# Efficient Semiparametric Estimation of European Climate Policy Effects

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

[a]Department of Economics, Universidad de Cantabria, Spain

[b]Department of Economics and Business, University of Ferrara, Italy

28th October 2023

**Abstract**

The European Union Emissions Trading System (EU ETS) has become the cornerstone of the European Union's strategy to decarbonize the economy and mitigate climate change. Following these objectives, the aim of this paper is to assess the impact of the price of carbon, which is linked to the European market of allowances, on carbon dioxide emissions. To do so, we propose an econometric model that extends the Environmental Kuznets Curve (EKC) model in several directions. First, the price of carbon, which is the policy variable, is introduced in the model in a nonparametric fashion; Second, we propose to use an interactive fixed effects approach to control for latent heterogeneities in both dimensions of panel data; Third, to allow for spatial dependence, we introduced spatially correlated errors. The extended EKC model poses various challenges for estimation. To cope with them, using a profile likelihood approach, we propose a Feasible Generalized Least Squares estimator of the parameters of interest. Furthermore, the policy effects curve is also efficiently estimated. The asymptotic properties of the estimators are shown and, based on these outcomes, we empirically evaluate the policy effects. Our approach yields significantly different and more meaningful results compared to those obtained using standard estimation techniques.

*Keywords*: Common factor, Cross-sectional dependence, Semiparametric regression, Generalized least squares, Panel data, EU ETS.

*JEL Classification*: C13, C14, C33, O3.

# 1 Introduction

During the last decades, the growing interest of citizens and governments in environmental degradation has led the world's major economies to take measures to mitigate the effects of global warming and climate change through intergovernmental negotiations and binding agreements such as the Kyoto Protocol or the Paris Agreement of 2015, among others. The reduction of greenhouse gas emissions has become one of the priority objectives of the countries that signed these agreements and the implementation of the European Union Emissions Trading System (EU ETS) in 2005 has become the cornerstone of the European Union's strategy to decarbonize the economy and mitigate climate change (Borghesi et al., 2020).

The literature investigating the effects of carbon pricing on economic and environmental performances has been developed for decades, focusing mainly on green taxation (Andersen and Ekins, 2009), with some recent developments (Metcalf and Stock, 2023). The other pillar of carbon pricing, i.e. the EU ETS, has witnessed relevant work since its origin in 2005.

As far as the effects of EU ETS on economic or environmental indicators such as $CO_2$, innovation, and GDP are concerned, three main research streams have been developed focusing on different outcomes: (i) Innovation, generally exploiting firm-level data of innovation activities (see Calel and Dechezlepretre, 2016); (ii) Economic performances, such as productivity, GDP, or investment (see Marin et al., 2018; Carratù et al., 2020; Koch and Themann, 2022 and Borghesi et al., 2020); and (iii) Environmental outcomes, generally using firms' data and focusing on carbon dioxide ($CO_2$) emissions as the dependent variable for the impact analysis. Furthermore, a recent paper by Colmer et al. (2022) has pointed out the existence of mitigation effects, and in Papież et al. (2022) decoupling of emissions from economic growth have been observed.

Nevertheless, despite the interesting results obtained in the previous papers, most of their conclusions have been undertaken at firm-level. Alternatively, there exist some papers that take a country-based macroeconomic data perspective (see Känzig and Konradt, 2023). Following this latter approach, in this paper we exploit a macroeconomic panel data set based on EU27 countries plus the UK, Iceland and Norway and develop a new approach to assess the impact of EU ETS on $CO_2$ emissions. In order to do so, we propose a partially linear Environmental Kuznets Curve (EKC) specification (see Millimet et al., 2003 and Baltagi et al., 2019, among others) where the key policy variable, which is invariant across units, is the price of carbon that is linked to the European market of allowances. Beyond the conditioning variables that appear in a standard EKC specification, it might be the case that some other unknown common factors could affect individual countries in a heterogeneous manner (see Mazzanti and Musolesi, 2013 and Musolesi and Mazzanti, 2014). To account for this issue we introduce an interactive fixed effects specification. Finally, following Rupasingha et al. (2004) we also account for spatial dependence.

It is important to note that, in principle, more flexible specifications, such as fully nonparametric models (Millimet et al., 2003; Azomahou et al., 2006; Musolesi and Mazzanti, 2014), could be of interest. However, in the case of moderate sample sizes, as the present one and quite often with macro data, this approach presents serious drawbacks. Indeed, the proposed specification is a partially linear model, which has a long tradition in both theoretical and applied econometrics (see Härdle et al., 2000 for a comprehensive review of the literature). The rationale of this model is that it represents a compromise between fully nonparametric models, which are limited in application because the rate of convergence is inversely related to the number of regressors (the so-called curse of dimensionality problem), and fully parametric models, which are often too much restrictive. The appeal of the semi-parametric partially linear model is that the parameters that appear in the linear component have a rate of convergence that exceeds that for fully nonparametric models while it also allows flexibility for a subset of the explanatory variables.

To summarize, the aim of this paper is to evaluate the impact of EU ETS on $CO_2$ emissions. To do so, we propose to estimate a partially linear EKC specification with both interactive fixed effects and spatially correlated errors. In order to do so, we use a macroeconomic EU panel data set. Since the seminal paper by Robinson (1988), several procedures have been developed for the estimation and inference of partially linear models. In the presence of interactive fixed effects, root-$NT$ consistency of the parameters of interest is obtained in Huang et al. (2021). Unfortunately, neither spatial dependence nor common policy variables that do not vary across individuals are allowed. Nor do they propose an estimator for the nonparametric component that is crucial in our case. As it is already well known, in this context, ignoring spatial correlation can render inefficient estimators and biased inference results. To cope with this problem, in this paper, we propose Feasible Generalized Least Squares (FGLS) $\sqrt{NT}$-consistent estimators of the parameters of interest. Furthermore, the nonparametric estimator of the policy effects is also efficiently estimated. The estimation technique is based on a profile likelihood approach (see Fan and Huang, 2005).

The rest of the paper is organized as follows. Section 2 introduces the model and discusses the estimation methods. Asymptotic properties of the proposed estimators are then established in Section 3, while Section 4 applies the proposed methodology to evaluate the effect of the EU ETS on $CO_2$ emissions. Finally, Section 5 concludes the paper. A Monte Carlo simulation study is conducted in the Appendix to demonstrate the finite sample performance of the proposed estimators. Furthermore, all mathematical proofs are relegated to the appendix.

# 2 Econometric model and estimation procedures

## 2.1 Model specification

Let us consider an EKC specification that has been widely employed in the econometric literature (Wagner, 2015; Wagner et al., 2020; Mazzanti and Musolesi, 2013), i.e.

$$co_2 = \alpha + \beta_1 gdp + \beta_2 gdp^2 + \beta_3 r\&d + error,$$

where $co_2$ refers to the level of $CO_2$ emissions per capita and $gdp$ stands for GDP per capita. Furthermore, we consider $r\&d$ as a proxy for the level of technology (see Costantini et al., 2013 and Baltagi et al., 2019) and it is also expressed in per capita terms. All variables are expressed in log terms.

Since the main aim of our work is to assess the impact of EU ETS on $CO_2$ emissions, following Costantini et al. (2013) and Cole et al. (2005) as a framework, we will introduce in the above model the environmental policy variable and specifically consider the carbon price, which arises from the market of allowances. This variable varies over time but not over cross-sectional units, so it can be introduced into the model as a common stochastic covariate, $z_t$, taken also in log form. The aggregate effect of EU ETS carbon pricing on $co_{2it}$ could be the result of economic mechanisms that involves composite and rather complex negative and positive effects on emissions. Indeed, on one side, negative effects on emission can be due to the EU ETS cap itself, the abatement-oriented induced innovation (Calel and Dechezlepretre, 2016; Calel, 2020) effect of targeted high emissions industrial sectors, 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 happens if carbon policies improve energy efficiency, which leads to an energy rebound effect and in turns it may produce a carbon rebound effect (Bolat et al., 2023). Consequently, 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. Furthermore, it can be also expected that both common price, $z_t$, and unobserved common factors, $f_t$, may produce a heterogeneous effect across units due to country-specific economic or technological features. To account for these issues, we consider the following model:

$$co_{2it} = \alpha_i + \beta_1 gdp_{it} + \beta_2 gdp_{it}^2 + \beta_3 r\&d_{it} + m_i(z_t) + \gamma_i' f_t + \epsilon_{it}, \tag{2.1}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where $f_t$ is a vector of unobserved common factors, $\gamma_i$ are the corresponding factor loadings, $\epsilon_{it}$ is the idiosyncratic error term, and $m_i(\cdot)$ is an unknown smooth function. A more flexible specification such as models with heterogeneous slopes (Mazzanti and Musolesi, 2013) could be of interest. However, they have a theoretical justification only for large $T$ and often underperform with respect to homogeneous estimators (Baltagi et al., 2004).

Just for notational convenience, let $y_{it} = co_{2it}$, $x_{it}' = (gdp_{it}, gdp_{it}^2, r\&d_{it})'$ and hence, for $i = 1, \ldots, N, \quad t = 1, \ldots, T$, the regression model (2.1) becomes

$$y_{it} \;=\; \alpha_i' d_t + x_{it}'\beta + m_i(z_t) + \gamma_i' f_t + \epsilon_{it}. \tag{2.2}$$

For the sake of generality, through the paper we define $d_t = (d_{1t}, d_{2t}, \ldots, d_{nt})'$ as a $n \times 1$ vector of deterministic components (such as time and seasonal dummies), $x_{it} \in I\!R^p$ is a vector of individual-specific explanatory variables on the $i$th cross-sectional at time $t$ and $z_t \in I\!R^q$ is a vector of observed continuous common stochastic covariates (policy effects), $m_i(\cdot)$ is an unknown smooth function to estimate, and $\beta$ is the unknown slope parameter to estimate. $f_t$ is a $r \times 1$ vector of unobserved common factors that are allowed to simultaneously affect all cross-section units, albeit with different degrees measured with the factor loadings, $\gamma_i$, and $\epsilon_{it}$ is an idiosyncratic error. Furthermore, spatial dependence is introduced by assuming that the $\epsilon_{it}$'s are conditionally correlated and heteroscedastic and the unobserved factors, $f_t$, are allowed to be correlated with the observed data $(x_{it}, z_t)$ through the following specification

$$x_{it} = A_i' d_t + g_i(z_t) + \Gamma_i' f_t + v_{it}, \tag{2.3}$$

where $A_i$ and $\Gamma_i$ are $p \times N$ and $p \times r$ 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.

## 2.2 Efficient estimation

In order to obtain efficient estimators for $\beta$ and $m_i(\cdot)$ we propose a standard two-step procedure. At the first stage, we will construct consistent estimators of both $\beta$ and $m_i(\cdot)$ and, based on them, we will compute an estimator for the variance-covariance matrix of the idiosyncratic error term. In the second stage, we will apply a standard FGLS technique to achieve asymptotic efficiency.

For the first stage, we will adapt the proposal in Pesaran (2006) and approximate the unobserved factors, $f_t$, by a suitable proxy that does not depend on an initial estimate of $\beta$ and $m_i(\cdot)$ [1]. More precisely, let $\overline{y}_{At}$ and $\overline{x}_{At}$ be the cross-sectional mean of $y_{it}$ and $x_{it}$, respectively, this suggest $f_t$ can be approximated by some linear function of $\lambda_t = (\overline{y}_{At}, \overline{x}_{At}, d_t)$, that is a $\ell \times 1$ vector, where $\ell = (1+p+n)$, plus a term $o_p(1)$. Hence, the following augmented regression is considered

$$y_{it} = \beta' x_{it} + m_i(z_t) + \delta_i' \lambda_t + \epsilon_{it} + o_p(1), \quad i = 1, \dots, N, \quad t = 1, \dots, T, \tag{2.4}$$

where $\delta_i$ is a $\ell \times 1$ vector of factor loadings. To clarify the structure of the model, we rewrite (2.4) into the following matrix form

$$Y_{i\cdot} = X_{i\cdot}\beta + m_i(Z) + \Lambda\delta_i + \epsilon_{i\cdot} + o_p(1), \tag{2.5}$$

where $Y_{i\cdot} \equiv (y_{i1}, \dots, y_{iT})'$ and $\epsilon_{i\cdot} \equiv (\epsilon_{i1}, \dots \epsilon_{iT})'$ are $T \times 1$ vectors, $X_{i\cdot} \equiv (x_{i1}, \dots, x_{iT})'$ is a $T \times p$ matrix, $\Lambda \equiv (\lambda_1, \dots, \lambda_T)'$ is a $T \times \ell$ matrix, and $m_i(Z) = (m_i(z_1), \dots, m_i(z_T))'$ is a $T \times 1$ vector.

By following a profile likelihood approach as proposed in Fan and Huang (2005), and assuming that $Z_z' K_{H_1}(z) Z_z$ is invertible, it can be shown that, for $i = 1, \dots, N$,

$$\widehat{m}_i(z, H_1) = \iota_1' (Z_z' K_{H_1}(z) Z_z)^{-1} Z_z' K_{H_1}(z) [Y_{i\cdot} - X_{i\cdot}\widehat{\beta} - \Lambda\widehat{\delta}_i], \tag{2.6}$$

where $\iota_1$ is a $(1+q) \times 1$ vector having 1 in the first entry and all other entries being 0, $Z_z$ is a $T \times (1+q)$ matrix whose $t$-th element is such as $Z_{z_t} = [1, (z_t - z)']$ for $z$ being a fixed point, and $K_{H_1}(z)$ is a $T \times T$ diagonal matrix such 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 $u$, it holds that $K_{H_1}(u) = |H_1|^{-1} \prod_{l=1}^{q} k(H_1^{-1} u_l)$ where $u = (u_1, \dots, u_q)'$ and $k(\cdot)$ is a univariate kernel function. Furthermore, if we write (2.6) in matrix notation then we can define a $T \times T$ smoothing matrix, $S$, as $\widehat{m}_i(Z, H_1) = S[Y_{i\cdot} - X_{i\cdot}\widehat{\beta} - \Lambda\widehat{\delta}_i]$. Note that $S$ depends only on the values of $z_t$ whose definition is apparent from (2.6).

In order to provide estimators for $\beta$ and $\delta_i$ we introduce some additional notation. Let $\widehat{Y} = (I_T - S)Y_{i\cdot}$, $\widehat{X}_{i\cdot} = (I_T - S)X_{i\cdot}$, and $\widehat{\Lambda} = (I_T - S)\Lambda$. Define the following matrices: $M_{\widehat{\Lambda}} = I_T - \widehat{\Lambda}(\widehat{\Lambda}'\widehat{\Lambda})^{-1}\widehat{\Lambda}'$ and $M_{\widehat{X}_i} = I_T - \widehat{X}_{i\cdot}(\widehat{X}_{i\cdot}'\widehat{X}_{i\cdot})^{-1}\widehat{X}_{i\cdot}'$, whereas $V_{\widehat{\Lambda}}$ is the $T \times (T - \ell)$ orthonormal eigenvector matrix of $M_{\widehat{\Lambda}}$ which corresponds to the eigenvalues of one. Furthermore, $M_{\widehat{\Lambda}}$

[1] It can be proved that $f_t$ can be well approximated by the cross-sectional averages of the observed variables, $y_{it}$ and $x_{it}$, assuming: (i) $rank(\Gamma^*) = r \leq (1+p)$ for sufficiently large $N$, where $\Gamma^* = E(\gamma_i, \Gamma_i) = (\gamma, \Gamma)$; (ii) $N^{-1}\sum_{i=1}^{N} v_{it} \xrightarrow{q.m.} 0$ and $N^{-1}\sum_{i=1}^{N} \epsilon_{it} \xrightarrow{q.m.} 0$ for each $t$ under rather weak conditions; (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 $z \in \mathcal{Z}$, where $\mathcal{Z}$ is the support of $z_t$.

and $M_{\widehat{X}_i}$ are idempotent matrices, which follows that $V_{\widehat{\Lambda}}V'_{\widehat{\Lambda}} = M_{\widehat{\Lambda}}$ and $V'_{\widehat{\Lambda}}V_{\widehat{\Lambda}} = I_{T-\ell}$, and similar notation and properties for $V_{\widehat{X}_i}$.

Then assuming that $\sum_{i=1}^{N}\widehat{X}'_{i\cdot}M_{\widehat{\Lambda}}\widehat{X}_{i\cdot}$ is invertible, and following again the procedure in Fan and Huang (2005) we obtain

$$\widehat{\beta} = \left(\sum_{i=1}^{N}\widehat{X}'_{i\cdot}M_{\widehat{\Lambda}}\widehat{X}_{i\cdot}\right)^{-1}\sum_{i=1}^{N}\widehat{X}'_{i\cdot}M_{\widehat{\Lambda}}\widehat{Y}_{i\cdot}, \tag{2.7}$$

$$\widehat{\delta}_i = \left(\widehat{\Lambda}'M_{\widehat{X}_i}\widehat{\Lambda}\right)^{-1}\widehat{\Lambda}^{\top}M_{\widehat{X}_i}\widehat{Y}_{i\cdot}. \tag{2.8}$$

Focusing now on the estimation of the mean of $m_i(\cdot)$, we propose the following nonparametric estimator

$$\widehat{\overline{m}}(z;H_1) = \iota'_1(Z'_zK_{H_1}(z)Z_z)^{-1}Z'_zK_{H_1}(z)\left[\overline{Y}_A - \overline{X}_A\widehat{\beta} - \Lambda\widehat{\overline{\delta}}\right], \tag{2.9}$$

where $\widehat{\overline{\delta}} = N^{-1}\sum_{i=1}^{N}\widehat{\delta}_i$, whereas $\overline{X}_A = (\overline{x}'_{A1},\ldots,\overline{x}'_{AT})'$ is a $T\times p$ matrix and $\overline{Y}_A = (\overline{y}_{A1},\ldots,\overline{y}_{AT})'$ is a $T$-dimensional vector.

In the next Section, we will show that (2.7)–(2.9) are consistent and asymptotically normal. However, since they completely ignore the information contained in the correlation structure of the $\epsilon_{it}$'s these estimators are inefficient. More precisely, we describe the correlation structure through the following assumption

**Assumption 2.1** *For $t = 1,\ldots,T$ and $i = 1,\ldots,N$, $E(\epsilon_{it}|z_t = z) = 0$. Furthermore, for $t = s$, $E\left(\epsilon_{\cdot t}\epsilon'_{\cdot t}|\, z_t\right) = \Omega_N\left(z_t\right)$, a $N\times N$ matrix, and for $t\neq s$, $E\left(\epsilon_{\cdot t}\epsilon'_{\cdot s}|\, z_t, z_s\right) = 0$. Let $\Omega(Z) = \mathrm{diag}_{t=1,\cdots,T}\left\{\Omega_N\left(z_t\right)\right\}$ and $\Omega_N\left(z_t\right) = \left\{\omega_{ij}\left(z_t\right)\right\}_{i,j=1,\ldots,N}$. The functions $\omega_{ij}(z)$ have uniformly bounded derivative of second order at $z$. Furthermore, the matrix $\Omega\left(Z\right)$, is nonsingular. Finally, $E(|\epsilon_{it}|^{\varsigma}|z_t = z) < \infty$, where $\varsigma > 4$.*

Note that for the sake of generality, in the previous assumption $\omega_{ij}(z)$ is considered as an unknown smooth function that needs to be estimated. Alternatively, it could be assumed that $\omega_{ij}(z)$ belongs to a family of parametric functions (see Soberon et al., 2022).

Now, in the second stage, we develop a procedure that enables us to obtain efficient estimators for $\beta$ and $\overline{m}(\cdot)$ by incorporating the information in Assumption 2.1. To obtain the FGLS estimator of $\beta$, note that

$$\widehat{Y}_{i\cdot} = \widehat{X}_{i\cdot}\beta + \widehat{\Lambda}\delta_i + \epsilon_{i\cdot} + o_p(1) + O_p(tr\{H_1^2\}). \tag{2.10}$$

Premultiplying both sides of (2.10) by $V'_{\widehat{\Lambda}}$ and stacking the resulting observations over

$NT$, we obtain

$$\widehat{Y}^* = \widehat{X}^*\beta + e^*, \tag{2.11}$$

where $\widehat{Y}^* = (I_N \otimes V_{\widehat{\Lambda}})'\widehat{Y}$ and $e^* = (I_N \otimes V_{\widehat{\Lambda}})'\left(\epsilon + o_p(1) + O_p(tr\{H_1^2\})\right)$ are $NT \times 1$ vectors, $\otimes$ is the Kronecker product, $\widehat{X}^* = (I_N \otimes V_{\widehat{\Lambda}})'\widehat{X}$ is a $NT \times p$ matrix, and $\epsilon$ is the $NT \times 1$ vector of the idiosyncratic error term. Now, noting that $\Omega_{e^*} \equiv Var(e^*|Z) = (I_N \otimes V'_{\widehat{\Lambda}})\Omega\,(Z)\,(I_N \otimes V_{\widehat{\Lambda}})$. Hence, premultiplying (2.11) by $\Omega_{e^*}^{-1/2}$ we get

$$\Omega_{e^*}^{-1/2}\widehat{Y}^* = \Omega_{e^*}^{-1/2}\widehat{X}^*\beta + \Omega_{e^*}^{-1/2}e^*, \tag{2.12}$$

and therefore the resulting Generalized Least Squares (GLS) estimator for $\beta$ is

$$\widehat{\beta}_{GLS} = \left[\widehat{X}'(I_N \otimes M_{\widehat{\Lambda}})\Omega^{-1}(Z)(I_N \otimes M_{\widehat{\Lambda}})\widehat{X}\right]^{-1}\widehat{X}'(I_N \otimes M_{\widehat{\Lambda}})\Omega^{-1}(Z)(I_N \otimes M_{\widehat{\Lambda}})\widehat{Y}. \tag{2.13}$$

However, this GLS estimator is infeasible since it depends on $\Omega\,(Z)$ that is generally unknown. To overcome it, we propose the following estimator for $\widehat{\Omega}\,(Z) = \text{diag}_{t=1,\cdots,T}\{\widehat{\Omega}_N(z_t)\}$

$$\widehat{\Omega}_N(z) = \frac{\sum_{t=1}^{T} K^*_{H_2}(z_t - z)\widehat{e}_{\cdot t}\widehat{e}_{\cdot t}{}'}{\sum_{t=1}^{T} K^*_{H_2}(z_t - z)}, \tag{2.14}$$

where $K^*(\cdot)$ is a nonnegative kernel function as the defined in (2.6), $H_2$ is a $q \times q$ symmetric and positive definite matrix, and $\widehat{e}_{\cdot t} = (\widehat{e}_{1t}, \ldots, \widehat{e}_{Nt})'$ is a $N \times 1$ vector of residuals defined as $\widehat{e}_{it} = y_{it} - x'_{it}\widehat{\beta} - \widehat{m}_i(z_t; H_1) - \widehat{\delta}'_i\lambda_t$. 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 (2.13) we get the Feasible Generalized Least Square Estimator,

$$\widehat{\beta}_{FGLS} = \left[\widehat{X}'(I_N \otimes M_{\widehat{\Lambda}})\widehat{\Omega}^{-1}(Z)(I_N \otimes M_{\widehat{\Lambda}})\widehat{X}\right]^{-1}\widehat{X}'(I_N \otimes M_{\widehat{\Lambda}})\widehat{\Omega}^{-1}(Z)(I_N \otimes M_{\widehat{\Lambda}})\widehat{Y}. \tag{2.15}$$

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

$$Y_{\cdot t} - X_{\cdot t}\beta - \Delta\lambda_t = \imath_N\overline{m}(z_t) + U_{\cdot t}, \tag{2.16}$$

where $Y_{\cdot t}$ and $U_{\cdot t}$ 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 $\Delta$ are $N \times p$ and $N \times \ell$ matrices, respectively. Following Lee and Robinson (2015) and Soberon et al. (2022), among others, and by imposing the identification condition $\varpi'\imath_N = 1$

to identify $\overline{m}(\cdot)$, we premultiply (2.16) by a given $N \times 1$ weight vector $\varpi$ obtaining

$$\varpi'(Y_{\cdot t} - X_{\cdot t}\beta - \Delta\lambda_t) = \overline{m}(z_t) + \varpi' U_{\cdot t}. \tag{2.17}$$

To estimate this regression model we choose $\varpi$ to minimize $Var(\varpi' U_{\cdot t}|z_t) = \varpi'\Phi_N(z_t)\varpi$, where $\Phi_N(z_t) = E(u_{\cdot t}u_{\cdot t}'|z_t) = \{\varphi_{ij}(z_t)\}_{i,j=1,\cdots,N}$. Hence, we deduce the optimal $\varpi = \varpi(z)$ obtaining $\varpi^*(z) = \arg\min_{\varpi} Var(\varpi' U_{\cdot t}|z_t) = (\imath_N'\Phi_N^{-1}(z)\imath_N)^{-1}\Phi_N^{-1}(z)\imath_N$. Replacing this latter result in (2.17) and following a similar procedure as in the previous section, the following GLS weighted local-least squares estimator for $\overline{m}(\cdot)$ is proposed,

$$\widehat{\overline{m}}_{GLS}(z; H_1, \varpi) = \iota_1'(Z_z' K_{H_1}(z) Z_z)^{-1} Z_z' K_{H_1}(z) \widetilde{Y}\varpi, \tag{2.18}$$

where $\widetilde{Y}$ is a $T \times N$ matrix whose $it$-th element is such as $\widetilde{y}_{it} = y_{it} - x_{it}'\beta - \lambda_t'\delta_i$. Finally, using the definition of $u_{it}$, and applying Assumption 2.1 we obtain $\varphi_{ij}(z) = \omega_{ij}(z) - [m_i(z) - \bar{m}(z)]^2$, for $i, j = 1, \cdots, N$.

Again this estimator is infeasible since $\beta$, $\delta_i$, and $\varpi$ are unknown, but following a similar procedure as in (2.14), with $\widehat{u}_{it} = y_{it} - x_{it}'\widehat{\beta} - \widehat{\delta}_i'\lambda_t - \widehat{\overline{m}}(z_t; H_1)$ 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

$$\widehat{\overline{m}}_{FGLS}(z; H_1, \widehat{\varpi}) = \iota_1'(Z_z' K_{H_1}(z) Z_z)^{-1} Z_z' K_{H_1}(z) \widehat{\widetilde{Y}}\widehat{\varpi}, \tag{2.19}$$

where $\widehat{\widetilde{Y}}$ is a $T \times N$ matrix whose $it$-th element is $\widehat{\widetilde{y}}_{it} = y_{it} - x_{it}'\widehat{\beta} - \lambda_t'\widehat{\delta}_i$ and

$$\widehat{\varpi} = \left(\imath_N'\widehat{\Phi}_N^{-1}(z)\imath_N\right)^{-1}\widehat{\Phi}_N^{-1}(z)\imath_N. \tag{2.20}$$

# 3 Asymptotic properties

In this section, we aim to derive the asymptotic properties of the proposed estimators. Firstly, we introduce some notations and assumptions that will be used throughout this article. Note that these assumptions are mostly inspired by those of Pesaran and Tosetti (2011), but are appropriately modified for the purposes of this paper. Later, we will present the main large sample properties of the proposed estimators.

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 $\mathcal{B}_{X^{(\varpi)}}(z) = E[X^{(\varpi)}|z_t = z]\rho_{z_t}(z)$ for $D \equiv (d_1, \ldots, d_T)'$ being a $T \times N$ matrix. We define $M_{\widetilde{G}} = I_T - \widetilde{G}(\widetilde{G}'\widetilde{G})^{-1}\widetilde{G}'$ as a $T \times T$ projection matrix, where $\widetilde{G} = (\widetilde{D}, \widetilde{F})$ is a $T \times (N + r)$ matrix.

**Assumption 3.1** *The* $(N + r + q) \times 1$ *vector of common components* $(d_t', f_t', z_t')'$ *is covariance stationary with absolute summable autocovariances, distributed independently of the individual-specific errors,* $\epsilon_{it}$ *and* $v_{it}$*, for all* $i$ *and* $t$*.*

**Assumption 3.2** *The individual-specific errors* $\epsilon_{it}$ *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* $i$*,* $\vartheta_{it}$ *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* $v_{it} = \sum_{\tau=0}^{\infty} S_{i\tau} S_{i\tau}' = \Sigma_{v_i} \leq C < \infty$*, where* $\Sigma_{v_i}$ *is a* $p \times p$ *positive definite matrix such that* $\sup_i \|\Sigma_{v_i}\|_2$ *and* $C$ *is some positive constant.*

**Assumption 3.3** *The unobserved factor loadings* $(\gamma_i, \Gamma_i)$ *are bounded, i.e.,* $\|\gamma_i\|_2 < C$ *and* $\|\Gamma_i\|_2 < C$*, for all* $i$*.*

**Assumption 3.4** *Let* $\Gamma^* = E(\gamma_i, \Gamma_i) = (\gamma, \Gamma)$*,* $Rank(\Gamma^*) = r \leq (p + 1)$*.*

**Assumption 3.5** *The* $p \times p$ *matrices* $(NT)^{-1} \sum_{i=1}^{N} \widetilde{X}_{i\cdot}' M_{\widehat{\Lambda}} \widetilde{X}_{i\cdot}$ *and* $(NT)^{-1} \sum_{i=1}^{N} \widetilde{X}_{i\cdot}' M_{\widetilde{G}} \widetilde{X}_{i\cdot}$ *exist and are non-singular. They also have finite second-order moments.*

**Assumption 3.6** *The probability density function of* $z_t$*,* $\rho_{z_t}(\cdot)$*, is continuous and bounded away from zero. Also,* $\rho_{z_t}(\cdot)$*,* $m_i(\cdot)$*, and* $\overline{m}(\cdot)$ *have bounded derivatives of order two in a neighborhood of* $z \in int(\mathcal{Z})$*.*

**Assumption 3.7** *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* $z$ *in the interior of* $\mathcal{Z}$*.*

**Assumption 3.8** $K(u) = \prod_{l=1}^{q} k(u_l)$ *is a product kernel, and the univariate kernel function* $k(\cdot)$ *is compactly supported and bounded such that* $\int k(u)du = 1$*,* $\int uu'k(u)du = \mu_2(K)I_q$*, and* $\int k^2(u)du = R(K)$*, where* $\mu_2(K) \neq 0$ *and* $R(K) \neq 0$ *are scalars and* $I_q$ *is a* $q \times q$ *identity matrix. All odd-order moments of* $k$ *vanish, that is* $\int u_1^{\imath_1}, \ldots, u_q^{\imath_q} k(u)du = 0$*, for all non-negative integers* $\imath_1, \ldots, \imath_q$ *such that their sum is odd.*

**Assumption 3.9** *Let* $c_{H_1} = tr\{H_1^2\} + (logT/T|H_1|)^{1/2}$*. The bandwidth matrix* $H_1$ *is symmetric and positive definite, where each element of* $H_1$ *tends to zero. As* $(N, T) \xrightarrow{j} \infty$*,* $\sqrt{N}c_{H_1}^2 \to 0$*,* $\sqrt{NT}c_{H_1}^2 \to 0$*,* $NT|H_1| \to \infty$*, and* $T|H_1| \to \infty$*.*

**Assumption 3.10** *For some* $\varsigma > 0$, $E[|\epsilon_{it}|^{(2+\varsigma)}]$ *exists and is bounded.*

Assumptions 3.1-3.4 are rather common assumptions concerning the individual-specific errors of $x_{it}$, common factors and rank condition (see Pesaran, 2006 or Pesaran and Tosetti, 2011 for further details). Assumption 3.5 is required to identify $\beta$. In addition, Assumptions 3.6-3.7 are standard smoothness and boundedness conditions on the density function and moment functionals. Assumptions 3.8-3.9 are kernel and bandwidth conditions quite common in the local linear literature and Assumption 3.10 is required for the Lyapunov condition. Note that the kernel function having a compact support in Assumption 3.8 is imposed for the sake of brevity and can be removed at the cost of lengthy proofs. Specifically, the Gaussian kernel is allowed.

In the following theorems, we focus on establishing the asymptotic normality of the above estimators under common weak dependence.

**Theorem 3.1** *Consider the panel data model (2.2) and (2.3), and suppose that Assumptions 2.1-3.9 hold. Then,* $\widehat{\beta}$ *and* $\widehat{\delta}_i$ *are consistent estimators for* $\beta$ *and* $\delta_i$*, 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* $\widetilde{X}$ *is a* $NT \times p$ *matrix,* $\Psi = \lim_{N,T\to\infty}(NT)^{-1}E\left[\widetilde{X}'\left(I_N \otimes M_{\widetilde{G}}\right)'\Omega\left(Z\right)\left(I_N \otimes M_{\widetilde{G}}\right)\widetilde{X}\right]$ *and* $Q = \lim_{N,T\to\infty}(NT)^{-1}\sum_{i=1}^{N} E\left(\widetilde{X}'_{i\cdot}M_{\widetilde{G}}\widetilde{X}_{i\cdot}\right)$. *are* $p \times p$ *matrices.*

In Theorem 3.1 is proved that the consistency problem related to the presence of unobserved common factors has been solved whether the rank condition of Assumption 3.4 holds and as $N$ and $T$ are sufficiently large. 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 taking into account the information of the correlation matrix $\Omega(Z)$. Note that similar conclusions can be obtained if the rank condition is violated (see Pesaran and Tosetti, 2011). The proof of this theorem is done in the Appendix.

Under the above assumptions, the following theorems give the asymptotic distribution of the nonparametric estimators, $\widehat{m}_i(\cdot; H_1)$ and $\widehat{\overline{m}}(\cdot; H_1)$, proposed for $m_i(\cdot)$ and $\overline{m}(\cdot)$, respectively.

**Theorem 3.2** *Consider the panel data model (2.2) and (2.3). Suppose that Assumptions 2.1-3.10 hold and that* $\sqrt{T|H_1|}tr\{H_1^2\} = O(1)$*. Given the* $\sqrt{NT}$*-consistency of* $\widehat{\beta}$ *and* $\sqrt{T}$*-consistency of* $\widehat{\delta}_i$*, as* $(N,T) \to \infty$,

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

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

The proof of Theorem 3.2 follows directly the proof of Theorem 2.1 in Ruppert and Wand (1994). In Theorem 3.2 it is shown that $\widehat{m}_i(\cdot; H_1)$ is consistent and asymptotically normal distributed with a rate of convergence of $\sqrt{T|H_1|}$, regardless of the rank condition assumption holds. Nevertheless, $\widehat{m}_i(z; H_1)$ completely ignores the information in the error term so it is subject to heteroscedasticity problems.

**Theorem 3.3** *Consider the panel data model (2.2) and (2.3). Suppose that Assumptions 2.1-3.10 hold and that $\sqrt{T|H_1|\nu_N^{-1}(z)}tr\{H_1^2\} = O(1)$. Given the $\sqrt{NT}$-consistency of $\widehat{\beta}$ and $\widehat{\widehat{\delta}}$, as $(N,T) \to \infty$,*

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

*where $\nu_N(z) = N^{-2}\imath_N' E(\epsilon_{\cdot t}\epsilon_{\cdot t}')\imath_N$ is a scalar term and $\mathcal{H}_{\overline{m}}(\cdot)$ is the Hessian matrix of $\overline{m}(\cdot)$.*

In Theorem 3.3 is shown that $\widehat{\widehat{\overline{m}}}(\cdot; H_1)$ is a consistent estimator of $\overline{m}(\cdot)$ as $N$ and $T$ are sufficiently large, but the variance of this nonparametric estimator exhibits a new element, i.e., $\nu_N(z)$, which reflects the strengthening of the spatial correlation and heteroscedasticity and depends directly on the particular specification of $\Omega(Z)$. Then, efficiency gains from pooling observations over the cross-section units as it is proposed in Pesaran (2006) are not achieved and efficient estimators could be obtained by taking into account 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, $\nu_N = O(N^{-1})$, 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. $\nu_N = O(1)$. Note that the proof of this theorem is done following a similar proof scheme as the corresponding for Theorem 2.2 in Soberon et al. (2022) in a different context, and it is therefore omitted.

In summary, it has been shown that, with the proposed estimation procedure, the bias problem related to the unobserved common factors has been solved. Furthermore, efficient estimators can be obtained by taking into account the spatial correlation and heteroscedasticity of the error term.

In order to obtain the asymptotic properties of the efficient estimators proposed in this paper, the following additional conditions are required

**Assumption 3.11** *The $p \times p$ matrices $(NT)^{-1}E\left(\widetilde{X}'(I_N \otimes M_{\widehat{\Lambda}})'\Omega^{-1}(Z)(I_N \otimes M_{\widehat{\Lambda}})\widetilde{X}\right)$ and $(NT)^{-1}E\left(\widetilde{X}'(I_N \otimes M_{\widetilde{G}})'\Omega^{-1}(Z)(I_N \otimes M_{\widetilde{G}})\widetilde{X}\right)$ exist and are non-singular. They also have finite second-order moments.*

**Assumption 3.12** *$K^*(u) = \prod_{l=1}^q k^*(u_l)$ 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 3.13** *The bandwidth matrix $H_2$ is symmetric and positive definite, where each element of $H$ tends to zero. As $(N,T) \xrightarrow{j} \infty$, $T|H_2| \to \infty$, $T|H_1|^2 = o(|H_2|)$, and $\frac{N^3}{T|H_1|} + \frac{Ntr\{H_2^2\}}{tr\{H_1^2\}} \to 0$.*

**Assumption 3.14** *The estimators $\widehat{\rho}(z)$ and $\widehat{\mathcal{H}}_{\overline{m}}(z)$, where $\mathcal{H}_m(\cdot) = \partial m(\cdot)/\partial z \partial z'$,*

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

*where $\widehat{\rho}(\cdot)$ and $\widehat{\mathcal{H}}_{\overline{m}}(\cdot)$ are consistent estimators of $\rho(\cdot)$ and $\mathcal{H}_{\overline{m}}(\cdot)$, respectively.*

Assumptions 2.1 and 3.12 together help to ensure that the bias of each element of the estimators of $\Omega(z)$ are $O_p(tr\{H_2^2\})$. Assumption 3.13 shows the relationship between $H_1$, $H_2$, $N$, and $T$. They are necessary to show the consistency of these efficient estimators. Assumption 3.14 is required to establish the asymptotic theory of the efficient estimators without involving too much technicality and simplify the proofs.

**Assumption 3.15** *Let $X^{\varpi} = (X_{1\cdot}, \ldots, X_{N\cdot})\varpi$ be a $T \times d$ matrix, the matrices $T^{-1}Z_z'K_{H_1}(z)X^{\varpi}$ and $T^{-1}Z_z'H_{H_1}(z)\Lambda$ exist.*

**Assumption 3.16** *For some $\varsigma > 0$, $E[|u_{it}|^{(2+\varsigma)}]$ exists and is bounded.*

**Assumption 3.17** *As $N \to \infty$, $\|\Phi_N^{-1}(z)\| + \frac{N\imath_N'\Phi_N^{-2}(z)\imath_N}{(\imath_N'\Phi_N^{-1}(z)\imath)^2} = O_p(1)$.*

Furthermore, to obtain an efficient estimator of the unknown function, Assumption 2.1 imposes the smoothness of the covariance function. Assumption 3.16 is necessary in order to check the Lyapunov condition for the CLT. Finally, Assumption 3.17 was discussed in detail in Robinson (2012) 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 3.4** *Consider the panel data model (2.2) and (2.3), and suppose that Assumptions 2.1-3.4, 3.6-3.9, and 3.11 hold. The GLS estimators for $\beta$ are consistent. If it is further assumed that $\sqrt{T}/N \to 0$ as $(N,T) \to \infty$,*

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

*where $Q_{\varpi} = \lim_{N,T\to\infty} \frac{1}{NT} E\left(\widetilde{X}'(I_N \otimes M_{\widetilde{G}})'\Omega^{-1}(Z)(I_N \otimes M_{\widetilde{G}})\widetilde{X}\right)$.*

**Theorem 3.5** *Consider the panel data model (2.2) and (2.3), and suppose that Assumptions 2.1-3.4, 3.6-3.9, 3.12-3.13, and 3.15-3.16 hold. If it is further assumed* $\sqrt{T|H_1|\nu_N^{(\varpi)-1}(z)}tr\{H_1^2\} = O(1)$*, as* $(N,T) \to \infty$*,*

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

*where* $\nu_N^{(\varpi)}(z) = (\imath_N'\Phi_N^{-1}(z)\imath_N)^{-1}$.

In Theorem 3.4 is shown that the asymptotic variance of $\widehat{\beta}_{GLS}$ has a sandwich structure. It can be observed also an efficiency gain in $\widehat{\beta}_{GLS}$ with respect to $\widehat{\beta}$. On its part, for $N$ and $T$ sufficiently large, in Theorem 3.5 is proved that the distribution of $\widehat{\widehat{m}}_{GLS}(z, H_1, \varpi)$ will be asymptotically normal if $N$ and $T$ are of the same order of magnitude (i.e., if $T/N \to \kappa$, where $\kappa$ 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 Robinson (2012) or Lee and Robinson (2015) for further details.

To finish the asymptotic analysis of the proposed estimators is necessary to show that both parametric and nonparametric FGLS estimators are asymptotically equivalent to their GLS counterparts.

**Theorem 3.6** *Consider the panel data model (2.2) and (2.3), and suppose that Assumptions 2.1-3.4, 3.6-3.9, and 3.13 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 3.7** *Consider the panel data model (2.2) and (2.3), and suppose that Assumptions 2.1-3.4, 3.6-3.9, and 3.12-3.17 hold. As* $(N,T) \to \infty$*,*

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

In Theorems 3.6-3.7 the asymptotically equivalence between the GLS and FGLS estimators is proved, so we can immediately arrive at the asymptotic results for $\widehat{\beta}_{FGLS}$ and $\widehat{\widehat{m}}_{FGLS}(z, H_1, \varpi)$ which has the same limiting distribution as $\widehat{\beta}_{GLS}$ and $\widehat{\widehat{m}}_{GLS}(z, H_1, \varpi)$, respectively. Finally note that a crucial condition to prove Theorem 3.7 is $\imath_N'\Phi_N^{-1}(z)\imath_N \geq N/\|\Phi_N(z)\|$, where $\|\Phi_N(z)\|$ denotes the square root of the largest eigenvalue of $\Phi_N(z)'\Phi_N(z)$. Therefore, we can conclude that if $\|\Phi_N(z)\|$ remains bounded, the variance rate of $\widehat{\widehat{m}}(z, H_1, \varpi)$ is $(NT|H_1|)^{-1}$.

# 4 Climate policy analysis: assessing the effects of EU ETS on $CO_2$ emissions

## 4.1 Data and variables

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 Purcharing 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 $z_t$, which is invariant across units, is the price of carbon, linked to the European Market of Allowances (EUA), the emission trading system that started being operational in 2005. Data on EUA are obtained from different sources Jiménez-Rodríguez (2019), 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. The annual carbon price used in our model is the average auction price in the primary market of all the transaction registered in a given year.

Figure 1 depicts the evolution of carbon price over time. The EU ETS was launched in 2005, marking a significant milestone as the first large-scale market for carbon emissions in the EU. In 2005, the price was about 18 €per tonne. Prior to its establishment, experiences in the United States had focused on regulating regional pollutants like NOx and SOx. The initial phase, a 3-year pilot program, aimed to create a functional market structure. During this period, the system concentrated on $CO_2$ emissions from power generators and energy-intensive industries, with most allowances allocated 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 determined a price for carbon emissions, albeit amidst considerable volatility, as noted by Ellerman and Joskow (2008). High market expectations were associated with the 2009 United Nations Climate Change Conference, commonly known as the Copenhagen Summit. However, the outcomes of the conference did not convey the policy outcomes that were expected. These unsatisfactory outcomes, combined with the global economic recession following the 2008 financial crisis, led to a sharp decline in carbon prices. The economic stagnation that followed the 2009 recession peak is reflected in the decline of carbon prices, indicating a period of uncertain climate policy commitments worldwide. However, 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 observed by

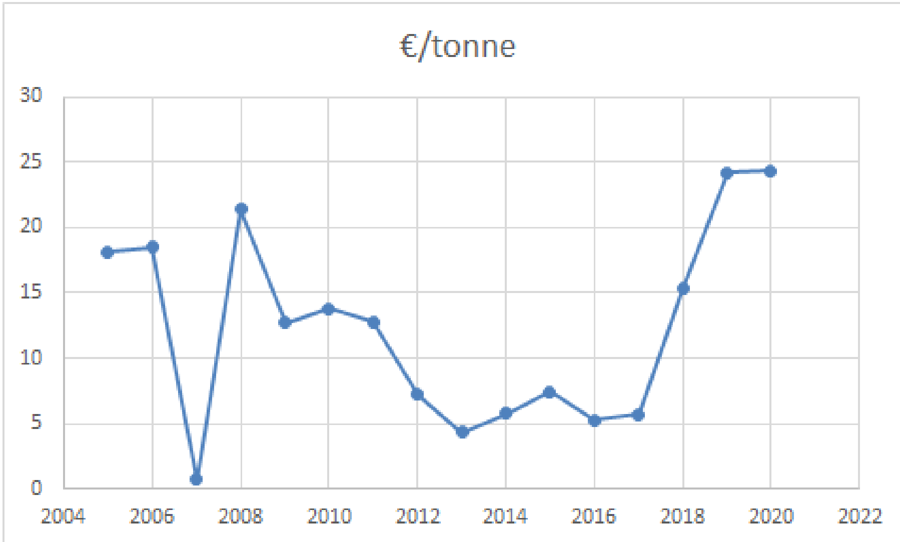


**Figure 1:** The evolution of carbon price over time.

Borghesi et al. (2016) and Ellerman et al. (2016). 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. It is also worth noting that descriptive statistics, available upon request, indicate that despite a certain degree of cross-country heterogeneity, European greenhouse gas emissions began decreasing significantly after 2007, coinciding with the recession. The recession and the subsequent uncertain stagnation phase showed a further declining trend in emissions. Lastly, a new significant decrease was observed over the last years in that dataset, i.e., 2017-2019, after a short period during which emissions rebounded upward (2014-2016). These descriptive figures also clearly highlight that common factors affecting emissions are a relevant feature of the data and they should be accounted for in the econometric model.

## 4.2 Estimation results

Building upon the discussion established in Section 2.1, we estimate (2.1) using the estimators proposed in Section 2. The results of the parametric component of the model are presented in Table 1. To evaluate the potential misspecification related to the carbon price, various specifications are considered (refer to columns $(i)$-$(v)$). 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 $z_t$ as an additional regressor. In column $(iii)$, the potential nonlinear effect of the carbon price is estimated within a fully parametric setting, employing a second-order polynomial function. All these parametric specifications are estimated using the CCEP estimator (Pesaran, 2006). Lastly, columns

($iv$)-($v$) present the results from the proposed semiparametric estimators: the first-stage and the second-stage estimator, respectively. The second-stage estimator accounts for spatial error dependence of unknown form and heteroscedasticity, while the first-stage does not.

According to the estimation results, the parametric specifications yield negative estimates for the coefficients associated with both $gdp_{it}$ and $gdp_{it}^2$, although these estimates are not statistically significant. Considering that our sample covers EU countries in very recent years, the finding of a negative elasticity concerning per capita GDP, which increases in magnitude with the rise of this variable, aligns with the original idea behind the EKC. This result is also consistent with respect to a substantial body of literature (Churchill et al., 2018). However, the lack of significance in these coefficients is unexpected and diverges from the existing literature. Note that $gdp_{it}$ is still not significant also when estimating a model that does not contain $gdp_{it}^2$.

As far as the effect of the technology variable is concerned, it is found that the estimated elasticity with respect to $R\&D$ expenditures is about -0.32. The literature, which is surveyed in Koçak and Ulucak (2019), is heterogeneous in terms of the adopted proxy and results. As for the proxy, while R&D expenditures is a common proxy for technology (Griliches, 1998) and is often employed (Fernandez et al., 2018), alternative proxies such as energy intensity (Baltagi et al., 2019) and process or product innovation (Costantini et al., 2013) are occasionally employed. While the conventional expectation is that technology would lead to a reduction in emissions, the results are mixed and sometimes show positive estimates. Notably, much of the existing literature has employed standard panel data specifications, neglecting strong cross-sectional dependence. For that reason, we have also estimated the models outlined in columns ($i$)-($iii$) without the multifactor component, specifically, by considering a one-way fixed effects model. Interestingly, the estimated elasticity concerning R&D expenditures decreases by approximately 40%, declining from -0.032 to around -0.02. Even more striking is the pronounced change in elasticity regarding GDP (detailed results available upon request). As Kapetanios et al. (2011) note within the common fully parametric framework, standard approaches that neglect common factors fail to identify $\beta$; instead, they yield an estimate of $\beta$ plus a term that is a function of the factor loadings $\gamma_i'$ and $\Gamma_i'$. Our results suggests that empirically, the bias arising from the omission of latent common factors is sizeable (see also Mazzanti and Musolesi, 2013). Finally, as far as the EU ETS policy variable is concerned, it is found that the associated coefficients of both the linear and the quadratic specification are very close to zero and are not statistically significant.

In summary, the parametric specifications in columns ($i$)-($iii$) yield unexpected results that should be reassessed by employing the proposed semiparametric estimators. This 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 bias not only with

**Table 1:** Fully parametric and semiparametric results.

| | Parametric | | | Semiparametric | |
|---|---|---|---|---|---|
| | (i) | (ii) | (iii) | (iv) | (v) |
| $gdp_{it}$ | $-0.606$ | $-0.607$ | $-0.607$ | $-3.560$ | $-2.630^{***}$ |
| | $(-0.465)$ | $(-0.465)$ | $(-0.465)$ | $(-1.956)$ | $(-9.298)$ |
| $gdp_{it}^2$ | $-0.113$ | $-0.113$ | $-0.113$ | $-0.542^{*}$ | $-0.412^{***}$ |
| | $(-0.638)$ | $(-0.638)$ | $(-0.637)$ | $(-2.227)$ | $(-10.372)$ |
| $r\&d_{it}$ | $-0.320^{***}$ | $-0.321^{***}$ | $-0.321^{***}$ | $-0.381^{***}$ | $-0.316^{***}$ |
| | $(-3.831)$ | $(-3.826)$ | $(-3.822)$ | $(-4.216)$ | $(-9.302)$ |
| $z_t$ | | 0.000 | $-0.000$ | | |
| | | (0.021) | $(-0.007)$ | | |
| $z_t^2$ | | | 0.000 | | |
| | | | (0.012) | | |
| $CD$ | 3.14 | 3.11 | 3.11 | 3.91 | 1.05 |
| | [0.002] | [0.002] | [0.002] | [0.000] | [0.293] |
| $CD_w$ | 2.41 | 2.41 | 2.41 | 2.12 | 0.79 |
| | [0.016] | [0.016] | [0.016] | [0.034] | [0.708] |
| $\widehat{a}$ | 0.802 | 0.802 | 0.802 | 0.728 | 0.528 |
| $\widehat{a}^*_{0.025}$ | 0.709 | 0.709 | 0.709 | 0.615 | 0.433 |
| $\widehat{a}^*_{0.975}$ | 0.896 | 0.895 | 0.895 | 0.841 | 0.624 |

Notes.
Columns (i)-(iii): CCEP (Pesaran, 2006).
Columns (iv) and (v): first-stage and second-stage semiparametric estimator.
**, **, *: Significance at 1%, 5%, 10%, respectively.
Standard errors within parentheses.
$CD$: CD test by Pesaran (2015, 2021).
$CD_w$: averaged weighted CD test by Juodis and Reese (2022).
p-values within square brackets.
$\widehat{a}$: bias-corrected version of $a$ given by equation (13) in Bailey et al. (2016).
$^*$95% level confidence bands.

respect to the estimated policy effect but also with respect to the estimated parameters of the standard ECK covariates. Furthermore, neglecting spatial error dependence may also seriously affects the results.

As far as the proposed semiparametric estimators are concerned (columns $(iv)$ and $(v)$), the results highlight the empirical importance of incorporating a nonparametric function for the EU ETS policy variable as well as addressing spatial error dependence. Firstly, the semiparametric first-stage estimator in column $(iv)$ exhibits notable changes as the parameters linked to $gdp_{it}$ and $gdp_{it}^2$ experience a substantial increase in absolute value, though their significance remains low. Additionally, it is observed a significant alteration in the estimated relationship between $co_{2it}$ and $z_t$, now being represented by a U-shaped function $\hat{m}(z_t)$ (refer to Figure 2).

Moving on to the semiparametric second-stage estimator in column $(v)$, the results appear more economically consistent. Both $gdp_{it}$ and $gdp_{it}^2$ become highly significant, indicating a statistically significant negative relation, which is consistent with the EKC framework. These findings also highlight a substantial underestimation of the parameters associated with $gdp_{it}$ and $gdp_{it}^2$ in the parametric model. Notably, a comparison between the two semiparametric estimators reveals significantly lower standard errors in the second-stage estimator. Crucially, the second-stage estimator presents a nonparametric function of the EU ETS policy effect that exhibits a negative shape for most of the domain, displaying in that region an estimated elasticity ranging from approximately $-0.06$ to $-0.08$. However, a threshold emerges for higher carbon price levels, where the estimated elasticity first becomes zero and then increases to 0.02. Overall, this finding aligns with expectations and suggests a credible result.

In our view, the reasons behind the observed positive effect that emerges for high levels of carbon prices are twofold. First, the EU ETS price was very erratic over 2005-2019, and in particular, the time periods associated with such high price levels are the initial 2005-2007 pilot phase and the latest years in the sample, 2018-2019. These two periods are very different; the 2005-2007 pilot phase was highly experimental as the EU ETS system was newly introduced. During this initial phase, when the European share of renewables was limited, markets predominantly relied on solid fossil fuels such as coal, which increased over 2005-2007. Second, the observed positive effect might be attributed to the composite effects we have previously discussed. The future inclusion of the most recent 2019-2023 price phase, where prices reached 100 €, could provide clearer evidence regarding this relationship. Also note that the elasticity of R&D expenditures remains remarkably stable across all specifications.

Finally, we also performed diagnostic checks on the residuals, specifically focusing on the issue of cross-sectional dependence. The $CD$ test developed by Pesaran (2021) is a widely adopted test, which is typically employed as a misspecification test in models that already account for cross-sectional dependence (Bailey et al., 2016; Ertur and Musolesi, 2017; Juodis and Reese, 2022). This test presents good small-sample properties and recent theoretical works have also provided additional insights that are useful from an empirical perspective. In particular, Pesaran (2015) demonstrated that the implicit null hypothesis of the $CD$ test is weak cross-sectional dependence in the most common cases. More precisely, let us define $\epsilon$ as a measure of the degree to which $T$ expands relative to $N$, as defined by $T = O(N^{\epsilon})$ for $0 < \epsilon \leq 1$ and $a$ being the exponent of cross-sectional dependence introduced in Bailey et al. (2016), which can take any value in the range $[0, 1]$. The values of $a$ in the range $[0, 1/2)$ correspond to different degrees of weak cross-sectional dependence, whereas values of $a$ in the range $(1/2, 1]$ relate to different degrees of strong cross-sectional dependence. Pesaran (2015) shows that the implicit null of the $CD$ test is given by $0 \leq a < (2 - \epsilon)/4$. Thus, for $\epsilon$ close to zero ($T$ almost fixed as $N \to \infty$), as it is roughly the case for the data used in this paper, such a null hypothesis is $0 \leq a < 1/2$, whereas in the case in which $\epsilon = 1$ ($N$ and $T \to \infty$ at the same rate), the implicit null of the $CD$ test is given by $0 \leq a < 1/4$. Moreover, Juodis and Reese (2022) demonstrate 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 time factors. To prevent erroneous rejection of the null hypothesis, they propose a modified test statistic, denoted as $CD_w$, which uses cross-section covariances instead of correlations. Additionally, these covariances are weighted using Rademacher distributed weights. To reduce the test's reliance on a specific set of random weights, they suggest averaging multiple $CD_w$ test statistics.

The $CD$ statistics (in Table 1) were 3.14, 3.11, 3.11 and 3.91 for specifications (i), (ii), (iii) and (iv), respectively. They are all highly statistically significant and strongly reject the null hypothesis, suggesting that the exponent of cross-sectional dependence, $a$, is in the range $[1/2, 1]$. Conversely, when considering the second-stage semiparametric estimator in column (v), the $CD$ statistic was equal to 1.05, so that the null was not rejected. We then employ the average $CD_w$ test, which basically confirms the results that were obtained with the standard $CD$ test. The only subtle distinction lies in the slight decrease of the test statistic when moving from the parametric specifications to the first-stage semiparametric model. Finally, to quantify the extent of CSD, we compute the bias-corrected version of $a$. As in Bailey et al. (2016), Holm's approach has been preferred over the Bonferroni procedure. These estimates, along with the 95% confidence bands, are also reported in Table 1. As for the fully parametric specifications, the exponent of cross-sectional dependence is estimated to be approximately 0.8 with 95% confidence bands lying above 0.5 and not including unity (0.71 and 0.9). It is worth to note that,similar to the findings in Ertur and Musolesi (2017), residuals obtained from a multifactor error regression model exhibit a lower

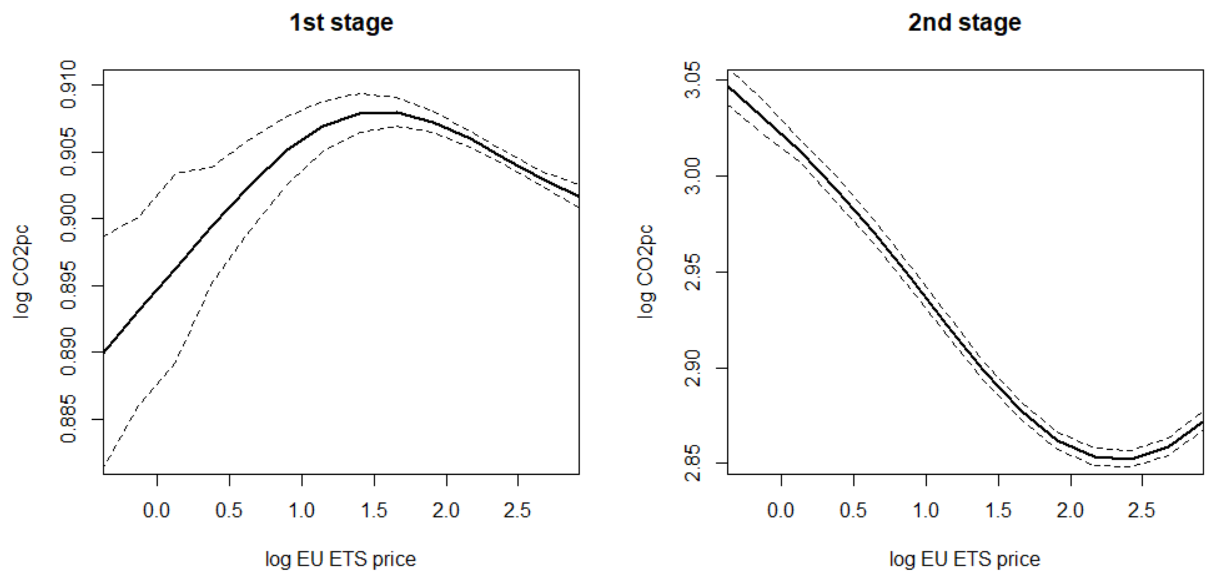


**Figure 2:** Estimated relation $CO_2$ - EU ETS carbon price.

degree of CSD compared to the variables incorporated into the model. For these variables, $\widehat{a}$ was approximately 1. When finally moving to the two semiparametric estimations, it is notable that $\widehat{a}$ decreases substantially. It is estimated at 0.73 for the first-stage estimator and further reduces to 0.53 for the second-stage estimator. Importantly, the 95% lower confidence band for the second-stage estimator now falls below 0.5. As for the interpretation of these results, it is worth emphasizing that according to Bailey et al. (2016) $a$ is identifiable only if $a > 1/2$ and that for values of $1/2 < a < 2/3$, the identification of $a$ is difficult, albeit theoretically possible. In summary, despite the above discussed caveats related to the $CD$ test and the identification of $a$, these results clearly indicate that employing the second-stage semiparametric estimator substantially reduces residual CSD.

# 5 Concluding remarks

In this paper, a new econometric model to assess the impact of the price of carbon, which is linked to the European market of allowances, on carbon dioxide emissions is proposed. Our proposal extends the EKC model in several directions. First, the price of carbon, which is introduced into the model as the policy variable, is incorporated nonparametrically. Secondly, we incorporate interactive fixed effects, and finally, we account for spatially correlated errors. To efficiently estimate the parameters of interest we have used a profile likelihood approach. The resulting Feasible Generalized Least Squares estimator of the parameters of interest has been shown to be $\sqrt{NT}$-consistent, asymptotically normal, and efficient. Furthermore, the policy effects curve is also efficiently estimated nonparametrically. Based on these outcomes, we have empirically evaluated the policy effects turning out that our results differ significantly from standard estimation techniques giving more consistent results. Importantly, the model proposed in this paper can be applied to many other empirical problems. Indeed, in a number of circumstances such as wage, cost, or production functions, parametric specifications for the main explanatory variables, which vary both over time and across units, are well established and build on economic theory. Conversely, a high degree of uncertainty concerning the functional form generally surrounds the impact of common covariates like energy shocks, global climatic conditions, or common policies on cross-sectional units such as individuals, households, firms, industries, or countries.

## Acknowledgements

This work is part of the I+D+i project Ref. TED2021-131763A-I00 financed by MCIN/AEI/10.13039/501100011033. 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 A. Monte Carlo experiments

In order to analyze the finite sample performance of the proposed estimators in this paper, 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 $\widehat{\overline{m}}(\cdot; H_1)$ (initial estimator), $\widehat{\overline{m}}_{GLS}(\cdot; H_2)$ (infeasible improved estimator), and $\widehat{\overline{m}}_{FGLS}(\cdot; H_2)$ (feasible improved estimator). Taking as benchmark Pesaran and Tosetti (2011), for all experiments we consider the following DGP based on Eq. (2.2)-(2.3):

$$
\begin{aligned}
y_{it} &= \alpha_i d_{1it} + x_{it}'\beta + m_i(z_t) + \gamma_{1i} f_{1t} + \gamma_{2i} f_{2t} + \epsilon_{it}, \\
x_{lit} &= a_{l1i} d_{1t} + a_{l2i} d_{2t} + g_{li}(z_t) + \gamma_{l1i} f_{1t} + \gamma_{l3i} f_{3t} + v_{lit},
\end{aligned}
$$

for $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$, $l = 1, 2$. In the above DGP, there are two individual-specific regressors ($x_{it} = (x_{1it}, x_{2it})'$), three observed common factors ($z_t, d_{1t}, d_{2t}$), and three unobserved common factors ($f_{1t}, f_{2t}, f_{3t}$). For $t = -49, \ldots, 0, 1, \ldots, T$ and $\rho = (0, 0.2, 0.5, 0.8)$, the observed common factors are generated as stationary AR(1) process:

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

whereas the unobserved common factors and individual-specific errors of $x_{it}$ are generated as stationary AR(1) process for $l = 1, 2, 3$:

$$
\begin{aligned}
& f_{lt} = \rho f_{l(t-1)} + u_{f_l,t}, \quad u_{f_l,t} \sim IIDN(0, 1-\rho^2), \quad f_{l,-50} = 0, \\
& v_{lit} = \rho_{v_l i} v_{li(t-1)} + v_{lit}, \quad v_{ilt} \sim IIDN(0, 1-\rho_{v_{l_i}}^2), \quad \rho_{v_{l_i}} \sim IIDU(0.05, 0.95), \quad v_{li,-50} = 0.
\end{aligned}
$$

Furthermore, the factor loadings of the observed common effects are generated as follows: $\alpha_i \sim IIDN(1, 1)$, $(a_{11i}, a_{21i}, a_{12i}, a_{22i}) \sim IIDN(0.5\imath_4, 0.5 I_4)$, where $\imath_4 = (1, 1, 1, 1)'$ and $I_4$ is the $4 \times 4$ identity matrix, and not change across replications. 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:

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

where $\gamma_{i1} \sim IIDN(1, 0.2)$, $\gamma_{i2} \sim IIDN(1, 0.2)$, so the rank condition is satisfied.

Furthermore, the heterogeneous unknown functions are generated such as $m_i(z_t) = (1/(1 + z_t^2)) + v_i$, $g_{1i}(z_t) = (1 + z_t^2) + v_{1i}$, and $g_{2i}(z_t) = sin(2z_t) + v_{2i}$, where $(v_i, v_{1i}, v_{2i}) \sim$

$IIDN(0, 0.04)$ are fixed across simulations. The idiosyncratic errors $\epsilon_{it}$ of $y_{it}$ are generated according to $\epsilon_{it} = b_i(z_t)\eta_t + \sqrt{0.5}\epsilon_{0it}$, where $b_i(z_t) = b_i z_t$, $b_i$ is generated as independent $N(0, 10)$ variates, kept fixed across replications, whereas $(\eta_t, \epsilon_{0it})$ are generated as independent Gaussian AR(1), with innovations having unit variance and using the $\rho$'s value as the autoregressive coefficient.

In each DGP, we consider the $(N, T)$ pairs $(N, T) = (100, 50), (100, 75), (150, 100)$, and $(150, 125)$. Also, we use 1000 replications and the Epanechnikov kernel functions. Because of the need for oversmoothing in the first stage, 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., $h_1 = 1.2h_2$, and three bandwidth values are used, i.e., $h_2 = 0.1, 0.3, 0.5$. Note that even though $(h_1 = 1.2h_2)$ does not imply oversmoothing asymptotically, in finite sample applications it effectively oversmooths.

For evaluation of the performance of our estimators, we use the bias and the root mean squared error (RMSE) for the slope parameters. In contrast, the squared root of the averaged squared error (RASE) is computed for the regression functions. In what follows we focus on the behavior of the estimators for $\beta_1$ since the results for $\beta_2$ are very similar and will not be reported. In particular, results for the experiments are summarized in Tables 2-3.

In the results of Table 2 it can be seen that the performance of the parametric estimators is very good. They display very small bias and their RMSEs decline steadily with increases in $N$ or $T$ and decreases of $\rho$, but they are a bit sensitive to the bandwidth selection. Furthermore, as was expected from the asymptotic properties, the estimator's efficiency is improved by taking the spatial dependence of the error term and heterogeneity into account and $\widehat{\beta}_F$ presents the best results.

Analyzing the results for the nonparametric estimators summarized in Table 3 it can be pointed out that the nonparametric procedure is robust to the bandwidth selection. Another important finding is that an increase in either $N$ or $T$ results in a decrease in the SM and SE of the RASE of the nonparametric function. Comparing the results from different strengths of serial dependence, $\rho$, we can see that the SM and SE of the RASE decrease as $\rho$ decreases. Finally, when $(N, T, \rho)$ are fixed, the GLS and FGLS nonparametric estimators that take into account the information contained in the error term (i.e., the spatial dependence and heterogeneity) give smaller RASEs than the initial estimator although the GLS estimator performs the best.

**Table 2:** Simulation results for the Bias (x100) and RMSE of the estimators for $\beta_1$.

| $\rho$ | $h_2$ | | N=100,T=50 | | N=100,T=75 | | N=150,T=100 | | N=150,T=125 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| 0 | 0.1 | $\widehat{\beta}$ | 0.041 | 0.189 | 0.759 | 0.180 | -0.370 | 0.132 | 0.127 | 0.125 |
| | | $\widehat{\beta}_G$ | -0.007 | 0.014 | -0.002 | 0.012 | 0.018 | 0.008 | 0.000 | 0.008 |
| | | $\widehat{\beta}_F$ | 0.044 | 0.182 | 0.747 | 0.177 | -0.360 | 0.130 | 0.128 | 0.123 |
| | 0.3 | $\widehat{\beta}$ | 0.515 | 0.206 | 1.125 | 0.182 | -0.167 | 0.138 | 0.457 | 0.128 |
| | | $\widehat{\beta}_G$ | 0.009 | 0.014 | -0.028 | 0.012 | -0.013 | 0.008 | 0.031 | 0.007 |
| | | $\widehat{\beta}_F$ | 0.452 | 0.192 | 0.011 | 0.175 | -0.164 | 0.132 | 0.430 | 0.124 |
| | 0.5 | $\widehat{\beta}$ | 0.041 | 0.189 | 0.759 | 0.180 | -0.370 | 0.132 | 0.127 | 0.125 |
| | | $\widehat{\beta}_G$ | -0.007 | 0.014 | -0.002 | 0.012 | 0.018 | 0.008 | 0.000 | 0.008 |
| | | $\widehat{\beta}_F$ | 0.044 | 0.182 | 0.747 | 0.177 | -0.360 | 0.130 | 0.127 | 0.123 |
| 0.2 | 0.1 | $\widehat{\beta}$ | 0.579 | 0.193 | 0.874 | 0.173 | 0.142 | 0.137 | 0.226 | 0.134 |
| | | $\widehat{\beta}_G$ | 0.008 | 0.016 | 0.046 | 0.012 | -0.060 | 0.009 | 0.005 | 0.008 |
| | | $\widehat{\beta}_F$ | 0.537 | 0.185 | 0.859 | 0.171 | 0.135 | 0.135 | 0.223 | 0.132 |
| | 0.3 | $\widehat{\beta}$ | 0.269 | 0.207 | 0.853 | 0.181 | -0.095 | 0.141 | 0.181 | 0.135 |
| | | $\widehat{\beta}_G$ | -0.036 | 0.016 | 0.018 | 0.012 | -0.055 | 0.008 | -0.015 | 0.008 |
| | | $\widehat{\beta}_F$ | 0.178 | 0.193 | 0.829 | 0.174 | -0.103 | 0.136 | 0.175 | 0.132 |
| | 0.5 | $\widehat{\beta}$ | 0.196 | 0.201 | 1.021 | 0.170 | -0.294 | 0.131 | 0.419 | 0.124 |
| | | $\widehat{\beta}_G$ | -0.035 | 0.015 | 0.029 | 0.012 | -0.056 | 0.008 | -0.014 | 0.008 |
| | | $\widehat{\beta}_F$ | 0.484 | 0.185 | 0.984 | 0.161 | -0.290 | 0.124 | 0.408 | 0.118 |
| 0.5 | 0.1 | $\widehat{\beta}$ | 0.772 | 0.259 | -0.231 | 0.220 | 0.605 | 0.166 | 0.426 | 0.165 |
| | | $\widehat{\beta}_G$ | -0.058 | 0.020 | 0.002 | 0.016 | -0.004 | 0.010 | 0.000 | 0.010 |
| | | $\widehat{\beta}_F$ | 0.712 | 0.248 | -0.214 | 0.215 | 0.589 | 0.163 | 0.415 | 0.163 |
| | 0.3 | $\widehat{\beta}$ | 0.596 | 0.282 | 0.039 | 0.229 | 1.002 | 0.172 | 0.536 | 0.163 |
| | | $\widehat{\beta}_G$ | -0.103 | 0.019 | -0.041 | 0.015 | -0.011 | 0.011 | 0.024 | 0.010 |
| | | $\widehat{\beta}_F$ | 0.474 | 0.260 | 0.050 | 0.219 | 0.916 | 0.164 | 0.526 | 0.157 |
| | 0.5 | $\widehat{\beta}$ | 0.320 | 0.274 | 0.006 | 0.222 | 0.917 | 0.164 | 0.711 | 0.153 |
| | | $\widehat{\beta}_G$ | -0.073 | 0.018 | -0.042 | 0.015 | -0.014 | 0.010 | 0.014 | 0.009 |
| | | $\widehat{\beta}_F$ | 0.206 | 0.247 | 0.015 | 0.209 | 0.803 | 0.153 | 0.684 | 0.145 |
| 0.8 | 0.1 | $\widehat{\beta}$ | -0.180 | 0.445 | -2.273 | 0.402 | 0.411 | 0.326 | -2.407 | 0.317 |
| | | $\widehat{\beta}_G$ | 0.073 | 0.027 | -0.029 | 0.024 | -0.026 | 0.016 | -0.079 | 0.015 |
| | | $\widehat{\beta}_F$ | -0.183 | 0.407 | -2.138 | 0.381 | 0.359 | 0.310 | -2.286 | 0.304 |
| | 0.3 | $\widehat{\beta}$ | 0.119 | 0.501 | -2.675 | 0.437 | 0.494 | 0.343 | -2.202 | 0.333 |
| | | $\widehat{\beta}_G$ | 0.051 | 0.029 | -0.064 | 0.024 | -0.028 | 0.016 | -0.068 | 0.015 |
| | | $\widehat{\beta}_F$ | 0.042 | 0.423 | -2.254 | 0.378 | 0.438 | 0.299 | -0.019 | 0.291 |
| | 0.5 | $\widehat{\beta}$ | 0.239 | 0.499 | -2.596 | 0.440 | 0.623 | 0.339 | -2.061 | 0.326 |
| | | $\widehat{\beta}_G$ | 0.038 | 0.028 | -0.060 | 0.023 | -0.020 | 0.016 | -0.069 | 0.015 |
| | | $\widehat{\beta}_F$ | 0.217 | 0.404 | -2.108 | 0.361 | 0.413 | 0.278 | -1.540 | 0.271 |

*Note*: $Bias = \frac{1}{1000}\sum_{\varphi=1}^{Q}(\widehat{\beta}_\varphi - \beta)$. $RMSE = \sqrt{\frac{1}{1000}\sum_{\varphi=1}^{Q}(\widehat{\beta}_\varphi - \beta)^2}$.

**Table 3:** Simulation results for RASE of the nonparametric estimators.

| $\rho$ | $h_2$ | | N=100,T=50 | | N=100,T=75 | | N=150,T=100 | | N=150,T=125 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | SM | SE | SM | SE | SM | SE | SM | SE |
| 0 | 0.1 | $\widehat{m}$ | 0.769 | 0.313 | 0.762 | 0.260 | 0.728 | 0.210 | 0.737 | 0.197 |
| | | $\widehat{\widehat{m}}_G$ | 0.378 | 0.036 | 0.390 | 0.029 | 0.399 | 0.026 | 0.406 | 0.024 |
| | | $\widehat{\widehat{m}}_F$ | 0.749 | 0.307 | 0.745 | 0.256 | 0.716 | 0.213 | 0.725 | 0.203 |
| | 0.3 | $\widehat{m}$ | 0.802 | 0.400 | 0.780 | 0.286 | 0.739 | 0.237 | 0.749 | 0.226 |
| | | $\widehat{\widehat{m}}_G$ | 0.399 | 0.039 | 0.413 | 0.033 | 0.420 | 0.029 | 0.427 | 0.027 |
| | | $\widehat{\widehat{m}}_F$ | 0.767 | 0.365 | 0.754 | 0.280 | 0.721 | 0.234 | 0.731 | 0.230 |
| | 0.5 | $\widehat{m}$ | 0.769 | 0.313 | 0.762 | 0.260 | 0.728 | 0.210 | 0.737 | 0.197 |
| | | $\widehat{\widehat{m}}_G$ | 0.378 | 0.036 | 0.390 | 0.029 | 0.399 | 0.026 | 0.406 | 0.024 |
| | | $\widehat{\widehat{m}}_F$ | 0.749 | 0.307 | 0.745 | 0.256 | 0.716 | 0.213 | 0.725 | 0.203 |
| 0.2 | 0.1 | $\widehat{m}$ | 0.778 | 0.297 | 0.735 | 0.246 | 0.726 | 0.226 | 0.721 | 0.212 |
| | | $\widehat{\widehat{m}}_G$ | 0.383 | 0.037 | 0.392 | 0.029 | 0.402 | 0.026 | 0.409 | 0.025 |
| | | $\widehat{\widehat{m}}_F$ | 0.757 | 0.287 | 0.719 | 0.248 | 0.714 | 0.223 | 0.709 | 0.217 |
| | 0.3 | $\widehat{m}$ | 0.798 | 0.347 | 0.756 | 0.273 | 0.727 | 0.244 | 0.724 | 0.225 |
| | | $\widehat{\widehat{m}}_G$ | 0.404 | 0.041 | 0.416 | 0.033 | 0.423 | 0.028 | 0.428 | 0.026 |
| | | $\widehat{\widehat{m}}_F$ | 0.765 | 0.316 | 0.727 | 0.269 | 0.709 | 0.241 | 0.708 | 0.227 |
| | 0.5 | $\widehat{m}$ | 0.778 | 0.297 | 0.735 | 0.246 | 0.726 | 0.226 | 0.721 | 0.212 |
| | | $\widehat{\widehat{m}}_G$ | 0.383 | 0.037 | 0.392 | 0.029 | 0.402 | 0.026 | 0.409 | 0.025 |
| | | $\widehat{\widehat{m}}_F$ | 0.757 | 0.287 | 0.719 | 0.248 | 0.713 | 0.223 | 0.709 | 0.217 |
| 0.5 | 0.1 | $\widehat{m}$ | 0.832 | 0.453 | 0.803 | 0.342 | 0.746 | 0.298 | 0.753 | 0.298 |
| | | $\widehat{\widehat{m}}_G$ | 0.395 | 0.041 | 0.406 | 0.035 | 0.412 | 0.030 | 0.418 | 0.026 |
| | | $\widehat{\widehat{m}}_F$ | 0.791 | 0.423 | 0.771 | 0.328 | 0.719 | 0.298 | 0.728 | 0.290 |
| | 0.3 | $\widehat{m}$ | 0.892 | 0.576 | 0.840 | 0.419 | 0.766 | 0.344 | 0.767 | 0.324 |
| | | $\widehat{\widehat{m}}_G$ | 0.414 | 0.043 | 0.426 | 0.037 | 0.433 | 0.031 | 0.436 | 0.028 |
| | | $\widehat{\widehat{m}}_F$ | 0.819 | 0.490 | 0.791 | 0.383 | 0.728 | 0.333 | 0.736 | 0.307 |
| | 0.5 | $\widehat{m}$ | 0.889 | 0.600 | 0.835 | 0.441 | 0.735 | 0.321 | 0.760 | 0.317 |
| | | $\widehat{\widehat{m}}_G$ | 0.412 | 0.045 | 0.425 | 0.039 | 0.432 | 0.033 | 0.436 | 0.030 |
| | | $\widehat{\widehat{m}}_F$ | 0.811 | 0.496 | 0.784 | 0.391 | 0.717 | 0.308 | 0.730 | 0.295 |
| 0.8 | 0.1 | $\widehat{m}$ | 1.496 | 1.529 | 1.497 | 1.397 | 1.362 | 1.163 | 1.304 | 0.962 |
| | | $\widehat{\widehat{m}}_G$ | 0.418 | 0.061 | 0.425 | 0.051 | 0.432 | 0.042 | 0.434 | 0.038 |
| | | $\widehat{\widehat{m}}_F$ | 1.360 | 1.333 | 1.352 | 1.224 | 1.218 | 1.024 | 1.159 | 0.830 |
| | 0.3 | $\widehat{m}$ | 1.762 | 1.999 | 1.710 | 1.728 | 1.582 | 1.616 | 1.480 | 1.219 |
| | | $\widehat{\widehat{m}}_G$ | 0.430 | 0.063 | 0.436 | 0.052 | 0.445 | 0.044 | 0.446 | 0.039 |
| | | $\widehat{\widehat{m}}_F$ | 1.570 | 1.709 | 1.515 | 1.492 | 1.389 | 1.329 | 1.311 | 1.116 |
| | 0.5 | $\widehat{m}$ | 1.764 | 2.050 | 1.718 | 1.814 | 1.579 | 1.694 | 1.478 | 1.268 |
| | | $\widehat{\widehat{m}}_G$ | 0.426 | 0.064 | 0.434 | 0.052 | 0.442 | 0.045 | 0.445 | 0.040 |
| | | $\widehat{\widehat{m}}_F$ | 1.566 | 1.721 | 1.531 | 1.616 | 1.394 | 1.368 | 1.320 | 1.169 |

*Note*: SM and SE are the sample mean and standard error, respectively, of the RASE of the estimators for the nonparametric function based on 1000 replications. $RASE(\widehat{m}(z)) = \sqrt{\frac{1}{T}(\widehat{\widehat{m}}(z_t) - \overline{m}(z_t))^2}$.

# Appendix B. Mathematical proofs and Lemmas

Before proceeding to the analysis of the main asymptotic properties of the proposed estimators, we first prove several lemmas that are used later in the proofs of the theorems. Remember that in the paper 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 $\widetilde{X}^{(\widehat{\varpi})} = X^{(\widehat{\varpi})} - \mathcal{B}_{X^{(\widehat{\varpi})}}(z)$, $\widetilde{X}^{(\varpi)} = X^{(\varpi)} - \mathcal{B}_{X^{(\varpi)}}(z)$, $\widetilde{X}^{(\widehat{\omega})} = X^{(\widehat{\omega})} - \mathcal{B}_{X^{(\widehat{\omega})}}(z)$, $\widetilde{X}^{(\omega)} = X^{(\varpi)} - \mathcal{B}_{X^{(\omega)}}(z)$, 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, $c_{H_1} = tr\{H_1^2\} + \{logT/T|H_1|\}^{1/2}$.

**Lemma 5.1** *Let $\overline{\varepsilon}_{At}$ be a composed error term defined as $\overline{\varepsilon}_{At} = (\overline{\epsilon}_{At} + \overline{v}'_{At}\beta, \overline{v}'_{At})'$. Under Assumptions 2.1 and 3.2, for each $t$, we have*

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

*b)* $Var(\overline{\varepsilon}_{At}) = O\left(\frac{1}{N}\right)$, *under weak dependence and* $Var(\overline{\varepsilon}_{At}) = O(1)$ *under strong dependence.*

**Proof of Lemma 5.1**: The proof of this lemma is straightforward from the proof of Lemma A1 in Pesaran and Tosetti (2011). This lemma guarantees that for any value of $z$, $\overline{\epsilon}_{At} \xrightarrow{q.m.} 0$ as $N \to \infty$ and the degree of spatial dependence of $\epsilon_{i\cdot}$ will be bounded by $\nu_N(z) = N^{-2}\imath'_N\Omega_N(z)\imath_N$, where $\imath_N$ is a $N \times 1$ vector of ones. Therefore, the results of this paper are valid for both types of spatial dependence.

∎

**Lemma 5.2** *Under Assumptions 3.1, 3.6-3.9, as $T \to \infty$ we have*

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

**Proof of Lemma 5.2**: This lemma can be proved in a similar way as in Theorem 2 in Hansen (2008) and it has been omitted for the sake of brevity.

∎

**Lemma 5.3** *Under Assumptions 2.1-3.4,*

*a)* $\frac{\overline{\varepsilon}'_{A\cdot}\overline{\varepsilon}_{A\cdot}}{T} = O(N^{-1})$.

*b)* $\frac{\widetilde{F}'\overline{\varepsilon}_{A\cdot}}{T} = O\left((NT)^{-1/2}\right)$ *and* $\frac{\widetilde{D}'\overline{\varepsilon}_{A\cdot}}{T} = O\left((NT)^{-1/2}\right)$.

c) $\frac{V_{i.}'\widetilde{D}}{T} = O\left(T^{-1/2}\right)$ and $\frac{V_{i.}'\widetilde{F}}{T} = O\left(T^{-1/2}\right)$.

d) $\frac{V_{i.}'\bar{\varepsilon}_{A.}}{T} = O\left(N^{-1}\right) + O\left((NT)^{-1/2}\right)$ and $\frac{\epsilon_{i.}'\bar{\varepsilon}_{A.}}{T} = O\left(N^{-1}\right) + O\left((NT)^{-1/2}\right)$.

**Proof of Lemma 5.3**: This lemma can be proved in a similar way as in Lemma 2 in Pesaran (2006) and it has been omitted for the sake of brevity.

∎

**Lemma 5.4** *Let $c_{H_1} = tr\{H_1^2\} + \{logT/T|H_1|\}^{1/2}$. Under Assumptions 2.1, 3.2, and 3.6-3.9, as $T \to \infty$, we have*

a) $\frac{\widehat{X}_{i.}'M_{\widehat{\Lambda}}\widehat{X}_{i.}}{T} = \frac{\widetilde{X}_{i.}'M_{\widetilde{G}}\widetilde{X}_{i.}}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p(c_{H_1})$ *uniformly over* $i$,

b) $\frac{\widehat{X}_{i.}'M_{\widehat{\Lambda}}(I_T - S)m_i(Z)}{T} = O_p(c_{H_1}^2)$ *uniformly over* $i$,

c) $\frac{\widehat{X}_{i.}'M_{\widehat{\Lambda}}\widehat{e}_{i.}}{T} = \frac{\widetilde{X}_{i.}'M_{\widetilde{G}}\epsilon_{i.}}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p(c_{H_1})$ *uniformly over* $i$,

d) $\frac{\widehat{X}_{i.}'M_{\widehat{\Lambda}}\widehat{F}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p(c_{H_1})$ *uniformly over* $i$,

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

**Proof of Lemma 5.4**: This lemma can be proved following similar reasoning as the proof of (A.12)–(A.14) in Pesaran and Tosetti (2011) and Lemma 3 in Cai et al. (2019) and it has been omitted for brevity.

∎

**Lemma 5.5** *Under Assumptions 3.1, 3.6, and 3.8-3.9, as $T \to \infty$, we have*

a) $(NT)^{-1}\widehat{X}'(I_N \otimes M_{\widehat{\Lambda}})\Omega^{-1}(Z)(I_N \otimes M_{\widehat{\Lambda}})\widehat{X} = (NT)^{-1}\widetilde{X}'(I_N \otimes M_{\widetilde{\Lambda}})\Omega^{-1}(Z)(I_N \otimes M_{\widetilde{\Lambda}})\widetilde{X} + o_p(1)$,

b) $(NT)^{-1}\widehat{X}'(I_N \otimes M_{\widehat{\Lambda}})\Omega^{-1}(Z)(I_N \otimes M_{\widehat{\Lambda}}) = (NT)^{-1}\widetilde{X}'(I_N \otimes M_{\widetilde{\Lambda}})\Omega^{-1}(Z)(I_N \otimes M_{\widetilde{\Lambda}}) + o_p(1)$.

**Proof of Lemma 5.5**: This lemma can be proved in a similar way as in Lemma 5.3 and it has been omitted for brevity.

∎

**Lemma 5.6** *Under Assumptions 2.1-3.9 and 3.12-3.13 and at $z$ such that $\rho_z(z) > 0$, as $T \to \infty$,*

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

*where* $R_{TH} = O_p\left(tr\{H_2^2\} + (T|H_1|)^{-1}\right)$.

**Proof of Lemma 5.6**: 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

$$\widehat{\omega}_{ij}(z) - \omega_{ij}(z) = \frac{\sum_{t=1}^T K^*_{H_2}(z_t - z)[\widehat{e}_{it}\widehat{e}_{jt} - \omega_{ij}(z)]}{\sum_{t=1}^T K^*_{H_2}(z_t - z)} = R^{(1)}_{ij} + R^{(2)}_{ij}, \tag{I.1}$$

where

$$R^{(1)}_{ij} = \frac{\sum_{t=1}^T K^*_{H_2}(z_t - z)[\epsilon_{it}\epsilon_{jt} - \omega_{ij}(z)]}{\sum_{t=1}^T K^*_{H_2}(z_t - z)}, \qquad R^{(2)}_{ij} = \frac{\sum_{t=1}^T K^*_{H_2}(z_t - z)[\widehat{e}_{it}\widehat{e}_{jt} - \epsilon_{it}\epsilon_{jt}]}{\sum_{t=1}^T K^*_{H_2}(z_t - z)}.$$

Hence, $R^{(1)}_{ij}$ 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

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

If we consider the bound of $R^{(2)}_{ij}$, we denote $\widetilde{g}_{1it} = X'_{it}(\widehat{\beta} - \beta)$, $\widetilde{g}_{2it} = (\delta_i - \widehat{\delta}_i)'\lambda_t$, 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 $\lambda_t$. Replacing this decomposition in $R^{(2)}_{ij}$ we are going to prove

$$\begin{aligned} R^{(2)}_{ij} &= T^{-1}\sum_{t=1}^T K^*_{H_2}(z_t - z)(I\!I_1 + I\!I_2 + o_p(1))/\widehat{\rho}(z) \\ &= O_p\left(tr\{H_1^4\} + (T|H_1|)^{-1} + (T^2|H_1||H_2|)^{-1/2}\right), \end{aligned} \tag{I.3}$$

where $\widehat{\rho}(z)$ is a nonparametric kernel estimator of $\rho_{z_t}(z)$ such as $\widehat{\rho}(z) = T^{-1}\sum_{t=1}^T K^*_{H_2}(z_t - z)$, $I\!I_1 = \epsilon_{it}\widetilde{g}_{1jt} + \epsilon_{it}\widetilde{g}_{2jt} + \widetilde{g}_{1it}\epsilon_{jt} + \widetilde{g}_{1it}\widetilde{g}_{1jt} + \widetilde{g}_{1it}\widetilde{g}_{2jt} + \widetilde{g}_{2it}\epsilon_{jt} + \widetilde{g}_{2it}\widetilde{g}_{2jt} + \widetilde{g}_{2it}\widetilde{g}_{1jt}$, and $I\!I_2 = \epsilon_{it}\widetilde{\xi}_{jt} + \widetilde{g}_{1it}\widetilde{\xi}_{jt} + \widetilde{g}_{2it}\widetilde{\xi}_{jt} + \widetilde{\xi}_{it}\epsilon_{jt} + \widetilde{\xi}_{it}\widetilde{g}_{1jt} + \widetilde{\xi}_{it}\widetilde{g}_{2jt} + \widetilde{\xi}_{it}\widetilde{\xi}_{jt}$. Under the assumptions stated in this paper, it can be proved that $\widehat{\rho}(z)$ is consistent following Lee and Robinson (2015), so $\frac{1}{\widehat{\rho}(z)} = \frac{1}{\rho_{z_t}(z) + o_p(1)} = O_p(1)$. Therefore, in order to prove (I.3) we only need to analyze $T^{-1}\sum_{t=1}^T K^*_{H_2}(z_t - z)(I\!I_1 + I\!I_2)$.

Following a similar proof scheme as in Lemma 5.2 and given the $\sqrt{NT}$-consistency of $\widehat{\beta}$ and the $\sqrt{T}$-consistency of $\widehat{\delta}_i$, it is easy to show

$$T^{-1}\sum_{t=1}^{T} K^*_{H_2}(z_t - z) I\!I_1 = o_p\left((NT)^{-1/2}\right) + o_p\left(T^{-1/2}\right). \tag{I.4}$$

Considering the bound of the second term of $R^{(2)}_{ij}$, there are two leading terms that 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 Hansen (2008) and Assumption 3.13,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} K^*_{H_2}(z_t - z)\widetilde{\xi}^2_{it} \\
&\le \sup_{|z|\le c_T}\left|T^{-1}\sum_{t=1}^{T} K^*_{H_2}(z_t - z) - \rho_{z_t}(z)\right| \sup_{|z|\le c_T} |\widehat{m}_i(z;H_1) - m_i(z)|^2 + |\rho_{z_t}(z)| \sup_{|z|\le c_T} |\widehat{m}_i(z;H_1) - m_i(z)|^2 \\
&= O_p\left(\sqrt{\frac{lnT}{T|H_2|}} + tr\{H_2^2\}\right) O_p\left(\delta_T^{-2}\left[\sqrt{\frac{lnT}{T|H_1|}} + tr\{H_1^2\}\right]^2\right) + O_p\left(\delta_T^{-2}\left[\sqrt{\frac{lnT}{T|H_1|}} + tr\{H_1^2\}\right]^2\right) \\
&= O_p\left(\frac{1}{T|H_1|} + tr\{H_1^4\}\right),
\end{aligned} \tag{I.5}$$

where $\delta_T = \inf_{|z|\le c_T}\rho_{z_t}(z) > 0$ and $c_T = ((lnT)^{1/q}T^{1/2\varsigma})$, for some $\varsigma > 0$.

Similarly, for the second leading term we have $E[T^{-1}\sum_t K_H(z_t - z)\epsilon_{it}\widetilde{\xi}_{it}] = 0$ and $Var[T^{-1}\sum_t K_H(z_t - z)\epsilon_{it}\widetilde{\xi}_{it}] = o_p\left(\frac{\delta_T^{-4}}{T|H_2|}\left(\sqrt{\frac{lnT}{T|H_1|}} + tr\{H_1^2\}\right)\right)$. Hence, the proof is done by replacing these results and (I.4)-(I.5) in (I.3). Replacing (I.2)-(I.3) in (I.1),

$$\max_{1\le i,j\le N} |\widehat{\omega}_{ij}(z) - \omega_{ij}(z)| = O_p(R_{TH}). \tag{I.6}$$

Finally, using (I.6) it is straightforward to show

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

as $N/T \to \kappa$, where $\kappa$ is a positive constant. Hence, the proof of the lemma is done.

∎

**Lemma 5.7** *Under Assumptions 3.6, 3.8, and 3.9 at z such that $\rho_z(z) > 0$, as $T \to \infty$,*

$$T^{-1} Z_z' K_{H_1}(z) Z_z = \begin{pmatrix} \rho_{z_t}(z) & \mu_2^q(K) H_1^2 D_\rho(z) \\ \mu_2^q(K) H_1^2 D_\rho(z) & H_1^2 \mu_2^q(K) \rho_{z_t}(z) \end{pmatrix} \{1 + O_p(c_{H_1})\}.$$

**Proof of Lemma 5.7**: The proof of this Lemma follows directly the proof of Theorem 2.1 in Ruppert and Wand (1994) and it has been omitted for brevity.

∎

**Lemma 5.8** *Under Assumptions 2.1-3.9 and 3.12-3.13 and at z such that $\rho_z(z) > 0$, as $T \to \infty$,*

$$\|\widehat{\varpi} - \varpi\| = O_p(N R_{TH}),$$

*where* $R_{TH} = O_p\left(tr\{H_2^2\} + (T|H_1|)^{-1}\right)$.

**Proof of Lemma 5.8**: In order to prove this lemma, $\|\widehat{\varpi} - \varpi\|$ can be rewritten as

$$\begin{aligned} \|\widehat{\varpi} - \varpi\| &\leq \left\| (\imath_N' \widehat{\Phi}_N(z)^{-1} \imath_N)^{-1} \widehat{\Phi}_N(z)^{-1} \imath_N - (\imath_N' \Phi_N(z)^{-1} \imath_N)^{-1} \Phi_N(z)^{-1} \imath_N \right\| \\ &= \left\| \imath_N' \widehat{\Phi}_N(z)^{-1} \right\| \left| (\imath_N' \widehat{\Phi}_N(z)^{-1} \imath_N)^{-1} - (\imath_N' \Phi_N(z)^{-1} \imath_N)^{-1} \right| \\ &+ \left\| (\imath_N' \Phi_N(z)^{-1} \imath_N)^{-1} \right\| \left\| \imath_N' \left( \widehat{\Phi}_N(z)^{-1} - \Phi_N(z)^{-1} \right) \right\|. \end{aligned} \tag{I.7}$$

Analyzing each of the above terms separately it is straightforward to show that, using the properties of $\Phi_N(z)$, we get $\|\imath_N' \widehat{\Phi}_N^{-1}(z)\| = O_p(\sqrt{N} \|\Phi_N^{-1}(z)\|) = O_p(\sqrt{N})$. Furthermore, using Assumption 3.17 we can prove

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

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

$$\left\|\imath_N'\widehat{\Phi}_N^{-1}(z)\right\|\left|\left(\imath_N'\widehat{\Phi}_N^{-1}(z)\imath_N\right)^{-1}-\left(\imath_N'\Phi_N^{-1}(z)\imath_N\right)^{-1}\right|=O_p\left(\frac{\|\widehat{\Phi}_N(z)-\Phi_N(z)\|}{\sqrt{N}}\right). \quad \text{(I.8)}$$

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

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

and using the above results it can be shown that the second term of (I.7) is bounded by

$$\begin{aligned}&\left\|(\imath_N'\Phi_N(z)^{-1}\imath_N)^{-1}\right\|\left\|\imath_N'\left(\widehat{\Phi}_N(z)^{-1}-\Phi_N(z)^{-1}\right)\right\| \\ &=\left\|\left(\imath_N'\Phi_N^{-1}(z)\imath_N\right)^{-1}\right\|O_p\left((\imath_N'\Phi_N^{-2}(z)\imath_N)^{1/2}\|\widehat{\Phi}_N(z)-\Phi_N(z)\|\right) \\ &=O_p\left((\imath_N'\Phi_N^{-1}(z)\imath_N)^{-1/2}\|\widehat{\Phi}_N(z)-\Phi_N(z)\|\right) \\ &=O_p((\imath_N'\Phi_N^{-1}(z)\imath_N)^{-1/2}NR_{TH}), \qquad \text{(I.9)}\end{aligned}$$

given that $(\imath_N'\Phi_N^{-1}(z)\imath_N)^{1/2}=O(1)$ and following a similar proof scheme as in Lemma 5.6 it is straightforward to show that $\|\widehat{\Phi}_N(z)-\Phi_N(z)\|=O_p(NR_{TH})$. Hence, plugging (I.8)-(I.9) in (I.7) the proof of the lemma is done.

∎

**Proof of Theorem 3.1**: Plugging (2.5) into (2.7) and rearranging terms we get

$$\widehat{\beta}-\beta=\left(\sum_{i=1}^N\widehat{X}_{i\cdot}'M_{\widehat{\Lambda}}\widehat{X}_{i\cdot}\right)^{-1}\sum_{i=1}^N\widehat{X}_{i\cdot}'M_{\widehat{\Lambda}}(I_T-S)[F\gamma_i+m_i(Z)+\epsilon_{i\cdot}+O_p(tr\{H^2\})+o_p(1)] \quad \text{(I.10)}$$

given that $M_{\widehat{\Lambda}}(I_T-S)D=0$ since $D\in\Lambda$. Note that in (I.10) 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 5.4 in (I.10), assuming that the rank condition holds and by the uniform boundedness assumption on $\gamma_i$, the expression to study is such as

$$\begin{aligned}\sqrt{NT}(\widehat{\beta}-\beta) &= \left(\frac{1}{NT}\sum_{i=1}^N\widetilde{X}_{i\cdot}'M_{\widetilde{G}}\widetilde{X}_{i\cdot}\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^N\widetilde{X}_{i\cdot}'M_{\widetilde{G}}\epsilon_{i\cdot}+O_p\left(\frac{\sqrt{T}}{N}\right)+O_p\left(\frac{1}{\sqrt{N}}\right) \\ &+ O_p\left(\sqrt{T}c_{H_1}^2\right). \qquad \text{(I.11)}\end{aligned}$$

Under Assumption 3.5 we can prove $\left((NT)^{-1}\sum_{i=1}^{N}\widetilde{X}_{i\cdot}'M_{\widetilde{G}}\widetilde{X}_{i\cdot}\right)^{-1} \to_p Q^{-1}$, where we define $Q=\lim_{N,T\to\infty}(NT)^{-1}\sum_{i=1}^{N}E\left(\widetilde{X}_{i\cdot}'M_{\widetilde{G}}\widetilde{X}_{i\cdot}\right)$ and $(NT)^{-1}\sum_{i=1}^{N}\widetilde{X}_{i\cdot}'M_{\widetilde{G}}\epsilon_{i\cdot} \to_p 0$, so the consistency of this estimator follows almost immediately. Furthermore, assuming $\sqrt{T}/N\to 0$ and $\sqrt{T}c_{H_1}^2\to 0$ as $(N,T)\to\infty$, we have

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

In order to obtain the asymptotic normality of $\widehat{\beta}$, we analyze the variance of the above expression for which we define $\widetilde{\mathcal{W}}_{i\cdot}'=\widetilde{X}_{i\cdot}'M_{\widetilde{G}}$ and $\widetilde{\mathcal{W}}_{\cdot t}=(\widetilde{\mathcal{W}}_{1t},\ldots,\widetilde{\mathcal{W}}_{Nt})'$ as $p\times T$ and $N\times p$ matrices, respectively. Then, by the law of iterated expectations, we can prove

$$\begin{aligned}
Var[\sqrt{NT}(\widehat{\beta}-\beta)] &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[Q^{-1}\widetilde{\mathcal{W}}_{i\cdot}'\epsilon_{i\cdot}\epsilon_{j\cdot}\widetilde{\mathcal{W}}_{j\cdot}Q^{-1}\right]\\
&= \frac{1}{NT}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[Q^{-1}\widetilde{\mathcal{W}}_{\cdot t}'E(\epsilon_{\cdot t}\epsilon_{\cdot s}|z_t)\widetilde{\mathcal{W}}_{\cdot s}Q^{-1}\right]\\
&= \frac{1}{NT}\sum_{t=1}^{T}E\left[Q^{-1}\widetilde{P}\widetilde{P}'Q^{-1}\right]
\end{aligned}$$

where $\widetilde{P}$ is a $p\times T$ matrix such as $\widetilde{P}=[\widetilde{\mathcal{W}}_{\cdot 1}'\Omega(Z)^{1/2},\ldots,\widetilde{\mathcal{W}}_{\cdot T}'\Omega(Z)^{1/2}]$. Therefore, using the above results we can conclude

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

where $\Psi=\lim_{N,T\to\infty}(NT)^{-1}E\left[\widetilde{X}'(I_N\otimes M_{\widetilde{G}})'\Omega(Z)(I_N\otimes M_{\widetilde{G}})\widetilde{X}\right]$ given that rearranging terms it is straightforward to show $\widetilde{P}\widetilde{P}'=\widetilde{X}'(I_N\otimes M_{\widetilde{G}})'\Omega(Z)(I_N\otimes M_{\widetilde{G}})\widetilde{X}$ and the proof of the theorem is done.

∎

**Proof of Theorem 3.4**: Let $D\equiv(d_1,\ldots,d_T)'$ and $F\equiv(f_1,\ldots,f_T)'$ are $T\times n$ and $T\times r$ matrices, respectively, and $\epsilon_{i\cdot}\equiv(\epsilon_{i1},\ldots,\epsilon_{iT})'$ 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 (2.13), we get

$$\begin{aligned}\widehat{\beta}_{GLS} &= \left[\widehat{X}'(I_N \otimes M_{\widehat{\Lambda}})\Omega^{-1}(Z)(I_N \otimes M_{\widehat{\Lambda}})\widehat{X}\right]^{-1} \widehat{X}'(I_N \otimes M_{\widehat{\Lambda}})\Omega^{-1}(Z)(I_N \otimes M_{\widehat{\Lambda}}) \\ &\times \left[X\beta + \sum_{\iota=1}^{r} F_\iota \otimes \gamma_\iota + \epsilon + O_p(c_{H_1}^2)\right], \qquad (I.12)\end{aligned}$$

where $F_\iota$ and $\gamma_\iota$ are $T \times 1$ and $N \times 1$ vectors, for $\iota = 1, \dots, r$, respectively, since it can be proved that, uniformly in $z$, $(NT)^{-1}\widehat{X}'(I_N \otimes M_{\widehat{\Lambda}})\Omega^{-1}(Z)(I_N \otimes M_{\widehat{\Lambda}})(I_T - S)m_i(Z) = O_p(c_{H_1}^2)$ by combining the proof scheme for Lemma 3 in Cai et al. (2019) and Lemma A.6 in Su and Jin (2010).

As it is quite common in this type of literature, in (I.12) is observed the direct dependence of $\widehat{\beta}$ of the observed and unobserved common factors (i.e. $z_t$ and $f_t$). Using Lemmas 5.1-5.5 it is straightforward to show

$$\begin{aligned}\widehat{\beta}_{GLS} - \beta &= \left[\frac{\widetilde{X}'(I_N \otimes M_{\widetilde{G}})\Omega^{-1}(Z)(I_N \otimes M_{\widetilde{G}})\widetilde{X}}{NT}\right]^{-1} \frac{\widetilde{X}'(I_N \otimes M_{\widetilde{G}})\Omega^{-1}(Z)(I_N \otimes M_{\widetilde{G}})\epsilon}{NT} \\ &+ O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{NT}\right) + O_p(c_{H_1}^2), \qquad (I.13)\end{aligned}$$

where $M_{\widetilde{G}} = I_T - \widetilde{G}(\widetilde{G}'\widetilde{G})^{-1}\widetilde{G}^{-1}$ is a $T \times T$ projection matrix, $\widetilde{G} = (\widetilde{D}, \widetilde{F})$ is a $T \times (n + r)$ matrix.

Under the assumptions of the theorem it can be shown that $(NT)^{-1}\widetilde{X}'(I_N \otimes M_{\widetilde{G}})\Omega^{-1}(Z)(I_N \otimes M_{\widetilde{G}})\widetilde{X} \xrightarrow{p} Q_\varpi$, where $Q_\varpi = {}_{N,T\to\infty}\left((NT)^{-1}\widetilde{X}'(I_N \otimes M_{\widetilde{G}})\Omega^{-1}(Z)(I_N \otimes M_{\widetilde{G}})\widetilde{X}\right)$. Hence, using this result and assuming $\sqrt{T}/N \to 0$ and $\sqrt{NT}c_{H_1}^2 \to 0$, as $(N, T) \to \infty$, we get

$$\sqrt{NT}(\widehat{\beta}_{GLS} - \beta) = Q_\varpi^{-1}\left(\frac{\widetilde{X}'(I_N \otimes M_{\widetilde{G}})\Omega^{-1}(Z)(I_N \otimes M_{\widetilde{G}})\epsilon}{\sqrt{NT}}\right) + o_p(1) \qquad (I.14)$$

and the proof of the theorem is done.

∎

**Proof of Theorem 3.5**: Plugging (2.17) in (2.18), a Taylor expansion leads to

$$\begin{aligned}&\sqrt{T|H_1|}(\widehat{\overline{m}}_{GLS}(z, H_1, \varpi) - \overline{m}(z)) - \sqrt{T|H_1|}\iota_1'(T^{-1}Z_z'K_{H_1}(z)Z_z)^{-1}Z_z'K_{H_1}(z)\left[\frac{1}{2}Q_{\overline{m}}(z) + R_{\overline{m}}(z)\right] \\ &= \sqrt{T|H_1|}\iota_1'(T^{-1}Z_z'K_{H_1}(z)Z_z)^{-1}Z_z'K_{H_1}(z)\widetilde{U}^{(\varpi)}, \qquad (I.15)\end{aligned}$$

where $Q_{\overline{m}}(z) = [(z_1 - z)'\mathcal{H}_{\overline{m}}(z)(z_1 - z), \dots, (z_T - z)'\mathcal{H}_{\overline{m}}(z)(z_T - z)]'$, $R_{\overline{m}}(z)$ is the residual term of the Taylor expansion, and $\widetilde{U}^{(\varpi)} \equiv (\varpi'U_{\cdot 1}, \dots, \varpi'U_{\cdot T})'$ is a $T \times 1$ vector. Using standard nonparametric techniques it can be proved that $\iota_1'(T^{-1}Z_z'K_{H_1}(z)Z_z)^{-1}Z_z'K_{H_1}(z)R_{\overline{m}}(z) = o_p(tr\{H_1^2\})$ and that the asymptotic bias of $\widehat{\overline{m}}_{GLS}(z;H_1,\varpi)$ is

$$\iota_1'(T^{-1}Z_z'K_H(z)Z_z)^{-1}Z_z'K_{H_1}(z)\left[\frac{1}{2}Q_{\overline{m}}(z)+R_{\overline{m}}(z)\right] = \frac{\mu_2^q(K)}{2}tr\{H_1^2\mathcal{H}_{\overline{m}}(z)\}+o_p(tr\{H_1^2\}). \quad \text{(I.16)}$$

Considering now the variance term of the right-hand side of (I.15) and using the law of iterated expectations,

$$\begin{aligned} T|H_1|Var\big(T^{-1}Z_z'K_{H_1}(z)\widetilde{U}^{(\varpi)}\big) &= T^{-1}|H_1|E[Z_z'K_{H_1}(z)E[\widetilde{U}^{(\varpi)}\widetilde{U}^{(\varpi)\prime}|z_t]K_{H_1}(z)Z_z] \\ &= \begin{pmatrix} \nu_N^{(\varpi)}(z)R^q(K)\rho_{z_t}(z) & O_p(|H_1|) \\ O_p(|H_1|) & H_1^2\nu_N^{(\varpi)}(z)R_2^q(K)\rho_{z_t} \end{pmatrix}, \end{aligned} \quad \text{(I.17)}$$

where $\nu_N^{(\varpi)}(z) = (\imath_N'\Phi_N^{-1}(z)\imath_N)$.

Therefore, using Lemma 5.7 and (I.17), by the Slutsky theorem, as $T\to\infty$,

$$Var\left[\sqrt{T|H_1|}\iota_1'(T^{-1}Z_z'K_{H_1}(z)Z_z)^{-1}Z_z'K_{H_1}(z)\widetilde{U}^{(\varpi)}\right] = \frac{R^q(K)\nu_N^{(\varpi)}(z)}{\rho_{z_t}(z)}. \quad \text{(I.18)}$$

Finally, the Lyapunov condition can be proved under Assumption 3.16, and the proof of the Theorem is completed.

∎

**Proof Theorem 3.6**: Following a similar reasoning as in the proof of Theorem 3.5 and denoting $\widehat{\Psi}_{NT} = \widetilde{X}'(I_T\otimes M_{\widetilde{G}})\widehat{\Omega}^{-1}(Z)(I_T\otimes M_{\widetilde{G}})\widetilde{X}$ and $\Psi_{NT} = \widetilde{X}'(I_T\otimes M_{\widetilde{G}})\Omega^{-1}(Z)(I_T\otimes M_{\widetilde{G}})\widetilde{X}$, it is easy to show

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

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

$$\begin{aligned}
\widehat{\beta}_{FLGS} - \widehat{\beta}_{GLS} &= \left(\widehat{\Psi}_{NT}^{-1} - \Psi_{NT}^{-1}\right)\widetilde{X}'(I_T \otimes M_{\widetilde{G}})\left[\widehat{\Omega}^{-1}(Z) - \Omega^{-1}(Z)\right](I_T \otimes M_{\widetilde{G}})\epsilon \\
&\quad + \left(\widehat{\Psi}_{NT}^{-1} - \Psi_{NT}^{-1}\right)\widetilde{X}'(I_T \otimes M_{\widetilde{G}})\Omega^{-1}(Z)(I_T \otimes M_{\widetilde{G}})\epsilon \\
&\quad + \Psi_{NT}^{-1}\widetilde{X}'(I_T \otimes M_{\widetilde{G}})\left[\widehat{\Omega}^{-1}(Z) - \Omega^{-1}(Z)\right](I_T \otimes M_{\widetilde{G}})\epsilon \\
&\quad + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{NT}\right) + O_p(c_{H_1}^2) \\
&= \mathit{II}_{g_1} + \mathit{II}_{g_2} + \mathit{II}_{g_3} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{NT}\right) + O_p(c_{H_1}^2), \qquad \text{(I.19)}
\end{aligned}$$

where the definitions of $\mathit{II}_{g_l}$, for $l = 1, 2, 3$, should be apparent from the context.

Now we are going to analyze the behaviour of the above elements 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 5.6 it is straightforward to show $\|\Omega^{-1}(Z) - \Omega^{-1}(Z)\| = O_p(NR_{TH})$, as $N/T \to \kappa$, where $\kappa$ is a positive constant. Therefore, to finish the proof is enough to show

$$\frac{1}{NT}\left(\widehat{\Psi}_{NT} - \Psi_{NT}\right) = o_p(1), \qquad \text{(I.20)}$$

$$\frac{1}{NT}\widetilde{X}'(I_N \otimes M_{\widetilde{G}})'\left[\Omega^{-1}(Z) - \Omega^{-1}(Z)\right](I_N \otimes M_{\widetilde{G}})\epsilon = O_p((NT)^{-1/2})O_p(NR_{TH}). \qquad \text{(I.21)}$$

Considering the proof of (I.20), it has the norm bounded by

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

using the fact that $(NT)^{-1}\widetilde{X}'(I_N \otimes M_{\widetilde{G}})\widetilde{X} = O_p(NR_{TH})$, $\|\Omega^{-1}(Z)\| = O_p(1)$, $\|\widehat{\Omega}(Z) - \Omega(Z)\| = O_p(NR_{TH})$ (see Lemma 5.6). Furthermore, Assumption 3.13 implies $Ntr\{H_2^2\}/tr\{H_1^2\} \to 0$ and $N(T|H_1|)^{-1} = o_p((NT|H_1|)^{-1/2}) = o_p((T|H_1|)^{-1/2})$, given that $N^3/(T|H_1|) \to 0$. Hence, $NR_{TH} = o((T|H_1|)^{-1/2} + tr\{H_1^2\})$ and (I.20) is proved.

Under similar reasoning, it can be shown that (I.21) has the norm bounded by

$$\begin{aligned}
&\|(NT)^{-1}\widetilde{X}'(I_N \otimes M_{\widetilde{G}})'[\widehat{\Omega}^{-1}(Z) - \Omega^{-1}(Z)](I_N \otimes M_{\widetilde{G}})\epsilon\| \\
&\leq (NT)^{-1/2}\|(NT)^{-1/2}\widetilde{X}'(I_N \otimes M_{\widetilde{G}})\epsilon\|\|\Omega_N^{-2}(Z)\|\|\widehat{\Omega}(Z) - \Omega\,(Z)\,\| \\
&= O_p((NT)^{-1/2})O_p(NR_{TH}).
\end{aligned}$$

Finally, using (I.20)-(I.21) in (I.19) it is straightforward to show

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

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 Theorem 3.7**: In order to prove this theorem, it can be written

$$\widehat{\widehat{m}}_{FLGS}(z; H_1, \widehat{\varpi}) - \widehat{\widehat{m}}_{GLS}(z; H_1, \varpi) \;\; = \;\; \imath_1'(Z_z' K_{H_1}(z) Z_z)^{-1} Z_z' K_{H_1}(z) \left(\widehat{\widetilde{Y}}\widehat{\varpi} - \widetilde{Y}\varpi\right), \tag{I.22}$$

where $\widehat{\widetilde{Y}}$ and $\widetilde{Y}$ are $T \times N$ matrices whose $it$-th elements are such as $\widehat{\widetilde{y}}_{it} = y_{it} - x_{it}'\widehat{\beta} - \lambda_t'\widehat{\delta}_i$ and $\widetilde{y}_{it} = y_{it} - x_{it}'\beta - \lambda_t'\delta_i$, respectively. Replacing (2.4) in (I.22) and rearranging terms, the final expression to analyze is such us

$$\begin{aligned} &|\widehat{\widehat{m}}_{FLGS}(z; H_1, \widehat{\varpi}) - \widehat{\widehat{m}}_{GLS}(z; H_1, \varpi)| \\ &\leq \imath_1' \left\| (Z_z' K_{H_1}(z) Z_z)^{-1} Z_z' K_{H_1}(z) \left[ U - \sum_{\varrho=1}^{p} X_\varrho(\widehat{\beta}_\varrho - \beta_\varrho) - \Lambda(\widehat{\delta} - \delta)' \right] \right\| \|\widehat{\varpi} - \varpi\|, \end{aligned} \tag{I.23}$$

where $U$ and $X_\varrho$ are $T \times N$ matrices and $\widehat{\delta}$ and $\delta$ are $N \times \ell$ matrices. From the results in Lemma 5.7, it is straightforward to show $\|T^{-1}Z_z'K_{H_1}(z)Z_z\| = O_p((T|H_1|)^{-1/2})$, whereas considering the behavior of the numerator term in (I.23), we have

$$\begin{aligned} &\left\| T^{-1} Z_z' K_{H_1}(z) \left[ U - \sum_{\varrho=1}^{p} X_\varrho(\widehat{\beta}_\varrho - \beta_\varrho) - \Lambda(\widehat{\delta} - \delta)' \right] \right\| \\ &\leq \|T^{-1}Z_z'K_{H_1}(z)U\| + \|T^{-1}Z_z'K_{H_1}(z)X\|\|\widehat{\beta} - \beta\| + \|T^{-1}Z_z'K_{H_1}(z)\Lambda\|\|\widehat{\delta} - \delta\|. \end{aligned} \tag{I.24}$$

Using the consistency result obtained previously for $\widehat{\beta}$, it can be shown $\|\widehat{\beta} - \beta\| = O_p((NT)^{-1/2})$ and, under a similar reasoning, it is straightforward to show $\|\widehat{\delta} - \delta\| = O_p(T^{-1/2})$. Following a similar reasoning as in Ruppert and Wand (1994) and using these results in (I.24), we can prove that $\|T^{-1}Z_z'K_{H_1}(z)X\|$ and $\|T^{-1}Z_z'K_{H_1}(z)\Lambda\|$ are $O_p((T|H_1|)^{-1/2})$. Using all these results in (I.24) and given that by Lemma 5.8 we get $\|\widehat{\varpi} - \varpi\| = O_p(NR_{TH})$

as $N/T \to \kappa$, where $\kappa$ is a positive constant, we have

$$\imath_1' \|T^{-1} Z_z' K_{H_1}(z) Z_z\| \|T^{-1} Z_z' K_{H_1}(z) X\| \|\widehat{\beta} - \beta\| = o_p(N R_{TH}), \tag{I.25}$$

$$\imath_1' \|T^{-1} Z_z' K_{H_1}(z) Z_z\| \|T^{-1} Z_z' K_{H_1}(z) \Lambda\| \|\widehat{\delta} - \delta\| = o_p(N R_{TH}). \tag{I.26}$$

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

$$\|T^{-1} Z_z' K_{H_1}(z) U\| = \begin{pmatrix} O_p(\|\Phi_N^{1/2}(z)\|(T|H_1|)^{-1/2}) \\ O_p(\|\Phi_N^{1/2}(z)\| tr\{H_1^2\}(T|H_1|)^{-1/2}) \end{pmatrix}, \tag{I.27}$$

given that, using the law of iterated expectations,

$$\begin{aligned} E\left\| T^{-1} \sum_{t=1}^{T} K_{H_1}(z_t - z) u_{\cdot t} \right\|^2 &= tr\left\{ T^{-2} \sum_{t=1}^{T} E[K_{H_1}^2(z_t - z) E(u_{\cdot t} u_{\cdot t}' | z_t)] \right\} \\ &= \frac{R^q(K)\rho_{z_t}(z)}{T|H_1|} \|\Phi_N^{1/2}(z)\| = O_p\left( \frac{\|\Phi_N^{1/2}(z)\|}{T|H_1|} \right) \end{aligned}$$

and $E\|T^{-1} \sum_{t=1}^{T} K_{H_1}(z_t - z)(z_t - z) u_{\cdot t}\|^2 = O_p\left( \frac{\|\Phi_N^{1/2}(z)\| tr\{H_1^2\}}{T|H_1|} \right)$. Hence, under a similar reasoning as in (I.25)-(I.26) and using Assumption 3.13 we have

$$\imath_1' \|T^{-1} Z_z' K_{H_1}(z) Z_z\| \|T^{-1} Z_z' K_{H_1}(z) U\| = o_p\left( \frac{\nu_N^{-1/2}(z)}{\sqrt{T|H_1|}} + tr\{H_1^2\} \right) \tag{I.28}$$

and plugging (I.25)-(I.26) and (I.28) in (I.23) the proof of the theorem is done.

∎

# References

Andersen, M. S. and P. Ekins (2009). *Carbon-energy taxation: lessons from Europe*. OUP Oxford.

Azomahou, T., F. Laisney, and P. N. Van (2006). Economic development and CO2 emissions: A nonparametric panel approach. *Journal of Public Economics 90*(6-7), 1347–1363.

Bailey, N., G. Kapetanios, and M. H. Pesaran (2016). Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics 31*(6), 929–960.

Baltagi, B. H., G. Bresson, and J.-M. Etienne (2019). Carbon dioxide emissions and economic activities: A mean field variational bayes semiparametric panel data model with random coefficients. *Annals of Economics and Statistics 134*, 43–77.

Baltagi, B. H., G. Bresson, and A. Pirotte (2004). Tobin q: forecast performance for hierarchical bayes, shrinkage, heterogeneous and homogeneous panel data estimators. *Empirical Economics 29*(1), 107–113.

Bolat, C. K., U. Soytas, B. Akinoglu, and S. Nazlioglu (2023). Is there a macroeconomic carbon rebound effect in eu ets? *Energy Economics 125*, 106879.

Borghesi, S., C. Franco, and G. Marin (2020). Outward foreign direct investment patterns of italian firms in the european union's emission trading scheme. *The Scandinavian Journal of Economics 122*(1), 219–256.

Borghesi, S., M. Montini, and A. Barreca (2016). *The European emission trading system and its followers: comparative analysis and linking perspectives*. Springer.

Cai, Z., Y. Fan, M. Lin, and J. Su (2019). Inferences for a partially varying coefficient models with endogenous regressors. *Journal of Business and Economic Statistics 37*(1), 158–170.

Calel, R. (2020). Adopt or innovate: Understanding technological responses to cap-and-trade. *American Economic Journal: Economic Policy 12*(3), 170–201.

Calel, R. and A. Dechezlepretre (2016). Environmental policy and directed technological change: evidence from the european carbon market. *Review of Economics and Statistics 98*(1), 173–191.

Carratù, M., B. Chiarini, and P. Piselli (2020). Effects of european emission unit allowance auctions on corporate profitability. *Energy Policy 144*, 111584.

Churchill, S. A., J. Inekwe, K. Ivanovski, and R. Smyth (2018). The environmental kuznets curve in the oecd: 1870–2014. *Energy economics 75*, 389–399.

Cole, M. A., R. J. Elliott, and K. Shimamoto (2005). Industrial characteristics, environmental regulations and air pollution: an analysis of the UK manufacturing sector. *Journal of Environmental Economics and Management 50*(1), 121–143.

Colmer, J., R. Martin, M. Muuls, and U. J. Wagner (2022). Does pricing carbon mitigate climate change? Firm-level evidence from the European Union emissions trading scheme. CEPR Discussion Paper No. DP16982.

Costantini, V., M. Mazzanti, and A. Montini (2013). Environmental performance, innovation and spillovers. evidence from a regional NAMEA. *Ecological Economics 89*, 101–114.

Ellerman, A. D. and P. L. Joskow (2008). *The European Union's emissions trading system in perspective*. Pew Center on Global Climate Change Arlington, VA.

Ellerman, A. D., C. Marcantonini, and A. Zaklan (2016). The european union emissions trading system: ten years and counting. *Review of Environmental Economics and Policy*.

Ertur, C. and A. Musolesi (2017). Weak and strong cross-sectional dependence: A panel data analysis of international technology diffusion. *Journal of Applied Econometrics 32*(3), 477–503.

Fan, J. and T. Huang (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli 11*, 1031–1057.

Fernandez, Y. F., M. F. Lopez, and B. O. Blanco (2018). Innovation for sustainability: the impact of r&d spending on co2 emissions. *Journal of cleaner production 172*, 3459–3467.

Griliches, Z. (1998). Patent Statistics as Economic Indicators: A Survey. In Z. Griliches (Ed.), *R&D and Productivity: The Econometric Evidence*, pp. 287–343. Chicago: University of Chicago Press.

Hansen, B. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory 24*, 726–748.

Härdle, W., H. Liang, and J. Gao (2000). *Partially Linear Models*. Heidelberg: Physica-Verlag.

Huang, B., Y. Sun, and S. Wang (2021). Estimation of partially linear panel data models with cross-sectional dependence. *Journal of Systems Science and Complexity 34*(6), 2219–2230.

Jiménez-Rodríguez, R. (2019). What happens to the relationship between EU allowances prices and stock market indices in Europe? *Energy Economics 81*, 13–24.

Juodis, A. and S. Reese (2022). The incidental parameters problem in testing for remaining cross-section correlation. *Journal of Business & Economic Statistics 40*(3), 1191–1203.

Känzig, D. R. and M. Konradt (2023). Climate policy and the economy: evidence from Europe's carbon pricing initiatives. Technical report, National Bureau of Economic Research.

Kapetanios, G., M. H. Pesaran, and T. Yamagata (2011). Panels with non-stationary multifactor error structures. *Journal of Econometrics 160*(2), 326–348.

Koçak, E. and Z. Ş. Ulucak (2019). The effect of energy r&d expenditures on co 2 emission reduction: estimation of the stirpat model for oecd countries. *Environmental Science and Pollution Research 26*, 14328–14338.

Koch, N. and M. Themann (2022). Catching up and falling behind: cross-country evidence on the impact of the EU ETS on firm productivity. *Resource and Energy Economics 69*, 101315.

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

Marin, G., M. Marino, and C. Pellegrin (2018). The impact of the of the european emission trading scheme on multiple measures of economic performance. *Environmental and Resource Economics 71*, 551–582.

Mazzanti, M. and A. Musolesi (2013). The heterogeneity of carbon kuznets curves for advanced countries: comparing homogeneous, heterogeneous and shrinkage/bayesian estimators. *Applied Economics 45*(27), 3827–3842.

Metcalf, G. E. and J. H. Stock (2023). The macroeconomic impact of Europe?s carbon taxes. *American Economic Journal: Macroeconomics 15*(3), 265–286.

Millimet, D. L., J. A. List, and T. Stengos (2003). The environmental kuznets curve: real progress or misspecified models? *Review of Economics and Statistics 85*(4), 1038–1047.

Musolesi, A. and M. Mazzanti (2014). Nonlinearity, heterogeneity and unobserved effects in the carbon dioxide emissions-economic development relation for advanced countries. *Studies in Nonlinear Dynamics and Econometrics 18*(5), 521–541.

Papież, M., S. Śmiech, and K. Frodyma (2022). Does the European Union energy policy support progress in decoupling economic growth from emissions? *Energy Policy 170*, 113247.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica 74*(4), 967–1012.

Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews 34*(6-10), 1089–1117.

Pesaran, M. H. (2021). General diagnostic tests for cross-sectional dependence in panels. *Empirical economics 60*(1), 13–50.

Pesaran, M. H. and E. Tosetti (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics 161*(2), 182–202.

Robinson, P. M. (1988). Root-n consistent semiparametric regression. *Econometrica 56*, 931–954.

Robinson, P. M. (2012). Nonparametric trending regression with cross-sectional dependence. *Journal of Econometrics 169*(1), 4–14.

Rupasingha, A., S. J. Goetz, D. L. Debertin, and A. Pagoulatos (2004). The environmental kuznets curve for US counties: A spatial econometric analysis with extensions. *Papers in Regional Science 83*(2), 407–424.

Ruppert, D. and M. P. Wand (1994). Multivariate locally weighted least squares regression. *Annals of Statistics 22*, 1346–1370.

Soberon, A., J. M. Rodriguez-Poo, and P. M. Robinson (2022). Nonparametric panel data regression with parametric cross-sectional dependence. *The Econometrics Journal 25*, 114–133.

Su, L. and S. Jin (2010). Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. *Journal of Econometrics 157*(1), 18–23.

Wagner, M. (2015). The environmental kuznets curve, cointegration and nonlinearity. *Journal of Applied Econometrics 30*(6), 948–967.

Wagner, M., P. Grabarczyk, and S. H. Hong (2020). 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*(1), 216–255.