SEEDS
Sustainability Environmental Economics and Dynamics Studies

Working Paper Series

A semiparametric panel data model with common factors and spatial dependence

by

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

01/2022

SEEDS is an interuniversity research centre. It develops research and higher education projects in the fields of ecological and environmental economics, with a special focus on the role of policy and innovation. Main fields of action are environmental policy, economics of innovation, energy economics and policy, economic evaluation by stated preference techniques, waste management and policy, climate change and development.

The SEEDS Working Paper Series are indexed in RePEc and Google Scholar.
Papers can be downloaded free of charge from the following websites:
http://www.sustainability-seeds.org/.

Enquiries:info@sustainability-seeds.org

SEEDS Working Paper 01/2022
January 2022
By Alexandra Soberon, Antonio Musolesi, Juan M. Rodriguez-Poo

A semiparametric panel data model with common factors and spatial dependence

Alexandra Soberon*

Department of Economics, University of Cantabria
and
Antonio Musolesi
Department of Economics and Business, University of Ferrara
and
Juan M. Rodriguez-Poo
Department of Economics, University of Cantabria

28th December 2021

Abstract

This paper proposes alternative estimation procedures for semiparametric panel data models that allow handling complex and relevant empirical problems simultaneously, namely (i) functional misspecification, by modelling stochastic observed common factors with a nonparametric function instead of assuming the usual parametric form; (ii) cross-sectional dependence arising simultaneously from common factors and spatial dependence; and iii) heterogeneous relations among variables. We then consider a more general panel data model with several types of cross-sectional dependence and obtain consistent and asymptotically normal estimators for both slope parameters and unknown functions by extending Pesaran's (2006) common correlated effect (CCE) approach to this semiparametric framework. Another methodological and empirical challenge is how to test for poolability and fully parametric functional form. For both cases, simple consistent test statistics are proposed and we show that they have limiting standard distributions under the null hypothesis. The theoretical findings are further supported for small samples via several Monte Carlo experiments, and an empirical application to the knowledge capital production function is conducted.

Keywords: Cross-sectional dependence, Nonparametric estimation, Common correlated effects estimator, Nonparametric test, Knowledge capital production function.

*This work is part of the I+D+i project ref. PID2019-105986GB-C22, funded by MCIN/AEI/10.13039/501100011033. The authors also gratefully acknowledge financial support from Research Project APIE 1/2015-17: "New methods for the empirical analysis of financial markets" of the Santander Financial Institute (SANFI) through the UCEIF Foundation, resolved by the University of Cantabria and funded through sponsorship from Banco Santander. Part of the work was done during visits by Alexandra Soberon to the University of Ferrara.

1 Introduction

In the last decade, with the increasing process of globalization and the growing importance of the economic and social interconnections across economic agents, the issue of cross-sectional dependence (CSD) in panel data models has become of crucial relevance from both a theoretical and an empirical point of view. While CSD has typically been introduced alternatively as a result of a finite number of unobservable (and/or observed) common factors (Pesaran, 2006; Bai, 2009) or in the consideration of spatial models (Lee and Yu, 2010), recent works highlight the importance of jointly modelling both forms of dependence (Holly et al., 2010; Pesaran and Tosetti, 2011; Bailey et al., 2016a; Shi and Lee, 2017). This is because although both factor and spatial models allow for CSD, the motivations underlying these models differ meaningfully. In the first approach, the unobserved factors are regarded as nuisance variables introduced to allow for CSD and to handle endogeneity due to unobservable features, whereby the explanatory variables are allowed to be correlated with the factors. The second approach is instead explicitly oriented toward modeling cross-sectional interactions and capturing spatial spillovers or analyzing the global spatial diffusion process of a random shock that occurs in a given cross-sectional unit (Debarsy et al., 2012). Moreover, while factor and spatial models are related to the concepts of strong and weak CSD, respectively, neglecting CSD may seriously affect empirical outcomes since cross-sectional independence of the errors cannot be guaranteed and traditional panel estimators may provide misleading inferences or inconsistent estimates (Pesaran, 2006; Chudik et al., 2011; Ertur and Musolesi, 2017).

One shortcoming of the above parametric models is that they are subject to strong restrictions among variables that may not be supported by the data. In such situations, the risk of misspecification may be high and alternative specifications such as nonparametric and semiparametric models may be more appealing. In recent years, these types of models have gained a lot of attention since they avoid inconsistent estimates by imposing relatively little restrictions on model structures (see Rodriguez-Poo and Soberon, 2017; Parmeter and Racine, 2018 for in-depth reviews). Nevertheless, while there is a relatively rich literature on linear panel data models with CSD (see Pesaran, 2006;

Holly et al., 2010; Bailey et al., 2016b, among others), few advances have been made in estimating nonparametric and/or semiparametric models with CSD. On the one hand, when common factors are allowed, Su and Jin (2012) and Huang (2013) consider a sieve and a local linear method, respectively, to estimate a nonparametric panel data model, Cai et al. (2020) develop a local linear procedure for functional-coefficient panel data models, and Dong et al. (2021) propose a sieve method for a varying-coefficient panel data model with a partially observed factor structure. On the other hand, when the CSD comes from spatial correlations, Robinson (2012), Lee and Robinson (2015), and Soberon et al. (2021) propose several local linear procedures that enable taking into account the CSD to achieve efficiency, whereas Gao et al. (2020) propose a semiparametric profile-likelihood estimation method to estimate the slopes and the individual-specific trending functions. However, to the best of our knowledge there is no nonparametric or semiparametric approach that can deal with both types of CSD simultaneously. Readers are referred to the survey in Xu et al. (2016) for more details on this topic.

In the first part of this paper we consider a more general semiparametric panel data model that enables the handling of complex empirical problems, namely (i) functional misspecification; (ii) CSD arising simultaneously from common factors and spatial dependence; and (iii) heterogeneous relations among variables. Andrews (2005), among others, highlights that assessing the effect of common factors (such as technological, institutional, environmental, and health factors) is of crucial relevance in many empirical cases. Specifically, one rationale for introducing such specification is that in a number of circumstances such as wage, cost, or production functions, parametric specifications for the main explanatory variables are well established and build on economic theory. However, there is generally a high degree of uncertainty surrounding the way in which common factors may affect the cross-sectional units (such as individuals, households, firms, industries, or countries). For example, allowing for flexible forms may be suitable when focusing on the effect of real common shocks (such as a decline in the aggregate demand) on productivity or economic growth, when estimating the effects of oil prices on wages, employment, or production activity (Keane and Prasad, 1996; Hamilton, 2003), or when assessing the effect of global policies such as the European Trading System, which sets a common carbon permit price across European countries. In all these situations and others, it is likely that a parametric specification would be an oversimplification of the underlying data generating process (DGP). To deal with this potential functional misspecification, we propose to model stochastic observed common factors with a nonparametric function, instead of assuming the usual parametric form. Note that adopting such a semiparametric model would also be extremely useful when the main interest lies in estimating the slope parameters since it avoids inconsistent estimation of these parameters from erroneously imposing a parametric form on the observed factors. At the same time, it also reduces the extent of the curse of dimensionality problem that a fully nonparametric model would face. Furthermore, with the aim of covering a variety of dependence settings, we propose a general form for the spatial correlation that does not impose a specific parametric spatial diffusion process—as is quite usual in this literature—and covers other well-known parametric spatial structures as special cases.

We first focus our attention on the consistent estimation of both homogeneous and heterogeneous slope parameters and unknown functions in the above model and also explore the asymptotic distributions of the resulting estimators under rather standard conditions. Later, we also deal with testing slope homogeneity and functional form. This can be very useful from an empirical point of view since, in practice, economic theory usually cannot tell whether the regression relationship is homogeneous or not, or whether a parametric functional form is preferred. There are indeed a number of studies on testing for slope homogeneity in panel data models (see Phillips and Sul (2003); Pesaran and Yamagata (2008); Su and Chen (2013), among others). However, this type of parametric test may falsely reject the null of poolability when the data are truly poolable because the functional form is misspecified. For all of these reasons, in the second part of the paper two alternative specification tests are proposed: a constancy test that enables assessing a linear parametric functional form against a semiparametric alternative, and a poolability slope test for a semiparametric specification. To the best of our knowledge, these tests are a novelty in the nonparametric/semiparametric CSD literature.

Finite-sample properties of the proposed estimators and test statistics are demonstrated via simulations. In both the case of heterogeneous and homogeneous slopes, the proposed estimators are robust to the presence of spatial dependence or fixed effects, while the test statistics perform well for small samples, based on critical values obtained from the bootstrap method. Finally, an empirical application to the knowledge capital production function of Griliches (1979) is conducted by exploiting an annual country-level balanced panel data set covering 24 OECD countries from 1971 to 2014, with the price of oil considered as an observed common factor. Our results highlight that we face a sizeable bias with respect to both the technological parameters and the effect of oil prices when estimating a fully parametric model.

The remainder of the paper is organized as follows. In Section 2, we set up the econometric model and provide a motivating example. In Section 3, we provide the main theoretical results. In Section 4, we derive the limiting results of the proposed test statistics. Section 5 revisits the empirical application. Finally, Section 6 offers some concluding remarks. Due to limitation of space, we move the rest of this article to an online supplementary material: Appendix S1 presents the technical assumptions required to obtain the main asymptotic properties of the estimators and tests statistics discussed in Sections 3 and 4. Appendix S2 justifies the extension to the CCE approach from a fully parametric model to a semiparametric setting. Appendix S3 presents an alternative bootstrap procedure to better approximate the finite sample null distribution of the test statistics. Appendix S4 defines the data and variables used in the empirical application. Appendix S5 includes the Monte Carlo simulation results. Appendix S6 presents preliminary lemmas and the proofs of the corresponding lemmas. Appendix S7 states the corresponding mathematical proofs to obtain the asymptotic properties of the proposed estimators and test statistics.

Notation: We adopt the following notation throughout the paper. For a real matrix $A \in \mathbb{M}^{m\times m}$, where $\mathbb{M}^{m\times m}$ is the space of real $m \times m$ matrices, we denote its Euclidean norm as $\|A\| = [tr(AA')]^{1/2}$ and its generalized inverse as A^{-1}. In addition, $\lambda_1^*(A) \geq \lambda_2^*(A) \geq \ldots \geq \lambda_m^*(A)$ are the eigenvalues of A and C is used for a fixed positive constant. The equation $a_n = O(b_n)$ states that the deterministic sequence a_n is at most of order b_n, $x_n = O_p(y_n)$ states that the vector of random variables x_n is at most of order y_n in probability, and $x_n = o_p(y_n)$ is of a smaller order than y_n in terms of probability. The operator $\xrightarrow{p}$ denotes convergence in probability, $\xrightarrow{d}$ denotes convergence in distribution, and $\xrightarrow{q.m.}$ denotes convergence in quadratic mean. All asymptotics are carried out under $(N, T) \xrightarrow{j} \infty$, which means that N and T tend to infinity jointly but in no particular order. Restrictions (if any) on the relative rates of convergence of N and T will be specified separately.

2 Model setting

2.1 Motivating example

Eberhardt et al. (2013) use data on 12 manufacturing industries in 10 countries to estimate the knowledge capital production function of Griliches (1979). In this framework, R&D capital stock is an additional input of a production function and accounts for knowledge capital:

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

where Y_{it} is value-added, L_{it} and K_{it} are standard labor and capital inputs, and R_{it} is knowledge capital. In addition, $c_t' = [d_t', z_t']$ is a vector of observed common factors, either deterministic (d_t, such as a time trend) or stochastic (z_t, such as oil prices), f_t is a vector of unobserved common factors, and ϵ_{it} is a zero mean error term. Following Griliches (1979), among many others, Eberhardt et al. (2013) adopt a Cobb–Douglas technology, $g(L_{it}, K_{it}, R_{it}) = L_{it}^{\beta_1} K_{it}^{\beta_2} R_{it}^{\beta_3}$, and taking logs they estimate the following specification:

$$y_{it} = \alpha_i c_t + \beta_1 l_{it} + \beta_2 k_{it} + \beta_3 r_{it} + \gamma_i' f_t + \epsilon_{it}, \tag{2.1}$$

where lowercase letters indicate that the variables are in log form. Eberhardt et al. (2013) use the

CCE approach of Pesaran (2006)—also considering the case in which the technological parameters β_1, β_2, and β_3 are heterogeneous across cross-sectional units—and finally obtain *"substantially lower coefficients for the R&D capital stock"*. This result contradicts most of the previous literature, and for this reason is worth reassessing.

In this paper, we propose a new semiparametric heterogeneous panel data model that extends the parametric one in (2.1) and which appears to be suitable for modeling the knowledge capital production function. Specifically, we propose the following specification:

$$y_{it} = \alpha_i d_t + \beta_{1i} l_{it} + \beta_{2i} k_{it} + \beta_{3i} r_{it} + m_i(z_t) + \gamma_i' f_t + \epsilon_{it}. \tag{2.2}$$

The first difference with respect to the parametric case (2.1) is that the stochastic observed common factors, denoted as z_t, enter the model via a nonparametric function $m_i(.)$ instead of entering the model linearly. A motivation for introducing such a specification is that, despite the well-known limitations of a Cobb–Douglas specification for $g(L_{it}, K_{it}, R_{it})$ (Ivaldi et al., 1996; Ma et al., 2020), it remains the cornerstone of a huge literature in empirical economics at all levels of aggregation (Doraszelski and Jaumandreu, 2013; Eberhardt et al., 2013; Ackerberg et al., 2015; Antonioli et al., 2021) because it builds on economic theory and generally provides a good approximation of the underlying data in a parsimonious setting. In contrast, there is generally little prior knowledge—theoretical or empirical—about the shape of the relation between the observed common factors and production at a sectoral level (or at a different level of aggregation). Moreover, erroneously imposing a parametric form for the factors may lead to biased estimates of the parameters of interest, β_{1i}, β_{2i}, and β_{3i}. Observed common factors may include oil price shocks that can affect production activity through their effects on production costs or measures of worldwide economic activity (Kilian, 2009). The heterogeneous effects of these factors may be the result of unit-specific technological constraints, for instance. More specifically, as far as the price of oil is concerned, while a large body of research has attempted to estimate its effects on economic activity, the fact that oil price shocks contribute directly to economic decline remains controversial as the empirical relation between oil prices and output has been very unstable across previous works. Since the pioneering work of Mork (1989), such instability has been attributed to a misspecification of the linear approximation of the relation between oil prices and economic activity. Thus, it has been argued that economic activity responds asymmetrically to positive and negative oil price shocks as rising oil prices should negatively affect aggregate economic activity more than decreasing oil prices should stimulate it (Lescaroux and Mignon, 2008), thus calling for the adoption of flexible specifications (Hamilton, 2003).

Second, it can be crucial to jointly model spatial interactions and common factors. While previous studies have estimated production functions and addressed the issue of CSD either by using a factor (Calderón et al., 2015) or a spatial approach (Parent and LeSage, 2008; Glass et al., 2016) separately, it can be argued that these two forms of dependence may coexist and should be considered simultaneously since they address different econometric issues and lead to different economic insights. Moreover, Millo (2019) provides evidence of spatially correlated residuals even after unobserved common factors are introduced in the model. It is therefore important to introduce spatial error dependence into the model. Commonly used spatial processes such as the spatial moving average (SMA) and the spatial autoregressive (SAR) model produce specific parametric forms for the so-called *spatial diffusion process* of a random shock (see Debarsy et al. (2012)). However, in the present empirical framework (as in many other empirical applications), a more flexible specification may be more appropriate. Moreover, a high degree of uncertainty may surround the specification of the spatial matrix. With the aim of covering a variety of dependence settings, we propose a more general form that does not impose a specific parametric spatial diffusion process and covers other well-known parametric spatial structures as special cases.

2.2 Econometric specification

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

$$y_{it} = \alpha_i' d_t + x_{it}'\beta_i + m_i(z_t) + u_{it}, \quad i = 1, \ldots, N, \quad t_i = 1, \ldots, T, \tag{2.3}$$

where $d_t = (d_{1t}, d_{2t}, \ldots, d_{nt})'$ is an $n \times 1$ vector of observed common effects (including deterministic regressors such as intercepts or seasonal dummies), x_{it} is a $p \times 1$ vector of observed explanatory variables, z_t is a $q \times 1$ vector of observed stochastic common effects, and $m_i(\cdot)$ is an unknown smooth function to estimate. In addition, u_{it} is a random error term which follows the following multi-factor structure

$$u_{it} = \gamma_i' f_t + \epsilon_{it}, \tag{2.4}$$

where $f_t = (f_{1t}, f_{2t}, \ldots, f_{rt})'$ is an $r \times 1$ vector of unobserved common factors, $\gamma_i = (\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{ir})'$ are factor loadings, and ϵ_{it} is an idiosyncratic error of y_{it} with zero mean, $E(\epsilon_{it}\epsilon_{jt}) = \omega_{ij}$ for $i \neq j$, and $E(\epsilon_{it}\epsilon_{js}) = 0$ when $i \neq j$ and $t \neq s$. Note that the number of unobserved factors, r, is assumed to be fixed relative to N—more specifically, $r < N$. Moreover, the common factors, f_t, simultaneously affect all cross-sectional units, albeit with different degrees, as measured by γ_i.

In general, the unobserved factors f_t are allowed to be correlated with the observed data (x_{it}, z_t, d_t). To allow for this possibility in a flexible way, this correlation can be modeled via the following fairly general semiparametric model:

$$x_{it} = A_i' d_t + g_i(z_t) + \Gamma_i' f_t + v_{it}, \tag{2.5}$$

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

In order to jointly model spatial dependence and unobserved common factors, we rely on the following assumption.

Assumption 2.1 (Spatial error dependence) *Let $\epsilon_{\cdot t} = (\epsilon_{1t}, \ldots, \epsilon_{Nt})'$ be an $N \times 1$ vector and the individual-specific errors follow an arbitrary form such as*

$$\epsilon_{\cdot t} = \Phi^{1/2} \eta_{\cdot t}, \tag{2.6}$$

where $\Phi^{1/2}$ is a given $N \times N$ matrix. In addition, for each i, η_{it} follows the linear stationary process with absolute summable autocovariances, $\eta_{it} = \sum_{s=0}^{\infty} a_{is}\zeta_{is}$, where $\zeta_{is} \sim iid(0,1)$ with finite fourth-order cumulants. In particular, $Var(\eta_{it}) = \sum_{s=0}^{\infty} a_{is}^2 = \sigma_{\eta_i}^2 \leq \overline{\sigma}_\eta^2 < \infty$ for all i and some constant $\overline{\sigma}_\eta^2$, where $\sigma_{\eta_i}^2 > 0$. Furthermore, $\Phi^{1/2}$ has bounded row and column norms.

Remark 2.1 ***Weak and strong dependence.*** *Although the literature does not provide a single definition of "weak" or "strong" dependence (Sarafidis, 2011; Chudik et al., 2011; Robinson, 2011; Ertur and Musolesi, 2017), it is worth noting that the type of dependence arising from the factor model and spatial errors depends on the adopted definition of weak/strong dependence and the limiting properties of averaged factor loadings (Chudik et al., 2011; Sarafidis and Wansbeek, 2012). A related and important concept is that of strong and weak factors (Chudik et al., 2011). Factor models are typically used in the literature to represent strong CSD processes. Conversely, commonly used parametric spatial models, under a standard set of regularity conditions, entail weak CSD regardless of the definition adopted. It is notable that the proposed general type of spatial model (2.6) may entail weak or strong dependence, as is discussed in Lemma 3.1.*

The first part of this paper is focused on estimating the slope coefficients β_i, smooth functions $m_i(\cdot)$, and their corresponding cross-sectional means, $\overline{\beta} = E(\beta_i)$ and $\overline{m}(\cdot) = E(m_i(\cdot))$, by controlling for various sources of CSD. Model (2.3)–(2.4) is sufficiently general and enables us to deal with several limitations in the existing literature. Firstly, the semiparametric specification of (2.5) avoids possible misspecification problems. The parametric specification $x_{it}'\beta_i$ enables us to incorporate some prior information coming from economic theory or past experience in many empirical frameworks, but (2.5) incorporates flexibility into the model, allowing the data to tailor the shape of the observed common factor component (i.e., z_t). Secondly, the above specification includes a variety of panel data models as special cases. Whether the observed common factors enter the model linearly, it renders the proposal in Pesaran (2006). Whether the coefficients α_i are treated as fixed or random, potential correlation with the regressors of the model must be considered, as in Bai (2003, 2009) or Bai and Ng (2001), among others. Furthermore, the proposed specification in (2.5) complements the contributions in Su and Jin (2012); Huang (2013); Cai et al. (2020); Gao et al. (2020) where both observed and unobserved factors are assumed to enter into the model parametrically, whereas a nonparametric relationship between the dependent variable and the common factors is allowed. Thirdly, potential misspecification problems related to the type of CSD are avoided by assuming that CSD may arise because of unobserved common effects and/or spillover effects due to spatial or local dependencies. Finally, the most widely used spatial models are covered by the general specification (2.6). These spatial models differ in the range of dependence implied by their covariance matrices, but under certain invertibility conditions they can all be written as special cases of (2.6). More precisely, the SAR model is obtained by imposing the following form $\Phi^{1/2}(\theta_0) = (I_N - \theta_0 W_N)^{-1}$, whereas the SMA process is given by $\Phi^{1/2}(\theta_0) = (I_N + \theta_0 W_N)$, where W_N is defined in the spatial econometrics literature as an $N \times N$ spatial weight matrix. Throughout the paper, we shall make use rather common assumptions that are presented in the Appendix S1.

3 Estimation procedure and asymptotic properties

The main objective of this section is to provide an estimation procedure to consistently estimate (2.3)–(2.4) under two different scenarios. In the first case, heterogeneous slope parameters and unknown functions are assumed and the estimation of their corresponding means (i.e., $\overline{m}(\cdot)$ and $\overline{\beta}$, respectively) is of interest. In the second case, homogeneous slope parameters are assumed, i.e., $\beta_i = \beta$. Further, their main asymptotic properties are analyzed.

3.1 Heterogeneous case

In Appendix S2, it is proved that the augmented method proposed in Pesaran (2006) can be extended to a semiparametric setting when multifactor structures and spatial processes are allowed jointly. Thus, f_t can be well approximated by a linear function of observed variables such as $\overline{y}_{At}, \overline{x}_{At}$, and d_t and propose the following augmented regression model for (2.3)–(2.4):

$$y_{it} = \beta_i' x_{it} + m_i(z_t) + \delta_i' \lambda_t + e_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{3.1}$$

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

Following Fan and Huang (2005), for any given β_i and δ_i we can write

$$y_{it} - x_{it}'\beta_i - \lambda_t'\delta_i = m_i(z_t) + e_{it}, \tag{3.2}$$

so the previous semiparametric model becomes a fully nonparametric model where the local linear regression technique can be applied to estimate the smooth unknown function. Hence, for z_t in a small neighborhood of z, one can estimate the functions of interest by minimizing the following weighted local least-squares problem:

$$\sum_{t=1}^{T} [(y_{it} - x_{it}'\beta_i - \lambda_t'\delta_i) - m_i(z) - (z_t - z)' D_{m_i}(z)]^2 K_H(z_t - z), \tag{3.3}$$

where H is a $q \times q$ matrix that is symmetric and positive definite and each $K(\cdot)$ is a nonnegative product kernel function such that for each u it holds that $K_H(u) = |H|^{-1} \prod_{\ell=1}^{q} k(H^{-1}u_\ell), \quad u =$

$(u_1, \ldots, u_q)'$, where $k(\cdot)$ is a univariate kernel function. Let $K_H(z) = diag\{K_H(z_z - z), \ldots, K_H(z_T - z)\}$ be a $T \times T$ matrix and $Z_z = [Z'_{z_1}, \ldots, Z'_{z_T}]'$ a $T \times (q+1)$ matrix, where $Z_{z_t} = [1, (z_t - z)]$. Assuming that $Z'_z K_H(z) Z_z$ is nonsingular, the solution to (3.3) for $m_i(\cdot)$ is

$$\widehat{m}_i(z, H) = \iota'_1 (Z'_z K_H(z) Z_z)^{-1} Z'_z K_H(z)(Y_{i\cdot} - X_{i\cdot}\beta_i - \Lambda\delta_i), \tag{3.4}$$

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

However, this estimator is infeasible since it depends on the unknown terms (β_i, δ_i). To overcome this, we use (3.4) to obtain a closed-form solution for these parametric estimators, plugging those estimators into (3.4). Let $m_i(Z) = (m_i(z_1), \ldots, m_i(z_T))'$ be a $T \times 1$ vector of the smooth unknown function. The corresponding estimator is of the form

$$\widehat{m}_i(Z, H) = \begin{pmatrix} \iota'_1 (Z'_z K_H(z_1) Z_z)^{-1} Z'_z K_H(z_1) \\ \vdots \\ \iota'_1 (Z'_z K_H(z_T) Z_z)^{-1} Z'_z K_H(z_T) \end{pmatrix} (Y_{i\cdot} - X_{i\cdot}\beta_i - \Lambda\delta_i) = S(Y_{i\cdot} - X_{i\cdot}\beta_i - \Lambda\delta_i), \tag{3.5}$$

where S is a smoothing matrix that depends only on the observations of z_t, $t = 1, \ldots, T$. Writing (3.2) in vectorial form and substituting (3.5) in (3.2), we get

$$\widehat{Y}_{i\cdot} = \widehat{X}_{i\cdot}\beta_i + \widehat{\Lambda}\delta_i + \widehat{e}_{i\cdot}, \tag{3.6}$$

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

In order to estimate β_i consistently, we use the idea of partitioned regression and define the projection matrix $M_{\widehat{\Lambda}} = I_T - \widehat{\Lambda}(\widehat{\Lambda}'\widehat{\Lambda})^{-1}\widehat{\Lambda}'$. Premultiplying both sides of (3.6) by $M_{\widehat{\Lambda}}$ and applying least squares to the resulting model, we get

$$\widehat{\beta}_{SCCE,i} = \left(\widehat{X}_{i\cdot}' M_{\widehat{\Lambda}} \widehat{X}_{i\cdot}\right)^{-1} \widehat{X}_{i\cdot}' M_{\widehat{\Lambda}} \widehat{Y}_{i\cdot}. \tag{3.7}$$

Similarly, a consistent estimator for δ_i is required, so we define the projection matrix $M_{\widehat{X}_i} = I_T - \widehat{X}_{i\cdot}(\widehat{X}_{i\cdot}'\widehat{X}_{i\cdot})^{-1}\widehat{X}_{i\cdot}'$. Then, premultiplying both sides of (3.6) yields

$$\widehat{\delta}_{SCCE,i} = \left(\widehat{\Lambda}' M_{\widehat{X}_i} \widehat{\Lambda}\right)^{-1} \widehat{\Lambda}' M_{\widehat{X}_i} \widehat{Y}_{i\cdot}, \tag{3.8}$$

and using (3.7)–(3.8) in (3.4), the feasible nonparametric estimator for $m_i(\cdot)$ is of the form

$$\widehat{m}_{CCE,i}(z,H) = \iota_1'(Z_z' K_H(z) Z_z)^{-1} Z_z' K_H(z)(Y_{i\cdot} - X_{i\cdot}\widehat{\beta}_i - \Lambda\widehat{\delta}_i). \tag{3.9}$$

Finally, if the parameters of interest are the cross-sectional means of β_i and $m_i(\cdot)$, we follow Pesaran and Smith (1995) and propose a SCCE mean group estimator (SCCEMG) for $\overline{\beta}$, which is a simple average of the individual SCCE estimators of β_i:

$$\widehat{\beta}_{SCCEMG} = \frac{1}{N}\sum_{i=1}^{N} \widehat{\beta}_{SCCE,i}, \tag{3.10}$$

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

$$\widehat{m}_{CCEMG}(z,H) = \frac{1}{N}\sum_{i=1}^{N} \widehat{m}_{CCE,i}(z,H). \tag{3.11}$$

Alternatively, we can make a generalization of the pooled estimator proposed in Pesaran (2006), obtaining the following semiparametric CCE pooled estimator for $\overline{\beta}$:

$$\widehat{\beta}_{SCCEP} = \left(\sum_{i=1}^{N} \widehat{X}_{i\cdot}' M_{\widehat{\Lambda}} \widehat{X}_{i\cdot}\right)^{-1} \sum_{i=1}^{N} \widehat{X}_{i\cdot}' M_{\widehat{\Lambda}} \widehat{Y}_{i\cdot}. \tag{3.12}$$

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

$$\widehat{m}_{CCEP}(z,H) = \iota_1'(Z_z'K_H(z)Z_z)^{-1}Z_z'K_H(z)\left[\overline{Y}_{A\cdot} - N^{-1}\sum_{i=1}^{N} X_{i\cdot}\widehat{\beta}_{SCCE,i} - N^{-1}\sum_{i=1}^{N}\Lambda\widehat{\delta}_{SCCE,i}\right]. \tag{3.13}$$

Note that a relevant feature of the above estimator is that given the fact that $\overline{m}(\cdot)$ is a nonparametric function of time-varying stochastic regressors, the pooled and mean group nonparametric estimators are exactly the same, i.e., $\widehat{m}_{CCEMG}(z,H) \equiv \widehat{m}_{CCEP}(z,H)$. Furthermore, before considering the main asymptotic properties of the proposed estimators, we first establish the following lemmas that will be key to determining the impact of the weak or strong CSD on the rate of convergence of the above estimators.

Lemma 3.1 *Let $\overline{\varepsilon}_{At}$ be a composed error term defined as in $\overline{\varepsilon}_{At} = N^{-1}\sum_{i=1}^{N}\varepsilon_{it}$, where $\varepsilon_{it} = ((\epsilon_{it} + v_{it}'\beta_i), v_{it})'$. Suppose that either $\|\beta_i\| < C$ or that the random coefficient assumption, Assumption $S1.4$ in the Appendix S1, holds. Under Assumption 2.1 and Assumption $S1.1$ in the supplementary material, for each t we have*

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

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

See Pesaran and Tosetti (2011) for a detailed proof of this lemma. For any process of form (2.6) this lemma guarantees $\overline{\varepsilon}_{At} \overset{q.m.}{\to} 0$, as $N \to \infty$, and the degree of cross-sectional dependence of $\varepsilon_{i\cdot}$ will be bounded by $\nu_N = N^{-2}\imath_N'\Phi\imath_N$, where $\imath_N$ is an $N \times 1$ vector of ones. If $\nu_N = O(N^{-1})$ is assumed, we get $Var(\overline{\varepsilon}_{At}) = O_p(N^{-1})$ which is analogous to the common weak dependence assumption in time series (see Chudik et al., 2011). For its part, boundedness Φ implies $\nu_N = O(1)$, so $Var(\overline{\varepsilon}_{At}) = O(1)$ and we are allowing "long-range cross-sectional dependence" (see Robinson, 2012).

Replacing (2.3)–(2.4) in (3.7) and rearranging terms, we obtain

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

since $\widehat{X}_{i\cdot}' M_{\widehat{\Lambda}} m_i(Z) = O_p(c_H^2)$ (see the proof of Theorem 3.1). In the above expression, the direct dependence of $\widehat{\beta}_{SCCE,i}$ on the unobserved factors F is observed, but in the following theorem it is shown that the T-distribution of this estimator will not depend on the factor loadings as $N \to \infty$, whereas ϵ_{it} is independently distributed from (x_{it}, z_t, d_t, f_t). Furthermore, to derive the asymptotic distribution of the above estimators, the following notation is required: $\widetilde{X}_{i\cdot} = X_{i\cdot} - \mathcal{B}_X(z)$, $\widetilde{\Lambda} = \Lambda - \mathcal{B}_\Lambda(z)$, $\widetilde{F} = F - \mathcal{B}_F(z)$, and $\widetilde{D} = D - \mathcal{B}_D(z)$, where $\mathcal{B}_X(z) = E(X_{i\cdot}|z_t = z)\rho_{z_t}(z)$, $\mathcal{B}_\Lambda(z) = E(\Lambda|z_t = z)\rho_{z_t}(z)$, $\mathcal{B}_F(z) = E[F|z_t = z]\rho_{z_t}(z)$, and $\mathcal{B}_D(z) = E[D|z_t = z]\rho_{z_t}(z)$ 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. The proof of the following theorems are given in the Appendix S7.

Theorem 3.1 *(*SCCE estimators*) Consider the panel data model (2.3) and (2.4), and suppose that Assumption 2.1 and Assumptions S*1.1*–S*1.6*(a), and S*1.7*–S*1.11 *in the Appendix S1 hold. As N and T tend to infinity (in no particular order),* $\widehat{\beta}_i$ *is a consistent estimator of* $\beta_{SCCE,i}$*. If it is further assumed that* $\sqrt{T}/N \to 0$ *as* $(N,T) \xrightarrow{j} \infty$,

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

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

In Theorem 3.1, it is shown that the finite T-distribution of $\widehat{\beta}_{SCCE,i}$ will be asymptotically normal if the rank condition of Assumption S1.5 holds and if N and T are of the same order of magnitude, namely, if $T/N \to C$, where C is a positive finite constant. Furthermore, similar asymptotic results can be obtained whether the rank condition is violated, but that is beyond the scope of this paper. See Pesaran and Tosetti (2011) for several examples of this.

Furthermore, consistent estimator of the asymptotic variance of $\widehat{\beta}_{SCCE,i}$ can be obtained using the Newey and West (1987)-type procedure, for example, and is given by

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

where $\widehat{\Omega}_{\epsilon_i} = T^{-1}\widehat{\epsilon}_{i\cdot}'\widehat{\epsilon}_{i\cdot}$ and $\widehat{\epsilon}_{it} = \widehat{Y}_{it} - \widehat{X}_{it}'\widehat{\beta}_i$.

Theorem 3.2 *(*Nonparametric SCCE estimators*) Consider the panel data model presented in (2.3) and (2.4) and suppose that Assumption 2.1 and Assumptions* $S1.1$–$S1.4$ *and* $S1.7$–$S1.11$ *in the Appendix S1 hold. Then,* $\widehat{m}_{CCE,i}(\cdot)$ *is a consistent estimator of* $m_i(\cdot)$*. If it is further assumed that* $\sqrt{T|H|}tr\{H^2\} = O(1)$*, as* $T \to \infty$*,*

$$\sqrt{T|H|}\left(\widehat{m}_{CCE,i}(z,H) - m_i(z) - \frac{1}{2}\mu_2^q(K)tr\{H^2\mathcal{H}_{m_i}(z)\}\right) \quad \stackrel{d}{\longrightarrow} \quad N\left(0, \frac{\sigma_{\eta_i}^2 R^q(K)}{\rho_{z_t}(z)}\right).$$

Theorem 3.2 reveals that rather standard results for local linear estimation are obtained as $T \to \infty$. In particular, $\widehat{m}_{CCE,i}(z,H)$ is consistent and asymptotically normally distributed with a rate of convergence of $\sqrt{T|H|}$, regardless of the rank condition assumption.

Theorem 3.3 *(*SCCEMG estimators*). Consider the panel data model presented in (2.3) and (2.4) and suppose that Assumption 2.1 and Assumptions* $S1.1$–$S1.6$*(b), and* $S1.7$–$S1.11$ *in the Appendix S1 hold. If it is further assumed that* $\sqrt{N}c_H^2 \to 0$*, as* $(N,T) \stackrel{j}{\longrightarrow} \infty$ *then*

$$\sqrt{N}(\widehat{\beta}_{SCCEMG} - \overline{\beta}) \quad \stackrel{d}{\longrightarrow} \quad N\left(0, \Omega_{\xi}\right).$$

Theorem 3.4 *(*SCCEP estimators*). Consider the panel data model presented in (2.3) and (2.4) and suppose that Assumption 2.1 and Assumptions* $S1.1$–$S1.6$*(c), and* $S1.7$–$S1.11$ *in the Appendix S1 hold. If it is further assumed that* $\sqrt{N}c_H^2 \to 0$*, as* $(N,T) \stackrel{j}{\longrightarrow} \infty$ *then*

$$\sqrt{N}(\widehat{\beta}_{SCCEP} - \overline{\beta}) \quad \stackrel{d}{\longrightarrow} \quad N\left(0, \Psi^{*-1}R^*\Psi^{*-1}\right),$$

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

The consistency problem related to the presence of unobserved common factors has been solved again. However, several results should be pointed out. On the one hand, standard spatial techniques are not required to deal with the efficiency issue related to the spatially correlated errors. As expected, efficiency gains from pooling observations over the cross-sectional units are achieved since the time-invariant variability of β_i in Assumption S1.4 dominates the other sources of randomness in the model (see Pesaran (2006) for further details), and these results are valid for any type of CSD (i.e., strong or weak dependence). On the other hand, the rate of convergence of $\widehat{\beta}_{SCCEMG}$ and $\widehat{\beta}_{SCCEP}$ is premultiplied by $\sqrt{N}$, rather than the usual $\sqrt{NT}$, if the rank condition in Assumption S1.5 in the Appendix S1 is satisfied. Similar results are obtained if this condition does not hold.

Furthermore, a consistent estimator of the asymptotic variance of the SCCEMG estimator is

$$\widehat{Asy.Var}(\widehat{\beta}_{SCCEMG}) = \frac{1}{N(N-1)}\sum_{i=1}^{N}(\widehat{\beta}_{SCCCE,i} - \widehat{\beta}_{SCCEMG})(\widehat{\beta}_{SCCCE,i} - \widehat{\beta}_{SCCEMG})',$$

whereas the corresponding estimator of the asymptotic variance of the SCCEP estimator is

$$\widehat{Asy.Var}(\widehat{\beta}_{SCCEP}) = \frac{1}{N}\Psi_{NT}^{-1}R_{NT}\Psi_{NT}^{-1},$$

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

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

A very nice feature of the above nonparametric variance estimators is that they avoid any misspecification problem related to the spatial weights matrix since they do not require a priori knowledge of the spatial weights matrix.

Theorem 3.5 *(*Nonparametric CCEMG and CCEP estimators*). Consider the panel data model presented in (2.3) and (2.4) and suppose that Assumption 2.1 and Assumptions* $S1.1$–$S1.4$ *and*

S1.7–S1.11 in the Appendix S1 hold. Given the $\sqrt{N}$*-consistency of* $\widehat{\beta}_{SCCEMG}$ *and* $\widehat{\delta}_{SCCEMG}$*, where* $\widehat{\delta}_{SCCEMG} = N^{-1}\sum_{i=1}^{N}\widehat{\delta}_i$*, the mean group estimator for* $\overline{m}(\cdot)$ *is consistent as* N *and* T *tend to infinity (in no particular order). If it is further assumed that* $\sqrt{NT|H|}tr\{H^2\} = O(1)$*, as* $(N,T) \xrightarrow{j} \infty$*,*

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

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

The proposed estimation technique developed for $\overline{m}(\cdot)$ enables solving the consistency problem again, but a relevant result is obtained in terms of variance. Unlike the parametric estimates, the variance of this nonparametric estimator exhibits a new element, i.e., ν_N, which reflects the strengthening of the spatial correlation and depends directly on the particular specification of Φ. Therefore, efficiency gains from pooling observations over the cross-section units are not achieved and more efficient estimators could be obtained by taking into account the spatial correlation. Note that this result is also obtained in Robinson (2012); Lee and Robinson (2015), and Soberon et al. (2021) in a different framework and two-step estimation procedures are proposed in order to obtain efficiency. Finally, the rate of convergence of this estimator depends on the rate of increase of ν_N, if any, as was noted in Lemma 3.1. If weak dependence is assumed, the rate of convergence is $\sqrt{NT|H|}$, whereas we get $\sqrt{T|H|}$ if strong dependence is allowed.

Finally, the optimal bandwidth order is $O_p(T^{-1/(4+q)})$, but we can be tempted to assume that the optimal bandwidth is $O_p((NT)^{-1/(4+q)})$. However, we are smoothing only in z_t over time series observations, so if we use the latter bandwidth we may be suggested a smaller bandwidth, especially when N is large. See Phillips and Wang (2021) for further details.

3.2 Homogeneous case

Considering now a restricted submodel of (2.3) in which homogeneous slopes are assumed, i.e., $\beta_i = \beta$, the homogeneous SCCE pooled (HSCCEP) and mean group (HSCCEMG) estimators proposed for β are exactly the same as the SCCEP and SCCEMG estimators in (3.10) and (3.12), respectively (i.e., $\widehat{\beta}_{HSCCEP} \equiv \widehat{\beta}_{SCCEP}$ and $\widehat{\beta}_{HSCCEMG} \equiv \widehat{\beta}_{SCCEMG}$), whereas the corresponding estimator for $m(\cdot)$ is of the form

$$\widetilde{m}_{HCCE}(z, H) = \iota_1'(Z_z' K_H(z) Z_z)^{-1} Z_z' K_H(z)[\overline{Y}_{A\cdot} - \overline{X}_{A\cdot}\widehat{\beta}_{HSCCE} - \Lambda\widehat{\delta}_{SCCEMG}], \quad (3.14)$$

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

The following lemmas can be proved following a similar reasoning as for Theorems 3.3–3.5. For the sake of brevity, the specific proofs have been omitted but are available upon request.

Theorem 3.6 *(*HSCCEMG estimators*). Consider the panel data model presented in (2.3) and (2.4) and suppose that Assumption 2.1 and Assumptions* $S1.1$*–*$S1.2$*,* $S1.5$*–*$S1.6$*(b), and* $S1.7$*–*$S1.11$ *in the Appendix S1 hold. Then,* $\widehat{\beta}_{HSCCEMG}$ *is a consistent estimator of* β*. If it is further assumed that* $\sqrt{T}c_H^2 \to 0$*, as* $(N,T) \xrightarrow{j} \infty$ *then*

$$\sqrt{T}\nu_N^{-1/2}(\widehat{\beta}_{HSCCEMG} - \beta) \quad \xrightarrow{d} \quad N\left(0, \overline{\sigma}_\eta^2 \Psi^{*-1}\right),$$

where $\Psi^* = \lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\Sigma_{v_i}\right)$.

Theorem 3.7 *(*HSCCEP estimators*). Consider the panel data model presented in (2.3) and (2.4) and suppose that Assumption 2.1 and Assumptions* $S1.1$*–*$S1.2$*,* $S1.5$*–*$S1.6$*(c), and* $S1.7$*–*$S1.11$ *in the Appendix S1 hold. Then,* $\widehat{\beta}_{HSCCEP}$ *is a consistent estimator of* β*. If it is further assumed that* $\sqrt{T}c_H^2 \to 0$*, as* $(N,T) \xrightarrow{j} \infty$ *then*

$$\sqrt{T}\nu_N^{-1/2}(\widehat{\beta}_{HSCCEP} - \beta) \quad \xrightarrow{d} \quad N\left(0, \overline{\sigma}_\eta^2 \Psi^{*-1}\right),$$

where $\Psi^* = \lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\Sigma_{v_i}\right)$.

Theorem 3.8 *(*Nonparametric HCCE estimators*). Consider the panel data model presented in (2.3) and (2.4) and suppose that Assumption 2.1 and Assumptions S*1.1*–S*1.2 *and S*1.7*–S*1.11 *in the Appendix S1 hold. Given the* $\sqrt{T}\nu_N^{-1/2}$*-consistency of* $\widehat{\beta}_{HSCCE}$ *and* $\widehat{\delta}_{SCCEMG}$ *and since* $\widehat{m}_{HCCE}(z, H)$ *is a consistent estimator of* $m(\cdot)$ *as N and T tend to infinity (in no particular order), if it is further assumed that* $\sqrt{T|H|}\nu_N^{-1/2} tr\{H^2\} = O(1)$*, as N and T tend to infinity*

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

From the results in these theorems, we get that $\widehat{\beta}_{HSCCEMG}$, $\widehat{\beta}_{HSCCEP}$, and $\widehat{m}_{HCCE}(z, H)$ are consistent estimators as N and T are sufficiently large. Nevertheless, the rate of convergence of these estimators depends on the rate of increase of ν_N, if any, as was noted in Lemma 3.1. On the one hand, the rate of convergence of the parametric estimators is $\sqrt{NT}$ and $\sqrt{T}$ under weak and strong CSD, respectively. On the other hand, the convergence of $\widehat{\beta}_{HSCCEP}$ is of order $(NT|H|)^{-1/2}$ under weak CSD and $(T|H|)^{-1}$ under strong CSD. Furthermore, the asymptotic properties for $\widehat{\beta}_{HSCCEMG}$ and $\widehat{\beta}_{HSCCEP}$ are exactly the same. In addition, the robust variance estimators proposed for $\widehat{\beta}_{SCCEP}$ and $\widehat{\beta}_{SCCEMG}$ in the heterogeneous case are still valid for the homogeneous specification since cross-sectional independence has been assumed for η_{it} in Assumption 2.1. However, robust estimates of these asymptotic variances could be obtained considering a generalized version of the Newey–West procedure that allows for spatial effects, but this is beyond the scope of this paper.

4 Hypothesis tests

In this section two simple consistent tests are proposed, combining the methodology of conditional moment tests and nonparametric estimation techniques. Furthermore, the asymptotic properties of both test statistics are analyzed using nondegenerate U-statistics theories.

4.1 Constancy test

We are interested in testing the null hypothesis (H_0^a) that the true model is fully parametric, versus the alternative hypothesis (H_1^a) that it is a semiparametric specification. Thus, we consider:

$$H_0^a : Pr[m(z_t) = \phi(z_t, \pi_0)] = 1 \quad \text{for some} \quad \pi_0 \in I\!R^q,$$

$$H_1^a : Pr[m(z_t) = \phi(z_t, \pi)] < 1 \quad \text{for some} \quad \pi \in I\!R^q,$$

where $\phi(\cdot)$ is a known function with ϕ being the $q \times 1$ ($q \geq 1$) unknown parameter.

Denote $\ddot{\epsilon}_{a,i\cdot} = M_\Lambda(Y_{i\cdot} - X_{i\cdot}\beta - \phi(Z, \pi))$ as a T-dimensional vector and $\rho_z(\cdot)$ as the density function of z_t. Under H_0^a, we have $E[\ddot{\epsilon}_{a,it} E(\ddot{\epsilon}_{a,it}|z_t)\rho_z(z_t)] = 0$ since $E(\ddot{\epsilon}_{a,it}|z_t) = E(\ddot{\epsilon}_{it}|z_t) + O_p\left(\sqrt{\frac{T}{N}}\right) = 0$, assuming $T/N \to 0$ as $(N, T) \to \infty$. Similarly, under H_1^a, $E[\ddot{\epsilon}_{a,it} E(\ddot{\epsilon}_{a,it}|z_t)\rho_z(z_t)] = E\{[E(\ddot{\epsilon}_{a,it}|z_t)]^2 \rho_z(z_t)\} > 0$ since $E(\ddot{\epsilon}_{a,i\cdot}|z_t) = E(M_\Lambda|z_t)(m(Z) - \phi(Z; \pi)) + o_p(1)$. Therefore, we can use the sample analogue of $E[\ddot{\epsilon}_{a,it} E(\ddot{\epsilon}_{a,it}|z_t)\rho_z(z_t)]$ to form a test, where the use of $\rho_z(\cdot)$ as a weight function avoids the usual random denominator problem (see Powell et al. (1989)).

The sample moment condition can be estimated by various nonparametric methods. However, when constructing the test statistic we opt to use the local constant least-squares estimator (i.e., Nadaraya–Watson) since the local linear estimator proposed previously complicates the asymptotic analysis of the test statistic (see Lin et al., 2014). Thus, the proposed test statistic is of the form

$$\widehat{I}^a = \frac{1}{NT|H_z|} \sum_{i=1}^{N} \sum_{j=1}^{N} \widehat{\ddot{\epsilon}}'_{a,i\cdot} K(Z, Z) \widehat{\ddot{\epsilon}}_{a,i\cdot}, \tag{4.1}$$

where $K(Z, Z)$ is a $T \times T$ matrix whose (t, s)th elements are $K(H_z^{-1}(z_t - z_s))$, and H_z is a $q \times q$ matrix of bandwidth. Note that to obtain a feasible test statistic and avoid further technical assumptions that would be required for the test statistic to converge to zero under the alternative (see Li and Wang (1998) for more details), we use the parametric residuals from the "mixed" regressions of the form $\widehat{\ddot{\epsilon}}_{a,i\cdot} = M_\Lambda \widehat{\epsilon}_{a,i\cdot}$, where $\widehat{\epsilon}_{a,it} = y_{it} - x'_{it}\widehat{\beta}_{SP} - \phi(z_t, \widehat{\pi})$, $\widehat{\beta}_{SP}$ is any semiparametric $\sqrt{NT}$-

consistent estimator of β based on the alternative model (see (Robinson, 1988) for example), and $\widehat{\pi}$ is the nonlinear least-squares estimator of π based on the null model. Furthermore, to avoid the asymptotically non-negligible center term of this double summation test, we remove the $(it) = (jt)$ elements in (4.1). Then, the diagonal elements of $K(\cdot)$ are all zeros and the final form of our test statistic is given by

$$\widehat{I}^a = \frac{1}{NT|H_z|} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \sum_{t=1}^{T} \sum_{s \neq t}^{T} \widehat{\ddot{\epsilon}}_{a,it} \widehat{\ddot{\epsilon}}_{a,js} K(z_t, z_s). \tag{4.2}$$

The proposed test statistic is of the form of a U-statistic, independently of the fact that the kernel function only depends on time series observations. Hence, we intend to employ the theories developed in Hall (1984) and de Jong (1987) for U-statistics in order to derive the asymptotic distribution of $\widehat{I}_n^a$. However, these CLTs cannot be applied directly due to the existence of CSD generated by the presence of unobserved common factors, f_t. Nevertheless, Assumption S1.13 in Appendix S1 enables us to overcome this drawback since it allows obtaining independence among different cross-sectional units conditional on both observable and unobservable common factors (see Cai et al., 2020).

The asymptotic null distribution of the test statistic and the consistency of the test under the alternative are given in the following two theorems.

Theorem 4.1 *Suppose that Assumptions S1.1-S1.3, S1.5, S1.8, and S1.12-S1.16 in the Appendix S1 hold. Under the null hypothesis H_0^a, we have*

$$J^a = \frac{NT|H_z|^{1/2} \widehat{I}_n^a}{\sqrt{\widehat{\Sigma}_a}} \quad \xrightarrow{d} \quad N(0,1), \tag{4.3}$$

where $\widehat{\Sigma}_a = \frac{2}{N^2T^2|H_z|} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \widehat{\ddot{\epsilon}}^2_{a,it} \widehat{\ddot{\epsilon}}^2_{a,js} K^2(z_t, z_s)$ is a consistent estimator of the asymptotic variance of $NT|H_z|^{1/2}\widehat{I}^a$, i.e., $\Sigma_a = 2\overline{\sigma}_\epsilon^4 E[\rho_z(z_t)] \int K^2(u)du$.

Theorem 4.2 *Suppose that Assumptions S1.1-S1.3, S1.5–S1.8, and S1.12-S1.16 in the Appendix*

S1 hold. Under the alternative hypothesis H_1^a, we have $Pr(J^a \geq M^a) \to 1$, as $(N,T) \xrightarrow{j} \infty$, where M^a is any non-stochastic, positive sequence such that $M^a = o(NT|H_z|^{1/2})$.

Theorem 4.1 shows that J^a has an asymptotic standard normal distribution under H_0^a at the rate $NT|H_z|^{1/2}$. Furthermore, this test statistic is user-friendly and easily computed since the estimation of the asymptotic variance Σ_a involves the use of the transformed residuals $\widehat{\ddot{\epsilon}}_{a,it}$, which does not require information on the unobserved factors. Theorem 4.2 means that the proposed test is a one-sided test. The null hypothesis H_0^a is rejected at a given significance level if J^a is greater than the corresponding critical value. In addition, J^a is a consistent test since the probability that the J^a test rejects H_0^a approaches one as $(N,T) \xrightarrow{j} \infty$ when H_0^a is false. Nevertheless, kernel-based nonparametric tests usually suffer substantial finite sample size distortions and converge very slowly to the asymptotic distributions. Then, a bootstrap procedure may be proposed to better approximate the finite sample null distribution of the J^a test. See Appendix S3 for further details.

4.2 Poolability test

We now consider a test for homogeneous versus heterogenenous slope parameters, i.e., does $\beta_i = \beta_j$ almost everywhere (a.e.) for all i and j? Hence, the null and alternative hypotheses are

$$H_0^b : \beta_i = \beta \quad \text{for some} \quad \beta_0 \in I\!R^d \quad \text{and all i;}$$
$$H_1^b : \beta_i \neq \beta_j \quad \text{for any} \quad \beta \in I\!R^d \quad \text{and some} \quad i \neq j.$$

Denote $\ddot{\epsilon}_{b,i\cdot} = (\ddot{\epsilon}_{b,i1}, \ldots, \ddot{\epsilon}_{b,iT})' = M_\Lambda(Y_{i\cdot} - X_{i\cdot}\beta - m_i(Z))$ as a T-dimensional vector and $\rho_{x,z}(x_{it}, z_t)$ as the joint p.d.f. of (x_{it}, z_t). The idea of the poolability test is as follows. Under H_0^b, it can be shown that $E[\ddot{\epsilon}_{b,it}E(\ddot{\epsilon}_{b,it}|x_{it}, z_t)\rho_{x,z}(x_{it}, z_t)] = 0$ a.e., while under H_1^b, $E[\ddot{\epsilon}_{b,it}E(\ddot{\epsilon}_{b,it}|x_{it}, z_t)\rho_{x,z}(x_{it}, z_t)] > 0$ a.e. Then, a consistent test for poolability can be constructed based on $E[\ddot{\epsilon}_{b,it}E(\ddot{\epsilon}_{b,it}|x_{it}, z_t)\rho_{x,z}(x_{it}, z_t)]$. However, $E(\ddot{\epsilon}_{b,it}|x_{it}, z_t)$ has to be estimated by an appropriate kernel estimator using some semi-parametric residuals to get a feasible test statistic and to overcome the random denominator problem in the nonparametric estimators. We opt to estimate a density-weighted version of $E[\ddot{u}_{b,it}E(\ddot{u}_{b,it}|x_{it},z_t)\rho_{x,z}(x_{it},z_t)]$ that is given by $E[\ddot{u}_{b,it}\rho_z(z_t)E(\ddot{u}_{b,it}\rho_z(z_t)|x_{it},z_t)\rho_{x,z}(x_{it},z_t)]$.

Let $\widehat{\ddot{\epsilon}}_{b,i\cdot} = M_\Lambda \widehat{\epsilon}_{b,i\cdot}$, where $\widehat{\epsilon}_{b,i\cdot} = Y_{i\cdot} - X_{i\cdot}\widehat{\beta}_{HSCCEP} - \widehat{m}_{CCE,i}(Z,H)$, the proposed test statistic is

$$\widehat{I}^b = \frac{1}{N^2T^2|H_x||H_z|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\widehat{\ddot{\epsilon}}_{b,it}\widehat{\ddot{\epsilon}}_{b,js}\widehat{\rho}_{z_t}\widehat{\rho}_{z_s}K(z_t,z_s)K(x_{it},x_{js}), \tag{4.4}$$

where $K(x_{it},x_{js}) = K(H_x^{-1}(x_{it}-x_{js}))$ is a non-negative product kernel function of the univariate kernel functions $k_x(\cdot)$. Note that in the above expression, the $(it) = (js)$ term has been dropped to remove a non-zero center term from $\widehat{I}^b$ under H_0^b. In addition, $\widehat{\rho}_{z_t}$ and $\widehat{\rho}_{z_s}$ are the estimators of the density functions $\rho_z(z_t)$ and $\rho_z(z_s)$, respectively, of the form $\widehat{\rho}_z(z_t) = \frac{1}{(T-1)|\mathcal{H}|}\sum_{s\neq t}^{T}K(\mathcal{H}^{-1}(z_t - z_s))$, where $\mathcal{H}$ is a $q \times q$ matrix of smoothing parameters that is symmetric and positive definite. Note that this estimator is common across sections and no further information is available in the cross-sectional observations to raise efficiency. The only requirement for the kernel density estimate $\widehat{\rho}_z(z_t)$ to converge to the true density $\rho_z(z_t)$ is $T|\mathcal{H}| \to \infty$ as $T \to \infty$.

The above test statistic is of the form of a degenerate U-statistic like the previous one, but we need additional conditions for the bandwidth parameter and density functions as the collected in Assumption S1.17 in Appendix S1. Following a similar reasoning as for the constancy test, in the following theorems it is shown that $\widehat{I}^b$ is asymptotically distributed standard normal under the null hypothesis H_0^b, while diverging to infinity at a rate arbitrarily close to the sample size under the alternative H_1^b. With this aim, we define $\psi(z) = E(x_{it}|z_t = z)$ and assume $\psi \in \mathcal{G}_\mu^4$.

Theorem 4.3 *Suppose that Assumptions S1.1-S1.3, S1.5, S1.8, S1.12-S1.14, and S1.17 in Appendix S1 hold. Under the null hypothesis H_0^b, we have*

$$J^b = \frac{NT|H_x|^{1/2}|H_z|^{1/2}\widehat{I}^b}{\sqrt{\widehat{\Sigma}_b}} \quad \xrightarrow{d} \quad N(0,1), \tag{4.5}$$

where $\widehat{\Sigma}_b = \frac{2}{N^2T^2|H_x||H_z|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\widehat{\ddot{\epsilon}}_{b,it}^2\widehat{\ddot{\epsilon}}_{b,js}^2\widehat{\rho}_{z_t}^2\widehat{\rho}_{z_s}^2K^2(x_{it},x_{js})K^2(z_t,z_s)$ *is a consistent es-*

timator of the asymptotic variance of $NT|H_x|^{1/2}|H_z|^{1/2}\widehat{I}^b$*, i.e.,*

$$\Sigma_b = 2\overline{\sigma}_\epsilon^4 \int K^2(u_1)du_1 \int K^2(u_2)du_2 E[\rho_z^4(z_t)\rho_{x,z}(x_{1t}, z_t)].$$

Theorem 4.4 *Suppose that Assumptions S1.1-S1.3, S1.5, S1.8, S1.12-S1.14, and S1.17 in Appendix S1 hold. Under the alternative hypothesis* H_1^b*, we have* $Pr(J^b \geq M^b) \to 1, \quad as \quad (N,T) \xrightarrow{j} \infty$*, where* M^b *is any non-stochastic, positive sequence such that* $M^b = o(NT|H_x|^{1/2}|H_z|^{1/2})$.

From Theorems 4.3–4.4, similar conclusions to those from Theorems 4.1–4.2 are obtained. Finally, a bootstrap method can be proposed to approximate the finite sample null distribution of J^b by extending the procedure proposed in Appendix S3 for this particular case and using the semiparametric residuals $\widehat{\epsilon}_{b,it} = y_{it} - x'_{it}\widehat{\beta}_{HSCCEP} - \widehat{m}_{CCE,i}(z_t, H)$ instead of $\widehat{\epsilon}_{a,it}$.

5 Empirical illustration

In this section, we provide an illustration of the usefulness of the proposed approach by estimating some variants of (2.2) using a new annual country-level balanced panel dataset covering 24 OECD countries from 1971 to 2014 (see Appendix S5). We also perform model selection by exploiting the above-discussed constancy and poolability tests and focus attention on the bias that may arise when erroneously estimating more constrained and misspecified models.

The estimation results are presented in Table 1. In column (i), we estimate the most constrained model similar to Eberhardt et al. (2013), i.e. the fully parametric CCE model with homogeneous slope parameters, $\beta_{1i} = \beta_1$, $\beta_{2i} = \beta_2$, $\beta_{3i} = \beta_3$, and without including the oil price index which has been built following Hamilton (1996) and Davis and Haltiwanger (2001). Then, in column (ii) we add a linear and homogeneous effect by imposing $m_i(z_t) = \phi z_t$ in (2.2). In column (iii), we allow for a possible nonlinear effect of the price of oil by considering a second-order polynomial function, $m_i(z_t) = \phi_1 z_t + \phi_2 z_t^2$. In column (iv), we consider the homogeneous pooled semiparametric estimator (HSCCEP), which relaxes the parametric specification for the observed common factor but maintains the homogeneity assumption, $m_i(z_t) = m(z_t)$, $\beta_{1i} = \beta_1$, $\beta_{2i} = \beta_2$, $\beta_{3i} = \beta_3$.

As far as the parametric CCE model (i) is concerned, the estimated output elasticities with respect to labor, capital, and R&D stock are 0.56, 0.29, and 0.03, respectively. These estimates are extremely similar to those obtained by Eberhardt et al. (2013) and Millo (2019). The inclusion of the oil price index (columns (ii)–(iii)) does not at all affect the estimated technological parameters. Also note that the estimated coefficient of the oil price index (ii) is very low in magnitude and is not significant, and that when the oil price index enters into the model with a second-order polynomial function (iii) the results do not improve as the two parameters associated with the oil price index are still non-significant.

Table 1: Fully parametric and semiparametric results

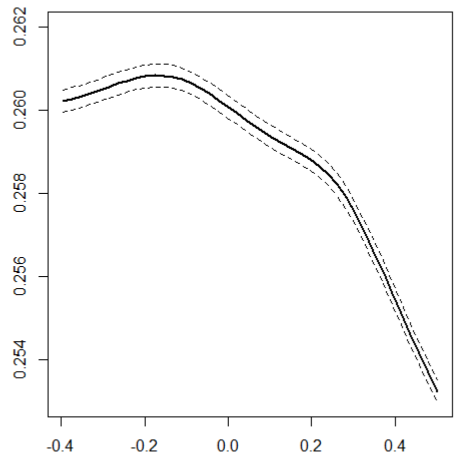
	CCEP			HSCCEP
$\ln L_{it}$	0.556***	0.556***	0.556***	0.620***
	(17.692)	(17.674)	(17.683)	(5.075)
$\ln K_{it}$	0.294***	0.294***	0.294***	0.266***
	(11.578)	(11.567)	(11.572)	(3.088)
$\ln R_{it}$	0.030***	0.030***	0.030***	0.055***
	(3.058)	(3.055)	(3.057)	(2.909)
oil		-9.694e-11	$-9.177e-11$	
		(-1.369e-08)	(-1.297e-08)	
oil^2			2.681e-10	
			(1.447e-08)	
SP test		0.346**	0.346**	
		(0.039)	(0.039)	
Pooling test				0.923
				(0.606)

Notes.
Estimators: CCEP: pooled estimator (Pesaran, 2006); HSCCEP: homogeneous semiparametric CCEP estimator. The t-values within brackets for the CCEP are constructed from White heteroskedasticity-robust standard errors.
Significance at the ***1% and **5% level.

Overall, these results are rather unsatisfactory as they suggest the existence of decreasing returns to scale, with an estimated elasticity of scale equal to 0.88, which may be implausibly low. More specifically, the non-significance of the oil price index as well as the very low coefficient of R&D stock contradict a huge amount of empirical literature.

We then perform the constancy test, and to address the finite sample size distortion we adopt the bootstrap procedure described in Appendix S3. The results of the test indicate a rejection of

Figure 1: Estimated output-oil price index relation in a semiparametric knowledge production function framework: nonparametric HCCE estimator, $\widehat{m}_{HCCE}(z, H)$



Notes: The dotted lines indicate the 95% pointwise confidence interval.

the fully parametric models (ii) and (iii) at a 5% significance level, in favor of the semiparametric alternative pooled model. Allowing for a nonparametric function $m(\cdot)$ may be important to avoid misspecification bias not only with respect to the estimated effect of the observed common factors—here, the oil price index—but also with respect to the estimation of the parameters of interest—here, the technological parameters in a Griliches-type production function. We find that this bias is empirically sizeable, as when estimating the semiparametric HSCCEP model in (iv) the results change substantially. Indeed, the returns to scale increase to 0.94 and the estimated output elasticity with respect to R&D increases to 0.055. This result appears to be consistent with the huge amount of literature surveyed by Hall et al. (2010). Moreover, and very importantly, as far the effect of oil price is concerned Figure 1 depicts the estimated functional relation, which is clearly nonlinear and confirms some previous complementary time series evidence of an asymmetric effect of oil prices on GDP (Hamilton, 2003).

Finally, we focus on the issue of poolability. Assuming homogeneity in the slope parameters and in the unknown functions could be questioned, as a theoretical foundation for heterogeneous slope parameters across countries can be found in the "new growth" literature, which argues that technology differs across countries (Brock and Durlauf, 2001; Durlauf et al., 2001). Moreover, the heterogeneous effect of the oil price may be the result of country-specific technological constraints. Consequently, in principle it may be relevant to consider estimators that are built under the assumption of heterogeneity both in the slope parameters, β_{1i}, β_{2i}, β_{3i}, and in the unknown function $m_i(z_t)$. We therefore perform the proposed poolability test, which clearly does not reject the null hypothesis, thus favoring the homogeneous specification (iv). This result is interesting and complements previous studies suggesting that despite the fact that heterogeneous estimators can be suitable in principle, according to both economic theory and to avoid an heterogeneity bias, homogeneous estimators can often be preferable in practice and can be more reliable because of their efficiency gains (see Baltagi et al. (2002, 2003)).

6 Conclusion

This paper proposes a new semiparametric heterogeneous panel data model to simultaneously handle complex and relevant empirical problems, namely (i) functional misspecification, by modelling stochastic observed common factors with a nonparametric function instead of assuming the usual parametric form; (ii) CSD arising simultaneously from common factors and spatial dependence, neither imposing a specific parametric spatial diffusion process nor requiring the specification of a given interaction matrix for the latter but, rather, it being directly derived from the data; iii) heterogeneous relations.

Asymptotic theory for alternative semiparametric estimators and specification testing procedures are provided, as well as Monte Carlo experiments and an empirical illustration. For the latter, we revisit the knowledge capital production function *à la* Griliches, for which the proposed constancy and poolability tests go toward the adoption of a semiparametric pooled specification. The empirical results highlight that relaxing the parametric specification for the observed factor (here, an oil price index) greatly affects the estimates—indicating a sizeable bias when estimating a fully parametric model—and ultimately, we obtain results that appear to be much more consistent with respect to economic theory.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

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

Antonioli, D., G. Gioldasis, and A. Musolesi (2021). Estimating a non-neutral production function. *Oxford Economic Papers 73*(2), 856–878.

Bai, J. (2003). Testing parametric conditional distributions of dynamic models. *Review of Economics and Statistics 85*, 531–549.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*(4), 1229–1279.

Bai, J. and S. Ng (2001). A consistent test for conditional symmetry in time-series models. *Journal of Econometrics 103*, 225–258.

Bailey, N., S. Holly, and M. H. Pesaran (2016a). A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. *Journal of Applied Econometrics 31*(1), 249–280.

Bailey, N., S. Holly, and M. H. Pesaran (2016b). A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. *Journal of Applied Econometrics 31*(1), 249–280.

Baltagi, B. H., G. Bresson, J. M. Griffin, and A. Pirotte (2003). Homogeneous, heterogeneous or shrinkage estimators? Some empirical evidence from french regional gasoline consumption. *Empirical Economics 28*(4), 795–811.

Baltagi, B. H., G. Bresson, and A. Pirotte (2002). Comparison of forecast performance for homogeneous, heterogeneous and shrinkage estimators: Some empirical evidence from us electricity and natural-gas consumption. *Economics Letters 76*(3), 375–382.

Brock, W. A. and S. N. Durlauf (2001). What have we learned from a decade of empirical research on growth? Growth empirics and reality. *The World Bank Economic Review 15*(2), 229–272.

Cai, Z., Y. Fang, and Q. Xu (2020). Testing capital asset pricing models using functional-coefficient panel data models with cross-sectional dependence. *Journal of Econometrics (in press)*.

Calderón, C., E. Moral-Benito, and L. Servén (2015). Is infrastructure capital productive? A dynamic heterogeneous approach. *Journal of Applied Econometrics 30*(2), 177–198.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal 14*(1), C45–C90.

Davis, S. J. and J. Haltiwanger (2001). Sectoral job creation and destruction responses to oil price changes. *Journal of Monetary Economics 48*(3), 465–512.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields 75*, 261–277.

Debarsy, N., C. Ertur, and J. P. LeSage (2012). Interpreting dynamic space–time panel data models. *Statistical Methodology 9*(1-2), 158–171.

Dong, C., J. Gao, and B. Peng (2021). Varying-coefficient panel data models with nonstationarity and partially observed factor structure. *Journal of Business & Economic Statistics 39*, 700–711.

Doraszelski, U. and J. Jaumandreu (2013). R&d and productivity: Estimating endogenous productivity. *Review of Economic Studies 80*(4), 1338–1383.

Durlauf, S. N., A. Kourtellos, and A. Minkin (2001). The local solow growth model. *European Economic Review 45*(4-6), 928–940.

Eberhardt, M., C. Helmers, and H. Strauss (2013). Do spillovers matter when estimating private returns to R&D? *Review of Economics and Statistics 95*(2), 436–448.

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.

Gao, J., K. Xia, and H. Zhu (2020). Heterogeneous panel data models with cross-sectional dependence. *Journal of Econometrics 219*, 329–353.

Glass, A. J., K. Kenjegalieva, and R. C. Sickles (2016). A spatial autoregressive stochastic frontier model for panel data with asymmetric efficiency spillovers. *Journal of Econometrics 190*(2), 289–300.

Griliches, Z. (1979). Issues in Assessing the Contribution of Research and Development to Productivity Growth. *The Bell Journal of Economics 10*(1), 92–116.

Hall, B. H., J. Mairesse, and P. Mohnen (2010). Chapter 24 - measuring the returns to R&D. In B. H. Hall and N. Rosenberg (Eds.), *Handbook of the Economics of Innovation, Volume 2*, Volume 2 of *Handbook of the Economics of Innovation*, pp. 1033–1082. North-Holland.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis 14*, 1–16.

Hamilton, J. D. (1996). This is what happened to the oil price-macroeconomy relationship. *Journal of Monetary Economics 38*(2), 215–220.

Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics 113*(2), 363–398.

Holly, S., M. Pesaran, and T. Yamagata (2010). A spatio-temporal model of house prices in the USA. *Journal of Econometrics 158*, 160–173.

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

Ivaldi, M., N. Ladoux, H. Ossard, and M. Simioni (1996). Comparing Fourier and translog specifications of multiproduct technology: Evidence from an incomplete panel of french farmers. *Journal of Applied Econometrics 11*(6), 649–667.

Keane, M. P. and E. S. Prasad (1996). The employment and wage effects of oil price changes: a sectoral analysis. *The Review of Economics and Statistics 78*(3), 389–400.

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review 99*(3), 1053–69.

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

Lee, L.-f. and J. Yu (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics 154*(2), 165–185.

Lescaroux, F. and V. Mignon (2008). On the influence of oil prices on economic activity and other macroeconomic and financial variables. *OPEC Energy Review 32*(4), 343–380.

Li, Q. and S. Wang (1998). A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics 87*, 145–165.

Lin, Z., Q. Li, and Y. Sun (2014). A consistent nonparametric test of parametric regression functional form in fixed effects panel data models. *Journal of Econometrics 178*, 167–179.

Ma, S., J. S. Racine, and A. Ullah (2020). Nonparametric estimation of marginal effects in regression-spline random effects models. *Econometric Reviews 39*(8), 792–825.

Millo, G. (2019). Private returns to R&D in the presence of spillovers, revisited. *Journal of Applied Econometrics 34*(1), 155–159.

Mork, K. A. (1989). Oil and the macroeconomy when prices go up and down: an extension of Hamilton's results. *Journal of Political Economy 97*(3), 740–744.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–08.

Parent, O. and J. P. LeSage (2008). Using the variance structure of the conditional autoregressive spatial specification to model knowledge spillovers. *Journal of Applied Econometrics 23*(2), 235–256.

Parmeter, C. and J. Racine (2018). Nonparametric estimation and inference for panel data models. In M. Tsionas (Ed.), *Panel Data Econometrics*, pp. 97–129. Academic Press.

Pesaran, M. and T. Yamagata (2008). Testing slope homogeneity in large panels. *Journal of Applied Econometrics 142*, 50–93.

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

Pesaran, M. H. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics 68*(1), 79–113.

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

Phillips, P. and D. Sul (2003). Dynamic panel estimation and homogeneity testing under cross section dependence. *The Econometrics Journal 6*, 217–259.

Phillips, P. and Y. Wang (2021). Functional coefficient panel modeling with communal smoothing covariates. *Journal of Econometrics*.

Powell, J. L., J. H. Stock, and T. M. Stoker (1989). Semiparametric estimation of index coefficients. *Econometrica 57*(6), 1403–1430.

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

Robinson, P. M. (2011). Asymptotic theory for nonparametric regression with spatial data. *Journal of Econometrics 165*(1), 5–19.

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

Rodriguez-Poo, J. M. and A. Soberon (2017). Nonparametric and semiparametric panel data models: Recent developments. *Journal of Economic Surveys 31*(4), 923–960.

Sarafidis, V. (2011). GMM estimation of short dynamic panel data models with error cross-sectional dependence. Munich Personal RePEc Archive.

Sarafidis, V. and T. Wansbeek (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews 31*(5), 483–531.

Shi, W. and L.-f. Lee (2017). Spatial dynamic panel data models with interactive fixed effects. *Journal of Econometrics 197*(2), 323–347.

Soberon, A., J. M. Rodriguez-Poo, and P. M. Robinson (2021). Nonparametric panel data regression with parametric cross-sectional dependence. *The Econometrics Journal. eprint*.

Su, I. and Q. Chen (2013). Testing heterogeneity in panel data models with interactive fixed effects. *Econometric Theory 29*(6), 1079–1135.

Su, J. and S. Jin (2012). Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics 169*, 34–47.

Xu, Q., Z. Cai, and Y. Fang (2016). Panel data models with cross-sectional dependence: a selective review. *Applied Mathematircs-A Journal of Chinese Universities 31*, 127–147.

Online Supplementary Appendices to "A semiparametric panel data model with common factors and spatial dependence"

Alexandra Soberon, Antonio Musolesi and Juan M. Rodriguez-Poo

28th December 2021

The supplementary file includes 7 sections: Appendix S1 presents the technical assumptions required to obtain the main asymptotic properties of the estimators and tests statistics discussed in Sections 3 and 4. Appendix S2 justifies the extension to the CCE approach from a fully parametric model to a semiparametric setting. Appendix S3 presents an alternative bootstrap procedure to better approximate the finite sample null distribution of the test statistics. Appendix S4 defines the data and variables used in the empirical application. Appendix S5 includes the Monte Carlo simulation results. Appendix S6 presents preliminary lemmas and the proofs of the corresponding lemmas. Appendix S7 states the corresponding mathematical proofs to obtain the asymptotic properties of the proposed estimators and test statistics.

S1 TECHNICAL ASSUMPTIONS

Assumption S1.1 (Individual-specific errors) *The individual-specific errors ϵ_{it} and v_{js} are distributed independently for all i, j, t, and s. For each i, v_{it} follows a linear stationary process with absolute summable autocovariances given by $v_{it} = \sum_{s=0}^{\infty} S_{is}\varsigma_{i,t-s}$, where for each i, ς_{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_{is} satisfy the condition $Var(v_{it}) = \sum_{s=0}^{\infty} S_{is}S'_{is} = \Sigma_{v_i}$, where Σ_{v_i} is a $p \times p$ positive definite matrix such that $\sup_i \|\Sigma_{v_i}\| < C$.*

Assumption S1.2 (Common effects) *The* $(n+r+q)\times 1$ *vector of common effects* $(d_t', f_t', z_t')'$ *is covariance stationary with absolute summable autocovariances, distributed independently of the individual-specific errors,* ϵ_{is} *and* v_{is}*, for all* i*,* t*, and* s*.*

Assumption S1.3 (Factor loadings) *The unobserved factor loadings* γ_i *and* Γ_i *are independently and identically distributed across* i*, and of the individual-specific errors* (v_{jt}, ϵ_{jt}) *and the common factors* (d_t, z_t, f_t) *for all* i *and* t*, with fixed means* γ *and* Γ*, respectively, and finite variances. In particular,*

$$\gamma_i = \gamma + \zeta_i, \quad \zeta_i \sim i.i.d.(0, \Omega_\zeta), \quad for \quad i = 1, 2, \ldots, N,$$

where Ω_ζ *is an* $r \times r$ *symmetric nonnegative definite matrix. In addition,* $\|\gamma\| < C$*,* $\|\Gamma\| < C$*, and* $\|\Omega_\zeta\| < C$ *for some positive constant* $C < \infty$*.*

Assumption S1.4 (Random slope coefficients) *The slope coefficients* β_i *follow the random coefficient model*

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

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

Assumption S1.5 (Rank condition) *Let* $\Gamma^* = E(\gamma_i, \Gamma_i) = (\gamma, \Gamma)$*,* $Rank(\Gamma^*) = r \leq (p+1)$*.*

Assumption S1.6 (Identification of β_i, $\overline{\beta}$, and β) *Consider the covariates contained in* $\lambda_t = (\overline{y}_{At}, \overline{x}_{At}, d_t)$ *and let* $M_{\widetilde{G}}$ *be defined as in (3.14). Then, the following conditions hold:*

(a) The $p \times p$ *matrices* $T^{-1}\widetilde{X}_{i\cdot}' M_{\widehat{\Lambda}} \widetilde{X}_{i\cdot}$ *and* $T^{-1}\widetilde{X}_{i\cdot} M_{\widetilde{G}} \widetilde{X}_{i\cdot}$ *exist and are nonsingular for all* i*. In addition, their corresponding inverse matrices have finite second-order moments for all* i*.*

(b) The matrix $\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\Sigma_{v_i}$ *exists and is nonsingular.*

(c) There exists T_0 *and* N_0 *such that for all* $T \geq T_0$ *and* $N \geq N_0$, $(T^{-1}\widetilde{X}_{i\cdot}' M_{\widehat{\Lambda}} \widetilde{X}_{i\cdot})$ *and* $(T^{-1}\widetilde{X}_{i\cdot}' M_{\widetilde{G}} \widetilde{X}_{i\cdot})$ *exist and are nonsingular for all* i. *In addition,* $\sup_i E\left\|\frac{\widetilde{X}_{i\cdot}' M_{\widetilde{G}} \widetilde{X}_{i\cdot}}{T}\right\| < C < \infty$.

Assumption S1.7 (Density function) *Let* $\{z_t\}_{t=1,\ldots,T}$ *be a set of i.i.d.* $I\!R^q$*-random variables, where* z_t *has a probability density function (p.d.f.) that satisfies* $0 < \rho_{z_t}(\cdot) < \infty$ *and is twice continuously differentiable in all its arguments with bounded second-order derivatives in a neighborhood of* $z \in int(\mathcal{Z})$.

Assumption S1.8 (Smoothness condition) *The map* $m(\cdot) : \mathcal{Z} \to I\!R$ *is Borel-measurable and twice continuously differentiable at* z *in the interior of* $\mathcal{Z}$ *with bounded derivatives.*

Assumption S1.9 (Kernel function) $K(u) = \prod_{l=1}^{q} k(u_l)$ *is a product kernel, and the univariate kernel function* $k(\cdot)$ *is compactly supported and bounded such that* $\int k(u)du = 1$, $\int uu'k(u)du = \mu_2(K)I_q$, *and* $\int k^2(u)du = R(K)$, *where* $\mu_2(K) \neq 0$ *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 S1.10 (Bandwidth) *Let* $c_H = tr\{H^2\} + \{logT/T|H|\}^{1/2}$. *The bandwidth matrix* H *is symmetric and positive definite, where each element of* H *tends to zero. As* $(N,T) \xrightarrow{j} \infty$, $\sqrt{N}c_H^2 \to 0$, $NT|H| \to \infty$, *and* $T|H| \to \infty$.

Assumption S1.11 (Lyapounov) *For some* $\varsigma > 0$, $E[|\epsilon_{it}|^{(2+\varsigma)}]$ *exists and is bounded.*

Assumptions S1.1–S1.2 impose covariance stationarity and moment conditions on common factors and error terms, as well some independence assumptions that are commonly used in the literature; see Pesaran (2006). Nevertheless, Assumption S1.2 can be relaxed to allow for unit roots in the common factors, along the lines set out in Kapetanios et al. (2011). Furthermore, Assumptions S1.3–S1.4 are quite standard in the common factor literature when the objective is the estimation of the means of the unknown parameters. Assumption S1.5 is the rank condition required by the CCE approach to show $\bar{\varepsilon}_{At} \overset{q.m.}{\to} 0$, but a similar result can be obtained using more complicated mathematical proofs, as noted in (Pesaran, 2006). Assumption S1.6 allows for the individual-specific regressors, x_{it}, to be random and correlated with the common factors. In addition, Assumptions S1.7–S1.8 are standard smoothness and boundedness conditions on the density and unknown functions. Condition S1.7 is a rather strong assumption in a context of time series variables, but we use it to simplify the mathematical proof. However, similar results are obtained under weaker dependence conditions such as the mixing processes. Furthermore, in several empirical studies this stronger condition may hold when some indexes based on first differences variables are used. Assumptions S1.9–S1.10 are kernel and bandwidth conditions quite common in the local linear literature. Note that the kernel function with a compact support in S1.9 is imposed for the sake of brevity of the proofs and can be removed at the cost of lengthy proofs. Finally, Assumption S1.11 is required for the central limit theorem (CLT).

In order to derive the main asymptotic distribution of the test statistic in Section 4, we use the definitions given in Robinson (1988) for the class of kernel functions $\mathcal{K}_l$ (where l is an integer) and the class of functions $\mathcal{G}_\mu^\alpha$, $\alpha > 0$, and $\mu > 0$. Note that the class of functions of $\mathcal{K}_l$ belongs to the second-order kernel and the l subscript denotes the tail property of the kernel functions $k \in \mathcal{K}_l$, not the order of kernels. In addition, the following regularity conditions are necessary:

Assumption S1.12 (Kernel and density functions) *The p.d.f. $\rho_z \in \mathcal{G}_\mu^\infty$ for some $\mu > 0$ and $k \in \mathcal{K}_2$.*

Assumption S1.13 (Conditional expectation) *Let $\mathcal{D} = \{(z_t, d_t, f_t)\}_{t=1}^T$ and $\lambda_{t,\kappa}$ be the κth element of λ_t. For $1 \leq \kappa_1$ and $\kappa_2 \leq \ell$, $E_\mathcal{D}|\lambda_{t_1,\kappa_1}\lambda_{t_2,\kappa_2}|^8 < \infty$, where $E_\mathcal{D}(\cdot)$ is the conditional expectation on $\mathcal{D}$.*

Assumption S1.14 (Individual-specific errors) *Let $\epsilon_{i\cdot} = (\epsilon_{1t}, \ldots, \epsilon_{Nt})'$, where $\{\epsilon_{i\cdot}\}_{i=1}^N$ are i.i.d. across i. For each i, ϵ_{it} follows a linear stationary process with absolute summable auto-*

covariances given by $\epsilon_{it} = \sum_{\tau=0}^{\infty} a_{i\tau}\vartheta_{i,t-\tau}$*, where* ϑ_{it} *are i.i.d. random variables with zero mean, unit variance, and finite fourth-order cumulants. In particular,* $Var(\epsilon_{it}) = \sigma^2_{\epsilon,i} \leq \overline{\sigma}^2_{\epsilon} < \infty$*, for all* i *and some constant* $\overline{\sigma}^2_{\epsilon}$*, where* $\sigma^2_{\epsilon,i} > 0$*.*

Assumption S1.15 (Identification of π) *Consider the covariates contained in* $\lambda_t = (\overline{y}_{At}, \overline{x}_{At}, d_t)$ *and let* $M_G = I_T - G(G'G)^{-1}G'$*, where* $G = (D, F)$ *is a* $T \times (n+r)$ *matrix. There exists* T_0 *and* N_0 *such that for all* $T \geq T_0$ *and* $N \geq N_0$*,* $(T^{-1}X'_{i\cdot}M_{\Lambda}X_{i\cdot})$ *and* $(T^{-1}X'_{i\cdot}M_G X_{i\cdot})$ *exist and are nonsingular for all* i*. In addition,* $\sup_i E\left\|\frac{X'_{i\cdot}M_G X_{i\cdot}}{T}\right\| < C < \infty$*.*

Assumption S1.16 (Bandwidth) *The bandwidth matrix* H_z *is symmetric and positive definite, where each element of* H_z *tends to zero. Moreover, as* $(N,T) \xrightarrow{j} \infty$*,* $T/N \to 0$ *and* $NT|H_z| \to \infty$*.*

Assumption S1.17 (Bandwidth) *(i) The bandwidth matrices* H_z*,* H_x*, and* $\mathcal{H}$ *are symmetric and positive definite, where each element of* H_z*,* H_x*, and* $\mathcal{H}$ *tends to zero. In addition, as* $(N,T) \xrightarrow{j} \infty$*,* $T/N \to 0$*,* $NT|H_x||H_z| \to \infty$*; (ii) furthermore,* $NTtr\{\mathcal{H}^2\}|H_x|^{1/2}|H_z|^{1/2} \to 0$ *and* $|H_x||H_z||\mathcal{H}|^{-2} \to 0$ *as* $(N,T) \xrightarrow{j} \infty$*; (iii)* $\rho_{z,x} \in \mathcal{G}^{\infty}_1$ *and* $k_x \in \mathcal{K}_2$*.*

At this point, let us make some remarks on Assumption S1.17. The conditions in (i) ensure that the kernel estimators involved are consistent. In addition, the conditions in (ii) are introduced to ensure that the limiting distribution of $NT|H_x|^{1/2}|H_z|^{1/2}\widehat{I}^b$ is centered correctly at zero under H^b_0. In other words, they ensure that the asymptotic mean squared error of the kernel estimator $\widehat{\widehat{\epsilon}}_{b,it}\widehat{\rho}_{z_t}$ is of a smaller order than $(NT|H_x|^{1/2}|H_z|^{1/2})^{-1}$, i.e., $(tr\{\mathcal{H}\}^2 + (NT|\mathcal{H}|)^{-1}) = o_p((NT|H_x|^{1/2}|H_z|^{1/2})^{-1})$ (see Fan and Li (1996) for more details). Finally, condition (iii) contains some smoothness assumptions about the joint p.d.f. $\rho_{z,x}(\cdot,\cdot)$.

S2 CCE MOTIVATION

Before proceeding to the analysis of the main asymptotic properties of the proposed estimators, we check whether the augmented method proposed in Pesaran (2006) can be extended to a semi-parametric setting when a multifactor structure and spatial processes are allowed jointly.

Let $\overline{y}_{At} \equiv N^{-1}\sum_{i=1}^{N} y_{it}$, $\overline{x}_{At} \equiv N^{-1}\sum_{i=1}^{N} x_{it}$, $\overline{\varepsilon}_{At} \equiv N^{-1}\sum_{i=1}^{N} \varepsilon_{it}$ be the corresponding cross-sectional means, where $\varepsilon_{it} = ((\epsilon_{it} + v_{it}'\beta_i), v_{it})'$. Applying cross-section averages on (2.3)-(2.5) implies

$$\begin{pmatrix} \overline{y}_{At} \\ \overline{x}_{At} \end{pmatrix} = \overline{B}_A' d_t + \overline{\mathcal{M}}(z_t) + \overline{C}_A' f_t + \overline{\varepsilon}_{At}, \tag{S2.1}$$

where $\overline{B}_A$, $\overline{C}_A$, and $\overline{\mathcal{M}}(z_t)$ are the sample averages over i of

$$B_i \equiv (\alpha_i, A_i)\begin{pmatrix} 1 & 0_p \\ \beta_i & I_p \end{pmatrix}, \quad C_i \equiv (\gamma_i, \Gamma_i)\begin{pmatrix} 1 & 0_p \\ \beta_i & I_p \end{pmatrix}, \quad \text{and} \quad \mathcal{M}_i(z_t) \equiv (m_i(z_t) \quad g_i'(z_t))\begin{pmatrix} 1 & 0_p \\ \beta_i & I_p \end{pmatrix},$$

respectively, I_p is an identity matrix or order p, and 0_p is a matrix of zeros of order p.

Premultiplying both sides of (S2.1) by $\overline{C}_A$ and solving for f_t, we get

$$f_t = \left(\overline{C}_A \overline{C}_A'\right)^{-1} \overline{C}_A \left[\begin{pmatrix} \overline{y}_{At} \\ \overline{x}_{At} \end{pmatrix} - \overline{B}_A' d_t - \overline{\mathcal{M}}(z_t) - \overline{\varepsilon}_{At}\right], \tag{S2.2}$$

provided that

$$Rank(\overline{C}_A) = r \le (p+1) \quad \text{for sufficiently large N.} \tag{S2.3}$$

As $N \to \infty$, $\overline{\mathcal{M}}(z_t) \xrightarrow{p} 0$ and $\overline{\varepsilon}_{At} \xrightarrow{q.m.} 0$ for each t under rather weak conditions, given that $N^{-1}\sum_i m_i(z_t) \xrightarrow{p} 0$, $N^{-1}\sum_i g_i(z_t) \xrightarrow{p} 0$, and $N^{-1}\sum_i \beta_i' g_i(z_t) \xrightarrow{p} 0$. Then, for sufficiently large N,

$$f_t - (\overline{C}_A \overline{C}_A')^{-1} \overline{C}_A \left[\begin{pmatrix} \overline{y}_{At} \\ \overline{x}_{At} \end{pmatrix} - \overline{B}_A' d_t\right] \xrightarrow{p} 0, \quad \text{as} \quad N \to \infty. \tag{S2.4}$$

This result suggests that the unobserved common factors f_t can be well approximated by a linear function of observed variables such as $\overline{y}_{At}, \overline{x}_{At}$, and d_t, so they can be used as observable proxies for f_t.

S3 A bootstrap procedure: constancy test

An alternative bootstrap procedure can be proposed based on the "fixed regressor bootstrap" given by Hansen (2000). We propose generating a bootstrapping sample using the observable covariates (x_{it}, z_t, d_t) and the residuals that are obtained using the principal component analysis (PCA) method to estimate the unobserved factors f_t and factor loadings γ_i. For the sake of simplicity, we only consider the case where $d_t = 1$, so the detailed bootstrap procedure that we propose is as follows:

i) Estimate the panel data model under H_0^a by adopting the estimation procedure proposed in Section 3 for the homogeneous slope case and obtain the residuals $\widehat{\epsilon}_{a,it} = y_{it} - x_{it}'\widehat{\beta}_{SP} - \phi(z_t, \widehat{\pi})$ and $\widehat{\ddot{\epsilon}}_{a,it} = M_\Lambda \widehat{\epsilon}_{a,i\cdot}$ for all i and t. Calculate J^a using (4.2)–(4.4).

ii) Estimate the unobservable common factors f_t and factor loadings γ_i using the PCA method. Denoting $\widehat{f}_t$ and $\widehat{\gamma}_i$ as the corresponding estimators for f_t and γ_i, respectively, estimate α_i using $\widehat{\alpha}_i = T^{-1}\sum_{t=1}^{T}(y_{it} - x_{it}'\widehat{\beta}_{SP} - \phi(z_t, \widehat{\pi}) - \widehat{\gamma}_i'\widehat{f}_t)$ for all i. Then, obtain the residuals $\check{\epsilon}_{a,it} = \widehat{\epsilon}_{a,it} - \widehat{\alpha}_i - \widehat{\gamma}_i'\widehat{f}_t$.

iii) Compute the wild-bootstrap errors from $\{\check{\epsilon}_{a,it}\}_{i=1,\ldots,N;t=1,\ldots,T}$ by $\check{\epsilon}_{a,it}^* = (\check{\epsilon}_{a,it} - \bar{\check{\epsilon}}_{a,i\cdot})\eta_{it}^*$, where η_{it}^* is generated from $IIDN(0,1)$ for all i and t, and $\bar{\check{\epsilon}}_{a,i\cdot} = T^{-1}\sum_{i=1}^{N}\check{\epsilon}_{a,it}$. Then, calculate y_{it}^* via $y_{it}^* = \widehat{\alpha}_i + x_{it}'\widehat{\beta}_{SP} + \phi(z_t, \widehat{\pi}) + \widehat{\gamma}_i'\widehat{f}_t + \check{\epsilon}_{a,it}^*$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. Call $\{(x_{it}, z_t, y_{it}^*)\}_{i=1,\ldots,N,t=1,\ldots,T}$ the bootstrap sample.

iv) Using the bootstrap sample, calculate the corresponding estimators $\widehat{\beta}_{SP}^*$ and $\widehat{\pi}^*$, which are the same as those obtained previously, but with y_{it}^* instead of y_{it}. Then, calculate the bootstrap residuals as $\widehat{\ddot{\epsilon}}_{a,i\cdot}^* = (\widehat{\ddot{\epsilon}}_{a,i1}^*, \ldots, \widehat{\ddot{\epsilon}}_{a,iT}^*)' = M_\Lambda \widehat{\epsilon}_{a,i\cdot}^*$ for all i, where $\widehat{\epsilon}_{a,it}^* = y_{it}^* - x_{it}'\widehat{\beta}_{SP}^* - \phi(z_t, \widehat{\pi}^*)$.

v) Calculate the bootstrap test statistic J^{a*}, where J^{a*} is calculated in the same way as J_n^a except that y_{it} is replaced by y_{it}^*. In other words, $J^{a*} = NT|H_z|^{1/2}\widehat{I}^{a*}/\sqrt{\widehat{\Sigma}_a^*}$, where

$$\widehat{I}^{a*} = \frac{1}{N^2T^2|H_z|}\sum_{i=1}^N\sum_{j\neq i}^N\sum_{t=1}^T\sum_{s=1}^T \widehat{\ddot{\epsilon}}^*_{a,it}\widehat{\ddot{\epsilon}}^*_{a,js}K(z_t,z_s) \quad \text{and} \quad \widehat{\Sigma}_a^* = \frac{1}{N^2T^2|H_z|}\sum_{i=1}^N\sum_{j\neq i}^N\sum_{t=1}^T\sum_{s=1}^T \widehat{\ddot{\epsilon}}^{*2}_{a,it}\widehat{\ddot{\epsilon}}^{*2}_{a,js}K^2(z_t,z_s).$$

vi) Repeat steps (i)–(v) for B times to get B bootstrapping test statistics $\{J_\tau^{a*}\}_{\tau=1}^B$. Calculate the bootstrapping p-value p^* via $p^* = B^{*-1}\sum_{\tau=1}^B \mathbb{1}(J_\tau^{a*} \geq J^a)$. Then, reject the null hypothesis of homogeneous slope parameters if the resulting p-value is smaller than the corresponding level of significance.

S4 MONTE CARLO EXPERIMENT

The purpose of this section is to illustrate the performance of the proposed estimators and test statistics in finite samples using simulated data. Firstly, we analyse the finite sample behavior of the parametric and nonparametric estimators proposed previously and later we assess the size and power performance of the proposed test statistics for given sample sizes.

Finite sample performance of estimators

With the aim of investigating the extend to which the proposed estimators capture the effects of various forms of cross-sectional dependence, we consider two alternative sets of experiments that involve different hypotheses on the data generating process (DGP). In the first DGP, we consider a semiparametric model with heterogeneous slope parameters, whereas in the second DGP homogeneous slope parameters are assumed.

For both experiments, we consider the following DGP

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

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

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

a) The factor loadings of the observed common factors are generated as $A_i' \sim IIDN(0.5\imath_2, 0.5I_2)$, where $\imath_2 = (1,1)'$ and I_2 is a 2×2 identity matrix, and $\alpha_i \sim IIDN(1,1)$, for $i = 1, \ldots, N$. As Pesaran and Tosetti (2011), α_i and A_i do not change across replications and $d_t = 1$.

b) The unobserved common factors are generated as independent stationary AR(1) processes with zero means and variances 1. More precisely, $s = 1, 2$, $f_{s,t} = 0.5 f_{s,(t-1)} + (1 - 0.5^2)^{1/2} \xi_{s,t}$, where $\xi_{s,t} \sim IIDN(0,1)$ across t, for $t = -49, \ldots, 0, 1, \ldots, T$.

c) The factor loadings $(\gamma_{1i}, \gamma_{2i})$ of the unobserved common factors in the y_{it} equation are generated as $\gamma_{1i} \sim IIDN(0,1)$ and $\gamma_{2i} \sim IIDN(0,1)$. Also, for the factor loadings of the unobserved common factors in the x_{it} equation we consider two different cases for $\Gamma_i = (\Gamma_{1i}, \Gamma_{2i})$ that we denote by A and B, respectively: $vec(\Gamma_i) = (\Gamma_{11,i}, \Gamma_{12,i}, \Gamma_{21,i}, \Gamma_{22,i})' \sim IIDN(\Gamma_\tau, I_4)$, $\quad \tau = A, B$. In case A, $\Gamma_A = (1,0,0,1)'$, so the rank condition in Assumption 3.1 is satisfied, whereas in Case B, $\Gamma_B = (1,1,0,0)'$, and the rank condition is not satisfied.

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

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

Note that the first 50 observations of v_{1it}, v_{2it}, f_{1t}, and f_{2t} are discarded. Further, two alternative assumptions on the slope coefficients are considered. In particular, heterogeneous slopes are assumed in DGP1 with $\beta_{s,i} = \beta_s + \psi_{s,i}$ where $\beta_s = 1$ and $\psi_{s,i} \sim IIDN(0, 0.04)$, for $i = 1, 2, \ldots, N$ and $s = 1, 2$ varying across replications, while homogeneous slope parameters are allowed in DGP2 with $\beta_{s,i} = 1$. Each experiment was replicated 1000 times for $N = 100, 140, 200$ and T to be either $(25, 50, 75)$. Also, the Epanechnikov kernel $k(u) = 0.75(1 - u^2)\mathbb{1}\{|u| \leq 1\}$ was used and we choose $H = I_q h_0$, where $h_0 = c_0 \widehat{\sigma}_z T^{-1/5}$ is the bandwidth term, $\widehat{\sigma}_z$ the sample standard deviation of the smoothing variable $\{z_t\}_{t=1,\ldots,T}$, and $c_0 = 2.34$.

For evaluation of the performance of our estimators, we use the bias and the root mean squared errors (RMSE) for the slope parameters, whereas the RMSE is computed for the regression functions. In what follows, we shall focus on β_1, since results for β_2 are very similar and will not be reported. Results for the full rank experiments (case A) and the rank deficient experiments (case B) are summarized in Tables 1-2 and Figures 1-2. More precisely, Table 1 and Figure 1 summarize the results for the heterogeneous slope case, while Table 2 and Figure 2 collect the results for the homogeneous slope setting.

Overall, the Monte Carlo results confirm the good performance of the proposed estimators in finite samples. More precisely, the SCCEMG and SCCEP estimators display very small biases and their RMSEs decline steadily with increases in N and/or T for the different experiments. Further, for the heterogeneous slope experiments the asymptotic efficiency of the SCCEMG estimators

Table 1: Small sample properties of the parametric estimators under slope heterogeneity

		Bias (x100)			RMSE (x100)			Bias (x100)			RMSE (x100)		
θ_0	$N\backslash T$	25	50	75	25	50	75	25	50	75	25	50	75
		Case A: Full rank						Case B: rank deficient					
CCEP													
0.3	100	0.057	0.040	0.012	2.771	2.619	1.944	0.129	0.035	0.003	2.588	2.656	1.921
	140	0.066	0.066	-0.018	2.108	1.617	1.381	0.027	0.045	0.013	1.896	1.459	1.228
	200	0.125	0.001	0.042	1.858	1.565	1.235	0.130	0.004	0.054	1.860	1.567	1.267
0.6	100	0.060	0.042	0.010	2.770	2.649	2.021	0.132	0.038	0.001	2.587	2.657	1.927
	140	0.067	0.069	-0.017	2.111	1.621	1.385	0.029	0.047	0.013	1.900	1.465	1.233
	200	0.127	0.002	0.043	1.858	1.565	1.236	0.081	-0.015	0.035	1.679	1.511	1.171
0.9	100	0.062	0.044	0.002	2.773	2.624	2.131	0.136	0.041	-0.07	2.591	2.660	2.103
	140	0.075	0.081	-0.012	2.174	1.692	1.470	0.043	0.060	0.013	1.962	1.548	1.339
	200	0.129	0.004	0.054	1.860	1.566	1.267	0.083	-0.013	0.040	1.679	1.512	1.211
CCEMG													
0.3	100	0.003	0.009	0.020	2.670	2.335	1.700	0.070	0.022	0.007	2.535	2.344	1.669
	140	0.080	0.067	-0.009	2.069	1.502	1.237	0.032	0.032	0.010	1.900	1.383	1.130
	200	0.092	0.000	0.023	1.788	1.434	1.099	0.096	0.003	0.032	1.790	1.436	1.136
0.6	100	0.004	0.011	0.018	2.667	2.335	1.703	0.071	0.025	0.005	2.533	2.344	1.674
	140	0.082	0.069	-0.009	2.070	1.504	1.241	0.034	0.033	0.010	1.903	1.386	1.135
	200	0.094	0.001	0.024	1.788	1.434	1.101	0.050	-0.007	0.002	1.670	1.360	1.045
0.9	100	0.006	0.014	0.008	2.669	2.338	1.865	0.072	0.029	-0.005	2.536	2.346	1.837
	140	0.091	0.075	-0.007	2.117	1.561	1.321	0.045	0.040	0.008	1.955	1.451	1.231
	200	0.100	0.003	0.032	1.790	1.436	1.136	0.053	-0.005	0.022	1.672	1.361	1.088

Table 2: Small sample properties of the parametric estimators under slope homogeneity

		Bias (x100)			RMSE (x100)			Bias (x100)			RMSE (x100)		
θ_0	$N\backslash T$	25	50	75	25	50	75	25	50	75	25	50	75
		Case A: Full rank						Case B: rank deficient					
CCEP													
0.3	100	0.053	0.022	0.064	3.098	2.878	2.104	-0.024	0.004	0.084	2.939	2.727	1.909
	140	-0.026	-0.041	0.139	2.417	1.758	1.551	0.035	-0.004	0.109	2.227	1.618	1.524
	200	0.053	0.022	0.064	2.069	1.704	1.378	0.038	0.036	0.042	1.751	1.527	1.283
0.6	100	0.054	0.023	0.069	3.093	2.879	2.112	-0.022	0.005	0.088	2.938	2.727	1.919
	140	-0.025	-0.039	0.139	2.417	1.759	1.551	0.038	-0.003	0.110	2.226	1.621	1.525
	200	0.067	0.046	0.062	2.326	2.098	1.992	0.038	0.036	0.042	1.751	1.527	1.283
0.9	100	0.055	0.024	0.127	3.092	2.882	2.314	-0.020	0.005	0.084	2.941	2.730	1.910
	140	-0.014	-0.034	0.150	2.473	1.836	1.643	0.057	-0.007	0.126	2.279	1.714	1.631
	200	0.069	0.045	0.065	2.078	1.705	1.409	0.040	0.034	0.046	1.758	1.529	1.300

relative to the SCCEP estimators is confirmed, although the differences between the two estimators are rather slight for relatively large samples. This general conclusion also holds on the rank deficient case.

Considering now the results in Figures 1-2 it can be noted that, as it was expected, in both experiments (i.e., homogeneous and heterogeneous slopes) the RMSEs of the CCEMG and HCCE nonparametric estimators shrink to zero as the sample size increases. Furthermore, all these results corroborate that both parametric and nonparametric estimates are robust to the presence of spatially correlated errors since their RMSEs are steady independently of the θ_0 value.

Figure 1: Boxplots of the RMSE values of the nonparametric estimates in 1000 independent simulations under slope heterogeneity.

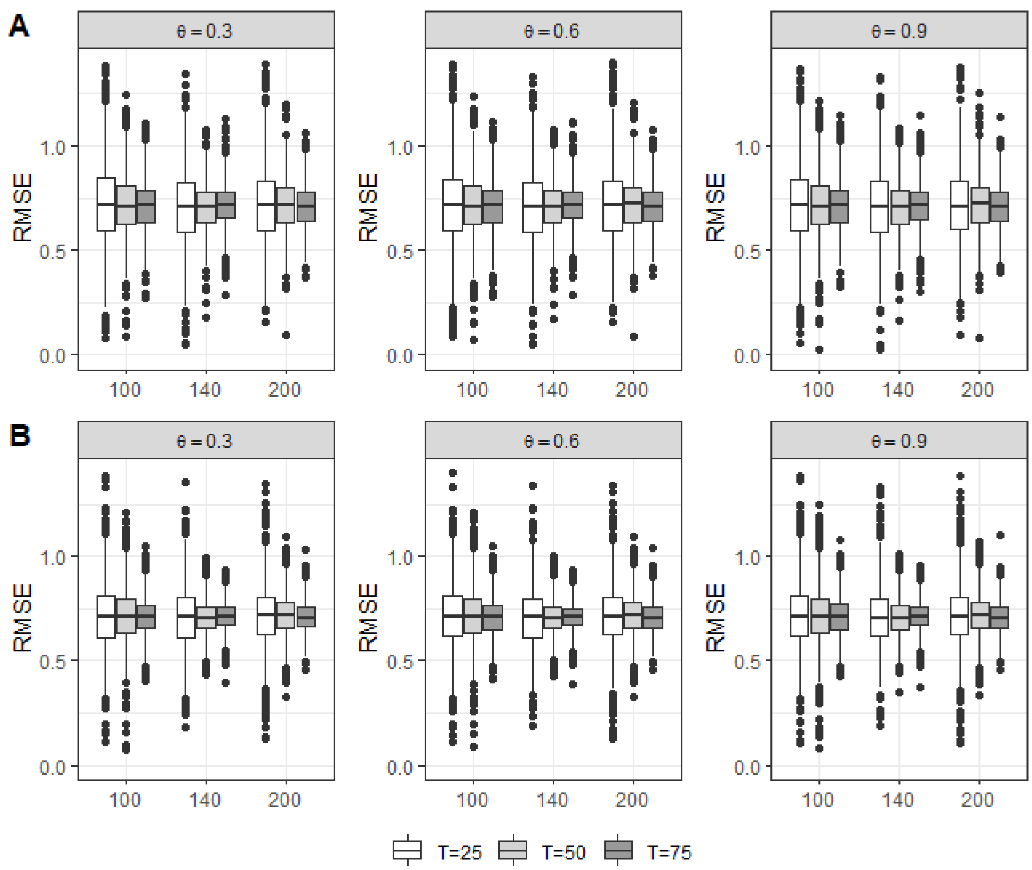
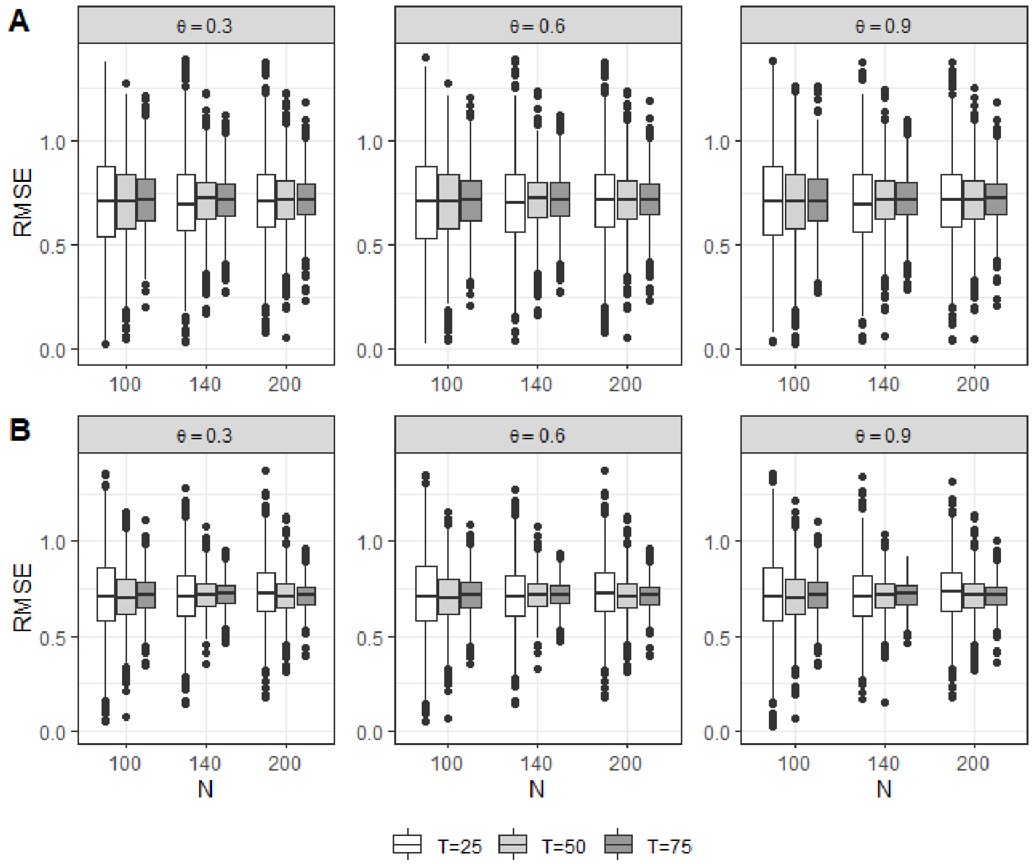


Figure 2: Boxplots of the RMSE values of the nonparametric estimates in 1000 independent simulations under slope homogeneity.



Finite sample performance of the consistency tests

Finite sample performance of the constancy test J^a: To assess the size and power of the test statistic J^a, one can consider the null hypothesis $H_0^a : m(z_t) = z_t'\pi$ versus the alternative $H_1^a : m(z_t) \neq z_t'\pi$, where $\pi = 1$. More precisely, the power of the test is evaluated under $m(z_t) = exp(z_t)/(1+exp(z_t))$ and we use $g_1(z_t) = (1 + \varphi_{1i})(1 + sin(10z_t))$ and $g_2(z_t) = (1 + \varphi_{2i})sin(2z_t)$, where φ_{1i} and φ_{2i} and $IIU(0, 0.1^2)$ random variables.

We consider the DGP defined previously where the factor loadings are generated to fulfill the rank condition (i.e., Case A). Also, the idiosyncratic errors ϵ_{it} are generated as stationary AR(1) processes: $\epsilon_{it} = \rho_{\epsilon_i}\epsilon_{i(t-1)} + \vartheta_{it}$, where $\rho_{\epsilon_i} \sim IIDU(0.05, 0.95)$ and ϑ are $i.i.d.$ $N(0, 1 - \rho_{\epsilon_i}^2)$ across i and t. In view of Assumption 4.4, we choose N to be (100, 120, 140) and set $T = (25, 50)$. The number of replications is 1000, and we use 200 bootstrap resamples for each replication.

Table 3 reports the actual proportion of rejections of our test under H_0^a based on the bootstrap p-values at significance levels 1%, 5%, and 10% , respectively. Also, to examine the sensitivity of the proposed test to the bandwidth selection, the bandwidth matrix H_z is taken as $H_z = I_q(c\widehat{\sigma}_z(NT)^{-1/5})$, where $c = 0.5, 0.8, 1.0$.

Table 3: Proportion of rejections under H_0^a based on bootstrap p-values.

		10%		5%		1%	
c	$N\backslash T$	25	50	25	50	25	50
0.50	100	0.125	0.044	0.072	0.027	0.031	0.008
	120	0.137	0.067	0.082	0.035	0.030	0.011
	140	0.155	0.079	0.099	0.045	0.052	0.016
0.80	100	0.153	0.061	0.098	0.030	0.039	0.008
	120	0.168	0.077	0.101	0.039	0.052	0.014
	140	0.195	0.097	0.134	0.048	0.071	0.021
1.00	100	0.181	0.078	0.105	0.041	0.045	0.011
	120	0.195	0.092	0.110	0.050	0.047	0.019
	140	0.218	0.116	0.148	0.060	0.080	0.018

In Table 3 we can see that the performance of the test statistic is relatively good since it tends to over-rejection when $T = 25$ but the results are close to the nominal sizes with the time dimension increases. Further, it can be noted that the statistic is sensitive to the choice of bandwidth and small values of c lead to under-rejection.

Considering now the power performance of our test statistic, in Table 4 it can be seen that the test statistic shows a good power performance in finite simple since the rejection rate increases as either N or T increases for any value of c, although the best results are obtained when $c = 1$.

Table 4: Proportion of rejections under H_1^a based on bootstrap p-values.

		10%		5%		1%	
c	$N\backslash T$	25	50	25	50	25	50
0.50	100	0.493	0.437	0.440	0.384	0.358	0.299
	120	0.498	0.397	0.448	0.346	0.377	0.294
	140	0.506	0.375	0.466	0.304	0.403	0.291
0.80	100	0.652	0.552	0.593	0.498	0.499	0.404
	120	0.654	0.524	0.587	0.475	0.517	0.407
	140	0.669	0.500	0.612	0.444	0.538	0.378
1.00	100	0.729	0.656	0.670	0.526	0.575	0.492
	120	0.723	0.638	0.686	0.596	0.583	0.497
	140	0.756	0.626	0.707	0.570	0.611	0.488

Finite sample performance of the pooling test J_n^b: To assess the size and power of the test statistic J_n^b we consider $\beta_{1,i} = \beta_{2,i} = 1$ for all i under the null hypothesis, and $\beta_{1i} = 2sin(\pi i/N) - 2/N$ and

$\beta_{2i} = 4cos(\pi i/N) + 4/N$ under the alternative hypothesis. Further, a similar DGP as the defined for the previous test statistic is considered, whereas for the selection of the bandwidths we take $H_x = I_p(c\widehat{\sigma}_x(NT)^{-1/5})$, $H_z = I_q(c\widehat{\sigma}_z T^{-1/5})$, and $\mathcal{H} = 1.3H_xH_z$, where $c = (0.5, 0.8, 1.0)$, so the conditions in Assumption S1.17 are fulfilled. In the experiment the number of replications is 1000, and we use 200 bootstrap resamples for each replication.

Table 5 reports the actual proportion of rejections of our test statistic under H_0^b, while Table 6 contains the power performance of our test. As it is expected from the nonparametric literature, this test statistic is very sensitive to the bandwidth choice. The smallest values of c (i.e. $c = 0.5$) leads to over-rejection, and the opposite is obtained for lager values such as $c = 1$. Nevertheless, in all cases the proportion of rejections ranges around the nominal sizes as the sample size increases, specially when $c = 0.8$. If we consider the performance of the test statistic under the alternative hypothesis, in Table 6 it can be seen that the rejection rate increases as either N or T increases corroborating the good performance of the test statistic.

Table 5: Proportion of rejections under H_0^b based on bootstrap p-values.

		10%		5%		1%	
c	$N\backslash T$	25	50	25	50	25	50
0.50	100	0.144	0.172	0.092	0.109	0.041	0.056
	120	0.138	0.154	0.087	0.103	0.040	0.047
	140	0.125	0.166	0.082	0.098	0.037	0.052
0.80	100	0.110	0.047	0.071	0.025	0.031	0.007
	120	0.130	0.080	0.093	0.045	0.033	0.014
	140	0.134	0.065	0.087	0.039	0.045	0.020
1.00	100	0.082	0.017	0.046	0.010	0.010	0.004
	120	0.093	0.026	0.054	0.010	0.024	0.002
	140	0.074	0.023	0.051	0.009	0.017	0.003

Table 6: Proportion of rejections under H_1^b based on bootstrap p-values.

		10%		5%		1%	
c	$N\backslash T$	25	50	25	50	25	50
0.50	100	0.106	0.174	0.049	0.096	0.009	0.032
	120	0.116	0.159	0.055	0.092	0.015	0.040
	140	0.117	0.177	0.052	0.108	0.015	0.041
0.80	100	0.318	0.511	0.229	0.459	0.113	0.240
	120	0.345	0.635	0.241	0.523	0.123	0.359
	140	0.333	0.697	0.232	0.595	0.136	0.413
1.00	100	0.395	0.417	0.291	0.312	0.157	0.163
	120	0.489	0.544	0.394	0.425	0.253	0.282
	140	0.508	0.625	0.418	0.510	0.251	0.355

S5 EMPIRICAL ILLUSTRATION: Data and variables

In order to estimate the *knowledge capital production function* by Griliches (1979) and recently revisited by Eberhardt et al. (2013), we build an annual country-level balanced panel data set covering 24 OECD countries from 1971 to 2014. We build a different dataset than Eberhardt et al. (2013) for two main reasons. The first reason is a computational one, as exploiting a balanced data set greatly facilitates the computations, even though the proposed estimators can be adapted to the case of unbalanced panels. Second, this is also useful to provide novel and complementary results.

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

Previous works have discussed what is an appropriate measure of oil shocks, and in particular, it has been suggested to use an oil shock measure that filters out both price declines and price increases (Hamilton, 1996). Following this line of reasoning, we construct an oil shock index in a similar way as Davis and Haltiwanger (2001). Our index equals the log of the ratio of the current crude real oil price divided by the average of the real prices in the previous 5 years.

S6 STATEMENTS AND PROOFS OF LEMMAS

Now we prove several lemmas that are used later in the proofs of the theorems. Remember that we denote $\widetilde{X}_{i\cdot} = X_{i\cdot} - \mathcal{B}_X(z)$, $\widetilde{\Lambda} = \Lambda - \mathcal{B}_\Lambda(z)$, $\widetilde{D} = D - \mathcal{B}_D(z)$, $\mathcal{G} = G - \mathcal{B}_G(z)$, 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, $c_H = tr\{H^2\} + \{logT/T|H|\}^{1/2}$.

Lemma 1 *Under Assumptions S1.2, S1.7–S1.10, as $T \to \infty$ we have*

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

Proof of Lemma 1: 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 2 *Under Assumptions S1.2 and S1.7–S1.10, as $T \to \infty$ we have*

$$T^{-1}\widehat{X}_{i\cdot}' M_{\widehat{\Lambda}} \widehat{X}_{i\cdot} \quad \xrightarrow{p} \quad T^{-1}\{X_{i\cdot} - \mathcal{B}_X(z)\}' M_{\widetilde{\Lambda}} \{X_{i\cdot} - \mathcal{B}_X(z)\}.$$

Proof of Lemma 2: Remember that $\widehat{X}_{i\cdot} = (X_{i\cdot} - S\overline{X}_{A\cdot})$. Focusing on the behavior of S and using Lemma 1 it can be shown that, uniformly in z,

$$\begin{aligned} T^{-1} Z_z' K_H(z) Z_z &= \begin{pmatrix} T^{-1}\sum_{t=1}^{T} K_H(z_t - z) & T^{-1}\sum_{t=1}^{T}(z_t - z)K_H(z_t - z) \\ T^{-1}\sum_{t=1}^{T}(z_t - z)'K_H(z_t - z) & T^{-1}\sum_{t=1}^{T}(z_t - z)(z_t - z)'K_H(z_t - z) \end{pmatrix} \\ &= \begin{pmatrix} \rho_{z_t}(z) & \mu_2^q(K)H^2 D_\rho(z) \\ \mu_2^q(K)H^2 D_\rho(z)' & H^2\mu_2^q(K)\rho_{z_t}(z) \end{pmatrix} \{1 + O_p(c_H)\}, \end{aligned} \tag{S6.1}$$

where $D_\rho(z) = \partial\rho(z)/\partial z'$. Following a similar reasoning,

$$T^{-1}Z_z'K_H(z)\overline{X}_{A\cdot} = \begin{pmatrix} T^{-1}\sum_{t=1}^{T}\overline{x}_{At}'K_H(z_t - z) \\ T^{-1}\sum_{t=1}^{T}(z_t - z)\overline{x}_{At}'K_H(z_t - z) \end{pmatrix} = \begin{pmatrix} \mathcal{B}_X(z) \\ O_p(H) \end{pmatrix}\{1 + O_p(c_H)\}. \quad \text{(S6.2)}$$

Similarly, $\widehat{\Lambda} = \widetilde{\Lambda}\{1 + O_p(c_H)\}$. Using this result and (S6.1)-(S6.2), by the Slutsky theorem it can be shown $S\overline{X}_{A\cdot} = \mathcal{B}_X(z)\{1 + O_p(c_{H_0})\}$, uniformly in z. Hence, it is proved that

$$T^{-1}\widehat{X}_{i\cdot}'M_{\widehat{\Lambda}}\widehat{X}_{i\cdot} = T^{-1}\{X_{i\cdot} - \mathcal{B}_X(z)\}'M_{\widetilde{\Lambda}}\{X_{i\cdot} - \mathcal{B}_X(z)\}\{1 + O_p(c_H)\} \quad \text{(S6.3)}$$

and by the law of large numbers, the result of the lemma holds.

∎

Lemma 3 *Suppose that either $\|\beta_i\| \leq C$ for each i or that the random coefficient assumption (i.e., Assumption S1.4) holds. Then, under Assumptions S1.7–S1.10,*

a) $\frac{\overline{\varepsilon}_{A\cdot}'\overline{\varepsilon}_{A\cdot}}{T} = O_p\left(\frac{1}{N}\right)$.

b) $\frac{\widetilde{F}'\overline{\varepsilon}_{A\cdot}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right)$ *and* $\frac{\widetilde{D}'\overline{\varepsilon}_{A\cdot}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right)$.

c) $\frac{V_{i\cdot}'\widetilde{D}}{T} = O_p\left(\frac{1}{T}\right)$ *and* $\frac{V_{i\cdot}'\widetilde{F}}{T} = O_p\left(\frac{1}{T}\right)$.

d) $\frac{V_{i\cdot}'\overline{\varepsilon}_{A\cdot}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right)$ *and* $\frac{\epsilon_{i\cdot}'\overline{\varepsilon}_{A\cdot}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right)$.

Proof of Lemma 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 4 *Under Assumptions S1.2 and S1.7–S1.10, as $T \to \infty$, we have*

$$T^{-1}\widehat{\Lambda}'M_{\widehat{X}_i}\widehat{\Lambda} \quad \xrightarrow{p} \quad T^{-1}\{\Lambda - \mathcal{B}_\Lambda(z)\}'M_{\widetilde{X}_i}\{\Lambda - \mathcal{B}_\Lambda(z)\}.$$

Proof of Lemma 4: This lemma can be proved in a similar way as in Lemma 2 and it has been omitted for the sake of brevity.

∎

Lemma 5 *Under Assumptions S1.2 and S1.7–S1.10 and assuming that the rank condition is satisfied, as* $T \to \infty$ *we have*

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

b) $\frac{\widetilde{X}'_{i.} M_{\widetilde{\Lambda}} \widetilde{F}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p(c_H),$

c) $\frac{\widetilde{X}'_{i.} M_{\widetilde{\Lambda}} \widetilde{e}_{i.}}{T} = \frac{\widetilde{X}'_{i.} M_{\widetilde{G}} \widetilde{e}_{i.}}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p(c_H),$

where $M_{\widetilde{G}} = I_T - \widetilde{G}(\widetilde{G}'\widetilde{G})^{-1}\widetilde{G}$ *and* $\widetilde{e}_{i.} = (I_T - S)e_{i.}$, *where* $e_{i.} = (e_{i1}, \ldots, e_{iT})'$ *is a* $T \times 1$ *vector of the composed error term* e_{it} *defined in (3.6).*

∎

Proof of Lemma 5: Considering the behavior of $T^{-1}\widetilde{X}'_{i.} M_{\widetilde{\Lambda}} \widetilde{X}_{i.}$ we have to analyze the following terms separately

$$\frac{\widetilde{X}'_{i.} M_{\widetilde{\Lambda}} \widetilde{X}_{i.}}{T} = \left(\frac{\widetilde{X}'_{i.}\widetilde{X}_{i.}}{T}\right) - \left(\frac{\widetilde{X}'_{i.}\widetilde{\Lambda}}{T}\right)\left(\frac{\widetilde{\Lambda}'\widetilde{\Lambda}}{T}\right)^{-1}\left(\frac{\widetilde{\Lambda}'\widetilde{X}_{i.}}{T}\right). \tag{S6.4}$$

Writing (2.5) in matrix form and using the Taylor expansion, $\widetilde{X}_{i.}$ can be written as

$$\widetilde{X}_{i.} = \widetilde{G}\Pi_i + V_{i.}. \tag{S6.5}$$

where $V_{i.} \equiv (v_{i1}, \ldots, v_{iT})'$, $\widetilde{G} = [\widetilde{D}, \widetilde{F}]$, and $\Pi_i = [A'_i, \Gamma'_i]$ are $T \times p$, $T \times (n+r)$ and $(n+r) \times p$ matrices, respectively. Similarly, using (S2.1), the elements $\widetilde{\Lambda}$ can be written as

$$\widetilde{\Lambda} \;=\; \widetilde{G}\overline{P}_A + \overline{\varepsilon}^*_A, \tag{S6.6}$$

where $\overline{\varepsilon}^*_A = (0_{T\times n}, \overline{\varepsilon}_A)$ is a $T \times (n+p+1)$ matrix, $\overline{\varepsilon}_A \equiv (\overline{\varepsilon}_{A1}, \ldots, \overline{\varepsilon}_{AT})'$ is a $T \times (1+p)$ matrix, $\widetilde{G} = [\widetilde{D}, \widetilde{F}]$ is a $T \times (n+r)$ matrix for $\widetilde{F} = F - E[F|z_t = z]\rho_{z_t}(z)$, whereas $\overline{P}_A$ is a $(n+r)\times(n+p+1)$ matrix such as

$$\overline{P}_A = \begin{pmatrix} I_N & \overline{B}_A \\ 0_{r\times N} & \overline{C}_A \end{pmatrix}.$$

Replacing (II.10)-(II.11) in (II.9) and using the results of Lemma 3,

$$\begin{aligned} \frac{\widetilde{X}_{i\cdot}\widetilde{\Lambda}}{T} &= \left(\frac{\widetilde{X}'_{i\cdot}\widetilde{G}}{T}\right)\overline{P}_A + \Pi'_i\left(\frac{\widetilde{D}\overline{\varepsilon}^*_A}{T}\right) + \Pi'_i\left(\frac{\widetilde{F}\overline{\varepsilon}^*_A}{T}\right) + \left(\frac{V'_{i\cdot}\overline{\varepsilon}^*_A}{T}\right) + O_p(c_H) \\ &= \left(\frac{\widetilde{X}'_{i\cdot}\widetilde{G}}{T}\right)\overline{P}_A + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p(c_H). \end{aligned} \tag{S6.7}$$

Similarly, it is straightforward to show

$$\begin{aligned} \frac{\widetilde{\Lambda}'\widetilde{\Lambda}}{T} &= \overline{P}'_A\left(\frac{\widetilde{G}'\widetilde{G}}{T}\right)\overline{P}_A + \left(\frac{\overline{\varepsilon}^{*\prime}_A\overline{\varepsilon}^*_A}{T}\right) + \left(\frac{\overline{\varepsilon}^{*\prime}_A\widetilde{D}}{T}\right)\overline{P}_A + \left(\frac{\overline{\varepsilon}^{*\prime}_A\widetilde{F}}{T}\right)\overline{P}_A + O_p(c_H) \\ &= \overline{P}'_A\left(\frac{\widetilde{G}'\widetilde{G}}{T}\right)\overline{P}_A + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p(c_H). \end{aligned} \tag{S6.8}$$

Defining $\widetilde{Q}_A = \widetilde{G}\overline{P}_A$ as a $T \times (1+p+n)$ matrix and replacing (II.12)-(II.13) in (II.9), we get

$$T^{-1}\widetilde{X}'_{i\cdot}M_{\widetilde{\Lambda}}\widetilde{X}_{i\cdot} = T^{-1}\widetilde{X}'_{i\cdot}M_{\widetilde{Q}}\widetilde{X}_{i\cdot} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p(c_H), \tag{S6.9}$$

where $M_{\widetilde{Q}} = I_T - \widetilde{Q}'_A(\widetilde{Q}'_A\widetilde{Q}_A)^{-1}\widetilde{Q}'_A$. Assuming that the rank condition is satisfied, $\widetilde{X}'_{i\cdot}M_{\widetilde{Q}}\widetilde{X}_{i\cdot}$ is simplified to $\widetilde{X}'_{i\cdot}M_{\widetilde{G}}\widetilde{X}_{i\cdot}$ and the proof of the result a) is completed. Following a similar reasoning the rest of the results of this lemma can be proved.

∎

Lemma 6 *Under H_0^a,*

$$\widehat{\Sigma}_a \quad \xrightarrow{p} \quad \Sigma_a,$$

where $\widehat{\Sigma}_a = \frac{2}{N^2T^2|H_z|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\widehat{\ddot{\epsilon}}_{a,it}^2\widehat{\ddot{\epsilon}}_{a,js}^2 K^2(z_t, z_s)$.

∎

Proof of Lemma 6: This lemma can be proved in a similar way as in Lemma A.2.3. in Cai et al. (2020) and it has been omitted for the sake of brevity

S7 MATHEMATICAL PROOFS

Proof of Theorem 3.1: Writing (2.3) and (2.4) in vectorial form and following Fan and Huang (2005) to control for $m_i(\cdot)$, it can be written

$$Y_{i\cdot} - S\overline{Y}_{A\cdot} = (I_T - S)D\alpha_i + (X_{i\cdot} - S\overline{X}_{A\cdot})\beta_i + (I_T - S)F\gamma_i + e_{i\cdot} + O_p(tr\{H^2\}), \qquad \text{(II.1)}$$

where $D \equiv (d_1, \ldots, d_T)'$ and $F \equiv (f_1, \ldots, f_T)'$ are $T \times n$ and $T \times r$ matrices, respectively, and $e_i = (e_{i1}, \ldots, e_{iT})'$ is a $T \times 1$ vector of the composed error term of the for $e_{it} = \epsilon_{it} + o_p(1)$.

Replacing (II.1) in (3.7) and rearranging terms,

$$\widehat{\beta}_{SCCE,i} - \beta_i \;=\; (\widehat{X}_{i\cdot}' M_{\widehat{\Lambda}} \widehat{X}_{i\cdot})^{-1}\widehat{X}_{i\cdot}' M_{\widehat{\Lambda}}(I_T - S)\left[F\gamma_i + m_i(Z) + e_{i\cdot} + O_p(tr\{H^2\})\right] \qquad \text{(II.2)}$$

given that $M_{\widehat{\Lambda}}(I_T - S)D = 0$ since $D \in \Lambda$. Then, it is shown that there is a direct dependence of $\widehat{\beta}_i$ of the unobserved and observed common factors (i.e., f_t and z_t).

Using Lemmas 2 and 5 in (II.2) and assuming that the rank condition holds, it yields

$$\widehat{\beta}_{SCCE,i} - \beta_i \;=\; \left(\frac{\widetilde{X}_{i\cdot}' M_{\widetilde{G}} \widetilde{X}_{i\cdot}}{T}\right)^{-1} \frac{\widetilde{X}_{i\cdot}' M_{\widetilde{G}} \epsilon_{i\cdot}}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p(c_H^2), \qquad \text{(II.3)}$$

given that, following a similar procedure as in Lemma 3 we can show that, uniformly in z, $T^{-1}\widehat{X}_{i\cdot}' M_{\widehat{\Lambda}} S m_i(Z) = T^{-1}\widehat{X}_{i\cdot}' M_{\widehat{\Lambda}} m_i(Z)\{1+O_p(c_H)\}$ so

$$\begin{aligned}\frac{\widehat{X}_{i\cdot}' M_{\widehat{\Lambda}}(I_T - S)m_i(Z)}{T} &= \frac{\widehat{X}_{i\cdot}' M_{\widehat{\Lambda}} m_i(Z)}{T}\{1+O_p(c_H)\} \\ &= \left(\frac{\widetilde{X}_{i\cdot} M_{\widetilde{G}} m_i(Z)}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}} + O_p(c_H)\right)\right)\{1+O_p(c_H)\} \\ &= O_p(c_H^2).\end{aligned}$$

Using a similar argument we get $T^{-1}\widehat{X}_{i\cdot}' M_{\widehat{\Lambda}}(I_T - S)\epsilon_{i\cdot} = T^{-1}\widehat{X}_{i\cdot}' M_{\widehat{\Lambda}}\epsilon_{i\cdot}\{1+O_p(c_H)\}$. Finally, the consistency of $\widehat{\beta}_i$ follows almost immediately from (II.3) since, under the assumptions of the theorem, $T^{-1}\widetilde{X}_{i\cdot}' M_{\widetilde{G}}\widetilde{X}_{i\cdot} \xrightarrow{p} \Sigma_{v_i}$ and $T^{-1}\widetilde{X}_{i\cdot}' M_{\widetilde{G}}\epsilon_{i\cdot} \xrightarrow{p} 0$. Finally, to derive the asymptotic distribution of $\widehat{\beta}_i$ we have

$$\sqrt{T}(\widehat{\beta}_{SCCE,i} - \beta_i) = \left(\frac{\widetilde{X}_{i\cdot}' M_{\widetilde{G}}\widetilde{X}_{i\cdot}}{T}\right)^{-1}\frac{\widetilde{X}_{i\cdot}' M_{\widetilde{G}}\epsilon_{i\cdot}}{\sqrt{T}} + O_p\left(\frac{\sqrt{T}}{N}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p(\sqrt{T}c_H^2) \quad \text{(II.4)}$$

and assuming $\sqrt{T}/N \to 0$ and $\sqrt{T}c_H^2 \to 0$ as $(N,T)\to\infty$,

$$\sqrt{T}(\widehat{\beta}_i - \beta_i) \xrightarrow{d} \Sigma_{v_i}^{-1}\left(\frac{\widetilde{X}_{i\cdot}' M_{\widetilde{G}}\epsilon_{i\cdot}}{\sqrt{T}}\right).$$

Then, the proof of the theorem is done.

∎

Proof of Theorem 3.2: We replace (2.3)-(2.4) in vectorial form in (3.4) and approximate the unobserved common factors by λ_t. Then, using the multivariate Taylor expansion it can be written

$$\begin{aligned}\widehat{m}_i(z,H) - m_i(z) &= \iota_1'(T^{-1}Z_z' K_H(z) Z_z)^{-1} T^{-1} Z_z' K_H(z)\left[\frac{1}{2}Q_{m_i}(z) - X_{i\cdot}(\widehat{\beta}_{SCCE,i} - \beta_i) - \Lambda(\widehat{\delta}_{SCCE,i} - \delta_i)\right. \\ &+ \left. e_{i\cdot} + O_p(tr\{H^2\})\right], \end{aligned} \quad \text{(II.5)}$$

where $Q_{m_i(z)} = [(z_1 - z)'\mathcal{H}_{m_i}(z)(z_1 - z), \ldots, (z_T - z)'\mathcal{H}_{m_i}(z)(z_T - z)]'$ and, using standard nonparametric techniques, it is straightforward to show $T^{-1}Z_z'K_H(z)R_{m_i(z)} = O_p(tr\{H^2\})$, where $R_{m_i}(z)$ is the residual term of the Taylor expansion.

Considering the asymptotic bias of $\widehat{m}_i(z, H)$ and given the $\sqrt{T}$-consistency of $\widehat{\beta}_{SCCE,i}$ and $\widehat{\delta}_{SCCE,i}$, following a similar reasoning as in the proof of Lemma 2 it is easy to show

$$E[\widehat{m}_i(z, H)] - m_i(z) = \frac{\mu_2^q(K)}{2} tr\{\mathcal{H}_{m_i}(z)H^2\} + O_p(tr\{H^2\}) + O_p\left(\frac{1}{\sqrt{T}}\right). \qquad \text{(II.6)}$$

Similarly, analyzing the corresponding variance term we have

$$\begin{aligned} T|H|Var(\widehat{m}_i(z, H)) &= T|H|\iota_1'(Z_z'K_H(z)Z_z)^{-1}Z_z'K_H(z)E(e_{i\cdot}e_{i\cdot}')K_H(z)Z_z(Z_z'K_H(z)Z_z)^{-1}\iota_1 \\ &= \frac{R^q(K)\sigma_{\eta_i}^2}{\rho_{z_t}(z)}\{1 + o_p(1)\}. \qquad \text{(II.7)} \end{aligned}$$

Finally, Assumption 2.1 recalls that ϵ_{it} depends on η_{it}, that is a stationary process independently distributed of the stationary processes X and G. Then, the CLT can be applied and the proof of the theorem is done.

∎

Proof of Theorem 3.3: Plugging (II.2) in (3.10) and using Assumption S1.4, by the multivariate Taylor theorem we get

$$\begin{aligned} \sqrt{N}(\widehat{\beta}_{SCCEMG} - \overline{\beta}) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\widehat{X}_{i\cdot}'M_{\widehat{\Lambda}}\widehat{X}_{i\cdot}}{T}\right)^{-1}\left(\frac{\widehat{X}_{i\cdot}'M_{\widehat{\Lambda}}\widehat{F}}{T}\right)\gamma_i \\ &+ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\widehat{X}_{i\cdot}'M_{\widehat{\Lambda}}\widehat{X}_{i\cdot}}{T}\right)^{-1}\left(\frac{\widehat{X}_{i\cdot}'M_{\widehat{\Lambda}}\widehat{e}_{i\cdot}}{T}\right), \qquad \text{(II.8)} \end{aligned}$$

where $\widehat{F} = (I_T - S)F$ and $\widehat{e}_{i\cdot} = (I_T - S)e_{i\cdot}$.

By Lemma 1, $T^{-1}\sum_{i=1}^{N}\widehat{X}_{i\cdot}'M_{\widehat{\Lambda}}\widehat{F} = T^{-1}\widetilde{X}_{i\cdot}'M_{\widetilde{\Lambda}}\widetilde{F} + O_p(N^{-1}) + O_p((NT)^{-1/2}) + O_p(c_H)$ and using the results in Lemma 2 it is easily seen that, for all bounded values of the factor loadings,

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\widehat{X}_{i\cdot}'M_{\widehat{\Lambda}}\widehat{X}_{i\cdot}}{T}\right)^{-1}\frac{\widetilde{X}_{i\cdot}'M_{\widetilde{G}}\widetilde{F}\gamma_i}{T}\xrightarrow{p}0 \quad \text{as} \quad N,T\to\infty.$$

Following a similar proof scheme as in the proof of Theorem 4.1 in Fan and Huang (2005) it can be proved that $T^{-1}\widetilde{X}_{i\cdot}M_{\widetilde{G}}\widetilde{e}_{i\cdot} = T^{-1}\widetilde{X}_{i\cdot}M_{\widetilde{G}}e_{i\cdot}\{1+o_p(1)\}$. Then, using the above results and Lemmas 1-2, the final expression to analyze is such as

$$\sqrt{N}(\widehat{\beta}_{SCCEMG}-\overline{\beta}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_i + \frac{1}{\sqrt{T}}\Delta_{NT} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p(\sqrt{N}c_H^2), \qquad \text{(II.9)}$$

where, assuming that the rank condition is satisfied, $\widetilde{X}_{i\cdot}'M_{\widetilde{Q}}\widetilde{X}_{i\cdot}$ is simplified to $\widetilde{X}_{i\cdot}'M_{\widetilde{G}}\widetilde{X}_{i\cdot}$ and $\Delta_{NT} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(\frac{V_{i\cdot}'M_{\widetilde{G}}V_{i\cdot}}{T}\right)^{-1}V_{i\cdot}'M_{\widetilde{G}}e_{it}$. Denoting $\mathcal{W}_{\cdot t}' = (\mathcal{W}_{1t},\ldots,\mathcal{W}_{Nt})$ and $\mathcal{W}_{it} \equiv (T^{-1}V_{i\cdot}'M_{\widetilde{G}}V_{i\cdot})^{-1}V_{i\cdot}'M_{\widetilde{G}}e_{it}$, Δ_{NT} can be written as

$$\Delta_{NT} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathcal{W}_{i\cdot}'e_{\cdot t}e_{\cdot s}'\mathcal{W}_{\cdot s} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathcal{W}_{it}e_{it} = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\mathcal{W}_{\cdot t}'e_{\cdot t}.$$

Under Assumption 2.1, $E(\Delta_{NT})=0$ and its variance satisfies

$$\begin{aligned} Var(\Delta_{NT}) &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{s=1}^{T}E(\mathcal{W}_{\cdot t}'e_{\cdot t}e_{\cdot s}'\mathcal{W}_{\cdot s}) = \frac{\overline{\sigma}_\eta^2}{NT}\sum_{t=1}E(\mathcal{W}_{\cdot t}'\Phi\mathcal{W}_{\cdot t}) \\ &\leq \frac{\overline{\sigma}_\eta^2}{NT}\sum_{t=1}^{T}E(\mathcal{W}_{\cdot t}'\mathcal{W}_{\cdot t})\lambda_{\max}^*(\Phi) \leq \frac{C\overline{\sigma}_\eta^2}{N}\sum_{i=1}^{N}\Sigma_{v_i}^{-1}, \end{aligned} \qquad \text{(II.10)}$$

where C is a constant term and $\lambda_{\max}^*(\cdot)$ is the maximum eigenvalue of Φ. Note that in order to obtain this result we resort to Bernstein (2005) (pp. 264 and 271). Under Assumption S1.6(a) it has been proved that $Var(\Delta_{NT})$ tends to a non-singular matrix with finite elements. Thus, replacing this result in (II.9) we get

$$\sqrt{N}(\widehat{\beta}_{SCCEMG}-\overline{\beta}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_i + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p(\sqrt{N}c_H^2)$$

which proves the theorem as N and T tends to infinity and $\sqrt{N}c_H^2 \to 0$.

■

Proof of Theorem 3.4: Replacing (II.1) in (3.12) and following a similar proof scheme as in the proof of Theorem 3.3 it is straightforward to show that, under Assumption S1.4,

$$\begin{aligned}\sqrt{N}(\widehat{\beta}_{SCCEP} - \overline{\beta}) &= \left(\frac{1}{NT}\sum_{i=1}^{N} \widetilde{X}_{i\cdot}' M_{\widetilde{G}} \widetilde{X}_{i\cdot}\right)^{-1} \left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \frac{\widetilde{X}_{i\cdot}' M_{\widetilde{G}} \widetilde{X}_{i\cdot}\xi_i}{T} + \frac{1}{\sqrt{T}} q_{NT}\right] \\ &+ O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p(\sqrt{N}c_H^2), \end{aligned} \tag{II.11}$$

where $q_{NT} = (NT)^{-1/2}\sum_{i=1}^{N} \widetilde{Q}_{i\cdot}' e_{i\cdot}$ and $\widetilde{Q}_{i\cdot} = M_{\widetilde{G}}\widetilde{X}_{i\cdot}$. Under Assumption 2.1 and using a similar reasoning as in (II.10), it can be shown that $E(q_{NT}) = 0$ and

$$\begin{aligned} Var(q_{NT}) &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{s=1}^{T} E(\widetilde{Q}_{\cdot t}' e_{\cdot t} e_{\cdot s}' \widetilde{Q}_{\cdot s}) = \frac{\overline{\sigma}_\eta^2}{NT}\sum_{t=1}^{T} E[\widetilde{Q}_{\cdot t}' \Phi \widetilde{Q}_{\cdot t}] \\ &= \frac{C\overline{\sigma}_\eta^2}{NT}\sum_{i=1}^{N} E(\widetilde{X}_{i\cdot}' M_{\widetilde{G}} \widetilde{X}_{i\cdot})\lambda_{max}^{*}(\Phi), \end{aligned}$$

which, under Assumption S1.6(b), tends to a finite, positive definite matrix.

Then, using the above results it follows that

$$\begin{aligned}\sqrt{N}(\widehat{\beta}_{SCCEP} - \overline{\beta}) &= \left(\frac{1}{N}\sum_{i=1}^{N} \frac{\widetilde{X}_{i\cdot}' M_{\widetilde{G}} \widetilde{X}_{i\cdot}}{T}\right)^{-1} \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \frac{\widetilde{X}_{i\cdot}' M_{\widetilde{G}} \widetilde{X}_{i\cdot}\xi_i}{T} \\ &+ O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p(\sqrt{N}c_H^2) \end{aligned} \tag{II.12}$$

which proves the theorem.

■

Proof of Theorem 3.5: Replacing (3.1) in (3.9), by the Taylor expansion of $\overline{m}(\cdot)$ we get

$$\begin{aligned}\widehat{m}_{CCEMG}(z,H)-\overline{m}(z) &= \iota_1'(Z_z'K_H(z)Z_z)^{-1}Z_z'K_H(z)\left[\frac{1}{2}Q_{\overline{m}}(z)+R_{\overline{m}}(z)-\frac{1}{N}\sum_{i=1}^{N}X_{i\cdot}(\widehat{\beta}_{SCCE,i}-\beta_i)\right.\\ &- \left.\frac{1}{N}\sum_{i=1}^{N}\Lambda(\widehat{\delta}_{SCCE,i}-\delta_i)+\frac{1}{N}\sum_{i=1}^{N}\epsilon_{i\cdot}+\varphi_{i\cdot}+o_p(1)\right],\end{aligned}$$

where $Q_{\overline{m}}(z)$ and $R_{\overline{m}}(z)$ are defined as in II.5 for $\overline{m}(\cdot)$, $\epsilon_{i\cdot}=(\epsilon_{i1},\ldots,\epsilon_{iT})'$ is a $T\times 1$ vector, and $o_p(1)$ captures possible approximation error for f_t. Again, using standard nonparametric techniques it can be proved that when $R_{\overline{m}}(z)$ is premultiplied by $e_1'(Z_z'K_H(z)Z_z)^{-1}Z_z'K_H(z)$ the resulting scalar is $o_p(tr\{H^2\})$.

Then, to analyze the asymptotic behavior of this estimator is enough to show

$$\begin{aligned}&\sqrt{T|H|}(\widehat{m}_{CCEMG}(z,H)-\overline{m}(z))-\sqrt{T|H|}\iota_1'(T^{-1}Z_z'K_H(z)Z_z)^{-1}(\mathcal{J}_1(z)-\mathcal{J}_2(z))\\ &=\sqrt{T|H|}\iota_1'(T^{-1}Z_z'K_H(z)Z_z)^{-1}\mathcal{J}_3\end{aligned}\qquad\text{(II.13)}$$

where

$$\begin{aligned}\mathcal{J}_1(z) &= T^{-1}Z_z'K_H(z)\left(\frac{1}{2}Q_{\overline{m}}(z)+R_{\overline{m}}(z)\right),\\ \mathcal{J}_2(z) &= T^{-1}Z_z'K_H(z)\left[N^{-1}\sum_{i=1}^{N}X_{i\cdot}(\widehat{\beta}_{SCCE,i}-\beta_i)+N^{-1}\sum_{i=1}^{N}\Lambda(\widehat{\delta}_{SCCE,i}-\delta_i)+o_p(1)\right],\\ \mathcal{J}_3(z) &= T^{-1}Z_z'K_H(z)\overline{\epsilon}_{A\cdot}.\end{aligned}$$

Using Lemma 2 it follows

$$\mathcal{J}_1(z)=\begin{pmatrix}\mu_2^q(K)tr(H^2\mathcal{H}_{\overline{m}}(z))\rho_{z_t}(z)+O_p(tr\{H^2\})\\ O_p(H^3)\end{pmatrix}.\qquad\text{(II.14)}$$

Following a similar proof scheme as in Theorem 3.3, it can be shown that $\widehat{\beta}_{SCCE,i}$ and $\widehat{\delta}_{SCCE,i}$ are $\sqrt{T}$-consistent estimators of β_i and δ_i, respectively. Hence, considering the behavior of $\mathcal{J}_2(z)$

it yields

$$\begin{aligned}\mathcal{J}_2(z) &= (NT)^{-1}\sum_{i=1}^{N} Z_z' K_H(z)\left[X_{i\cdot}(\widehat{\beta}_{SCCE,i}-\beta_i)+\Lambda(\widehat{\delta}_{SCCE,i}-\delta_i)+o_p(1)\right] \\ &= O_p\left(T^{-1/2}\right). \qquad \text{(II.15)}\end{aligned}$$

Then, using (S6.1), (II.14)-(II.15) and the Slutsky's theorem, the asymptotic bias of $\widehat{m}_{CCEMG}(z,H)$ is such as

$$\iota_1'(T^{-1}Z_z'K_H(z)Z_z)^{-1}(\mathcal{J}_1(z)-\mathcal{J}_2(z)) = \frac{\mu_2^q(K)}{2}tr\{H^2\mathcal{H}_{\overline{m}}(z)\}+O_p(tr\{H^2\})+O_p\left(\frac{1}{\sqrt{T}}\right). \text{(II.16)}$$

In order to obtain the asymptotic variance of the right-hand side of (II.13), we have to analyze the behavior of $\mathcal{J}_3(z)$. Let $\Omega_{u^*} = E(\overline{\epsilon}_{A\cdot}\overline{\epsilon}_{A\cdot}')$ be a $T\times T$ matrix and using Assumptions 2.1 and S1.10, by the law of iterated expectations we get

$$\begin{aligned}T|H|Var(\mathcal{J}_3(z)) &= T^{-1}|H|E[Z_z'K_H(z)\Omega_\epsilon K_H(z)Z_z] \\ &= \begin{pmatrix} \overline{\sigma}_\eta^2\nu_N R^q(K)\rho_{z_t}(z) & O_p(H) \\ O_p(H) & \overline{\sigma}_\eta^2 H^2\nu_N R_2^q(K)\rho_{z_t}\end{pmatrix}(1+o_p(1)). \qquad \text{(II.17)}\end{aligned}$$

Therefore, using (S6.1) and (II.17), by the Slutsky theorem, as $T\to\infty$,

$$T|H|Var((T^{-1}Z_z'K_H(z)Z_z)^{-1}\mathcal{J}_3(z)) = \frac{\nu_N\overline{\sigma}_\eta^2 R^q(K)}{\rho_{z_t}(z)}(1+o_p(1)). \qquad \text{(II.18)}$$

Finally, to complete the proof of this theorem the Lyapounov condition has to be checked. Let $\psi = |H|^{1/2}\imath_{(1+q)}' Z_z'K_H(z)\overline{\epsilon}_{A\cdot}$, where $\imath_{(1+q)}$ is a $(1+q)\times 1$ unitary vector. Under Assumption 2.1 and using the above results, $E(\psi)=0$ and $Var(\sum_{t=1}^T \psi_t) = (T/N)^{(2+\varsigma)/2}$. By Minkowsky's inequality,

$$E|\psi_t|^{(2+\varsigma)} \leq C_1 E|\psi_{a,t}|^{(2+\varsigma)} + C_2 E|\psi_{b,t}|^{(2+\varsigma)}$$

where $\psi_{a,t} = |H|^{1/2}\bar{\epsilon}_{At}K_H(z_t - z)$ and $\psi_{b,t} = |H|^{1/2}(z_t - z)\bar{\epsilon}_{At}K_H(z_t - z)$. Analyzing each of these terms separately and using Assumptions S1.10–S1.11, it can be proved $E|\psi_{a,t}| = O_p(N^{-(1+\varsigma)}|H|^{-\varsigma/2})$ and $E|\psi_{b,t}| = O_p(N^{-(1+\varsigma)}|H|^{2-\varsigma(q-2)/2})$. Hence, using there results we can conclude

$$\left(\frac{N}{T}\right)^{(2+\varsigma)/2}\sum_{t=1}^{T} E|\overline{\psi}_t|^{(2+\varsigma)} \leq C(NT|H|)^{-\varsigma/2},$$

so the Lyapounov condition is proved and Theorem 3.5 holds.

∎

Proof Theorem 4.1: Under H_0^a, $\widehat{\ddot{\epsilon}}_{a,it}$ can be written as $\widehat{\ddot{\epsilon}}_{a,i\cdot} = M_\Lambda(\epsilon_{i\cdot} - X_{i\cdot}(\widehat{\beta}_{SP} - \beta) - Z(\widehat{\pi} - \pi) + F\gamma_i + D\alpha_i)$. Following the same reasoning as in the proof of Lemma A.1.2(d)-(f) in Cai et al. (2020), it can be shown that $M_\Lambda F\gamma_i = O_p\left(\sqrt{\frac{T}{N}}\right) = o_p(1)$ as $T/N \to 0$. Also, $M_\Lambda D\alpha_i = 0$ since $D \in \Lambda$. Therefore, one can show easily that

$$\begin{aligned} \widehat{I}^a &= \frac{1}{N^2T^2|H_z|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\widehat{\ddot{\epsilon}}'_{a,i\cdot}K(Z,Z)\widehat{\ddot{\epsilon}}_{a,i\cdot} \\ &= \widehat{I}^{a*}(1 + o_p(1)), \end{aligned} \tag{II.19}$$

where $\widehat{I}^{a*} = \frac{1}{N^2T^2|H_z|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\widetilde{\ddot{\epsilon}}'_{a,i\cdot}K(Z,Z)\widetilde{\ddot{\epsilon}}_{a,i\cdot}$ and $\widetilde{\ddot{\epsilon}}_{a,i\cdot} = M_\Lambda\epsilon_{i\cdot} - M_\Lambda X_{i\cdot}(\widehat{\beta}_{SP} - \beta) - M_\Lambda Z(\widehat{\pi} - \pi)$. Hence, the statistic to study is such as

$$\begin{aligned} \widehat{I}^{a*} &= \frac{1}{N^2T^2|H_z|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\epsilon'_{i\cdot}M'_\Lambda K(Z,Z)M_\Lambda\epsilon_{i\cdot} - \frac{2}{N^2T^2|H_z|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\epsilon'_{i\cdot}M'_\Lambda K(Z,Z)M_\Lambda(X_{j\cdot}(\widehat{\beta}_{SP} - \beta) + Z(\widehat{\pi} - \pi)) \\ &+ \frac{1}{N^2T^2|H_z|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}(X_{i\cdot}(\widehat{\beta}_{SP} - \beta) + Z(\widehat{\pi} - \pi))'M'_\Lambda K(Z,Z)M_\Lambda(X_{j\cdot}(\widehat{\beta}_{SP} - \beta) + Z(\widehat{\pi} - \pi)) \\ &= I\!I^a_{n_1} - 2I\!I^a_{n_2} + I\!I^a_{n_3}, \end{aligned} \tag{II.20}$$

where the definitions of $I\!I^a_{n_l}$, $l = 1, 2, 3$, should be apparent from the context. Using $\widehat{\beta}_{SP} - \beta = O_p((NT)^{-1/2})$ (see Theorem 3.6-3.7 under weak CSD) and $\widehat{\pi} - \pi = O_p((NT)^{-1/2})$ (see Theorem 4 in Pesaran (2006)), it is straightforward to show that under H_0^a, $I\!I^a_{n_1} = O_p((NT)^{-1})$, $I\!I^a_{n_2} =$

$O_p((NT)^{-1})$, and $I\!I^a_{n3} = O_p((NT|H_z|)^{-1})$. Hence, $I\!I^a_{n3}$ is the leading term of $\widehat{I}^{a*}$ under H^a_0. Using $M_\Lambda = I_T - P_\Lambda$, where $P_\Lambda = \Lambda(\Lambda'\Lambda)^{-1}\Lambda'$, we can split up $I\!I^a_{n,3}$ as follows

$$\begin{aligned} I\!I^a_{n3} &= \frac{1}{N^2T^2|H_z|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\epsilon'_{i.}K(Z,Z)\epsilon_{j.} + \frac{1}{N^2T^2|H_z|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\epsilon'_{i.}P'_\Lambda K(Z,Z)P_\Lambda\epsilon_{j.} \\ &- \frac{2}{N^2T^2|H_z|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\epsilon'_{i.}P'_\Lambda K(Z,Z)\epsilon_{j.} \\ &= I\!I^a_{n3,1} + I\!I^a_{n3,2} - 2I\!I^a_{n3,3}. \end{aligned} \tag{II.21}$$

Let (c_{κ_1,κ_2}) be the (κ_1,κ_2) element of $(\Lambda'\Lambda)^{-1}$ for $1 \leq \kappa_1$ and $\kappa_2 \leq \ell$, it can be proved that $T^{-1}\Lambda'\Lambda - E(\lambda_t\lambda_t') = O_p(T^{-1/2})$. Then, following a similar reasoning as in Lemma A.2.2 in Cai et al. (2020) it can be shown that $NT|H_z|^{1/2}I\!I^a_{n3,2} = o_p(1)$ and $NT|H_z|^{1/2}I\!I^a_{n3,3} = o_p(1)$, so it yields

$$NT|H_z|^{1/2}I\!I^a_{n3} = NT|H_z|^{1/2}I\!I^a_{n3,1} + o_p(1). \tag{II.22}$$

To derive the asymptotic distribution of $\widehat{I}^a$ we can write $TI\!I^a_{n3,1} = \left(\frac{N-1}{N}\right)U^a_n$, where

$$U^a_n = \frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=1+i}^{N}H^a_n(\chi^a_i,\chi^a_j),$$

$H^a_n(\chi^a_i,\chi^a_j) = T^{-1}\sum_{t=1}^{T}\sum_{s\neq t}\epsilon_{it}\epsilon_{js}K_H(z_t - z_s)$, and $\chi^a_i = (\epsilon_{i\cdot}, Z)$ is a symmetric, centered and degenerate variable.

Note that despite the fact that U^a_n possesses the form of a U-statistic, we cannot resort to the classical CLT for the second-order degenerate U-statistic (see Hall (1984)) due to the existence of CSD. However, χ^a_i is *i.i.d.* across i conditional on both the observable and unobservable common factors (i.e. $\mathbb{D} = \{(d_t, z_t, f_t) : t = 1, \ldots, T\}$. Therefore, we can apply Theorem 1 in Hall (1984) to derive the asymptotic normality of U^a_n since straightforward calculations yield

(i) $E_{\mathbb{D}}[H^a_n(\chi^a_1,\chi^a_2)|\chi^a_1] = 0$,

(ii) $\frac{E_{\mathbb{D}}[\{H^a_n(\chi^a_1,\chi^a_2)\}^2 + N^{-1}E_{\mathbb{D}}[\{H^a_n(\chi^a_1,\chi^a_2)\}^4]]}{\{E_{\mathbb{D}}[\{H^a_n(\chi^a_1,\chi^a_2)\}^2]\}^2} = \frac{O_p(|H_z|^{-1}) + O_p(N^{-1}|H_z|^{-2})}{O_p(|H_z|^{-2})} \to 0$ provided that $|H_z| \to 0$ and

$N|H_z| \to \infty$ as $N \to \infty$,

where $H_n^a(\chi_1^a, \chi_2^a) = E_{\mathbb{D}}[H_n^a(\chi_3^a, \chi_1^a) H_n^a(\chi_3^a, \chi_2^a) | \chi_1^a, \chi_2^a]$.

Then, applying Hall's (1984) CLT, one can show that under both H_0^a and H_1^a,

$$\frac{NT I\!I_{n_3,1}^a}{2E_{\mathbb{D}}[\{H_n^a(\chi_1^a, \chi_2^a)\}^2]} \quad \xrightarrow{d} \quad N(0,1), \tag{II.23}$$

where, under the assumptions of the theorem, it can be shown

$$\begin{aligned} E_{\mathbb{D}}[\{H_n^a(\chi_1^a, \chi_2^a)\}^2] &= \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s \neq t} E_{\mathbb{D}} \left[\epsilon_{1t}^2 \epsilon_{2s}^2 K_H^2(z_t - z_s) \right] (1 + o_p(1)) \\ &= \frac{\overline{\sigma}_\epsilon^4}{|H_z|} E[\rho_z(z_1)] \int K^2(u) du (1 + o_p(1)). \end{aligned} \tag{II.24}$$

Using (II.20)-(II.24) in (II.19), it is shown that

$$NT|H_z|^{1/2} \widehat{I}^a = [NT|H_z|^{1/2} I_{n_3,1}^a + O_p(|H_z|^{1/2}) + o_p(1)](1 + o_p(1)).$$

Hence, we can conclude that

$$NT|H_z|^{1/2} \widehat{I}^a / \sqrt{\Sigma_a} \quad \xrightarrow{d} \quad N(0,1), \tag{II.25}$$

where $\Sigma_a = 2\overline{\sigma}_\epsilon^2 E[\rho_z(z_1)] \int K^2(u) du$. Finally, in Lemma 6 it is shown the asymptotic equivalence between $\widehat{\Sigma}_a$ and Σ_a. Then, the proof of the theorem is done.

∎

Proof Theorem 4.2: Assumptions S1.14–S1.15 ensure $\widehat{\pi} = \overline{\pi} + O_p((NT)^{-1/2})$, where $\overline{\pi}$ is the probability limit of $\widehat{\pi}$. In addition, under H_1^a we have $\widehat{\ddot{\epsilon}}_{a,i\cdot} = M_\Lambda(\epsilon_{i\cdot} - X_{i\cdot}(\widehat{\beta}_{SP} - \beta) + [m(Z) - \phi(Z, \overline{\pi})] + F\gamma_i + D\alpha_i$. Following a similar reasoning as in the proof of Theorem 4.1 and replacing $\widehat{\pi}$ by $\overline{\pi}$, we get $\widehat{\ddot{\epsilon}}_{a,i\cdot} = M_\Lambda(\epsilon_{i\cdot} - X_{i\cdot}(\widehat{\beta}_{SP} - \beta) + [m(Z) - \phi(Z, \pi)]) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\sqrt{\frac{T}{N}}\right)$ since

$M_\Lambda D = 0$ and $M_\Lambda F = O_p\left(\sqrt{\frac{T}{N}}\right)$. Letting $\mathcal{M}(Z,\overline{\pi}) = m(Z) - \phi(Z,\overline{\pi})$, as $T/N \to 0$ and $NT \to \infty$ it can be shown that the test statistic to study is such as

$$\widehat{I}^a = \left(I\!I^a_{n_3} + I\!I^a_{n_4}\right)(1 + o_p(1)), \tag{II.26}$$

where $I\!I^a_{n_4} = \frac{1}{T^2|H_z|}\sum_{t=1}^T\sum_{s=1}^T \mathcal{M}(z_t,\overline{\pi})\mathcal{M}(z_s,\overline{\pi})K(z_t,z_s)$. Note that $I\!I^a_{n_3}$ is the same term analyzed previously, so we can conclude $I\!I^a_{n_3} = O_p((NT|H_z|^{1/2})^{-1})$. Considering now the behaviour of $I\!I^a_{n_4}$, we get

$$\begin{aligned}
E[I\!I^a_{n_4}] &= \frac{1}{T^2|H_z|}\sum_{t=1}^T\sum_{s\neq t}^T \mathcal{M}(z_t,\overline{\pi})\mathcal{M}(z_s,\overline{\pi})K\left(\frac{z_t - z_s}{H_z}\right)\rho(z_t)\rho(z_s)dz_t dz_s \\
&= \frac{1}{T^2}\sum_{t=1}^T\sum_{s\neq t}^T \int \mathcal{M}(z_t,\overline{\pi})\mathcal{M}(z_t - H_z u,\overline{\pi})K(u)\rho(z_t)\rho(z_t - H_z u)dz_t du \\
&= \frac{1}{T^2}\sum_{t=1}^T\sum_{s\neq t}^T \int [\mathcal{M}(z_t,\overline{\pi})]^2 K(u)\rho^2(z_t)dz_t du + o_p(1) \\
&= E\{[\mathcal{M}(z_1,\overline{\pi})]^2\}\rho(z_1)] + o_p(1) = C_a + o_p(1),
\end{aligned} \tag{II.27}$$

where $C_a = E\{[\mathcal{M}(z_1,\overline{\pi})]^2\rho(z_1)\} > 0$ is a positive constant. Also, it is straightforward to show $E[\{T^{-2}|H_z|^{-1}\sum_{t=1}^T\sum_{s\neq t}^T \mathcal{M}(z_t,\overline{\pi})\mathcal{M}(z_s,\overline{\pi})K(z_t,z_s)\}^2] = O_p(|H_z|^{-2}) = o_p(NT)$. Then, it follows from Lemma 3.1 in Powell et al. (1989) that $I\!I^a_{n_4} = C_a + o_p(1)$ or, equivalently, $I\!I_{n_4} \xrightarrow{p} C_a > 0$. Therefore, using these results we can conclude that, under H_1^a, $\widehat{I}^a = C_a + O_p\left(\frac{1}{NT}\right) + O_p\left(\frac{1}{NT|H_z|^{1/2}}\right)$.

Finally, following a similar reasoning as in the proof of Theorem 2.2. in Lin et al. (2014) it is straightforward to show that $\widehat{\Sigma}_a$ converges to a finite positive constant under H_1^a (i.e., $\widehat{\Sigma}_a \xrightarrow{p} \Sigma_a$ as $(N,T) \to \infty$). Then, combining the results above together, it is proved that, under the alternative hypothesis, J^a diverges to $(+\infty)$ at the rate of $NT|H_z|^{1/2}$. This complete the proof of the Theorem.

∎

Proof Theorem 4.3: For the sake of simplicity we denote $W_{i\cdot} = (w'_{i1},\ldots,w'_{iT})'$ as a $T\times(p+q)$ matrix of covariates, whose tth element is such as $w_{it} = (x_{it}, z_t)$, and $\widehat{\rho}_z = (\widehat{\rho}_z(z_1),\ldots,\widehat{\rho}_z(z_T))$ is a

$T \times T$ diagonal matrix. Then, the test statistic $\widehat{I}^b$ can be rewritten as

$$\widehat{I}^b = \frac{1}{N^2T^2|H_z||H_x|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\widehat{\ddot{\epsilon}}'_{b,i\cdot}\widehat{\rho}'_z K(W_{i\cdot}, W_{j\cdot})\widehat{\rho}_z\widehat{\ddot{\epsilon}}_{b,j\cdot}, \tag{II.28}$$

where $K(W_{i\cdot}, W_{j\cdot})$ is a $T\times T$ diagonal matrix of the form $K(W_{i\cdot}, W_{j\cdot}) = diag(K(W_{i1}, W_{j1}), \ldots, K(W_{iT}, W_{jT}))$.

Following a similar reasoning as in (II.19), it is straightforward to show that, under H_0^b, $\widehat{\ddot{\epsilon}}_{b,i\cdot} = M_\Lambda[\epsilon_{i\cdot} + (m_i(Z) - \widehat{m}_{CCE,i}(Z))] - M_\Lambda X_{i\cdot}(\widehat{\beta}_{HSCCEP} - \beta) + o_p(1)$, as $T/N \to 0$. Then, the test statistic to be analyzed is such as

$$\begin{aligned}
\widehat{I}^b &= \frac{1}{N^2T^2|H_z||H_x|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}[\epsilon_{i\cdot} + (m_i(Z) - \widehat{m}_{CCE,i}(Z))]'M'_\Lambda\widehat{\rho}'_z K(W_{i\cdot}, W_{j\cdot})\widehat{\rho}_z M_\Lambda[\epsilon_{j\cdot} + (m_j(Z) - \widehat{m}_{CCE,j}(Z))] \\
&+ \frac{1}{N^2T^2|H_z||H_x|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}(\widehat{\beta}_{HSCCEP} - \beta)'X'_{i\cdot}M'_\Lambda\widehat{\rho}'_z K(W_{i\cdot}, W_{j\cdot})\widehat{\rho}_z M_\Lambda X_{j\cdot}(\widehat{\beta}_{HSCCEP} - \beta) \\
&- \frac{2}{N^2T^2|H_z||H_x|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}[\epsilon_{i\cdot} + (m_i(Z) - \widehat{m}_{CCE,i}(Z))]'M'_\Lambda\widehat{\rho}'_z K(W_{i\cdot}, W_{j\cdot})\widehat{\rho}_z M_\Lambda X_{j\cdot}(\widehat{\beta}_{HSCCEP} - \beta) + o_p(1) \\
&= I\!I^b_{n_1} + I\!I^b_{n_2} - 2I\!I^b_{n_3} + o_p(1),
\end{aligned} \tag{II.29}$$

where the definitions of $I\!I^b_{n_l}$, for $l = 1, 2, 3$, should be apparent from the context.

Focusing on the behaviour of $I\!I^b_{n_1}$, it can be written

$$I\!I^b_{n_1} = I\!I^b_{n_1,1} + I\!I^b_{n_1,2} - 2I\!I^b_{n_1,3}, \tag{II.30}$$

where

$$\begin{aligned}
I\!I^b_{n_1,1} &= \frac{1}{N^2T^2|H_z||H_x|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\epsilon'_{i\cdot}M'_\Lambda\widehat{\rho}'_z K(W_{i\cdot}, W_{j\cdot})\widehat{\rho}_z M_\Lambda\epsilon_{j\cdot}, \\
I\!I^b_{n_1,2} &= \frac{1}{N^2T^2|H_z||H_x|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}[\widehat{m}_{CCE,i}(Z) - m_i(Z)]'M'_\Lambda\widehat{\rho}'_z K(W_{i\cdot}, W_{j\cdot})\widehat{\rho}_z M_\Lambda[\widehat{m}_{CCE,j}(Z) - m_j(Z)], \\
I\!I^b_{n_1,3} &= \frac{1}{N^2T^2|H_z||H_x|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\epsilon'_{i\cdot}M'_\Lambda\widehat{\rho}'_z K(W_{i\cdot}, W_{j\cdot})\widehat{\rho}_z M_\Lambda[\widehat{m}_{CCE,j}(Z) - m_j(Z)].
\end{aligned}$$

Analyzing each element of $I\!I^b_{n_1}$ separately, it can be proved

$$NT|H_x|^{1/2}|H_z|^{1/2} I\!I^b_{n_1} = \frac{1}{NT|H_x|^{1/2}|H_z|^{1/2}} \sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T} \epsilon_{it}\epsilon_{js}\widehat{\rho}_z(z_t)\widehat{\rho}_z(z_s)K(z_t,z_s)K(x_{it},x_{js}) + o_p(1), \quad \text{(II.31)}$$

given that, following a similar reasoning as in (II.21), it is straightforward to show

$$\begin{aligned} NT|H_z|^{1/2}|H_x|^{1/2} I\!I^b_{n_1,1} &= \frac{1}{NT|H_x|^{1/2}|H_z|^{1/2}} \sum_{i=1}^{N}\sum_{j\neq i}^{N} \epsilon'_{i\cdot}(I_T - P_\Lambda)'\widehat{\rho}'_z K(W_{i\cdot}, W_{j\cdot})\widehat{\rho}_z(I_T - P_\Lambda)\epsilon_{i\cdot} \\ &= \frac{1}{NT|H_x|^{1/2}|H_z|^{1/2}} \sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T} \epsilon_{it}\epsilon_{js}\widehat{\rho}_z(z_t)\widehat{\rho}_z(z_s)K(z_t,z_s)K(x_{it},x_{js}) + o_p(1). \end{aligned}$$

Also, considering the behavior of $I\!I^b_{n_1,2}$ and $I\!I^b_{n_1,3}$, we have to deal with an extremely tedious proof since there are many summations. Nevertheless, using a symmetry argument and under similar reasoning as above, it can be shown $I\!I^b_{n_1,2} = o_p\left((NT|H_x|^{1/2}|H_z|^{1/2})^{-1}\right)$ and $I\!I^b_{n_1,3} = o_p\left((NT|H_x|^{1/2}|H_z|^{1/2})^{-1}\right)$ following a similar procedure as in the proof of Proposition A1 and A4 in Fan and Li (1996), respectively.

Focusing now on the behavior of $I\!I^b_{n_3}$ and under similar reasoning as in (II.21), we get

$$\begin{aligned} I\!I^b_{n_3} &= \frac{1}{N^2T^2|H_x||H_z|} \sum_{i=1}^{N}\sum_{j\neq i}^{N} [\epsilon_{i\cdot} + (m_i(Z) - \widehat{m}_{CCE,i}(Z))]'M'_\Lambda\widehat{\rho}'_z K(W_{i\cdot}, W_{j\cdot})\widehat{\rho}_z M_\Lambda X_{j\cdot}(\widehat{\beta}_{HSCCEP} - \beta) \\ &= I\!I^b_{n_3,1}(\widehat{\beta}_{HSCCEP} - \beta) + o_p\left(\frac{1}{NT|H_x|^{1/2}|H_z|^{1/2}}\right), \qquad \text{(II.32)} \end{aligned}$$

where $I\!I^b_{n_3,1} = (NT|H_x|^{1/2}|H_z|^{1/2})^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}[\epsilon_{it}+(m_i(z_t)-\widehat{m}_{CCE,i}(z_t))]X'_{js}\widehat{\rho}_{z_t}\widehat{\rho}_{z_s}K(x_{it},x_{js})K(z_t,z_s)$. Note that in order to show $I\!I^b_{n_3} = o_p\left((NT|H_x|^{1/2}|H_z|^{1/2})^{-1}\right)$ we do not require the conditions in Robinson (1988) or Fan et al. (1995). More precisely, under the conditions of Theorem 4.3 in this paper it can be shown $\widehat{\beta}_{HSCCEP} - \beta = O_p\left((NT)^{-1/2} + (NT|\mathcal{H}|)^{-1} + tr\{\mathcal{H}\}^4\right)$ (see Fan et al. (1995) for a detailed proof). Using this result in (II.32) and comparing the terms in $I\!I^b_{n_3,1}$ and the terms in $I\!I^b_{n_1}$, it is obvious that $I\!I^b_{n_3,1}$ has at most an element of order $O_p((NT|H_x|^{1/2}|H_z|^{1/2})^{-1})$. Hence, we can conclude that $I\!I^b_{n_3}$ has the order of $O_p(I\!I^b_{n_3,1})O_p((NT)^{-1/2} + (NT|\mathcal{H}|)^{-1} + tr\{\mathcal{H}\}^4) =$

$o_p((NT|H_x|^{1/2}|H_z|^{1/2})^{-1})$. Similarly, we define $\mathbb{I}^b_{n2,1}$ from $\mathbb{I}^b_{n_2} = (\widehat{\beta}_{HSCCEP} - \beta)'\mathbb{I}^b_{n2,1}(\widehat{\beta}_{HSCCEP} - \beta) + o_p((NT|H_x|^{1/2}|H_z|^{1/2})^{-1})$. Under similar reasoning as above, it can be shown that $E|\mathbb{I}^b_{n2,1}| = O(1)$, so we can conclude $\mathbb{I}^b_{n_2}$ is of order $(\widehat{\beta}_{HSCCEP} - \beta)'(\widehat{\beta}_{HSCCEP} - \beta) = O_p([(NT)^{-1/2} + (NT|\mathcal{H}|)^{-1} + tr\{\mathcal{H}\}^4]^2) = o_p((NT|H_x||H_z|)^{-1})$.

Using all the above results in (II.29) we get

$$NT|H_x|^{1/2}|H_z|^{1/2}\widehat{I}^b = NT|H_x|^{1/2}|H_z|^{1/2}\mathbb{I}^b_n + o_p(1) \tag{II.33}$$

where $\mathbb{I}^b_n$ can be rewritten in terms of a U-statistic to derive the asymptotic distribution of the proposed test statistic $\widehat{I}^b_n$ obtaining

$$\begin{aligned}
\mathbb{I}^b_n &= \frac{1}{N^2T^2|H_x||H_z|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\epsilon_{it}\epsilon_{js}\widehat{\rho}_{z_t}\widehat{\rho}_{z_s}K(x_{it},x_{js})K(z_t,z_s)\\
&= \frac{1}{N^2T^4|H_x||H_z||\mathcal{H}|^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{s_1\neq t_1}^{T}\sum_{s_2\neq t_2}^{T}\epsilon_{it_1}\epsilon_{js_1}K(x_{it_1},x_{js_1})K(z_{t_1},z_{s_1})K(z_{t_1},z_{t_2})K(z_{s_1},z_{s_2}) + \widetilde{I}^b_nR\\
&= \widetilde{I}^b_nU + \widetilde{I}^b_nR,
\end{aligned} \tag{II.34}$$

where $\widetilde{I}^b_nU$ denotes the case where all four temporal subscripts (i.e., t_1, t_2, s_1, s_2) are different and $\widetilde{I}^b_nR$ denotes the sum of the remaining terms. Rewriting $\widetilde{I}^b_nU$ in terms of a U-statistic, it yields $T\widetilde{I}^b_nU = \left(\frac{N-1}{N}\right)U^b_n$, where

$$U^b_n = \frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=1+i}^{N}H^b_n(\chi^b_i,\chi^b_j), \tag{II.35}$$

$H^b_n(\chi_i,\chi_j) = (T^3|H_x||H_z||\mathcal{H}|^2)^{-1}\sum_{t_1\neq t_2\neq s_1\neq s_2}^{T}\epsilon_{it_1}\epsilon_{js_1}K(x_{it_1},x_{js_1})K(z_{t_1},z_{s_1})K(z_{t_1},z_{t_2})K(z_{s_1},z_{s_2})$ and $\chi^b_i = (\epsilon_{i\cdot}, X_{i\cdot}, Z)$. Again, conditional on $\mathbb{D}$, χ^b_i is $i.i.d.$ across i, so we can apply Theorem 1 in Hall (1984) to derive the asymptotic normality of $\widetilde{\mathbb{I}}^b_nU$ since, following a similar proof scheme as in Theorem 3.1 in Cai et al. (2020), it can be show

(i) $E_{\mathbb{D}}[H^b_n(\chi^b_1,\chi^b_2)|\chi^b_1] = 0$.

(ii) $\frac{E_{\mathbb{D}}[\{H_n^b(\chi_1^b,\chi_2^b)\}^2+N^{-1}E_{\mathbb{D}}[\{H_n^b(\chi_1^b,\chi_2^b)\}^4]]}{\{E_{\mathbb{D}}[\{H_n^b(\chi_1^b,\chi_2^b)\}^2]\}^2} = \frac{O_p(|H_x|^{-1}|H_z|^{-1})+O_p(N^{-1}|H_z|^{-2}|H_x|^{-2}|\mathcal{H}|^{-4})}{O_p(|H_z|^{-2}|H_x|^{-2}|\mathcal{H}|^4)} \to 0$, provided that $N \to \infty$ and Assumption 4.6,

where $H_n^b(\chi_1^b, \chi_2^b) = E_{\mathbb{D}}[H_n^b(\chi_3^b, \chi_1^b)H_n^b(\chi_3^b, \chi_2^b)|\chi_1^b, \chi_2^b]$. Further, by extending the proof scheme in the Proposition A.2 in Fan and Li (1996) to this particular scheme it is straightforward to show $\widetilde{I}_n^b R_n = O_p((NT|H_x|^{1/2}|H_z|^{1/2})^{-1})$. Then, by Hall's CLT, one can show that under both H_0^b and H_1^b,

$$\frac{NT I\!\!I_n^b}{2E_{\mathbb{D}}[\{H_n^b(\chi_1^b, \chi_2^b)\}^2]} \quad \xrightarrow{d} \quad N(0,1) \tag{II.36}$$

where, under the assumptions of the theorem, it can be shown

$$\begin{aligned} E_{\mathbb{D}}[\{H_n^b(\chi_1^b, \chi_2^b)\}^2] &= \frac{1}{T^2|H_x|^2|H_z|^2|\mathcal{H}|^4}\sum_{t=1}^{T}\sum_{s\neq t}^{T} E_{\mathbb{D}}[\epsilon_{1t}^2\epsilon_{2s}^2 K^2(x_{1t}, x_{2s})K^2(z_t, z_s)\widehat{\rho}_z^2(z_t)\widehat{\rho}_z^2(z_s)(1+o_p(1)) \\ &= \frac{\overline{\sigma}_\epsilon^4}{|H_x||H_z|}E[\rho_{x,z}(x_{1t}, z_t)\rho_z^4(z_t)] \int K^2(u_1)du_1 \int K^2(u_2)du_2(1+o_p(1)). \end{aligned} \tag{II.37}$$

Therefore, using (II.35)-(II.37) in (II.34) it is shown $NT|H_x|^{1/2}|H_z|^{1/2}\widehat{I}^b = [NT|H_x|^{1/2}|H_z|^{1/2} I\!\!I_n^b + o_p(1)$ so we can conclude

$$NT|H_x|^{1/2}|H_z|^{1/2}\widehat{I}_n^b/\sqrt{\Sigma_b} \quad \xrightarrow{d} \quad N(0,1), \tag{II.38}$$

where $\Sigma_b = 2\overline{\sigma}_\epsilon^4 E[\rho_{x,z}(x_{1t}, z_t)\rho_z^4(z_t)] \int K^2(u_1)du_1 \int K^2(u_2)du_2$.

Finally, in order to verify the asymptotic equivalence between $\widehat{\Sigma}_b$ and Σ_b, we follow a similar proof scheme as in Lemma 6. From (II.37) it can be verified that $E\{E_{\mathbb{D}}[H_n^b(\chi_i^b, \chi_j^b)]^2\} = o_p(NT)$ so the Lemma 3.1 in Powell et al. (1989) can be used. Hence, we can conclude that $\widehat{\Sigma}_b = \Sigma_b + o_p(1)$ and the proof of the theorem is done.

∎

Proof Theorem 4.4: We denote $\widetilde{\Pi}(x_{it}) = x_{it}'(\beta_i - \beta)$. Following a similar reasoning as in the proof of Theorem 4.2 and using the facts that $\widehat{m}_{CCE,i}(z_t) - m_i(z_t) = O_p((T|H_z|)^{-1/2})$ and $\widehat{\beta}_{HSCCEP} - \beta = O_p((NT)^{-1/2} + (NT|\mathcal{H}|)^{-1} + tr\{\mathcal{H}\}^4)$, under H_1^b we get $\widehat{\widehat{\epsilon}}_{b,i\cdot} = M_\Lambda(\epsilon_{i\cdot} + \widetilde{\Pi}(X_{i\cdot})) + o_p(1)$, as $T/N \to 0$, $T|H_z| \to \infty$, as $(N,T) \to \infty$. Then, the test statistic to be analyzed is such as

$$\begin{aligned}
\widehat{I}^b &= \frac{1}{N^2T^2|H_x||H_z|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\epsilon_{i\cdot}'M_\Lambda'\widehat{\rho}_z'K(W_{i\cdot},W_{j\cdot})\widehat{\rho}_z M_\Lambda\epsilon_{j\cdot} \\
&+ \frac{1}{N^2T^2|H_x||H_z|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\epsilon_{i\cdot}'M_\Lambda'\widehat{\rho}_z'K(W_{i\cdot},W_{j\cdot})\widehat{\rho}_z M_\Lambda\widetilde{\Pi}(X_{j\cdot}) \\
&+ \frac{1}{N^2T^2|H_x||H_z|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\widetilde{\Pi}(X_{i\cdot})'M_\Lambda'\widehat{\rho}_z'K(W_{i\cdot},W_{j\cdot})\widehat{\rho}_z M_\Lambda\widetilde{\Pi}(X_{j\cdot}) + o_p(1) \\
&= I\!I_{n_4}^b + I\!I_{n_5}^b + I\!I_{n_6}^b, \qquad \text{(II.39)}
\end{aligned}$$

where the definitions of $I\!I_{n_r}^b$, for $r = 4,5,6$, should be apparent from the context.

Following a similar reasoning as for $I\!I_{n_1}^b$ it can be proved that $I\!I_{n_4}^b$ is of order $O_p((NT|H_x|^{1/2}|H_z|^{1/2})^{-1})$ whereas $I\!I_{n_5}^b$ is $o_p((NT|H_x|^{1/2}|H_z|^{1/2})^{-1})$. Then, analyzing the behaviour of $I\!I_{n_6}^b$ and using a similar proof scheme as in (II.21) and by the law of large numbers, we get

$$\begin{aligned}
I\!I_{n_6}^b &= \frac{1}{N^2T^2|H_x||H_z|}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\widetilde{\Pi}(X_{it})\widetilde{\Pi}(X_{js})\widehat{\rho}_z(z_t)\widehat{\rho}_z(z_s)K(x_{it},x_{js})K(z_t,z_s) + o_p(1) \\
&= \frac{1}{T^2|H_x||H_z|}\sum_{t=1}^{T}\sum_{s\neq t}^{T}E[\widetilde{\Pi}(X_{it})\widetilde{\Pi}(X_{js})\widehat{\rho}_z(z_t)\widehat{\rho}_z(z_s)K(x_{it},x_{js})K(z_t,z_s)] + o_p(1) \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\int\widetilde{\Pi}(X_{it})^2K(u_1)K(u_2)K(v_1)K(v_2)du_1du_2dv_1dv_2\rho_{x,z}^2(x_{it},z_t)\rho_z^2(z_t)dx_{it}dz_t + o_p(1) \\
&= E[\widetilde{\Pi}(X_{it})^2\rho_{x,z}(x_{it},z_t)\rho_z^2(z_t)] + o_p(1) = C_b + o_p(1), \qquad \text{(II.40)}
\end{aligned}$$

where $C_b = E[\widetilde{\Pi}(X_{it})^2\rho_{x,z}(x_{it},z_t)\rho_z^2(z_t)] > 0$ is a positive constant. Also, it can be shown that $Var[I\!I_{n_6}^b] = O_p(|H_z|^{-2}|H_x|^{-2}) = o_p(NT)$. Then, it follows from Lemma 3.1 in Powell et al. (1989) that $I\!I_{n_6}^b \xrightarrow{p} C_b > 0$. Using all these results in (II.39) we can conclude that, under H_1^b, $\widehat{I}^b = C_b + O_p\left(\frac{1}{NT|H_x|^{1/2}|H_z|^{1/2}}\right) + o_p\left(\frac{1}{NT|H_x|^{1/2}|H_z|^{1/2}}\right)$. Finally, following a similar reasoning as in Lemma 6 we can prove $\widehat{\Sigma}_b = \Sigma_b + o_p(1)$ under H_1^b. This complete the proof of the theorem.

References

Bernstein, D. (2005). *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*. Princeton University Press.

Cai, Z., Y. Fang, and Q. Xu (2020). Testing capital asset pricing models using functional-coefficient panel data models with cross-sectional dependence. *Journal of Econometrics (in press)*.

Coe, D. T. and E. Helpman (1995). International r&d spillovers. *European Economic Review 39*(5), 859–887.

Coe, D. T., E. Helpman, and A. W. Hoffmaister (2009). International r&d spillovers and institutions. *European Economic Review 53*(7), 723–741.

Davis, S. J. and J. Haltiwanger (2001). Sectoral job creation and destruction responses to oil price changes. *Journal of Monetary Economics 48*(3), 465–512.

Eberhardt, M., C. Helmers, and H. Strauss (2013). Do spillovers matter when estimating private returns to R&D? *Review of Economics and Statistics 95*(2), 436–448.

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.

Fan, J., Q. Li, and T. Stengos (1995). Root-n-consistent semiparametric regression with conditional heteroscedastic disturbances. *Journal of Quantitative Economics 11*, 229–240.

Fan, Y. and Q. Li (1996). Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica 64*, 865–90.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015). The next generation of the Penn World Table. *American Economic Review 105*(10), 3150–82.

Griliches, Z. (1979). Issues in Assessing the Contribution of Research and Development to Productivity Growth. *The Bell Journal of Economics 10*(1), 92–116.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis 14*, 1–16.

Hamilton, J. D. (1996). This is what happened to the oil price-macroeconomy relationship. *Journal of Monetary Economics 38*(2), 215–220.

Hansen, B. (2000). Testing for structural change in conditional moments. *Journal of Econometrics 97*, 93–115.

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

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

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

Lin, Z., Q. Li, and Y. Sun (2014). A consistent nonparametric test of parametric regression functional form in fixed effects panel data models. *Journal of Econometrics 178*, 167–179.

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

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

Powell, J. L., J. H. Stock, and T. M. Stoker (1989). Semiparametric estimation of index coefficients. *Econometrica 57*(6), 1403–1430.

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