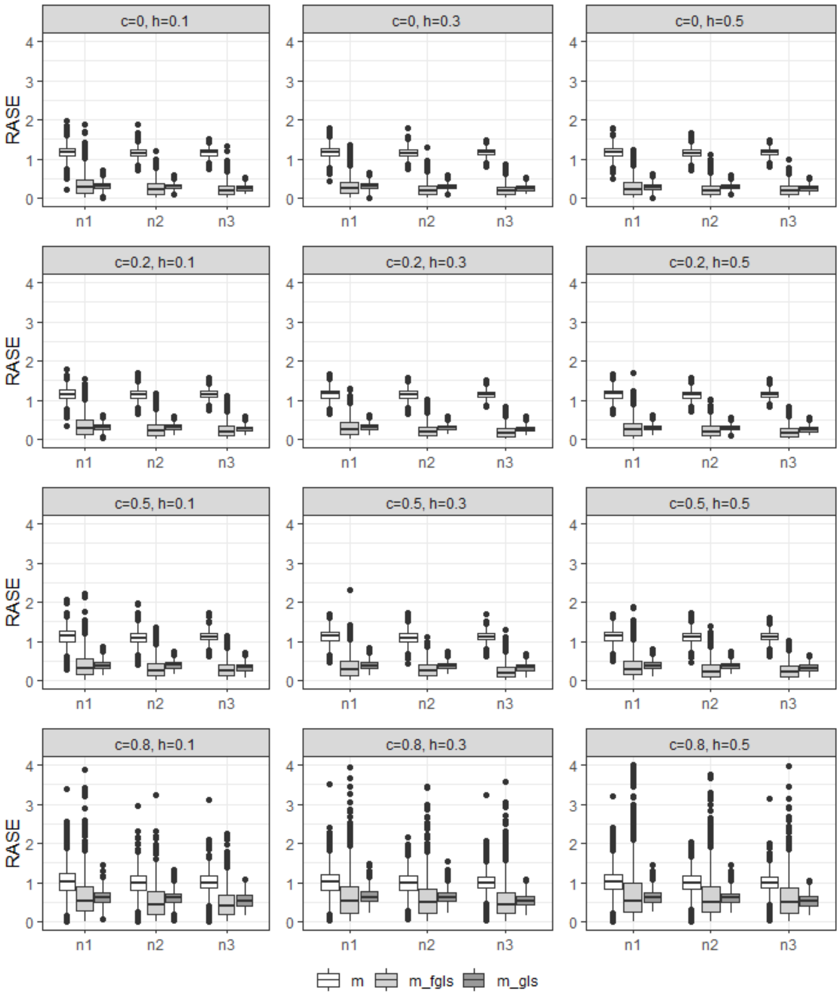

Fig. 1: Boxplots of RASE for the nonparametric estimators in 1,000 simulations where m, m_fgls, and m_gls are the initial estimator, feasible improved estimator, and infeasible estimator, respectively, and three sample sizes are considered (i.e., $n1 \in \{125, 75\}$, $n2 \in \{150, 100\}$, and $n3 \in \{175, 125\}$).

has been widely studied since its launch in 2005. Research has explored its impact on economic and environmental indicators, identifying three main research areas. The first one focuses on innovation activities, often exploiting firm-level data [13]. The second one examines economic performance, analyzing productivity, GDP, and investment, with notable contributions from [9, 14, 32, 34], among others. Lastly, the third one centers on environmental outcomes, particularly CO_2 emissions, with recent studies such as those by [17] highlighting mitigation effects, and [38] observing the decoupling of emissions from economic growth.

However, despite the interesting findings in prior studies, many of their conclusions have been drawn at the firm level. To offer fresh insights, we opt for a country-based macroeconomic perspective as in [29].

5.2. Model specification

Let us consider an augmented EKC specification such as

$$co_{2it} = \alpha_i^\top d_t + \beta_1 gdp_{it} + \beta_2 gdp_{it}^2 + \beta_3 r\&d_{it} + m_i(z_t) + \gamma_i^\top f_t + \epsilon_{it}, \tag{18}$$

for $i \in \{1, \ldots, N\}$ and $t \in \{1, \ldots, T\}$, where co_{2it} refers to the level of CO_2 emissions per capita of country i at time t, gdp_{it} stands for GDP per capita, and $r\&d_{it}$ is considered as a proxy for the level of technology [6, 52, 53] and it is

Table 2: Relative MSE: $MSE(\widehat{m}_{GLS}(z;H_1))/MSE(\widehat{m}(z;H_1))$ and $MSE(\widehat{m}_{FGLS}(z;H_1))/MSE(\widehat{m}(z;H_1))$, where $MSE(\widehat{m}(z)) = Bias^2(\widehat{m}(z)) + \mathrm{Var}(\widehat{m}(z))$. This expression is used for the three nonparametric estimators considered in this paper, i.e., $\widehat{m}(z;H_1)$, $\widehat{m}_{GLS}(z;H_1)$, and $\widehat{m}_{FGLS}(z;H_1)$.

		ρ				ρ			
z	h_2	0	0.2	0.5	0.8	0	0.2	0.5	0.8
					N=125,T=75				
0.25	0.1	0.065	0.063	0.077	0.132	0.172	0.183	0.238	0.526
	0.3	0.058	0.055	0.069	0.126	0.143	0.145	0.190	0.431
	0.5	0.045	0.043	0.056	0.107	0.135	0.139	0.186	0.560
0.50	0.1	0.025	0.025	0.037	0.081	0.153	0.163	0.224	0.593
	0.3	0.023	0.023	0.034	0.082	0.104	0.115	0.157	0.546
	0.5	0.020	0.020	0.031	0.076	0.093	0.105	0.160	0.723
0.75	0.1	0.004	0.003	0.007	0.031	0.216	0.216	0.273	0.865
	0.3	0.002	0.002	0.006	0.032	0.100	0.116	0.152	0.684
	0.5	0.002	0.002	0.006	0.034	0.083	0.106	0.151	0.833
					N=150,T=100				
0.25	0.1	0.061	0.063	0.080	0.139	0.141	0.145	0.203	0.389
	0.3	0.054	0.056	0.072	0.129	0.114	0.119	0.157	0.333
	0.5	0.041	0.043	0.057	0.110	0.109	0.117	0.154	0.382
0.50	0.1	0.024	0.025	0.038	0.086	0.101	0.103	0.154	0.414
	0.3	0.022	0.024	0.036	0.084	0.068	0.076	0.120	0.467
	0.5	0.020	0.021	0.033	0.079	0.048	0.052	0.080	0.615
0.75	0.1	0.002	0.003	0.006	0.034	0.147	0.136	0.206	0.662
	0.3	0.001	0.002	0.006	0.034	0.071	0.071	0.109	0.641
	0.5	0.001	0.002	0.007	0.037	0.062	0.067	0.103	0.780
					N=175,T=125				
0.25	0.1	0.033	0.035	0.044	0.044	0.151	0.155	0.188	0.188
	0.3	0.029	0.030	0.039	0.039	0.127	0.123	0.155	0.155
	0.5	0.021	0.023	0.030	0.030	0.121	0.121	0.154	0.154
0.50	0.1	0.012	0.013	0.019	0.019	0.092	0.096	0.136	0.136
	0.3	0.011	0.012	0.018	0.018	0.068	0.066	0.106	0.106
	0.5	0.009	0.010	0.016	0.016	0.035	0.035	0.063	0.063
0.75	0.1	0.001	0.001	0.003	0.003	0.093	0.099	0.135	0.135
	0.3	0.001	0.001	0.002	0.002	0.048	0.050	0.081	0.081
	0.5	0.001	0.001	0.002	0.002	0.045	0.046	0.081	0.081

also expressed in per capita terms. All variables are expressed in natural logarithms and $d_t = 1$.

Since the aim of this study is to assess the impact of EU ETS on CO_2 emissions, following [16, 18] we introduce in (18) the environmental policy variable (i.e., the carbon price which arises from the market of allowances). This variable is introduced as a common stochastic covariate, z_t, taken in logarithm form. However, while the EKC formulation has theoretical bases and is consistent with a huge amount of the literature, a high degree of uncertainty surrounds the shape (and sign) of the policy effect and there are no ex-ante theoretical or empirical reasons to impose a specific parametric relation between co_{2it} and the price of polluting, z_t. Indeed, the aggregate effect of EU ETS carbon pricing on co_{2it} could be the result of economic mechanisms that involve composite and rather complex negative and positive effects on emissions. On the one side, negative effects on emission can be due to several reasons: The EU ETS cap itself, the abatement-oriented induced innovation effect of targeted high emissions industrial sectors [12, 13], or the diffusion and adoption of those innovations throughout the economy by inter sector links and value chains. On the other side, positive effects can be related to the scale effect of production, which might also be more pronounced through the competitiveness effect of process and product innovations that are generated by the policy. A positive effect on emission can be also due to a carbon rebound effect that may happen if carbon policies improve energy efficiency, which leads to an energy rebound effect, and in turn, it may produce a carbon rebound effect as was noted in [8]. On these grounds, we opt to introduce z_t through a nonparametric function and it is expected that both common price and unobserved common factors, f_t, may produce a heterogeneous effect across units due to country-specific economic or technological features. The rest of the components of the model have been already defined in Section 2.

Finally, it is worth noting that more flexible specifications, such as fully nonparametric models [36, 37] or models with heterogeneous slopes [35], could be of interest in principle. However, moderate sample sizes like that in the present paper and often encountered in macroeconomic data, pose significant challenges. Fully nonparametric models, for instance, suffer from the curse of dimensionality problem, while heterogeneous panel data models are theoretically justified only for large T and often underperform compared to homogeneous estimators [7].

5.3. *Data, variables and preliminary analysis*

Our data is derived from official sources and covers EU27 countries plus the UK, Iceland, and Norway over the period 2005-2019. GDP expressed in Purchasing Power Parity (PPP), population, and R&D (GERD) are derived from EUROSTAT. The CO_2 series is provided by EUROSTAT as well. We opt for CO_2 series accounted by EUROSTAT because it includes all the emitting sectors and indirect CO_2 emissions and is reported in thousands of tonnes. The key policy variable z, which is invariant across units, is the price of carbon (in logarithms), linked to the European Market of Allowances (EUA). Data on EUA are obtained from both the International Carbon Action Partnership (https://icapcarbonaction.com/en/ets-prices) and Sendeco (https://www.sendeco2.com/it/prezzi-co2). Data on ETS auctions are registered daily and the annual carbon price used is the average auction price in the primary market of all the transactions registered in a given year.

The EU ETS was launched in 2005. In 2005, the price was about 18 €per tonne. The initial phase, a 3-year pilot program (2005-2007), aimed to create a functional market structure. During this period, the system targeted CO_2 emissions from power generators and energy-intensive industries, allocating most allowances to businesses without charge. This phase, characterized as a period of 'learning by doing', laid the foundation for the subsequent phases of the EU ETS. Phase 1 of the EU ETS successfully established a carbon emissions price, albeit experiencing volatility, as highlighted by [19], notably marked by a significant decline in carbon prices in 2007. The second phase, initiated in 2008, was marked by high market expectations, likely stemming from anticipation surrounding the 2009 United Nations Climate Change Conference, possibly contributing to the price increase observed in 2008. However, the outcomes of the conference did not convey the policy outcomes that were expected. These unsatisfactory outcomes, coupled with the global economic recession following the 2008 financial crisis, are mirrored in the decline of carbon prices, signaling a period of uncertain climate policy commitments worldwide. Subsequently, between 2014 and 2016, promising signs of economic recovery emerged. This progress was further reinforced by the pivotal 2015 Paris Agreement, outlining global commitments to reduce carbon emissions. This significant policy development likely led to an increase in CO_2 prices within the EU, as documented by [10, 20]. The 2.6% GDP growth in 2017, resulting from enhanced policy commitments following the Paris Agreement in 2015, may explain the rise in prices observed from 2017 to 2019, when the price reached its maximum level of 24.2 €in 2019.

Before discussing the estimation results, it is worth noting that the stationarity of the observed covariate, z_t, is fundamental for valid estimation. When examining the logarithm of the carbon price, it exhibits a rather smooth time pattern over the period except for a negative break observed in 2007. Additionally, the series displays a constant dispersion around its mean. To check formally the stationarity of z_t, we conducted a series of statistical tests (i.e., traditional unit root tests, including the Augmented Dickey-Fuller (ADF), Phillips-Perron, Breitung's nonparametric test, and Bierens HOAC tests). However, given the potential size distortion issues associated with unit root tests on time series data of moderate sample sizes, we supplemented these tests by simulating p-values using an AR(1) Gaussian model and employing the wild bootstrap method with 2000 replications. Also, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and Bierens-Guo tests are considered to test the null hypothesis of stationarity against the alternative of a unit root (see [48] for detailed references). All the above tests indicate that z_t exhibits stationarity.

Finally, the dependent variable in (18), co_{2it}, exhibits a decreasing evolution over time, evolving around a linear trend for most of the countries under consideration. It is also found to be stationary according to the PANIC [3] and PANICCA [44] panel unit roots tests (detailed results are available upon request).

5.4. *Estimation results*

Building upon the discussion established in the previous subsections, we compute the quantities of interest in model (18) using the estimators for β that were proposed in this paper. To assess the sensitivity of the empirical results related to the functional form of the policy variable (carbon price), estimators of a fully parametric model are presented in Table 3, columns (i)-(iii). Furthermore, in Table 3, columns (iv)-(v), we provide the semiparametric

Table 3: Fully parametric and semiparametric results. Columns (i)-(iii): CCEP Pesaran [39]; Columns (iv) and (v): first-stage and second-stage semiparametric estimator. Absolute t-statistics in parentheses and p-value in square brackets. ***, **, * are the significance level at 1%, 5%, 10%, respectively. The bandwidth term has been chosen following the Silverman's rule of thumb. *Diagnostics*: CD: CD test by Pesaran [40] and Pesaran [41]. CD_w: averaged weighted CD test by Juodis and Reese [28]. Averaging reduces the test's reliance on a specific set of random weights. $\widehat{a}$: bias-corrected version of a given by equation (13) in Bailey et al. [4]. *95% level confidence bands.

	Parametric			Semiparametric	
	(i)	(ii)	(iii)	(iv)	(v)
gdp_{it}	-0.606	-0.607	-0.607	-3.402*	-2.386***
	(0.465)	(0.465)	(0.465)	(1.873)	(8.649)
gdp_{it}^2	-0.113	-0.113	-0.113	-0.515**	-0.373***
	(0.638)	(0.638)	(0.637	(2.114)	(9.733)
$r\&d_{it}$	-0.321***	-0.321***	-0.321***	-0.387***	-0.288***
	(3.831)	(3.826)	(3.822)	(4.342)	(9.529)
z_t		1.32e-04	-1.81e-04		
		(0.021)	(0.007)		
z_t^2			9.70e-05		
			(0.012)		
CD	3.14	3.11	3.11	3.55	1.49
	[0.002]	[0.002]	[0.002]	[0.00]	[0.14]
CD_w	2.41	2.41	2.41	4.31	0.51
	[0.016]	[0.016]	[0.016]	[0.00]	[0.61]
$\widehat{a}$	0.80	0.80	0.80	0.78	0.59
$\widehat{a}^*_{0.025}$	0.71	0.71	0.71	0.68	0.43
$\widehat{a}^*_{0.975}$	0.90	0.90	0.90	0.88	0.76

estimates for the parameters of interest. Finally, in Fig. 2 we represent the curve estimates for the effect of the policy variable.

In column (i), a fully parametric model that does not contain the carbon price is considered. Subsequently, in column (ii), a linear effect of the carbon price is introduced by including z as an additional regressor. In column (iii), the potential nonlinear effect of the carbon price is estimated by employing a second-order polynomial function. All these specifications are estimated using the CCEP estimator proposed in [39] by controlling for unobserved individual effects, i.e., by setting $d_t = 1$. Lastly, columns (iv)-(v) present the results from the estimators proposed in this paper. In column (iv) the estimates for β are obtained using expression (6) and in column (v) we represent the efficient estimates (FGLS) of the β that have been computed using expression (12) from Section 3.

According to the estimation results, the fully parametric specifications yield negative and nonsignificant estimates for the coefficients associated with both GDP and GDP^2. Considering that our sample covers EU countries in very recent years, the finding of a negative elasticity concerning per capita GDP, which decreases in magnitude with the rise of this variable, aligns with the original idea behind the EKC. This result is consistent with a substantial body of literature as in [15], but the statistical insignificance of the GDP variable is an unexpected result.

As far as the effect of the technology variable is concerned, when the policy variable z_t is omitted, the estimated elasticity of $R\&D$ expenditures is significant and about -0.32 and this result does not change when the carbon price variable is included. The literature, which is surveyed in [31], is heterogeneous in terms of the adopted proxy and results. While R&D expenditures is a common proxy for technology [23] and is often employed as in [22], alternative proxies such as energy intensity [6] and process or product innovation [18] are occasionally employed. While the conventional expectation is that technology would lead to a reduction in emissions, the results from the literature are mixed and sometimes show positive estimates. Finally, as far as the EU ETS policy variable is concerned, it is not statistically significant for the different fully parametric specifications and we do not observe conclusive results.

In summary, the fully parametric specifications in columns (i)-(iii) yield unexpected results that should be reassessed by employing the proposed semiparametric estimators. This is because allowing for a nonparametric function $m_i(\cdot)$ instead of imposing a parametric specification for the policy variable may be important to avoid a misspecification error that might lead to inconsistent estimators, not only for the estimated policy effect but also with respect the impact of the standard EKC covariates. Moreover, efficiency improvements resulting from exploiting the correlation structure of the error term can significantly impact both the inference and estimation.

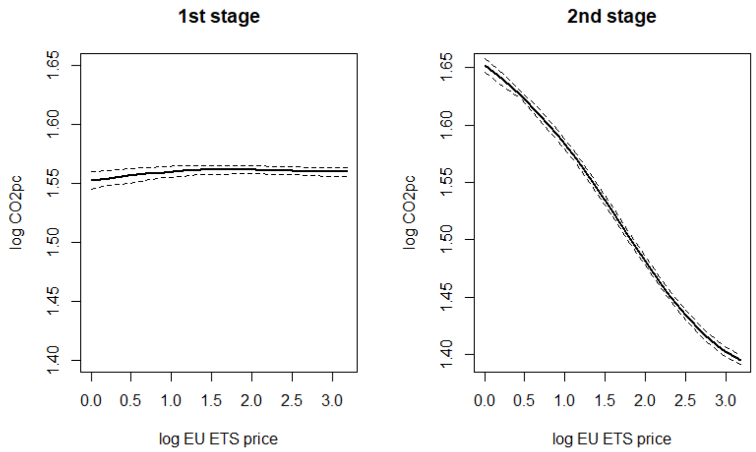

Fig. 2: Estimated relation CO_2 - EU ETS carbon price, where the left panel represents the initial estimator $\widehat{\widehat{m}}(\cdot)$ and the right panel represents the FGLS estimated curve, $\widehat{\widehat{m}}_{FGLS}(\cdot)$. The thick lines denote the estimated curves while dotted lines are the 95% pointwise confidence intervals.

As far as the proposed semiparametric estimators are concerned (i.e., columns (*iv*) and (*v*)), the results highlight the empirical importance of using consistent and efficient estimators for the quantities of interest. Parametric estimates differ significantly, particularly those related to the GDP variables, with the estimated elasticity of co_{2it} with respect to gdp_{it} increasing significantly in absolute value. Additionally, the standard error estimates from the semiparametric estimators are smaller than the fully parametric ones, with this efficiency improvement particularly notable for the efficient estimator. Overall, the parametric results using the efficient estimator (column (*v*)) appear more consistent both from the economic and statistical points of view. Both gdp and gdp^2 become highly significant, indicating a statistically significant negative relationship that is the expected result in the EKC framework. Furthermore, by comparing the nonparametric curve estimates (see Fig. 2) is easily seen the empirical relevance of using more efficient estimates. When the structure of the spatial correlation is ignored, misleading conclusions are obtained about the effect of EU ETS on CO_2. Indeed, the inefficient curve estimate is very flat compared with the most efficient one, which exhibits a nonlinear negative shape. Note that the introduction of the spatial correlation structure may affect the shape of the curve according to equations (8) and (17).

The impact of carbon pricing on CO_2 emissions can vary depending on a country's economic specialization, green technological intensity, and history of environmental policies. It is, therefore, worthwhile to examine the estimated nonparametric function for some countries (see Fig. 3). Taking three key examples from Europe's diverse economic systems, we observe a negative effect in the UK. This aligns with the UK's more market-oriented capitalism, which provides a favorable environment for pricing mechanisms. Additionally, the UK was an early adopter of emissions trading, launching its own market in 2002, earlier than the EU, before the two markets later converged. Germany, on the other hand, shows a bell-shaped relationship between carbon pricing and emissions. This aligns with the country's progress toward green technological leadership over recent decades, supported by increasingly stringent environmental policies. As a highly industrialized nation, Germany's innovations may have created a conducive environment for emissions reduction through industry-driven innovation, in response to the EU ETS policy. In contrast, Spain, with a relatively more service-based economy, may have experienced less innovation and diffusion in response to carbon pricing, resulting in a less effective policy. This raises the question, which future research could address, of how to effectively target non-industrial sectors in climate policies, as these sectors make up the largest share of GDP in advanced economies.

Finally, it is interesting to investigate the presence of CSD in the residuals. The CD test developed in [41] is a widely adopted test, which is typically employed as a misspecification test in models that already account for CSD [5, 21, 28]. This test presents good small-sample properties and recent theoretical works have provided additional insights that are useful from an empirical perspective. In particular, [40] demonstrates that the null hypothesis of

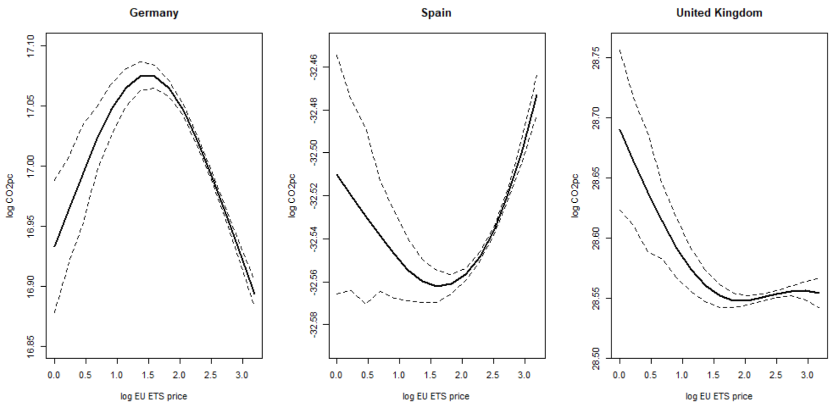

Fig. 3: Heterogeneous estimated relation CO_2 - EU ETS carbon price for Germany, Spain, and United Kingdom. The thick lines denote the estimated curve obtained using the initial heterogeneous estimator, $\widehat{m}_i(\cdot)$, while dotted lines are the 95% pointwise confidence intervals.

the CD test is weak cross-sectional dependence in the most common cases. More precisely, for T almost fixed and $N \to \infty$, such a null hypothesis is $0 \leq a < 1/2$, where a is the exponent of cross-sectional dependence introduced in [5]. Moreover, [28] shows that the CD test statistic is biased for any fixed T and becomes divergent as $T \to \infty$ when the CD test is applied to residuals obtained from a regression model containing common factors. They propose a modified test statistic, denoted as CD_w, where cross-sectional covariances, which are employed instead of correlations, are weighted using Rademacher distributed weights.

The CD statistics in Table 3 for specifications (i), (ii), (iii), and (iv) are all highly statistically significant and strongly reject the null hypothesis. That suggests that the exponent of cross-sectional dependence, a, is in the range $[1/2, 1]$. Conversely, when considering the efficient semiparametric estimator in column (v), the CD statistic is equal to 1.49, so the null is not rejected. We then employ the average CD_w test, which confirms the results that are obtained with the standard CD test. Finally, to quantify the extent of CSD, we compute the bias-corrected version of a. As in [5], Holm's approach has been preferred over the Bonferroni procedure. These estimates, along with the 95% confidence bands, are also reported in Table 3. In the fully parametric specifications, as well as in the semiparametric inefficient one, the exponent of CSD is estimated to be close to 0.8, with 95% confidence bands lying above 0.5 and not including unity. It is worth noting that, similar to the findings in [21], residuals obtained from a multifactor error regression model exhibit a lower degree of CSD compared to the variables incorporated into the model. For these variables, $\widehat{a}$ was approximately 1. When finally moving to the semiparametric efficient estimation, it is notable that $\widehat{a}$ decreases substantially, reducing to 0.59, with the 95% lower confidence band now falling below 0.5. See [5] for a detailed discussion on identifying a. In summary, these results indicate that employing the efficient semiparametric estimator significantly reduces residual CSD.

6. Conclusions

In this paper, we have considered efficiency improvements in a partially linear panel data model that accounts for possible nonlinear effects of common covariates and allows for CSD arising simultaneously from common factors and spatial dependence. We proposed a GLS-type estimator that accounts for this dependence structure and studied potential gains in the rate of convergence. A key theoretical finding is that exploiting the correlation structure for efficiency improvements can notably affect both inference and estimation. This conclusion was supported by both Monte Carlo simulations and an empirical analysis assessing the impact of carbon pricing within the European Union Emission Trading System on carbon dioxide emissions. Our empirical findings suggest that the proposed efficient semiparametric estimator not only yields more meaningful estimates from an economic perspective, but also significantly enhances inference and reduces residual CSD.

Acknowledgments

The authors are grateful for all the suggestions made by the Editor, Associate Editor and referees. They have contributed to improving the quality of the paper. The usual disclaimer applies. 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. Furthermore, Alexandra Soberon and Juan M. Rodriguez-Poo also acknowledge financial support from the I+D+i project Ref. PID2019-105986GB-C22, funded by MCIN/AEI/ 10.13039/501100011033. Part of the work was done during visits by Alexandra Soberon to the University of Ferrara.

Appendix

Before proceeding to the analysis of the main asymptotic properties of the proposed estimators, we first present several lemmas that are used later to prove the main results of the paper. Remember that we denote $\widetilde{X}_{i\cdot} = X_{i\cdot} - \mathcal{B}_X(z)$, $\widetilde{\Lambda} = \Lambda - \mathcal{B}_\Lambda(z)$, $\widetilde{D} = D - \mathcal{B}_D(z)$, $\widetilde{G} = G - \mathcal{B}_G(z)$, where $\mathcal{B}_X(z) = E[X_{i\cdot}|z_t = z]\rho_{z_t}(z)$, $\mathcal{B}_\Lambda(z) = E[\Lambda|z_t = z]\rho_{z_t}(z)$, $\mathcal{B}_D(z) = E[D|z_t = z]\rho_{z_t}(z)$, $\mathcal{B}_G(z) = E[G|z_t = z]\rho_{z_t}(z)$. Also, we define $\widetilde{X}^{(\widehat{\varpi})} = X^{(\widehat{\varpi})} - \mathcal{B}_{X^{(\widehat{\varpi})}}(z)$, $\widetilde{X}^{(\varpi)} = X^{(\varpi)} - \mathcal{B}_{X^{(\varpi)}}(z)$, $\widetilde{X}^{(\widehat{\omega})} = X^{(\widehat{\omega})} - \mathcal{B}_{X^{(\widehat{\omega})}}(z)$, and $\widetilde{X}^{(\omega)} = X^{(\varpi)} - \mathcal{B}_{X^{(\omega)}}(z)$, where $\mathcal{B}_{X^{(\widehat{\varpi})}} = E[X^{(\widehat{\varpi})}|z_t = z]$, $\mathcal{B}_{X^{(\varpi)}} = E[X^{(\varpi)}|z_t = z]$, $\mathcal{B}_{X^{(\widehat{\omega})}} = E[X^{(\widehat{\omega})}|z_t = z]$, $\mathcal{B}_{X^{(\omega)}} = E[X^{(\omega)}|z_t = z]$. Similar notation for $\widetilde{Y}^{(\widehat{\varpi})}$, $\widetilde{Y}^{(\varpi)}$, $\widetilde{Y}^{(\widehat{\omega})}$, $\widetilde{Y}^{(\omega)}$. Also, $c_{H_1} = tr(H_1^2) + (\ln T/T|H_1|)^{1/2}$.

Lemma 1. Denote $\overline{\varepsilon}_{At} = (\overline{\epsilon}_{At} + \overline{v}_{At}^\top\beta, \overline{v}_{At}^\top)^\top$ as an error term. Under Assumptions 1 and 3, for each t, we have

(i) $E(\overline{\varepsilon}_{At}) = 0$.

(ii) $\mathrm{Var}(\overline{\varepsilon}_{At}) = O\left(N^{-1}\right)$, under weak dependence and $\mathrm{Var}(\overline{\varepsilon}_{At}) = O(1)$ under strong dependence.

Proof of Lemma 1: Let $\overset{q.m.}{\to}$ be the convergence in quadratic mean (or mean squared error), the proof of this lemma is straightforward from the proof of Lemma A1 in [42]. This lemma guarantees that for any value of z, $\overline{\epsilon}_{At} \overset{q.m.}{\to} 0$ as $N \to \infty$ and the degree of spatial dependence of $\epsilon_{i\cdot}$ will be bounded by $\nu_N(z) = N^{-2}\iota_N^\top\Omega_N(z)\iota_N$, where ι_N is a $N \times 1$ vector of ones. Then, the results of this paper are valid for both types of spatial dependence. □

Lemma 2. Under Assumptions 2, 7-10, as $T \to \infty$ we have

$$\sum_{\|z\|\le c_{H_1}} \left| \frac{1}{T}\sum_{t=1}^{T} \left[K_{H_1}(z_t - z)x_{it} - E\left\{K_{H_1}(z_t - z)x_{it}\right\}\right] \right| = O_p\left(\sqrt{\frac{\ln T}{T|H_1|}}\right).$$

Proof of Lemma 2: This lemma can be proved as in Theorem 2 in [25] and it has been omitted for brevity. □

Lemma 3. Under Assumptions 1-5,

(i) $T^{-1}\overline{\varepsilon}_{A\cdot}^\top\overline{\varepsilon}_{A\cdot} = O\left(N^{-1}\right)$.

(ii) $T^{-1}\widetilde{F}^\top\overline{\varepsilon}_{A\cdot} = O\left((NT)^{-1/2}\right)$ and $T^{-1}\widetilde{D}^\top\overline{\varepsilon}_{A\cdot} = O\left((NT)^{-1/2}\right)$.

(iii) $T^{-1}V_{i\cdot}^\top\widetilde{D} = O\left(T^{-1/2}\right)$ and $T^{-1}V_{i\cdot}^\top\widetilde{F} = O\left(T^{-1/2}\right)$.

(iv) $T^{-1}V_{i\cdot}^\top\overline{\varepsilon}_{A\cdot} = O\left(N^{-1}\right) + O\left((NT)^{-1/2}\right)$ and $T^{-1}\epsilon_{i\cdot}^\top\overline{\varepsilon}_{A\cdot} = O\left(N^{-1}\right) + O\left((NT)^{-1/2}\right)$.

Proof of Lemma 3: This lemma can be proved as in Lemma 2 in [39] and it has been omitted for brevity. □

Lemma 4. Let $c_{H_1} = tr(H_1^2) + (\ln T/T|H_1|)^{1/2}$. Under Assumptions 1, 3, and 7-10, as $T \to \infty$, we have

(i) $T^{-1}\widehat{X}_{i\cdot}^\top M_{\widehat{\Lambda}}\widehat{X}_{i\cdot} = T^{-1}\widetilde{X}_{i\cdot}^\top M_{\widetilde{G}}\widetilde{X}_{i\cdot} + O_p\left(N^{-1}\right) + O_p\left((NT)^{-1/2}\right) + O_p(c_{H_1})$ uniformly over i,

(ii) $T^{-1}\widehat{X}_{i\cdot}^\top M_{\widehat{\Lambda}}(I_T - S)m_i(Z) = O_p(c_{H_1}^2)$ uniformly over i,

(iii) $T^{-1}\widehat{X}_{i\cdot}^\top M_{\widehat{\Lambda}}\widehat{e}_{i\cdot} = T^{-1}\widetilde{X}_{i\cdot}^\top M_{\widetilde{G}}\epsilon_{i\cdot} + O_p\left(N^{-1}\right) + O_p\left((NT)^{-1/2}\right) + O_p(c_{H_1})$ uniformly over i,

(iv) $T^{-1}\widehat{X}_{i\cdot}^{\top} M_{\widehat{\Lambda}} \widehat{F} = O_p\left(N^{-1}\right) + O_p\left((NT)^{-1/2}\right) + O_p(c_{H_1})$ uniformly over i,

where $M_{\widetilde{G}} = I_T - \widetilde{G}(\widetilde{G}^{\top}\widetilde{G})^{-1}\widetilde{G}^{\top}$.

Proof of Lemma 4: This lemma is proved following similar reasoning as the proof of (A.12)–(A.14) in [42] and Lemma 3 in [11] and it has been omitted for brevity. □

Lemma 5. Under Assumptions 2, 7, and 9-10, as $T \to \infty$, we have

(i) $(NT)^{-1}\widehat{X}^{\top}\left(I_N \otimes M_{\widehat{\Lambda}}\right)\Omega^{-1}(Z)\left(I_N \otimes M_{\widehat{\Lambda}}\right)\widehat{X} = (NT)^{-1}\widetilde{X}^{\top}\left(I_N \otimes M_{\widetilde{\Lambda}}\right)\Omega^{-1}(Z)\left(I_N \otimes M_{\widetilde{\Lambda}}\right)\widetilde{X} + o_p(1)$,

(ii) $(NT)^{-1}\widehat{X}^{\top}\left(I_N \otimes M_{\widehat{\Lambda}}\right)\Omega^{-1}(Z)\left(I_N \otimes M_{\widehat{\Lambda}}\right) = (NT)^{-1}\widetilde{X}^{\top}\left(I_N \otimes M_{\widetilde{\Lambda}}\right)\Omega^{-1}(Z)\left(I_N \otimes M_{\widetilde{\Lambda}}\right) + o_p(1)$.

Proof of Lemma 5: This lemma is proved in a similar way as in Lemma 3 and it has been omitted for brevity. □

Lemma 6. Let $R_{TH} = O_p\left\{tr(H_2^2) + (T|H_1|)^{-1}\right\}$ and $\rho_z(z) > 0$. Under Assumptions 1-10 and 13-14, as $T \to \infty$,

$$\max_{1\le i,j\le N}\left|\widehat{\omega}_{ij}(z) - \omega_{ij}(z)\right| = O_p(R_{TH}), \quad \text{and} \quad \left\|\widehat{\Omega}_N(z) - \Omega_N(z)\right\| = O_p(NR_{TH}).$$

Proof of Lemma 6: Denote by $\widehat{\omega}_{ij}(z)$ and $\omega_{ij}(z)$ the (ij)-th element of $\widehat{\Omega}_N(z)$ and $\Omega_N(z)$, respectively, we can write

$$\widehat{\omega}_{ij}(z) - \omega_{ij}(z) = \frac{\sum_{t=1}^{T} K_{H_2}^{*}(z_t - z)\left\{\widehat{e}_{it}\widehat{e}_{jt} - \omega_{ij}(z)\right\}}{\sum_{t=1}^{T} K_{H_2}^{*}(z_t - z)} = R_{ij}^{(1)} + R_{ij}^{(2)}, \tag{19}$$

where $R_{ij}^{(1)} = \sum_{t=1}^{T} K_{H_2}^{*}(z_t - z)\{\epsilon_{it}\epsilon_{jt} - \omega_{ij}(z)\}/\sum_{t=1}^{T} K_{H_2}^{*}(z_t - z)$ and $R_{ij}^{(2)} = \sum_{t=1}^{T} K_{H_2}^{*}(z_t - z)\left\{\widehat{e}_{it}\widehat{e}_{jt} - \epsilon_{it}\epsilon_{jt}\right\}/\sum_{t=1}^{T} K_{H_2}^{*}(z_t - z)$. Hence, $R_{ij}^{(1)}$ is the estimation error of the usual Nadaraya-Watson estimator of the conditional expectation of $E(\epsilon_{it}\epsilon_{jt}|z_t = z)$ and, under the assumptions given in the paper, it is straightforward to show that

$$\left|R_{ij}^{(1)}\right| = O_p\left\{tr(H_2^2) + (T|H_2|)^{-1/2}\right\}. \tag{20}$$

If we consider the bound of $R_{ij}^{(2)}$, we denote $\widetilde{g}_{1it} = X_{it}^{\top}\left(\widehat{\beta} - \beta\right)$, $\widetilde{g}_{2it} = \left(\delta_i - \widehat{\delta}_i\right)^{\top}\lambda_t$, and $\widetilde{\xi}_{it} = \{m_i(z_t) - \widehat{m}_i(z_t; H_1)\}$. Hence, $\widehat{e}_{it}$ can be expressed as $\widehat{e}_{it} = \epsilon_{it} + \widetilde{g}_{1it} + \widetilde{g}_{2it} + \widetilde{\xi}_{it} + o_p(1)$, where $o_p(1)$ captures possible approximation error for replacing f_t by the proxy's vector λ_t. Replacing this decomposition in $R_{ij}^{(2)}$ we are going to prove

$$R_{ij}^{(2)} \quad = \quad T^{-1}\sum_{t=1}^{T} K_{H_2}^{*}(z_t - z)\left\{\boldsymbol{I}_1 + \boldsymbol{I}_2 + o_p(1)\right\}/\widehat{\rho}(z) = O_p\left\{tr(H_2^2) + (T|H_1|)^{-1}\right\}, \tag{21}$$

where $\widehat{\rho}(z)$ is a nonparametric kernel estimator of $\rho_{z_t}(z)$ such as $\widehat{\rho}(z) = T^{-1}\sum_{t=1}^{T} K_{H_2}^{*}(z_t - z)$, $\boldsymbol{I}_1 = \epsilon_{it}\widetilde{g}_{1jt} + \epsilon_{it}\widetilde{g}_{2jt} + \widetilde{g}_{1it}\epsilon_{jt} + \widetilde{g}_{1it}\widetilde{g}_{1jt} + \widetilde{g}_{1it}\widetilde{g}_{2jt} + \widetilde{g}_{2it}\epsilon_{jt} + \widetilde{g}_{2it}\widetilde{g}_{2jt} + \widetilde{g}_{2it}\widetilde{g}_{1jt}$, and $\boldsymbol{I}_2 = \epsilon_{it}\widetilde{\xi}_{jt} + \widetilde{g}_{1it}\widetilde{\xi}_{jt} + \widetilde{g}_{2it}\widetilde{\xi}_{jt} + \widetilde{\xi}_{it}\epsilon_{jt} + \widetilde{\xi}_{it}\widetilde{g}_{1jt} + \widetilde{\xi}_{it}\widetilde{g}_{2jt} + \widetilde{\xi}_{it}\widetilde{\xi}_{jt}$. Under the assumptions stated in this paper, it can be proved that $\widehat{\rho}(z)$ is consistent following [33], so $\widehat{\rho}(z)^{-1} = \{\rho_{z_t}(z) + o_p(1)\}^{-1} = O_p(1)$. Therefore, in order to prove (21) we only need to analyze $T^{-1}\sum_{t=1}^{T} K_{H_2}^{*}(z_t - z)(\boldsymbol{I}_1 + \boldsymbol{I}_2)$.

Following a similar proof scheme as in Lemma 2 and given the $\sqrt{NT}$-consistency of $\widehat{\beta}$ and the $\sqrt{T}$-consistency of $\widehat{\delta}_i$, it is easy to show

$$T^{-1}\sum_{t=1}^{T} K_{H_2}^{*}(z_t - z)\boldsymbol{I}_1 = o_p\left\{(NT)^{-1/2}\right\} + o_p\left(T^{-1/2}\right). \tag{22}$$

Considering the bound of the second term of $R_{ij}^{(2)}$, two leading terms have to be analyzed separately since the other elements are asymptotically negligible using the consistency results of $\widehat{\beta}$ and $\widehat{\delta}_i$. For the first one, we obtain the following result using Theorems 6 and 10 in [25] and Assumption 14,

$$\frac{1}{T}\sum_{t=1}^{T} K_{H_2}^{*}(z_t - z)\widetilde{\xi}_{it}\epsilon_{it} = o_p\left\{tr(H_1^2) + \frac{1}{\sqrt{T|H_1|}}\right\}, \tag{23}$$

since by the law of iterated expectations (LIE) we get $E\left[T^{-1}\sum_{t=1}^{T} K^*_{H_2}(z_t - z)\widetilde{\xi}_{it}\epsilon_{it}\right] = 0$ and

$$E\left[\frac{1}{T}\sum_{t=1}^{T} K^*_{H_2}(z_t - z)\widetilde{\xi}_{it}\epsilon_{it}\right]^2 \le \frac{1}{T^2}\sum_{t=1}^{T} E\left[K^*_{H_2}(z_t - z)\epsilon_{it}^2\right]\sum_{|z|\le c_T}\left|\widehat{m}_i(z;H_1) - m_i(z)\right|^2 = O_p\left(\frac{\delta_T^4}{T|H_2|}\left\{\frac{\ln T}{T|H_1|} + tr(H_1^4)\right\}\right),$$

where $\delta_T = \inf_{|z|\le c_T}\rho_{z_t}(z) > 0$ and $c_T = \left\{(\ln T)^{1/q}T^{1/2\varsigma}\right\}$, for some $\varsigma > 0$. Similarly, it can be proved that the term $T^{-1}\sum_t K_H(z_t - z)\widetilde{\xi}_{it}^2$ is bounded by $T^{-1}\sum_t K_H(z_t - z)\widetilde{\xi}_{it}^2 = O_p\left(\delta_T^{-2}\left\{\ln T/T|H_1| + tr(H_1^4)\right\}\right)$. Hence, the proof is done by replacing these results and (22)-(23) in (21). Plugging (20)-(21) in (19),

$$\max_{1\le i,j\le N}\left|\widehat{\omega}_{ij}(z) - \omega_{ij}(z)\right| = O_p(R_{TH}). \tag{24}$$

Finally, using (24) and let κ be a positive constant, as $N/T \to \kappa$ it is easy to show

$$\left\|\widehat{\Omega}_N(z) - \Omega_N(z)\right\| \le \left[\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\widehat{\omega}_{ij}(z) - \omega_{ij}(z)\right\}\right]^{1/2} = O_p(NR_{TH}).$$

□

Lemma 7. *Under Assumptions 7, 9, and 10 at z such that $\rho_z(z) > 0$, as $T \to \infty$,*

$$T^{-1}Z_z^\top K_{H_1}(z)Z_z = \begin{pmatrix} \rho_{z_t}(z) & \mu_2^q(K)H_1^2 D_\rho(z) \\ \mu_2^q(K)H_1^2 D_\rho(z) & H_1^2\mu_2^q(K)\rho_{z_t}(z)\end{pmatrix}\left\{1 + O_p(c_{H_1})\right\}.$$

Proof of Lemma 7: The proof of this Lemma follows directly the proof of Theorem 2.1 in [47] and it has been omitted for brevity. □

Lemma 8. Let $R_{TH} = O_p\left\{tr(H_2^2) + (T|H_1|)^{-1}\right\}$ and $\rho_z(z) > 0$. Under Assumptions 1-10 and 13-14, as $T \to \infty$,

$$\left\|\widehat{\varpi} - \varpi\right\| = O_p(NR_{TH}).$$

Proof of Lemma 8: In order to prove this lemma, $\|\widehat{\varpi} - \varpi\|$ can be rewritten as

$$\begin{aligned}\|\widehat{\varpi} - \varpi\| &\le \left\|\left\{\iota_N^\top\widehat{\Phi}_N(z)^{-1}\iota_N\right\}^{-1}\widehat{\Phi}_N(z)^{-1}\iota_N - \left\{\iota_N^\top\Phi_N(z)^{-1}\iota_N\right\}^{-1}\Phi_N(z)^{-1}\iota_N\right\| \\ &= \left\|\iota_N^\top\widehat{\Phi}_N(z)^{-1}\right\|\left|\left\{\iota_N^\top\widehat{\Phi}_N(z)^{-1}\iota_N\right\}^{-1} - \left\{\iota_N^\top\Phi_N(z)^{-1}\iota_N\right\}^{-1}\right| + \left\|\left(\iota_N^\top\Phi_N(z)^{-1}\iota_N\right)^{-1}\right\|\left\|\iota_N^\top\left\{\widehat{\Phi}_N(z)^{-1} - \Phi_N(z)^{-1}\right\}\right\|. \end{aligned} \tag{25}$$

Analyzing each of the above terms separately it is straightforward to show that, using the properties of $\Phi_N(z)$, we get $\left\|\iota_N^\top\widehat{\Phi}_N^{-1}(z)\right\| = O_p\left(\sqrt{N}\left\|\Phi_N^{-1}(z)\right\|\right) = O_p\left(\sqrt{N}\right)$. Furthermore, using Assumption 18 we can prove

$$\begin{aligned}&\left|\left\{\iota_N^\top\widehat{\Phi}_N(z)\iota_N\right\}^{-1} - \left\{\iota_N^\top\Phi_N^{-1}(z)\iota_N\right\}^{-1}\right| \le \left|\frac{\iota_N^\top\widehat{\Phi}_N^{-1}(z)\left\{\widehat{\Phi}_N^{-1}(z) - \Phi_N^{-1}(z)\right\}\Phi_N^{-1}(z)\iota_N}{\left\{\iota_N^\top\widehat{\Phi}_N^{-1}(z)\iota_N\right\}\left\{\iota_N^\top\Phi_N^{-1}(z)\iota_N\right\}}\right| \\ &\le O_p\left(\frac{\iota_N^\top\Phi_N^{-2}(z)\iota_N}{\left\{\iota_N^\top\Phi_N^{-1}(z)\iota_N\right\}^2}\left\|\widehat{\Phi}_N(z) - \Phi_N(z)\right\|\right) = O_p\left(\frac{\|\Phi_N^{-1}(z)\|}{N}\left\|\widehat{\Phi}_N(z) - \Phi_N(z)\right\|\right) = O_p\left(\frac{\left\|\widehat{\Phi}_N(z) - \Phi_N(z)\right\|}{N}\right).\end{aligned}$$

Therefore, using the above results it can be proved that the first element of $\left\|\widehat{\varpi} - \varpi\right\|$ is bounded by

$$\left\|\iota_N^\top\widehat{\Phi}_N^{-1}(z)\right\|\left|\left\{\iota_N^\top\widehat{\Phi}_N^{-1}(z)\iota_N\right\}^{-1} - \left\{\iota_N^\top\Phi_N^{-1}(z)\iota_N\right\}^{-1}\right| = O_p\left(\frac{\left\|\widehat{\Phi}_N(z) - \Phi_N(z)\right\|}{\sqrt{N}}\right). \tag{26}$$

On its part, considering the behavior of $\left\|\iota_N^\top\left\{\widehat{\Phi}_N^{-1}(z)-\Phi_N^{-1}(z)\right\}\right\|$ it can be shown

$$\begin{aligned}\left\|\iota_N^\top\left\{\widehat{\Phi}_N^{-1}(z)-\Phi_N^{-1}(z)\right\}\right\| &= \left\|\iota_N^\top\widehat{\Phi}_N^{-1}(z)\left\{\widehat{\Phi}_N(z)-\Phi_N(z)\right\}\Phi_N^{-1}(z)\right\| \leq \left\|\iota_N^\top\Phi_N^{-1}(z)\right\|\left\|\widehat{\Phi}_N(z)-\Phi_N(z)\right\|\left\|\widehat{\Phi}_N^{-1}(z)\right\| \\ &= O_p\left(\left\{\iota_N^\top\Phi_N^{-2}(z)\iota_N\right\}^{1/2}\left\|\widehat{\Phi}_N(z)-\Phi_N(z)\right\|\right).\end{aligned}$$

and using the above results it can be shown that the second term of (25) is bounded by

$$\begin{aligned}&\left\|\left\{\iota_N^\top\Phi_N(z)^{-1}\iota_N\right\}^{-1}\right\|\left\|\iota_N^\top\left\{\widehat{\Phi}_N(z)^{-1}-\Phi_N(z)^{-1}\right\}\right\| = \left\|\left\{\iota_N^\top\Phi_N^{-1}(z)\iota_N\right\}^{-1}\right\| O_p\left(\left\{\iota_N^\top\Phi_N^{-2}(z)\iota_N\right\}^{1/2}\left\|\widehat{\Phi}_N(z)-\Phi_N(z)\right\|\right) \\ &= O_p\left(\left\{\iota_N^\top\Phi_N^{-1}(z)\iota_N\right\}^{-1/2}\left\|\widehat{\Phi}_N(z)-\Phi_N(z)\right\|\right) = O_p\left(\left\{\iota_N^\top\Phi_N^{-1}(z)\iota_N\right\}^{-1/2} NR_{TH}\right), \qquad (27)\end{aligned}$$

given that $\left\{\iota_N^\top\Phi_N^{-1}(z)\iota_N\right\}^{1/2} = O(1)$ and following a similar proof scheme as in Lemma 6 it is straightforward to show that $\left\|\widehat{\Phi}_N(z)-\Phi_N(z)\right\| = O_p(NR_{TH})$. Hence, plugging (26)-(27) in (25) the proof of the lemma is done. □

Proof of Theorem 1: Plugging (1) into (6) and rearranging terms we get

$$\widehat{\beta}-\beta = \left(\sum_{i=1}^N \widehat{X}_{i\cdot}^\top M_{\widehat{\Lambda}}\widehat{X}_{i\cdot}\right)^{-1}\sum_{i=1}^N \widehat{X}_{i\cdot}^\top M_{\widehat{\Lambda}}(I_T-S)[F\gamma_i+m_i(Z)+\epsilon_{i\cdot}+O_p\{tr(H^2)\}+o_p(1)] \qquad (28)$$

given that $M_{\widehat{\Lambda}}(I_T-S)D=0$, since $D\in\Lambda$. Note that in (28) it can be seen that $\widehat{\beta}$ exhibits a direct dependence of the unobserved common factors (i.e., z_t and f_t).

Using Lemma 4 in (28), assuming that the rank condition holds, and by the uniform boundedness assumption on γ_i, the expression to study is

$$\sqrt{NT}(\widehat{\beta}-\beta) = \left(\frac{1}{NT}\sum_{i=1}^N \widetilde{X}_{i\cdot}^\top M_{\widetilde{G}}\widetilde{X}_{i\cdot}\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^N \widetilde{X}_{i\cdot}^\top M_{\widetilde{G}}\epsilon_{i\cdot} + O_p\left(\frac{\sqrt{T}}{N}\right)+O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\sqrt{T}c_{H_1}^2\right). \qquad (29)$$

Under Assumption 6 we can prove $\left\{(NT)^{-1}\sum_{i=1}^N \widetilde{X}_{i\cdot}^\top M_{\widetilde{G}}\widetilde{X}_{i\cdot}\right\}^{-1}\xrightarrow{p} Q^{-1}$, where $Q=\lim_{N,T\to\infty}(NT)^{-1}\sum_{i=1}^N E\left(\widetilde{X}_{i\cdot}^\top M_{\widetilde{G}}\widetilde{X}_{i\cdot}\right)$ and $(NT)^{-1}\sum_{i=1}^N \widetilde{X}_{i\cdot}^\top M_{\widetilde{G}}\epsilon_{i\cdot}\xrightarrow{p} 0$, so the consistency of this estimator follows almost immediately. Further, assuming $\sqrt{T}/N\to 0$ and $\sqrt{T}c_{H_1}^2\to 0$ as $(N,T)\to\infty$, we have

$$\sqrt{NT}(\widehat{\beta}-\beta) = Q^{-1}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^N \widetilde{X}_{i\cdot}^\top M_{\widetilde{G}}\epsilon_{i\cdot}\right)+o_p(1).$$

To obtain the asymptotic normality of $\widehat{\beta}$, we analyze the variance of the above expression and define $\widetilde{\mathcal{W}}_{i\cdot}^\top = \widetilde{X}_{i\cdot}^\top M_{\widetilde{G}}$ and $\widetilde{\mathcal{W}}_{\cdot t} = \left(\widetilde{\mathcal{W}}_{1t},\ldots,\widetilde{\mathcal{W}}_{Nt}\right)^\top$ as $p\times T$ and $N\times p$ matrices, respectively. Then, by the LIE, we can prove

$$\begin{aligned}Var\left[\sqrt{NT}(\widehat{\beta}-\beta)\right] &= \frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N E\left[Q^{-1}\widetilde{\mathcal{W}}_{i\cdot}^\top\epsilon_{i\cdot}\epsilon_{j\cdot}\widetilde{\mathcal{W}}_{j\cdot}Q^{-1}\right] \\ &= \frac{1}{NT}\sum_{t=1}^T\sum_{s=1}^T E\left[Q^{-1}\widetilde{\mathcal{W}}_{\cdot t}^\top E(\epsilon_{\cdot t}\epsilon_{\cdot s}|z_t)\widetilde{\mathcal{W}}_{\cdot s}Q^{-1}\right] = \frac{1}{NT}\sum_{t=1}^T E\left[Q^{-1}\widetilde{P}\widetilde{P}^\top Q^{-1}\right],\end{aligned}$$

where $\widetilde{P}$ is a $p\times T$ matrix such as $\widetilde{P}=\left[\widetilde{\mathcal{W}}_{\cdot 1}^\top\Omega(Z)^{1/2},\ldots,\widetilde{\mathcal{W}}_{\cdot T}^\top\Omega(Z)^{1/2}\right]$. Hence, using the above results in (29) we get

$$\sqrt{NT}(\widehat{\beta}-\beta) \quad \xrightarrow{d} \quad N\left(0, Q^{-1}\Psi Q^{-1}\right),$$

where $\Psi = \lim_{N,T\to\infty}(NT)^{-1}E\left[\widetilde{X}^\top\left(I_N\otimes M_{\widetilde{G}}\right)^\top\Omega(Z)\left(I_N\otimes M_{\widetilde{G}}\right)\widetilde{X}\right]$ given that rearranging terms it is straightforward to show $\widetilde{P}\widetilde{P}^\top = \widetilde{X}^\top\left(I_N\otimes M_{\widetilde{G}}\right)^\top\Omega(Z)\left(I_N\otimes M_{\widetilde{G}}\right)\widetilde{X}$ and the proof of the theorem is done. □

Proof of Theorem 4: Let $D \equiv (d_1, \dots, d_T)^\top$ and $F \equiv (f_1, \dots, f_T)^\top$ are $T \times n$ and $T \times r$ matrices, respectively, and $\epsilon_{i\cdot} \equiv (\epsilon_{i1}, \dots, \epsilon_{iT})^\top$ is a $T \times 1$ vector, it can be written $\widehat{Y}_{i\cdot} = (I_T - S)\{D\alpha_i + X_{i\cdot}\beta + m_i(Z) + F\gamma_i\} + \epsilon_{i\cdot} + O_p\{tr\{H_1^2\}\}$. Using the fact that $M_{\widehat{\Lambda}}(I_T - S)D = 0$, since $D \in \Lambda$, and assuming that the rank condition holds. If we stack the resulting expression over NT observations and replace $\widehat{Y}_{i\cdot}$ in (10), we get

$$\widehat{\beta}_{GLS} = \left\{\widehat{X}^\top \left(I_N \otimes M_{\widehat{\Lambda}}\right)\Omega^{-1}(Z)\left(I_N \otimes M_{\widehat{\Lambda}}\right)\widehat{X}\right\}^{-1} \widehat{X}^\top \left(I_N \otimes M_{\widehat{\Lambda}}\right)\Omega^{-1}(Z)\left(I_N \otimes M_{\widehat{\Lambda}}\right)\left\{X\beta + \sum_{\iota=1}^{r} F_\iota \otimes \gamma_\iota + \epsilon + O_p\left(c_{H_1}^2\right)\right\}, \tag{30}$$

where F_ι and γ_ι are $T \times 1$ and $N \times 1$ vectors, for $\iota \in \{1, \dots, r\}$, respectively, since it can be proved that, uniformly in z, $(NT)^{-1}\widehat{X}^\top \left(I_N \otimes M_{\widehat{\Lambda}}\right)\Omega^{-1}(Z)\left(I_N \otimes M_{\widehat{\Lambda}}\right)(I_T - S)m_i(Z) = O_p\left(c_{H_1}^2\right)$ by combining the proof scheme for Lemma 3 in [11] and Lemma A.6 in [51].

As it is quite common in this type of literature, in (30) is observed the direct dependence of $\widehat{\beta}$ of the observed and unobserved common factors (i.e., z_t and f_t). Using Lemmas 1-5 it is straightforward to show

$$\begin{aligned}\widehat{\beta}_{GLS} - \beta &= \left\{\frac{\widetilde{X}^\top \left(I_N \otimes M_{\widetilde{G}}\right)\Omega^{-1}(Z)\left(I_N \otimes M_{\widetilde{G}}\right)\widetilde{X}}{NT}\right\}^{-1} \frac{\widetilde{X}^\top \left(I_N \otimes M_{\widetilde{G}}\right)\Omega^{-1}(Z)\left(I_N \otimes M_{\widetilde{G}}\right)\epsilon}{NT} \\ &\quad + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{NT}\right) + O_p\left(c_{H_1}^2\right),\end{aligned} \tag{31}$$

where $M_{\widetilde{G}} = I_T - \widetilde{G}\left(\widetilde{G}^\top\widetilde{G}\right)^{-1}\widetilde{G}^{-1}$ is a $T \times T$ projection matrix, $\widetilde{G} = \left(\widetilde{D}, \widetilde{F}\right)$ is a $T \times (n + r)$ matrix.

Under the assumptions of the theorem we get $(NT)^{-1}\widetilde{X}^\top \left(I_N \otimes M_{\widetilde{G}}\right)\Omega^{-1}(Z)\left(I_N \otimes M_{\widetilde{G}}\right)\widetilde{X} \xrightarrow{p} Q_\varpi$, where

$$Q_\varpi = \lim_{N,T\to\infty}\left\{(NT)^{-1}\widetilde{X}^\top \left(I_N \otimes M_{\widetilde{G}}\right)\Omega^{-1}(Z)\left(I_N \otimes M_{\widetilde{G}}\right)\widetilde{X}\right\}.$$

Hence, using this result in (31) and assuming $\sqrt{T}/N \to 0$ and $\sqrt{NT}c_{H_1}^2 \to 0$, as $(N, T) \to \infty$, we get

$$\sqrt{NT}(\widehat{\beta}_{GLS} - \beta) = Q_\varpi^{-1}\left\{\frac{\widetilde{X}^\top \left(I_N \otimes M_{\widetilde{G}}\right)\Omega^{-1}(Z)\left(I_N \otimes M_{\widetilde{G}}\right)\epsilon}{\sqrt{NT}}\right\} + o_p(1).$$

□

Proof of Theorem 5: Plugging (14) in (16), a Taylor expansion leads to

$$\begin{aligned}&\sqrt{T|H_1|}\left\{\widehat{\overline{m}}_{GLS}(z, H_1, \varpi) - \overline{m}(z)\right\} - \sqrt{T|H_1|}\iota_1^\top\left\{T^{-1}Z_z^\top K_{H_1}(z)Z_z\right\}^{-1} Z_z^\top K_{H_1}(z)\left\{\frac{1}{2}Q_{\overline{m}}(z) + R_{\overline{m}}(z)\right\} \\ &= \sqrt{T|H_1|}\iota_1^\top\left\{T^{-1}Z_z^\top K_{H_1}(z)Z_z\right\}^{-1} Z_z^\top K_{H_1}(z)\widetilde{U}^{(\varpi)},\end{aligned} \tag{32}$$

where $Q_{\overline{m}}(z) = \left[(z_1 - z)^\top\mathcal{H}_{\overline{m}}(z)(z_1 - z), \dots, (z_T - z)^\top\mathcal{H}_{\overline{m}}(z)(z_T - z)\right]^\top$, $R_{\overline{m}}(z)$ is the residual term of the Taylor expansion, and $\widetilde{U}^{(\varpi)} \equiv (\varpi^\top U_{\cdot 1}, \dots, \varpi^\top U_{\cdot T})^\top$ is a $T \times 1$ vector. Using standard nonparametric techniques it can be proved that $\iota_1^\top\left(T^{-1}Z_z^\top K_{H_1}(z)Z_z\right)^{-1} Z_z^\top K_{H_1}(z)R_{\overline{m}}(z) = o_p\left\{tr(H_1^2)\right\}$ and that the asymptotic bias of $\widehat{\overline{m}}_{GLS}(z; H_1, \varpi)$ is

$$\iota_1^\top\left\{T^{-1}Z_z^\top K_H(z)Z_z\right\}^{-1} Z_z^\top K_{H_1}(z)\left\{\frac{1}{2}Q_{\overline{m}}(z) + R_{\overline{m}}(z)\right\} = \frac{\mu_2^q(K)}{2}tr\left\{H_1^2\mathcal{H}_{\overline{m}}(z)\right\} + o_p\left\{tr(H_1^2)\right\}. \tag{33}$$

Let $\nu_N^{(\varpi)}(z) = \left\{\iota_N^\top\Phi_N^{-1}(z)\iota_N\right\}$ and consider now the variance term of the right-hand side of (32), by the LIE we get

$$\begin{aligned}T|H_1|\text{Var}\left[T^{-1}Z_z^\top K_{H_1}(z)\widetilde{U}^{(\varpi)}\right] &= T^{-1}|H_1|E\left[Z_z^\top K_{H_1}(z)E\left(\widetilde{U}^{(\varpi)}\widetilde{U}^{(\varpi)\top}|z_t\right)K_{H_1}(z)Z_z\right] \\ &= \begin{pmatrix} \nu_N^{(\varpi)}(z)R^q(K)\rho_{z_t}(z) + o(1) & O(|H_1|) \\ O(|H_1|) & H_1^2\nu_N^{(\varpi)}(z)R_2^q(K)\rho_{z_t} + o(H_1^2)\end{pmatrix}.\end{aligned} \tag{34}$$

Using Lemma 7 and (34), by the Slutsky theorem, as $T \to \infty$,

$$\mathrm{Var}\left[\sqrt{T|H_1|}\iota_1^\top \left\{T^{-1}Z_z^\top K_{H_1}(z)Z_z\right\}^{-1} Z_z^\top K_{H_1}(z)\widetilde{U}^{(\varpi)}\right] = \frac{R^q(K)\nu_N^{(\varpi)}(z)}{\rho_{z_t}(z)}. \quad (35)$$

Finally, the Lyapunov condition can be proved under Assumption 17, so using (33) and (35) the proof of the Theorem is completed. □

Proof of Theorem 6: Denote $\widehat{\Psi}_{NT} = \widetilde{X}^\top \left(I_T \otimes M_{\widetilde{G}}\right)\widehat{\Omega}^{-1}(Z)\left(I_T \otimes M_{\widetilde{G}}\right)\widetilde{X}$ and $\Psi_{NT} = \widetilde{X}^\top \left(I_T \otimes M_{\widetilde{G}}\right)\Omega^{-1}(Z)\left(I_T \otimes M_{\widetilde{G}}\right)\widetilde{X}$ and following a similar reasoning as in the proof of Theorem 5, it is easy to show

$$\begin{aligned}
\widehat{\beta}_{FGLS} - \beta &= \widehat{\Psi}_{NT}^{-1}\widetilde{X}^\top \left(I_T \otimes M_{\widetilde{G}}\right)\widehat{\Omega}^{-1}(Z)\left(I_T \otimes M_{\widetilde{G}}\right)\epsilon + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{NT}\right) + O_p\left(c_{H_1}^2\right),\\
\widehat{\beta}_{GLS} - \beta &= \Psi_{NT}^{-1}\widetilde{X}^\top \left(I_T \otimes M_{\widetilde{G}}\right)\Omega^{-1}(Z)\left(I_T \otimes M_{\widetilde{G}}\right)\epsilon + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{NT}\right) + O_p\left(c_{H_1}^2\right).
\end{aligned}$$

Using $\{a_1a_2 - b_1b_2 = (a_1 - b_1)(a_2 - b_2) + (a_1 - b_1)b_2 + b_1(a_2 - b_2)\}$ over the above results and rearranging terms,

$$\begin{aligned}
\widehat{\beta}_{FLGS} - \widehat{\beta}_{GLS} &= \left(\widehat{\Psi}_{NT}^{-1} - \Psi_{NT}^{-1}\right)\widetilde{X}^\top \left(I_T \otimes M_{\widetilde{G}}\right)\left\{\widehat{\Omega}^{-1}(Z) - \Omega^{-1}(Z)\right\}\left(I_T \otimes M_{\widetilde{G}}\right)\epsilon\\
&\quad + \left(\widehat{\Psi}_{NT}^{-1} - \Psi_{NT}^{-1}\right)\widetilde{X}^\top \left(I_T \otimes M_{\widetilde{G}}\right)\Omega^{-1}(Z)\left(I_T \otimes M_{\widetilde{G}}\right)\epsilon + \Psi_{NT}^{-1}\widetilde{X}^\top \left(I_T \otimes M_{\widetilde{G}}\right)\left\{\widehat{\Omega}^{-1}(Z) - \Omega^{-1}(Z)\right\}\left(I_T \otimes M_{\widetilde{G}}\right)\epsilon\\
&\quad + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{NT}\right) + O_p\left(c_{H_1}^2\right)\\
&= \boldsymbol{I}_{g_1} + \boldsymbol{I}_{g_2} + \boldsymbol{I}_{g_3} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{NT}\right) + O_p\left(c_{H_1}^2\right), \quad (36)
\end{aligned}$$

where the definitions of $\boldsymbol{I}_{g_j}$, for $j = 1, 2, 3$, should be apparent from the context and they have to be analyzed separately.

Given that $\widehat{\varpi}_{ij}(z)$ and $\varpi_{ij}(z)$ are the (ij)th element of $\widehat{\Omega}^{-1}(Z)$ and $\Omega(Z)$, respectively, and using Lemma 6 it is easy to show $\left\|\Omega^{-1}(Z) - \Omega^{-1}(Z)\right\| = O_p(NR_{TH})$, as $N/T \to \kappa$, where κ is a positive constant. Therefore, to finish the proof is enough to show

$$\frac{1}{NT}\left(\widehat{\Psi}_{NT} - \Psi_{NT}\right) = o_p(1), \quad (37)$$

$$\frac{1}{NT}\widetilde{X}^\top \left(I_N \otimes M_{\widetilde{G}}\right)^\top \left\{\Omega^{-1}(Z) - \Omega^{-1}(Z)\right\}(I_N \otimes M_{\widetilde{G}})\epsilon = O_p\left\{(NT)^{-1/2}\right\}O_p(NR_{TH}). \quad (38)$$

Considering the proof of (37), it has the norm bounded by

$$\begin{aligned}
&\left\|(NT)^{-1}\widetilde{X}^\top \left(I_N \otimes M_{\widetilde{G}}\right)^\top \left\{\widehat{\Omega}^{-1}(Z) - \Omega^{-1}(Z)\right\}\left(I_N \otimes M_{\widetilde{G}}\right)\widetilde{X}\right\| \le \left\|(NT)^{-1}\widetilde{X}^\top \left(I_N \otimes M_{\widetilde{G}}\right)\widetilde{X}\right\|\left\|\widehat{\Omega}^{-1}(Z) - \Omega^{-1}(Z)\right\|\\
&\le \left\|(NT)^{-1}\widetilde{X}^\top \left(I_N \otimes M_{\widetilde{G}}\right)\widetilde{X}\right\|\left\|\widehat{\Omega}^{-1}(Z)\left\{\widehat{\Omega}(Z) - \Omega(Z)\right\}\Omega^{-1}(Z)\right\| \le \left\|(NT)^{-1}\widetilde{X}^\top \left(I_N \otimes M_{\widetilde{G}}\right)\widetilde{X}\right\|\left\|\Omega^{-2}(Z)\right\|\left\|\widehat{\Omega}(Z) - \Omega(Z)\right\|\\
&= O_p(NR_{TH}),
\end{aligned}$$

using the fact that $(NT)^{-1}\widetilde{X}^\top \left(I_N \otimes M_{\widetilde{G}}\right)\widetilde{X} = O_p(1)$, $\left\|\Omega^{-1}(Z)\right\| = O_p(1)$, $\|\widehat{\Omega}(Z) - \Omega(Z)\| = O_p(NR_{TH})$ (see Lemma 6). Furthermore, following a similar reasoning as in [45] (proof of Theorem 6) and using Assumption 14, it can be proved that $NR_{TH} = o\left\{(NT|H_1|)^{-1/2} + tr(H_1^2)\right\}$ and the proof of (37) is done.

Similarly, it can be shown that the norm of (38) is bounded by

$$\begin{aligned}
&\left\|(NT)^{-1}\widetilde{X}^\top \left(I_N \otimes M_{\widetilde{G}}\right)^\top \left\{\widehat{\Omega}^{-1}(Z) - \Omega^{-1}(Z)\right\}\left(I_N \otimes M_{\widetilde{G}}\right)\epsilon\right\|\\
&\le (NT)^{-1/2}\left\|(NT)^{-1/2}\widetilde{X}^\top \left(I_N \otimes M_{\widetilde{G}}\right)\epsilon\right\|\left\|\Omega_N^{-2}(Z)\right\|\left\|\widehat{\Omega}(Z) - \Omega(Z)\right\| = O_p\left\{(NT)^{-1/2}\right\}O_p(NR_{TH}).
\end{aligned}$$

Finally, using (37)-(38) in (36) it is straightforward to show

$$\sqrt{NT}\left(\widehat{\beta}_{FGLS} - \widehat{\beta}_{GLS}\right) = O_p(1) + O_p\left(\sqrt{\frac{T}{N}}\right) + O_p\left(\sqrt{NT}c_{H_1}^2\right)$$

and given that $T/N \to 0$ and $\sqrt{NT}c_{H_1}^2 \to 0$, as $(N, T) \to \infty$, the proof of the theorem is done. □

Proof of Theorem 7: To prove this theorem, it can be written

$$\widehat{\widehat{m}}_{FLGS}(z; H_1, \widehat{\varpi}) - \widehat{\widehat{m}}_{GLS}(z; H_1, \varpi) \quad = \quad \iota_1^\top \left\{ Z_z^\top K_{H_1}(z) Z_z \right\}^{-1} Z_z^\top K_{H_1}(z) \left\{ \widehat{\widetilde{Y}}\widehat{\varpi} - \widetilde{Y}\varpi \right\}, \tag{39}$$

where $\widehat{\widetilde{Y}}$ and $\widetilde{Y}$ are $T \times N$ matrices whose it-th elements are such as $\widehat{\widetilde{y}}_{it} = y_{it} - x_{it}^\top \widehat{\beta} - \lambda_t^\top \widehat{\delta}_i$ and $\widetilde{y}_{it} = y_{it} - x_{it}^\top \beta - \lambda_t^\top \delta_i$, respectively. Replacing (3) in (39) and rearranging terms, the final expression to analyze is such us

$$\left| \widehat{\widehat{m}}_{FLGS}(z; H_1, \widehat{\varpi}) - \widehat{\widehat{m}}_{GLS}(z; H_1, \varpi) \right| \le \iota_1^\top \left\| \left\{ Z_z^\top K_{H_1}(z) Z_z \right\}^{-1} Z_z^\top K_{H_1}(z) \left\{ U - \sum_{\varrho=1}^{p} X_\varrho \left(\widehat{\beta}_\varrho - \beta_\varrho \right) - \Lambda \left(\widehat{\delta} - \delta \right)^\top \right\} \right\| \left\| \widehat{\varpi} - \varpi \right\|, \tag{40}$$

where U and X_ϱ are $T \times N$ matrices and $\widehat{\delta}$ and δ are $N \times \ell$ matrices. From the results in Lemma 7, it is easy to show $\|T^{-1} Z_z^\top K_{H_1}(z) Z_z\| = O_p\left((T|H_1|)^{-1/2} \right)$, whereas considering the behavior of the numerator term in (40), we have

$$\begin{aligned} \left\| T^{-1} Z_z^\top K_{H_1}(z) \left\{ U - \sum_{\varrho=1}^{p} X_\varrho \left(\widehat{\beta}_\varrho - \beta_\varrho \right) - \Lambda \left(\widehat{\delta} - \delta \right)^\top \right\} \right\| \quad &\le \quad \|T^{-1} Z_z^\top K_{H_1}(z) U\| + \|T^{-1} Z_z^\top K_{H_1}(z) X\| \left\| \widehat{\beta} - \beta \right\| \\ &+ \quad \|T^{-1} Z_z^\top K_{H_1}(z) \Lambda\| \left\| \widehat{\delta} - \delta \right\|. \end{aligned} \tag{41}$$

Using the consistency result obtained previously for $\widehat{\beta}$, it can be shown $\left\| \widehat{\beta} - \beta \right\| = O_p((NT)^{-1/2})$ and, under similar reasoning, it is straightforward to show $\left\| \widehat{\delta} - \delta \right\| = O_p(T^{-1/2})$. Following a similar reasoning as in [47] and using these results in (41), we can prove that $\|T^{-1} Z_z^\top K_{H_1}(z) X\|$ and $\|T^{-1} Z_z^\top K_{H_1}(z) \Lambda\|$ are $O_p((T|H_1|)^{-1/2})$. Using all these results in (41) and given that by Lemma 8 we get $\|\widehat{\varpi} - \varpi\| = O_p(NR_{TH})$ as $N/T \to \kappa$, where κ is a positive constant, we have

$$\iota_1^\top \left\| T^{-1} Z_z^\top K_{H_1}(z) Z_z \right\| \left\| T^{-1} Z_z^\top K_{H_1}(z) X \right\| \left\| \widehat{\beta} - \beta \right\| = o_p(NR_{TH}), \tag{42}$$

$$\iota_1^\top \left\| T^{-1} Z_z^\top K_{H_1}(z) Z_z \right\| \left\| T^{-1} Z_z^\top K_{H_1}(z) \Lambda \right\| \left\| \widehat{\delta} - \delta \right\| = o_p(NR_{TH}). \tag{43}$$

Focusing now on the behavior of $\|T^{-1} Z_z^\top K_{H_1}(z) U\|$ and using the Markov's inequality, it can be proved

$$\|T^{-1} Z_z^\top K_{H_1}(z) U\| = \begin{pmatrix} O_p\left(\left\| \Phi_N^{1/2}(z) \right\| (T|H_1|)^{-1/2} \right) \\ O_p\left(\left\| \Phi_N^{1/2}(z) \right\| tr(H_1^2) (T|H_1|)^{-1/2} \right) \end{pmatrix}, \tag{44}$$

given that by the LIE,

$$\begin{aligned} E \left\| T^{-1} \sum_{t=1}^{T} K_{H_1}(z_t - z) u_{\cdot t} \right\|^2 \quad &= \quad tr\left\{ T^{-2} \sum_{t=1}^{T} E\left[K_{H_1}^2(z_t - z) E(u_{\cdot t} u_{\cdot t}^\top | z_t) \right] \right\} \\ &= \quad \frac{R^q(K) \rho_{z_t}(z)}{T|H_1|} \left\| \Phi_N^{1/2}(z) \right\| = O_p\left(\frac{\left\| \Phi_N^{1/2}(z) \right\|}{T|H_1|} \right) \end{aligned}$$

and $E \left\| T^{-1} \sum_{t=1}^{T} K_{H_1}(z_t - z)(z_t - z) u_{\cdot t} \right\|^2 = O_p\left(\left\| \Phi_N^{1/2}(z) \right\| tr(H_1^2)/T|H_1| \right)$. Hence, under a similar reasoning as in (42)-(43) and using Assumption 14 and (44) we have

$$\iota_1^\top \left\| T^{-1} Z_z^\top K_{H_1}(z) Z_z \right\| \left\| T^{-1} Z_z^\top K_{H_1}(z) U \right\| = o_p\left\{ \frac{\nu_N^{-1/2}(z)}{\sqrt{NT|H_1|}} + tr(H_1^2) \right\}. \tag{45}$$

Finally, plugging (42)–(45) in (40) the proof of the theorem is done. □

References

[1] M. S. Andersen, P. Ekins, Carbon-Energy Taxation: Lessons from Europe, OUP Oxford, 2009.
[2] D. W. K. Andrews, Cross-section regression with common shocks, Econometrica 73 (2005) 1551–1585.
[3] J. Bai, S. Ng, A panic attack on unit roots and cointegration, Econometrica 72 (2004) 1127–1177.
[4] N. Bailey, S. Holly, M. H. Pesaran, A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence, Journal of Applied Econometrics 31 (2016) 249–280.
[5] N. Bailey, G. Kapetanios, M. H. Pesaran, Exponent of cross-sectional dependence: Estimation and inference, Journal of Applied Econometrics 31 (2016) 929–960.
[6] B. H. Baltagi, G. Bresson, J.-M. Etienne, Carbon dioxide emissions and economic activities: A mean field variational bayes semiparametric panel data model with random coefficients, Annals of Economics and Statistics 134 (2019) 43–77.
[7] B. H. Baltagi, G. Bresson, A. Pirotte, Tobin q: forecast performance for hierarchical bayes, shrinkage, heterogeneous and homogeneous panel data estimators, Empirical Economics 29 (2004) 107–113.
[8] C. K. Bolat, U. Soytas, B. Akinoglu, S. Nazlioglu, Is there a macroeconomic carbon rebound effect in EU ETS?, Energy Economics 125 (2023) 106879.
[9] S. Borghesi, C. Franco, G. Marin, Outward foreign direct investment patterns of Italian firms in the European Union's emission trading scheme, The Scandinavian Journal of Economics 122 (2020) 219–256.
[10] S. Borghesi, M. Montini, A. Barreca, The European emission trading system and its followers: Comparative analysis and linking perspectives, Springer, 2016.
[11] Z. Cai, Y. Fan, M. Lin, J. Su, Inferences for a partially varying coefficient models with endogenous regressors, Journal of Business and Economic Statistics 37 (2019) 158–170.
[12] R. Calel, Adopt or innovate: Understanding technological responses to cap-and-trade, American Economic Journal: Economic Policy 12 (2020) 170–201.
[13] R. Calel, A. Dechezlepretre, Environmental policy and directed technological change: Evidence from the European carbon market, Review of Economics and Statistics 98 (2016) 173–191.
[14] M. Carratù, B. Chiarini, P. Piselli, Effects of european emission unit allowance auctions on corporate profitability, Energy Policy 144 (2020) 111584.
[15] S. A. Churchill, J. Inekwe, K. Ivanovski, R. Smyth, The environmental Kuznets curve in the OECD: 1870–2014, Energy Economics 75 (2018) 389–399.
[16] M. A. Cole, R. J. Elliott, K. Shimamoto, Industrial characteristics, environmental regulations and air pollution: An analysis of the UK manufacturing sector, Journal of Environmental Economics and Management 50 (2005) 121–143.
[17] J. Colmer, R. Martin, M. Muuls, U. J. Wagner, Does pricing carbon mitigate climate change? Firm-level evidence from the European Union emissions trading scheme, 2022. CEPR Discussion Paper No. DP16982.
[18] V. Costantini, M. Mazzanti, A. Montini, Environmental performance, innovation and spillovers. Evidence from a regional NAMEA, Ecological Economics 89 (2013) 101–114.
[19] A. D. Ellerman, P. L. Joskow, The European Union's emissions trading system in perspective, Pew Center on Global Climate Change Arlington, VA, 2008.
[20] A. D. Ellerman, C. Marcantonini, A. Zaklan, The European Union emissions trading system: Ten years and counting, Review of Environmental Economics and Policy 10 (2016).
[21] C. Ertur, A. Musolesi, Weak and strong cross-sectional dependence: A panel data analysis of international technology diffusion, Journal of Applied Econometrics 32 (2017) 477–503.
[22] Y. F. Fernandez, M. F. Lopez, B. O. Blanco, Innovation for sustainability: The impact of R&D spending on CO2 emissions, Journal of Cleaner Production 172 (2018) 3459–3467.
[23] Z. Griliches, Patent Statistics as Economic Indicators: A Survey, in: Z. Griliches (Ed.), R&D and Productivity: The Econometric Evidence, University of Chicago Press, Chicago, 1998, pp. 287–343.
[24] J. D. Hamilton, What is an oil shock?, Journal of Econometrics 113 (2003) 363–398.
[25] B. Hansen, Uniform convergence rates for kernel estimation with dependent data, Econometric Theory 24 (2008) 726–748.
[26] C. Hsiao, A. Tahmiscioglu, Estimation of dynamic panel data models with both individual and time-specific effects, Journal of Statistical and Planning Inference 138 (2008) 2698–2721.
[27] X. Huang, Nonparametric estimation in large panels with cross-sectional dependence, Econometric Reviews 32 (2013) 754–777.
[28] A. Juodis, S. Reese, The incidental parameters problem in testing for remaining cross-section correlation, Journal of Business & Economic Statistics 40 (2022) 1191–1203.
[29] D. R. Känzig, M. Konradt, Climate policy and the economy: evidence from Europe's carbon pricing Initiatives, Technical Report, National Bureau of Economic Research, 2023.
[30] M. P. Keane, E. S. Prasad, The employment and wage effects of oil price changes: A sectoral analysis, The Review of Economics and Statistics (1996) 389–400.
[31] E. Koçak, Z. Ş. Ulucak, The effect of energy R&D expenditures on CO2 emission reduction: estimation of the STIRPAT model for OECD countries, Environmental Science and Pollution Research 26 (2019) 14328–14338.
[32] N. Koch, M. Themann, Catching up and falling behind: Cross-country evidence on the impact of the EU ETS on firm productivity, Resource and Energy Economics 69 (2022) 101315.
[33] J. Lee, P. Robinson, Panel nonparametric regression with fixed effects, Journal of Econometrics 188 (2015) 346–362.
[34] G. Marin, M. Marino, C. Pellegrin, The impact of the of the european emission trading scheme on multiple measures of economic performance, Environmental and Resource Economics 71 (2018) 551–582.
[35] M. Mazzanti, A. Musolesi, The heterogeneity of carbon Kuznets curves for advanced countries: Comparing homogeneous, heterogeneous and shrinkage/bayesian estimators, Applied Economics 45 (2013) 3827–3842.

[36] D. L. Millimet, J. A. List, T. Stengos, The environmental Kuznets curve: Real progress or misspecified models?, Review of Economics and Statistics 85 (2003) 1038–1047.
[37] A. Musolesi, M. Mazzanti, Nonlinearity, heterogeneity and unobserved effects in the carbon dioxide emissions-economic development relation for advanced countries, Studies in Nonlinear Dynamics and Econometrics 18 (2014) 521–541.
[38] M. Papież, S. Śmiech, K. Frodyma, Does the European Union energy policy support progress in decoupling economic growth from emissions?, Energy Policy 170 (2022) 113247.
[39] M. H. Pesaran, Estimation and inference in large heterogeneous panels with a multifactor error structure, Econometrica 74 (2006) 967–1012.
[40] M. H. Pesaran, Testing weak cross-sectional dependence in large panels, Econometric Reviews 34 (2015) 1089–1117.
[41] M. H. Pesaran, General diagnostic tests for cross-sectional dependence in panels, Empirical Economics 60 (2021) 13–50.
[42] M. H. Pesaran, E. Tosetti, Large panels with common factors and spatial correlation, Journal of Econometrics 161 (2011) 182–202.
[43] P. Phillips, D. Sul, Dynamic panel estimation and homogeneity testing under cross section dependence, The Econometrics Journal 6 (2003) 217–259.
[44] S. Reese, J. Westerlund, Panicca: Panic on cross-section averages, Journal of Applied Econometrics 31 (2016) 961–981.
[45] P. M. Robinson, Nonparametric trending regression with cross-sectional dependence, Journal of Econometrics 169 (2012) 4–14.
[46] A. Rupasingha, S. J. Goetz, D. L. Debertin, A. Pagoulatos, The environmental Kuznets curve for US counties: A spatial econometric analysis with extensions, Papers in Regional Science 83 (2004) 407–424.
[47] D. Ruppert, M. P. Wand, Multivariate locally weighted least squares regression, Annals of Statistics 22 (1994) 1346–1370.
[48] A. Soberon, A. Musolesi, J. Rodriguez-Poo, A semiparametric panel data model with common factors and spatial dependence, Oxford Bulletin of Economics and Statistics 86 (2024) 905–927.
[49] A. Soberon, J. M. Rodriguez-Poo, P. M. Robinson, Nonparametric panel data regression with parametric cross-sectional dependence, The Econometrics Journal 25 (2022) 114–133.
[50] J. Su, S. Jin, Sieve estimation of panel data models with cross section dependence, Journal of Econometrics 169 (2012) 34–47.
[51] L. Su, S. Jin, Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models, Journal of Econometrics 157 (2010) 18–23.
[52] M. Wagner, The environmental Kuznets curve, cointegration and nonlinearity, Journal of Applied Econometrics 30 (2015) 948–967.
[53] M. Wagner, P. Grabarczyk, S. H. Hong, Fully modified OLS estimation and inference for seemingly unrelated cointegrating polynomial regressions and the environmental Kuznets curve for carbon dioxide emissions, Journal of Econometrics 214 (2020) 216–255.

"Efficient estimation of a partially linear panel data model with cross-sectional dependence"

Author Statement:

Alexandra Soberon: Methodology, Software, Writing-Original draft preparation, Supervision, Funding acquisition. **Juan Manuel Rodriguez-Poo**: Conceptualization, Methodology; Writing-Reviewing and Editing. **Antonio Musolesi**: Data curation, Formal análisis. **Massimiliano Mazzanti**: Data curation, Formal analysis