Monash University

Monash Business School

DEPARTMENT OF ECONOMETRCS
AND BUSINESS STATISTICS

ISSN 1440-771X

WORKING PAPER SERIES

Mean Group Instrumental Variable Estimation of Time-Varying Large Heterogenous Panels with Endogenous Regressors

Yu Bai, Massimiliano Marcellino and George Kapetanios

Working Paper No. 13/23
June 2023

Mean group instrumental variable estimation of time-varying large heterogeneous panels with endogenous regressors

Yu Bai[1 a], Massimiliano Marcellino[b], George Kapetanios[c]

[a]*Monash University*
[b]*Bocconi University, IGIER, CEPR, Baffi Carefin and BIDSA*
[c]*King's College London*

Abstract

The large heterogeneous panel data models are extended to the setting where the heterogenous coefficients are changing over time and the regressors are endogenous. Kernel-based non-parametric time-varying parameter instrumental variable mean group (TVP-IV-MG) estimator is proposed for the time-varying cross-sectional mean coefficients. The uniform consistency is shown and the pointwise asymptotic normality of the proposed estimator is derived. A data-driven bandwidth selection procedure is also proposed. The finite sample performance of the proposed estimator is investigated through a Monte Carlo study and an empirical application on multi-country Phillips curve with time-varying parameters.

Keywords: Large heterogeneous panels, Non-Parametric Methods, Time-varying parameters, Mean Group estimator

JEL classification: C14, C26, C51

---

[1]Corresponding author: Yu Bai, Department of Econometrics and Business Statistics, Monash University, 900 Dandenong Rd, Caulfield East, VIC 3145. Email: yu.bai1@monash.edu.

1. Introduction

Since the study by Pesaran and Smith (1995), large heterogeneous panel data models have received a lot of attention in both theoretical work and practical applications. Double-index panel data models enable researchers to explore both dynamic information over the time-span and heterogeneity over cross-sections, which may be difficult to examine by applying purely cross sectional or time series models. It is now quite common to have panels with both large cross-sectional units (N) and time-series periods (T) and it has been found that neglected heterogeneity may lead to misleading inferences in empirical applications (see, for instance, Haque et al. (1999)).

Various methods have been proposed to identify and handle structural change in econometric models. As parameter instability is pervasive (see e.g. Stock and Watson (1996)), allowing coefficients to vary over time would offer benefits for flexible modeling of the true relationship between economic and financial variables. One stand of the vast and growing literature on dealing with parameter instability allows for a smooth evolution of parameters without specifying the form of parameter time variation. The true time-varying parameters can be either smooth deterministic functions of time, as in Robinson (1991) and Chen and Hong (2012), or persistent stochastic processes, as in Giraitis et al. (2014, 2018, 2021). These papers have provided theoretical, Monte Carlo and empirical results to justify their estimation methods and showed that they indeed perform very well in finite samples. The approach has also been extended to panel data models. Chen and Huang (2018) propose methods to estimate and test smooth structural changes in panel data models with exogenous regressors and homogenous time-varying coefficients.

When the parameters of interests are coefficients attached to endogenous variables, endogeneity bias invalidates least square estimation and instrumental variable (IV) estimation comes to play a role. A usual assumption made when carrying out IV estimation is that the parameters in the entertained model are constant over time. This assumption is clearly restrictive, because relations between economic variables as well as instruments and endogenous variables may vary over time. Recently, some papers have attempted to develop estimation methods in the time-series IV framework that account for the possible presence of parameter instability with smooth structural change. Chen (2015) extends the framework of deterministic smooth evolution of parameters to the IV case. Giraitis et al. (2021) propose non-parametric kernel-based estimation and inferential theory for time-varying IV regression. The true time-varying parameters are assumed to be either deterministic sequences, or random processes. There is also limited but growing attention in the panel data literature to models allowing for endogenous regressors and parameter instability. Baltagi et al. (2019) develop an estimation procedure for large heterogeneous panels with endogenous regressors and cross-sectional dependence (Castagnetti et al. (2019) and Pesaran (2006)) in the structural change context. Jiang and Kurozumi (2023) propose a test to detect whether break points are common in heterogeneous panel data models when T is large relative to N.

The aim of this article is to extend the literature of large heterogenous panels to the setting where the heterogenous coefficients are changing over time and the regressors are endogenous. The model includes standard random coefficient panel data models (Hsiao and Pesaran (2008)) as a special case in which coefficients remain constant over time. The model also extends the work of Chen and Huang (2018) by allowing cross-sectional heterogeneity of the coefficients and endogenous regressors. In addition, Chen and Huang (2018) follows the standard assumption in the nonparametric modeling of time-varying parameters literature by assuming the time variation is very smooth, but the degree of smoothness is generally unknown. Our framework allows a much wider range of smoothness in parameter variation compared to Chen and Huang (2018).

We propose a time-varying parameter mean group instrumental variable (TVP-IV-MG) estimator for the time-varying cross-sectional mean coefficients. The estimator is obtained by simply taking cross-sectional averages of the non-parametric kernel-based estimators proposed in Giraitis et al. (2021) in the time-series context. We show the uniform consistency and derive the pointwise asymptotic normality of the estimator under very mild conditions. It is shown that the pointwise asymptotic expansion of the estimator takes a simple form, which implies that our approach is broadly applicable, even in the correlated random coefficient setting (Hsiao et al. (2019)). As our approach requires selection of bandwidth parameters which are related to the unknown degree of the smoothness of time-varying parameters, we also propose a data-driven bandwidth selection procedure. The finite sample performance of the proposed estimator is evaluated in a Monte Carlo study. We evaluate the biases and coverage probabilities of the estimator. The results are encouraging and clearly show the superiority of the data-driven bandwidth selection procedure compared to the rule-of-thumb approach.

Finally, we provide an empirical application to explore in practice the use of our TVP-IV-MG estimator. We estimate panel versions of time-varying hybrid Phillips curves with 19 Eurozone countries over the period 2000M1–2019M12. We find that the Eurozone Phillips curve is alive after 2004, but gradually flattening after 2010 and with the forward looking component generally dominating the backward looking one. The time series estimation based on aggregated Eurozone data delivers similar point estimates, but with much wider confidence intervals. This shows that our panel methods are more efficient than the time-series approach.

The remainder of this paper is organized as follows. Section 2 describes our framework, the TVP-IV-MG estimator, derives the related theoretical results and introduces the data-driven bandwidth selection procedure. In Section 3 we evaluate our proposed estimator and the bandwidth selection procedure in a Monte Carlo study. Section 4 presents an empirical application related to multi-country Phillips curves. Section 5 concludes the paper. The appendix contains definition of notations, technical lemma and proofs of the main results.

2. Theoretical considerations

2.1. Model and the estimator

Giraitis et al. (2021) introduced a non-parametric time-varying instrumental variable (IV) estimation method in the time-series context. The main innovation of their work is to show the uniform consistency and asymptotic normality of kernel-based estimator in the presence of time-varying coefficients, which evolve as either smooth deterministic functions of (scaled) time or persistent stochastic processes. We would like to extend their results in the context of large heterogeneous panel data models.

Let us consider the following model:

$$y_{it} = x_{it}'\beta_{it} + u_{it}, \tag{1}$$

$$x_{it} = \Psi_{it}'z_{it} + v_{it}, \quad i = 1, 2, \cdots, N, \quad t = 1, 2, \cdots, T, \tag{2}$$

where y_{it} denotes the explained variable, x_{it} is a $k \times 1$ dimensional vector of explanatory variables, one of them being a constant (e.g., $x_{1,it} = 1$), u_{it} is the disturbance term, $z_{it} = (z_{1,it}, z_{2,it}, \cdots, z_{p,it})'$ is a $p \times 1$ ($p \geqslant k$) vector of instruments, and $v_{it} = (v_{1,it}, v_{2,it}, \cdots, v_{k,it})'$ is a $k \times 1$ vector of error terms. $\Psi_{it} = (\psi_{\ell_1\ell_2,it})$ is a $p \times k$ random parameter matrix and β_{it} is a $k \times 1$ vector of random structural parameters of interests. We assume that for each i, the endogenous variables x_{it} are correlated with u_{it}, but there exist exogenous instruments z_{it} which are uncorrelated with u_{it} and v_{it} but are correlated with x_{it}.

Remark 1. An important issue relates to the modelling of fixed effects. This is addressed within our framework in two closely related but distinct ways. Firstly, one can simply set $x_{1,it} = 1$. Then, $\beta_{1,it}$ is the fixed effect which, in analogy to the rest of the regressor coefficients, can slowly vary over t. An alternative way is to let the model be either $y_{it} = \alpha_i + x_{it}'\beta_{it} + u_{it}$ or $y_{it} = \alpha_{it} + x_{it}'\beta_{it} + u_{it}$, and then simply regress separately y_{it} and x_{it} on a constant, obtaining the residuals $\tilde{y}_{it}$ and $\tilde{x}_{it}$ and then proceed with the model $\tilde{y}_{it} = \tilde{x}_{it}'\beta_{it} + u_{it}$. The preliminary regression on the constant can be either time varying, if the fixed effect, α_{it}, is assumed to change slowly over time, in a similar fashion to β_{it}, or not, if the effect is assumed fixed over time (α_i). It is straightforward to show that this preliminary step does not affect the subsequent theoretical analysis.

The objective of this paper is to construct consistent estimators of the cross-section time-varying mean coefficients: $\beta_{0,t} = E(\beta_{it})$ and derive asymptotic normality for the estimator of $\beta_{0,t}$. As in Giraitis et al. (2021), the individual time-varying IV estimation is obtained by the kernel type estimates

$$\hat{\beta}_{it}^{IV} = \Big(\sum_{j=1}^{T} b_{j,t}(H)\hat{\Psi}_{ij}'z_{ij}x_{ij}'\Big)^{-1}\Big(\sum_{j=1}^{T} b_{j,t}(H)\hat{\Psi}_{ij}'z_{ij}y_{ij}\Big), \tag{3}$$

where $\hat{\Psi}_{it}$ is the first-stage time-varying OLS estimator

$$\hat{\Psi}_{it} = \Big(\sum_{j=1}^{T} b_{j,t}(L)z_{ij}z_{ij}'\Big)^{-1}\Big(\sum_{j=1}^{T} b_{j,t}(L)z_{ij}x_{ij}'\Big). \tag{4}$$

(3) and (4) are computed with kernel weights

$$b_{j,t}(H) = K\Big(\frac{|j-t|}{H}\Big), \;\; b_{j,t}(L) = K\Big(\frac{|j-t|}{L}\Big), \tag{5}$$

where the bandwidth parameters H and L are such that $H = o(T)$, $L = o(T)$ as $T \to \infty$ and L can be different from H. Both H and L are assumed to satisfy another condition:

$$c_1 T^{1/(\theta/4-1)+\delta} \leqslant H, L \leqslant c_2 T^{1-\delta} \tag{6}$$

for some $\theta > 4$ as defined in Assumption 2.1(i), $c_1, c_2 > 0$ and $\delta > 0$ is arbitrarily small. $K(x)$ defined in (5) is a non-negative continuous function with either bounded or unbounded support satisfying

$$|K(x)| \leqslant C(1 + x^{\nu})^{-1}, \quad |(d/dx)K(x)| \leqslant C(1 + x^{\nu})^{-1}, \tag{7}$$

for some $C > 0$ and $\nu \geqslant 2$. Examples include $K(x) = \frac{1}{2}I\{|x| \leqslant 1\}$ (uniform kernel), $K(x) = \frac{3}{4}(1 - x^2)I\{|x| \leqslant 1\}$ (Epanechnikov kernel) and $K(x) \propto \exp(-cx^{\alpha})$ with $c > 0$, $\alpha > 0$ (Gaussian kernel).

Then, the time-varying instrumental variable mean group (TVP-IV-MG) estimator is defined as a simple average of the individual time-varying IV estimators

$$\hat{\beta}^{IV}_{MG,t} = \frac{1}{N}\sum_{i=1}^{N} \hat{\beta}^{IV}_{i,t}, \tag{8}$$

which generalizes the OLS-type MG estimator for random coefficient panel data with constant mean coefficients, as in Pesaran and Smith (1995).

2.2. *Asymptotic properties*

Before presenting the asymptotic properties of the estimator (8), we outline assumptions on x_{it}, z_{it}, u_{it}, v_{it} and time-varying parameters β_{it}, Ψ_{it}.

Assumption 2.1. *For each i, elements in x_{it}, z_{it}, u_{it}, v_{it} have the properties:*

(i) There exists $\theta > 8$ such that uniformly over ℓ, t

$$E\big|x_{\ell,it}\big|^{\theta}, \; E\big|u_{\ell,it}\big|^{\theta}, \; E\big|z_{\ell,it}\big|^{\theta}, \; E\big|v_{\ell,it}\big|^{\theta} \leqslant C < \infty \tag{9}$$

(ii) $\forall \ell, i, t$, $(x_{\ell,it} - Ex_{\ell,it})$, $(u_{\ell,it})$, $(z_{\ell,it} - Ez_{\ell,it})$, and $(v_{\ell,it})$ are strong-mixing processes with mixing coefficients $\alpha_k^{i,j}$ satisfying

$$\alpha_k^{i,j} \leqslant c_{i,j}\phi_{i,j}^k, \;\; k \geqslant 1 \tag{10}$$

for some $0 < \phi_{i,j} < 1$ and $c_{i,j} > 0$, where $j = \{x, u, z, v\}$;

(iii) For each i, the matrix $\Sigma^i_{zz,t} = E[z_{it}z'_{it}]$ *is such that* $\max_{t\geqslant 1}\left\|(\Sigma^i_{zz,t})^{-1}\right\|_{sp} < \infty$.

Assumption 2.2. $(x_i, z_i, u_i, v_i, e_i, \Psi_i)$ *are mutually independently distributed over i, where* $e_i = (e_{i1}, e_{i2}, \cdots, e_{iT})'$, $u_i = (u_{i1}, u_{i2}, \cdots, u_{iT})'$, $v_i = (v_{i1}, v_{i2}, \cdots, v_{iT})'$, $x_i = (x'_{i1}, x'_{i2}, \cdots, x'_{iT})'$, $z_i = (z'_{i1}, z'_{i2}, \cdots, z'_{iT})'$, $\Psi_i = (\Psi'_{i1}, \Psi'_{i2}, \cdots, \Psi'_{iT})'$.

Assumption 2.3. *For each i, elements in* $\Psi_{it} = (\psi_{\ell_1\ell_2,it})$, $\ell_1 = 1, 2, \cdots, p$, $\ell_2 = 1, 2, \cdots, k$, *are random processes that satisfy the following smoothness condition:*

$$\left|\psi_{\ell_1\ell_2,it} - \psi_{\ell_1\ell_2,is}\right| \leqslant \left(\frac{|t-s|}{T}\right)^{\gamma_1} q_{\ell_1\ell_2,i,ts}, \quad t, s = 1, 2, \cdots, T \tag{11}$$

for some $0 < \gamma_1 \leqslant 1$ *and the distribution of variables* $X = \psi_{\ell_1\ell_2,it}, q_{\ell_1\ell_2,i,ts}$ *has a thin tail* $\mathcal{E}(\alpha)$*:*

$$\mathbb{P}(|X| \geqslant \omega) \leqslant \exp(-c_1|\omega|^{\alpha}), \quad \omega > 0, \tag{12}$$

for $c_1 > 0$, $\alpha > 0$ *which does not depend on* ℓ_1, ℓ_2, t, s *and* T.

Assumption 2.4. *The coefficients* β_{it} *follow the random coefficient model:*

$$\beta_{it} = \beta_{0,t} + e_{it}, \quad i = 1, 2, \cdots, N, \quad t = 1, 2, \cdots, T, \tag{13}$$

where

(i) $(\beta_{0,t}) = (E(\beta_{it}))$ *are the sequences of cross-section time-varying non-random mean coefficients of the processes* (β_{it})*;*

(ii) Elements in $\beta_{0,t} = (\beta_{\ell,0,t})$ *are uniformly bounded in t and satisfy the following smoothness condition:*

$$\left|\beta_{\ell,0,t} - \beta_{\ell,0,s}\right| \leqslant C\left(\frac{|t-s|}{T}\right)^{\gamma_2}, \quad t, s = 1, 2, \cdots, T \tag{14}$$

for some $0 < \gamma_1 \leqslant 1$ *and the positive constant C does not depend on* ℓ, t, s *and* T*;*

(iii) For each i, elements in the random part $e_{it} = (e_{\ell,it})$ *satisfy the following smoothness condition:*

$$\left|e_{\ell,it} - e_{\ell,is}\right| \leqslant \left(\frac{|t-s|}{T}\right)^{\gamma_3} r_{\ell,i,ts}, \quad t, s = 1, 2, \cdots, T, \tag{15}$$

for some $0 < \gamma_3 \leqslant 1$ *and the distribution of each variable in* $X_e = e_{\ell,it}, r_{\ell,i,ts}$ *has a thin tail* $\mathcal{E}(\alpha)$ *as defined in* (12).

Assumption 2.5. *For each i, the following conditions hold* *(i)* $E[u_{it}|z_{it}] = 0$, $E[v_{it}|z_{it}] = 0$;

(ii) *The matrix* $\Sigma^i_{\Psi zz\Psi,t} = \Psi'_{it}\Sigma^i_{zz,t}\Psi_{it}$ *is such that* $\max_{t\geqslant 1} \left\|(\Sigma^i_{\Psi zz\Psi,t})^{-1}\right\|_{sp} \leqslant \nu_i < \infty$ *a.s..*

Assumption 2.6. *The bandwidth parameters* H, L *also satisfy* $H = o\Big(L/(\log T)^{\max(1,2/\alpha)}\Big)$, $L = o(T^{\gamma_1})$.

Assumption 2.1(i) imposes some moment conditions on regressors, instruments and error terms. Assumption 2.1(ii) consists of strong mixing conditions to control temporal dependence, which are weaker than the conditions imposed in Chen (2015) and Chen and Huang (2018), since we allow (u_{it}), (v_{it}), $(z_{it}u_{it})$ and $(z_{it}v_{it})$ to be serially correlated sequences. Assumptions 2.1(i)-(ii) are necessary to apply the exponential inequalities provided in Dendramis et al. (2021) for weighted sums with (possible) random scaling to establish uniform consistency rate and derive pointwise asymptotic normality (See Theorem 2.1). Assumption 2.1(iii) parallels Assumption 1(iii) in Giraitis et al. (2021). Assumption 2.2 implies that all the regressors, error terms and random coefficients are cross-sectionally independent. A similar assumption is also imposed in the literature, e.g., Li et al. (2011).

Assumptions 2.3 and 2.4 are related to model parameters. For each i, Assumption 2.3 imposes exactly the same assumption in Giraitis et al. (2021) when coefficients are random. Regarding to Assumption 2.4, as in the random coefficient panel data model literature (see Hsiao and Pesaran, 2008), we assume an additive structure of β_{it} ((13)) by decomposing it into a deterministic component $\beta_{0,t}$ and a zero-mean random component e_{it}. We are interested in estimation and inference of the cross-sectional time-varying deterministic mean coefficients $\beta_{0,t}$. (14) implies that $\beta_{\ell,0,t} = g_\ell(t/T)$, where the deterministic function $g_\ell(\cdot) : [0,1] \longrightarrow \mathbb{R}$ is Hölder continuous with exponent γ_2. This is more general than the existing literature (see Chen and Huang, 2018), in which valid inference relies on the assumption that $g_\ell(\cdot)$ is smooth: $g_\ell(\cdot)$ has bounded second order derivative, so a subset of cases when $\gamma_2 = 1$. The assumption on the random part e_{it} ((15)) is similar to the one for Ψ_{it}, with a possibly different smoothness parameter. These conditions imply that, for each i, $(\psi_{\ell k,it})_t$ and $(e_{\ell,it})_t$ are persistent stochastic processes with bounded variation. For example, consider an array of random processes (bounded random walk) defined as $e_{\ell,it} = \frac{1}{\sqrt{T}}u_{\ell,it}$, where $u_{\ell,it}$ are random walk processes, satisfying Assumption 2.4(iii). As shown in Lemma 1 of Dendramis et al. (2021), Assumption 2.4(iii) is satisfied if $\omega_{\ell,it} = u_{\ell,it} - u_{\ell,i,t-1}$ is α-mixing and has a thin tail. Assumption 2.4(ii) can also be viewed that $(\beta_{\ell,0,t})$ is a realization of a random process with smoothness parameter γ_2. Other allowable processes are discussed by Dendramis et al. (2021), Giraitis et al. (2014) and Giraitis et al. (2018).

Remark 2. (13) *is a generalization in the literature of random coefficient panel data models. If* $\beta_{0,t} = \beta_0$ *and* $e_{it} = e_i$, *for all* t, *the model simplifies to the standard random coefficient model settings with time-invariant coefficients (see Hsiao and Pesaran, 2008). Horváth and Trapani (2016) generalize the specification in Hsiao and Pesaran (2008) by allowing the heterogenous part to vary over time, where* e_{it} *is assumed to be i.i.d. across* t *for each* i *and mean coefficients* β *are still constant over time.*

Remark 3. *Throughout the paper, we focus on the case when the first stage coefficients Ψ_{it} are random. Giraitis et al. (2021) also consider the case when they are deterministic. It is straightforward to show that our results follow accordingly for the deterministic case.*

Assumption 2.5 contains standard conditions to ensure that the model defined in (1)-(2) is identified. As explained in Chen (2015), these conditions allow dynamic regression models with serial correlation and conditional heteroscedasticity of unknown form. Assumption 2.6 parallels Assumption 4 in Giraitis et al. (2021) to ensure uniform consistency of the first-stage estimates of Ψ_{it} for each i.

Define

$$r_{T,H,\gamma} = (\frac{H}{T})^{\gamma} + \sqrt{\frac{\log T}{H}}, \quad \bar{r}_{T,H,\gamma,\alpha} = \sqrt{\frac{\log T}{H}} + (\frac{H}{T})^{\gamma} \log^{1/\alpha} T. \tag{16}$$

In the next theorem, we establish uniform consistency rates and asymptotic distributions for (8) (at the interior points).

Theorem 1. *Under Assumptions 2.1–2.6, assume that the bandwidth parameters H, L satisfy* (6) *and the kernel function $K(x)$ satisfies* (7)*. Then, as $(N, T) \longrightarrow \infty$ we have the following:*

(i) Uniform consistency: $\hat{\beta}^{IV}_{MG,t}$ has the property

$$\max_{t=1,2,\cdots,T} \left\| \hat{\beta}^{IV}_{MG,t} - \beta_{0,t} \right\| = O_p\left(\frac{\bar{r}_{T,L,\gamma_1,\alpha} + \left(\log T \right)^{1/\alpha} r_{T,H,\gamma_2}}{\sqrt{N}} + \frac{\left(\log T \right)^{1/\alpha}}{\sqrt{N}} (\frac{H}{T})^{\gamma_3} + \frac{\log^{2/\alpha} T}{\sqrt{N}} \right),$$

where $\bar{r}_{T,L,\gamma_1,\alpha}$ and r_{T,H,γ_2} are defined as in (16)*;*

(ii) Asymptotic normality: Suppose that $(\frac{H}{T})^{\gamma_2} = o(N^{-1/2})$, for $t = \lfloor Tr \rfloor$, $0 < r < 1$,

$$\sqrt{N} \left(\Sigma_{e,t} \right)^{-1/2} \left(\hat{\beta}^{IV}_{MG,t} - \beta_{0,t} \right) \xrightarrow{d} N(0, I_k),$$

where $\Sigma_{e,t}$ is given by

$$\Sigma_{e,t} = \lim_{N \to \infty} Var\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} \right).$$

Further more, $\hat{\Sigma}_{e,t} \xrightarrow{p} \Sigma_{e,t}$, where

$$\hat{\Sigma}_{e,t} = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{\beta}^{IV}_{it} - \hat{\beta}^{IV}_{MG,t} \right) \left(\hat{\beta}^{IV}_{it} - \hat{\beta}^{IV}_{MG,t} \right)'. \tag{17}$$

It is worth mentioning that, as implied by Theorem 1(ii), the pointwise asymptotic expansion of the estimator (as $(N,T) \longrightarrow \infty$) takes a simple form

$$\sqrt{N}\left(\hat{\beta}^{IV}_{MG,t} - \beta_{0,t}\right) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} + o_p(1).$$

To derive these results, we have not made any dependence assumptions on (e_{it}) with x_{it}. This implies that our framework does not rule out the correlated random coefficient model. Note that, in standard random coefficient panel data model when $e_{it} = e_i$, MG estimator is inconsistent if the assumption $E(e_i|x_{it}) = 0$ is violated. However, as shown in the Appendix B, the key assumption imposed to derive the results in Theorem 1 is the smoothness condition (15) in Assumption 2.4(iii). This implies that our TVP-IV-MG estimator is still valid even if $E(e_i|x_{it}) \neq 0$. We will investigate more on this in the Monte Carlo study (see Section 3).

Remark 4. *Following Chen (2015) and Giraitis et al. (2021), our model* (1) *does not explicitly allow for exogenous regressors. However, (weakly) exogenous regressors (g_{it} satisfying $E\left[u_{it}|g_{it}\right] = 0$ for each i) can be added. We just need to modify* (3) *by augmenting $\hat{\Psi}'_{i,j}z_{ij}$ with g_{it}:*

$$\hat{\beta}^{IV}_{it} = \Big(\sum_{j=1}^{T} b_{j,t}(H)[\hat{\Psi}'_{ij}z_{ij} \;\; g_{ij}]x'_{ij}\Big)^{-1}\Big(\sum_{j=1}^{T} b_{j,t}(H)[\hat{\Psi}'_{ij}z_{ij} \;\; g_{ij}]y_{ij}\Big).$$

The weak exogeneity condition $E\left[u_{it}|g_{it}\right] = 0$ also implies that the dynamic panel data model is not ruled out ($y_{i,t-1}$ as a regressor). Following Hsiao et al. ([1999), the MG estimator remains valid in the case of dynamic panel data models and proofs follow exactly as in the static case. The finite sample performance of the estimator, with intercept, lagged y_{it} and exogenous regressor added to the model (1)*, will be evaluated in Section 3.*

2.3. Bandwidth selection

The use of th estimator $\hat{\beta}_{MG,t}$ makes it necessary to choose the bandwidth parameters H and L. As stated in Assumption 2.6, the choice of L requires the knowledge of γ_1. γ_3 also appears in the uniform consistency rate. In addition, asymptotic normality results also require $(\frac{H}{T})^{\gamma_2} = o(N^{-1/2})$. This implies that $\sqrt{N}(\frac{H}{T})^{\gamma_2} \to 0$ when both $(N,T) \longrightarrow \infty$. This means that the choice of the bandwidth H should depend on the relative rate of N, T and the degree of smoothness of $(\beta_{0,t})$, which is γ_2. As in Giraitis et al. (2014), Giraitis et al. (2018), Giraitis et al. (2021), H can be selected via a simple rule-of-thumb approach by setting $H = T^{\overline{\alpha}}$, for some $0 < \overline{\alpha} < 1$. Then, the condition simplifies to $\sqrt{N}/T^{\gamma_2(1-\overline{\alpha})} \to 0$. Giraitis et al. (2018) and Giraitis et al. (2021) consider a special case when $\gamma_2 = 0.5$. In their simulations, it is found that setting $H = \sqrt{T}$ leads to overall best finite sample performance. Now the condition $\sqrt{N}/T^{\gamma_2(1-\overline{\alpha})} \to 0$ simplifies further to $\sqrt{N}/T^{0.25} \to 0$ as $(N,T) \longrightarrow \infty$. A practically meaningful implication is that T has to diverge at a faster rate than N.

Since the degree of smoothness is generally unknown in practice, we hereby propose a data-driven bandwidth selection procedure for both H and L, which is adapted from Sun et al. (2009) and is called leave-one-unit-out cross-validation (CV) method. The idea is to remove $\{(z_{it}, x_{it}, y_{it})\}_t$ for each i and construct the TVP-IV-MG estimator by $\hat{\beta}^{i,IV}_{MG,t} = \frac{1}{N}\sum_{j\neq i}\hat{\beta}^{IV}_{j,t}$, given (H, L). Then, the optimal bandwidth parameters are chosen such that they minimize the mean squared error of the form $\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - x_{it}'\hat{\beta}^{i,IV}_{MG,t})^2$. We will compare these two procedures in the Monte Carlo study.

3. Monte Carlo study

In this section, we conduct Monte-Carlo experiments to evaluate the finite sample performance of the TVP-IV-MG estimator (8). To be consistent with the empirical application, we generate data using the model:

$$y_{it} = \alpha_{it} + \rho_{it}y_{i,t-1} + \beta_{1,it}x_{1,it} + \beta_{2,it}x_{2,it} + u_{it}, \quad i = 1, 2, ..., N, \;\; t = 1, 2, ..., T,$$

which contains a time-varying intercept, first-order lag and two regressors (one of them being endogenous). The endogenous regressor is generated according to

$$x_{2,it} = \Psi_{it}'z_{it} + \alpha_{2,it}e_{2,it} + v_{it}, \quad i = 1, 2, ..., N, \;\; t = 1, 2, ..., T, \tag{18}$$

where two instruments $z_{it} = (z_{1,it}, z_{2,it})'$ and the exogenous regressor $x_{1,it}$ are generated from AR(1) processes:

$$\overline{x}_{it} = \rho\overline{x}_{i,t-1} + s_{it}, \quad i = 1, 2, ..., N, \;\; t = 1, 2, ..., T,$$

for $\overline{x}_{it} = z_{1,it}, z_{2,it}, x_{1,it}$. ρ is generated from a uniform distribution on the support $[-0.99, 0.99]$ and $s_{it} \sim \mathcal{N}(0, 1)$. We introduce time-varying correlation between u_{it} and v_{it} by specifying them as

$$u_{it} = \alpha_{1,it}q_{1,it} + q_{2,it}, \quad v_{it} = \alpha_{1,it}q_{1,it} + q_{3,it},$$

where $q_{j,it}$, $j = 1, 2, 3$ are generated independently from $N(0, 1)$.

For the time-varying parameters $\alpha_{it}, \rho_{it}, \beta_{1,it}, \beta_{2,it}, \alpha_{1,it}, \Psi_{1,it}, \Psi_{2,it}$, they are generated according to

$$\begin{aligned}
&\beta_{1,it} = \beta_{0,1,t} + e_{1,it}, \beta_{2,it} = \beta_{0,2,t} + e_{2,it},\\
&\Psi_{1,it} = \Psi_{0,1,t} + \Upsilon_{1,it}, \Psi_{2,it} = \Psi_{0,2,t} + \Upsilon_{2,it},\\
&\alpha_{it} = \alpha_{0,t} + \iota_{it}, \alpha_{1,it} = \alpha_{0,1,t} + \iota_{1,it}, \alpha_{2,it} = \alpha_{0,2,t} + \iota_{2,it},\\
&\rho_{it} = \rho_{0,t} + \varepsilon_{it},
\end{aligned}$$

where elements in $X^{(1)}_{MC} = \{\beta_{0,1,t}, e_{1,it}, \beta_{0,2,t}, e_{2,it}, \Psi_{0,1,t}, \Psi_{0,2,t}, \Upsilon_{1,it}, \Upsilon_{2,it}, \alpha_{0,t}, \alpha_{0,1,t}, \alpha_{0,2,t}, \iota_{it}, \iota_{1,it}, \iota_{2,it}\}$ are generated from the scaled random walk processes, such that $X_{\ell,t} = \xi_{\ell,t}/\sqrt{t}$, for $\xi_{\ell,t} - \xi_{\ell,t-1} \overset{i.i.d.}{\sim} N(0, 1)$,

$t = 1, 2, \cdots, T$. For $\rho_{0,t}$ and ε_{it}, they are all generated according to $\rho_{0,t} = \rho_1 a_{1,t} / \max_{1 \leqslant t \leqslant T} \left| a_{1,t} \right|$ and $\varepsilon_{it} = \rho_2 a_{2,it} / \max_{1 \leqslant t \leqslant T} \left| a_{2,it} \right|$, $(a_{1,t})$ and $a_{2,it}$ are all generated from standard random walk processes with innovations being drawn from $\mathcal{N}(0, 1)$. We set $\rho_1 = 0.5$, $\rho_2 = 0.49$ to ensure stationarity. Notice that we have introduced correlated random coefficients by adding $e_{2,it}$ in (18). Motivated by the requirement that T must be larger than N for asymptotic normality, we choose each (N, T) pair with $N = 10, 20, 50$ and $T = 100, 200, 500$. Each experiment was replicated 1,000 times.

To compute the TVP-IV-MG estimator, we consider Gaussian kernel $K(x) = \exp(-x^2/2)$, uniform kernel $K(x) = \frac{1}{2} I\{|x| \leqslant 1\}$ and Epanechnikov kernel $K(x) = \frac{3}{4}(1 - x^2) I\{|x| \leqslant 1\}$. For bandwidth parameters H, L, since our data generating process (DGP) is for the case when $\gamma_1 = \gamma_2 = \gamma_3 = 1/2$ ((see Giraitis et al., 2014)), we consider either the rule-of-thumb approach by setting $H = L = T^{0.5}$ or the CV method proposed in section 2.3 over a discrete grid T^b, where $b \in \{0.3, 0.4, 0.5, 0.6, 0.7\}$.

The global performance of the estimators for $\beta_t = \{\alpha_{0,t}, \rho_{0,t}, \beta_{0,1,t}, \beta_{0,2,t}\}$ is evaluated by both the average of median absolute deviations (MADs), $\frac{1}{M} \sum_{r=1}^{M} med_{t=1,2,\cdots,T} \left| \hat{\beta}_{r,t} - \beta_{r,0,t} \right|$, and 95% coverage rates, $(\hat{\beta}_t - 1.96 \, \mathrm{std}(\hat{\beta}_t), \hat{\beta}_t + 1.96 \, \mathrm{std}(\hat{\beta}_t))$, computed over the sample period, $t = H + 1, \cdots, T - H$. The standard errors are obtained as in (17).

Table 1 reports the average MAD and coverage probability of the TVP-IV-MG estimator for the Gaussian kernel. A number of comments can be made. First, for both bandwidth selection procedures, both cross-sectional and time series dimensions are useful to reduce the bias of the estimates. All values become smaller as (N, T) increase. Second, for each N, the coverage probability increases with T. However, it decreases with N for a given T. This is far more evident from the rule-of-thumb bandwidth selection procedure. As explained in section 2.3, pointwise asymptotic normality requires T to be relatively larger than N. As N gets closer to T, the condition is likely to be violated so the coverage probability deteriorates. Finally, except for a few cases for the estimates of α_t when (N, T) are relatively small, it is clear that the CV approach delivers better finite sample performance compared to the rule-of-thumb approach, with lower average MAD and higher coverage probability.

Tables 2 and 3 report the average MAD and coverage probability of the TVP-IV-MG estimator for the Epanechnikov kernel and uniform kernel, respectively. The results are overall similar to the ones obtained from the Gaussian kernel. However, MADs are generally higher and coverage probabilities are larger. This is likely due to the fact that unlike Gaussian kernel, not all observations are used for the Epanechnikov kernel and the uniform kernel. Hence, overall, the TVP-IV-MG estimator with Gaussian kernel might be more efficient in finite samples.

Table 1: Small sample properties of the TVP-IV-MG estimator: Average MAD and coverage probability

(N, T)	100	200	500	100	200	500	100	200	500	100	200	500
		α_t			ρ_t			$\beta_{1,t}$			$\beta_{2,t}$	
	MAD ($H = L = T^{0.5}$)											
10	0.369	0.323	0.315	0.110	0.101	0.094	0.303	0.275	0.269	0.292	0.287	0.269
20	0.298	0.268	0.243	0.096	0.090	0.078	0.242	0.219	0.205	0.248	0.222	0.207
50	0.247	0.220	0.187	0.089	0.081	0.069	0.200	0.175	0.153	0.206	0.181	0.159
	Coverage probability ($H = L = T^{0.5}$)											
10	0.835	0.845	0.852	0.719	0.739	0.755	0.815	0.838	0.850	0.813	0.820	0.847
20	0.807	0.822	0.836	0.654	0.666	0.711	0.786	0.811	0.835	0.765	0.798	0.829
50	0.723	0.728	0.767	0.496	0.525	0.574	0.669	0.713	0.767	0.650	0.697	0.751
		α_t			ρ_t			$\beta_{1,t}$			$\beta_{2,t}$	
	MAD (CV)											
10	0.396	0.347	0.325	0.086	0.077	0.070	0.301	0.274	0.267	0.285	0.281	0.268
20	0.304	0.268	0.242	0.069	0.061	0.052	0.225	0.205	0.191	0.229	0.212	0.198
50	0.238	0.202	0.171	0.057	0.048	0.039	0.169	0.145	0.129	0.181	0.158	0.142
	Coverage probability (CV)											
10	0.887	0.899	0.898	0.835	0.853	0.872	0.867	0.886	0.891	0.865	0.871	0.885
20	0.898	0.906	0.917	0.813	0.837	0.879	0.875	0.893	0.905	0.854	0.871	0.894
50	0.881	0.881	0.911	0.720	0.762	0.823	0.821	0.856	0.892	0.786	0.824	0.866

Note: Results presented in this Table are based on Gaussian kernel. ”CV” refers to leave-one-unit-out cross-validation (CV) method for bandwidth selection proposed in section 2.3. See Section 3 for details on the computations of mean absolute deviation (MAD) and coverage probability.

Table 2: Small sample properties of the TVP-IV-MG estimator: Average MAD and coverage probability

(N, T)	100	200	500	100	200	500	100	200	500	100	200	500
		α_t			ρ_t			$\beta_{1,t}$			$\beta_{2,t}$	
	MAD ($H = L = T^{0.5}$)											
10	0.413	0.346	0.315	0.089	0.080	0.075	0.307	0.275	0.266	0.293	0.282	0.267
20	0.322	0.274	0.235	0.072	0.064	0.057	0.232	0.209	0.194	0.238	0.215	0.199
50	0.258	0.211	0.173	0.060	0.053	0.046	0.179	0.153	0.136	0.190	0.164	0.145
	Coverage probability ($H = L = T^{0.5}$)											
10	0.885	0.892	0.887	0.832	0.841	0.844	0.864	0.878	0.879	0.867	0.870	0.878
20	0.894	0.895	0.898	0.808	0.816	0.836	0.869	0.878	0.884	0.856	0.866	0.882
50	0.876	0.859	0.873	0.718	0.729	0.747	0.809	0.829	0.852	0.796	0.817	0.841
		α_t			ρ_t			$\beta_{1,t}$			$\beta_{2,t}$	
	MAD (CV)											
10	0.747	0.556	0.460	0.129	0.098	0.083	0.453	0.348	0.309	0.380	0.337	0.305
20	0.630	0.453	0.372	0.110	0.081	0.067	0.360	0.263	0.225	0.327	0.264	0.235
50	0.572	0.366	0.281	0.095	0.065	0.051	0.287	0.192	0.158	0.281	0.210	0.178
	Coverage probability (CV)											
10	0.924	0.921	0.915	0.889	0.888	0.890	0.909	0.908	0.904	0.899	0.889	0.898
20	0.943	0.939	0.938	0.899	0.891	0.897	0.932	0.928	0.927	0.908	0.907	0.912
50	0.954	0.945	0.947	0.883	0.866	0.875	0.934	0.929	0.933	0.897	0.899	0.905

Note: Results presented in this Table are based on Epanechnikov kernel. ”CV” refers to leave-one-unit-out cross-validation (CV) method for bandwidth selection proposed in section 2.3. See Section 3 for details on the computations of mean absolute deviation (MAD) and coverage probability.

Table 3: Small sample properties of the TVP-IV-MG estimator: Average MAD and coverage probability

(N,T)	100	200	500	100	200	500	100	200	500	100	200	500
		α_t			ρ_t			$\beta_{1,t}$			$\beta_{2,t}$	
	MAD ($H = L = T^{0.5}$)											
10	0.424	0.353	0.321	0.096	0.086	0.080	0.315	0.280	0.270	0.307	0.294	0.274
20	0.335	0.288	0.246	0.080	0.071	0.062	0.243	0.217	0.200	0.257	0.227	0.207
50	0.278	0.228	0.185	0.069	0.060	0.052	0.195	0.165	0.144	0.212	0.180	0.156
	Coverage probability ($H = L = T^{0.5}$)											
10	0.873	0.882	0.879	0.811	0.819	0.826	0.850	0.866	0.870	0.860	0.861	0.871
20	0.875	0.878	0.882	0.779	0.785	0.807	0.845	0.858	0.869	0.841	0.854	0.870
50	0.847	0.830	0.847	0.674	0.683	0.703	0.769	0.792	0.823	0.778	0.796	0.821
		α_t			ρ_t			$\beta_{1,t}$			$\beta_{2,t}$	
	MAD (CV)											
10	0.620	0.505	0.432	0.108	0.091	0.080	0.388	0.326	0.299	0.360	0.330	0.304
20	0.516	0.412	0.349	0.091	0.074	0.062	0.304	0.247	0.219	0.304	0.259	0.233
50	0.463	0.330	0.265	0.077	0.059	0.048	0.240	0.182	0.155	0.259	0.208	0.178
	Coverage probability (CV)											
10	0.916	0.918	0.914	0.882	0.887	0.892	0.896	0.903	0.901	0.898	0.894	0.901
20	0.935	0.934	0.936	0.887	0.889	0.904	0.917	0.921	0.922	0.909	0.911	0.919
50	0.943	0.937	0.942	0.864	0.863	0.883	0.910	0.914	0.924	0.895	0.903	0.916

Note: Results presented in this Table are based on uniform kernel. ”CV” refers to leave-one-unit-out cross-validation (CV) method for bandwidth selection proposed in section 2.3. See Section 3 for details on the computations of mean absolute deviation (MAD) and coverage probability.

4. Empirical application

In this section, we consider an empirical application on modeling inflation dynamics. We estimate a multi-country version of the hybrid Phillips curve (Gali and Gertler (1999)):

$$\pi_{it} = c_{it} + \gamma_{it}\pi_{i,t-1} + \alpha_{it}u_{it} + \rho_{it}\pi^e_{i,t+1} + \nu_{it},\ i = 1, \cdots, N, t = 1, \cdots, T, \tag{19}$$

where π_{it} is the inflation rate, $\pi^e_{i,t+1}$ is the inflation expectation, and u_{it} is the unemployment rate. An intercept term c_{it} is there to capture the slow moving component in inflation.

We use monthly data for 19 Eurozone countries: Austria, Belgium, Cyprus, Estonia, Finland, France, Germany, Greece, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Portugal, Slovenia, Slovakia and Spain, over the period 2000M1–2019M12. Inflation is measured as the percentage changes of the Harmonised index of consumer prices (HICP). The (total) unemployment rate (UR) is obtained from the EU Labour Force Survey (EU-LFS). Both variables are seasonally adjusted and are downloaded from Eurostat.

Let $\epsilon_{it} = \rho_{it}(\pi^e_{i,t+1} - \pi_{i,t+1}) + \nu_{it}$, we can also write (19) as

$$\pi_{it} = c_{it} + \gamma_{it}\pi_{i,t-1} + \alpha_{it}u_{it} + \rho_{it}\pi_{i,t+1} + \epsilon_{it}. \tag{20}$$

It is clear that, since $\pi^e_{i,t+1}$ is not observed, we have to replace it by $\pi_{i,t+1}$ and endogeneity arises. We use an intercept, three lags of inflation and two lags of unemployment as instruments: $(1, \pi_{i,t-2}, \pi_{i,t-3}, \pi_{i,t-4}, u_{i,t-1}, u_{i,t-2})'$. Motivated by the Monte Carlo study and the fact that we do not know a priori the degree of the smoothness of those coefficients in the empirical application, the Gaussian kernel is used and bandwidth parameters H, L are selected according to the leave-one-unit-out cross-validation (CV) method described in section 2.3. Both H and L are parameterized over a discrete grid $T^{\bar{b}}$, where $\bar{b} \in \{0.3, 0.35, \cdots, 0.8, 0.85\}$.

Figure 1 provides empirical estimates of c_t, γ_t, α_t and ρ_t, which are the time-varying cross-sectional mean coefficients estimates of the associated parameters in Eq. (19). The solid blue lines show the point estimates and the blue shaded areas show the 95% pointwise confidence intervals, which are computed as $(\hat{\beta}_t - 1.96\,\mathrm{std}(\hat{\beta}_t), \hat{\beta}_t + 1.96\,\mathrm{std}(\hat{\beta}_t))$, where $\mathrm{std}(\hat{\beta}_t)$ are obtained as in (17). First, it is worth mentioning that the bandwidth parameters selected by CV method are $H = T^{0.65}$, $L = T^{0.3}$, indicating that the time-varying parameters in (19) are likely to be smoother than the coefficients in the first-stage regression. Let us start by commenting on α_t, which measures the unemployment inflation tradeoff and is the key parameter of interest. Interestingly, the coefficient is insignificant at the start of the sample period (a few years after the Eurozone was established), becomes significant after 2004 and remains roughly stable over the remaining sample period. After 2010, there is mild evidence that the magnitude of the coefficient starts to gradually decline over time. This finding is consistent with studies on U.S. data, such as Del Negro et al. (2020), who also provide evidence and explanation on the flattening Phillips Curve. Turning to the intercept term c_t, we see that the estimates remain roughly stable over time. For ρ_t and γ_t, the patterns are very similar: they first increase until 2010 and decline afterwards. However, the magnitude of those estimates indicates that forward-looking features in Eurozone inflation dominate the back-looking ones.

As a robustness check, we collect aggregated Eurozone time-series data (HICP and UR) over the same period 2000M1–2019M12 and estimate the time-series version of the hybrid Phillips curve as in Gali and Gertler (1999). We use the TVP-IV estimator proposed in Giraitis et al. (2021) and the same set of instruments $(1, \pi_{t-2}, \pi_{t-3}, \pi_{t-4}, u_{t-1}, u_{t-2})'$. The Gaussian kernel is used to compute the estimator. The bandwidth parameters H, L are selected again according to the leave-one-out cross-validation (CV) method over a discrete grid $T^{\bar{b}}$, where $\bar{b} \in \{0.3, 0.35, \cdots, 0.8, 0.85\}$. Empirical estimates are provided in Figure 2. Although the estimates are quantitatively similar, the confidence intervals are much wider compared to the panel approach. This implies that panel estimates are more efficient compared to the time-series approach.

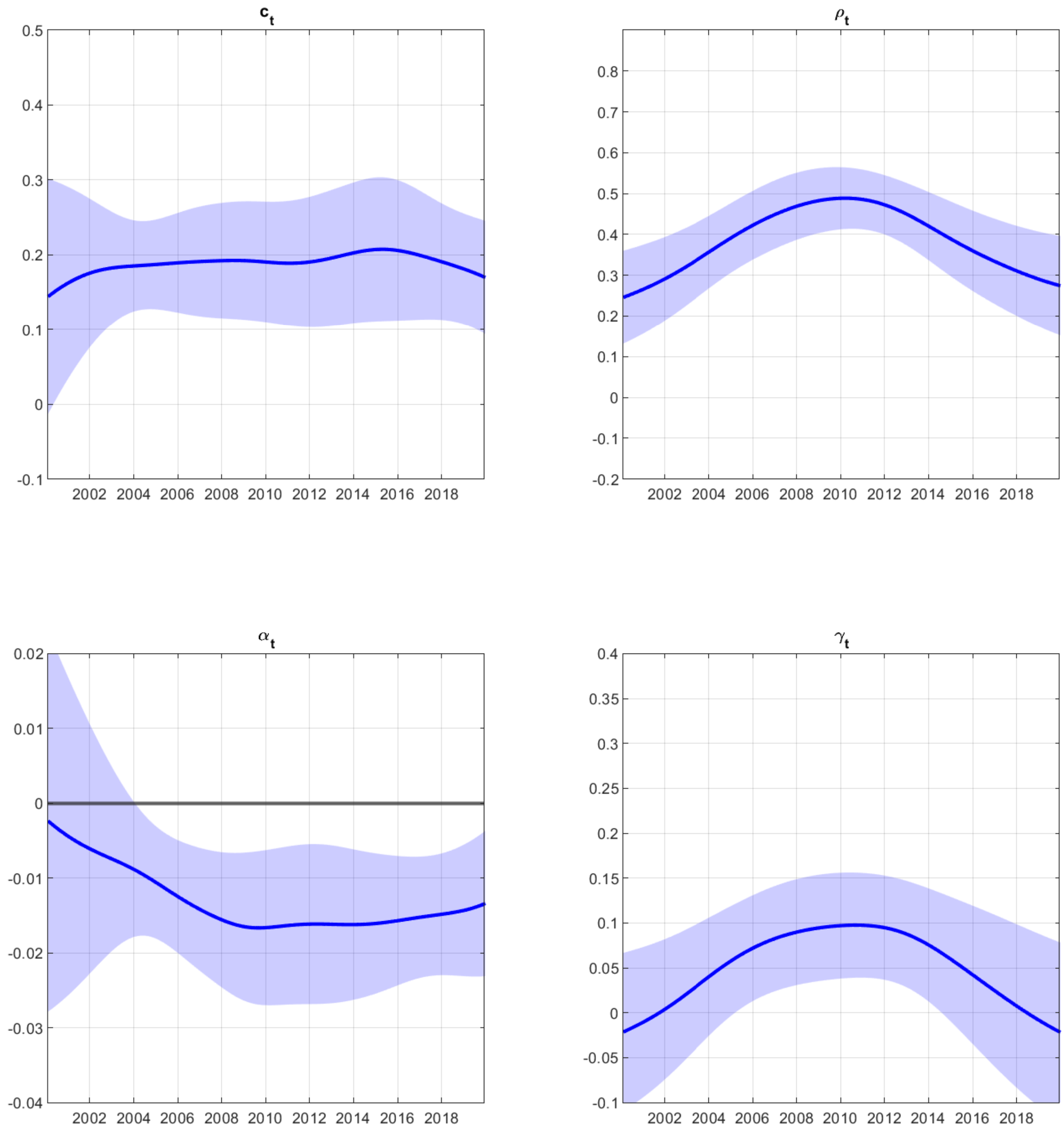

Figure 1: Empirical results for multi-country hybrid Phillips curve in the Eurozone. Shaded areas show the 95% pointwise confidence intervals, which are computed as $(\hat{\beta}_t - 1.96\,\mathrm{std}(\hat{\beta}_t), \hat{\beta}_t + 1.96\,\mathrm{std}(\hat{\beta}_t))$, where $\mathrm{std}(\hat{\beta}_t)$ are obtained as in (17).

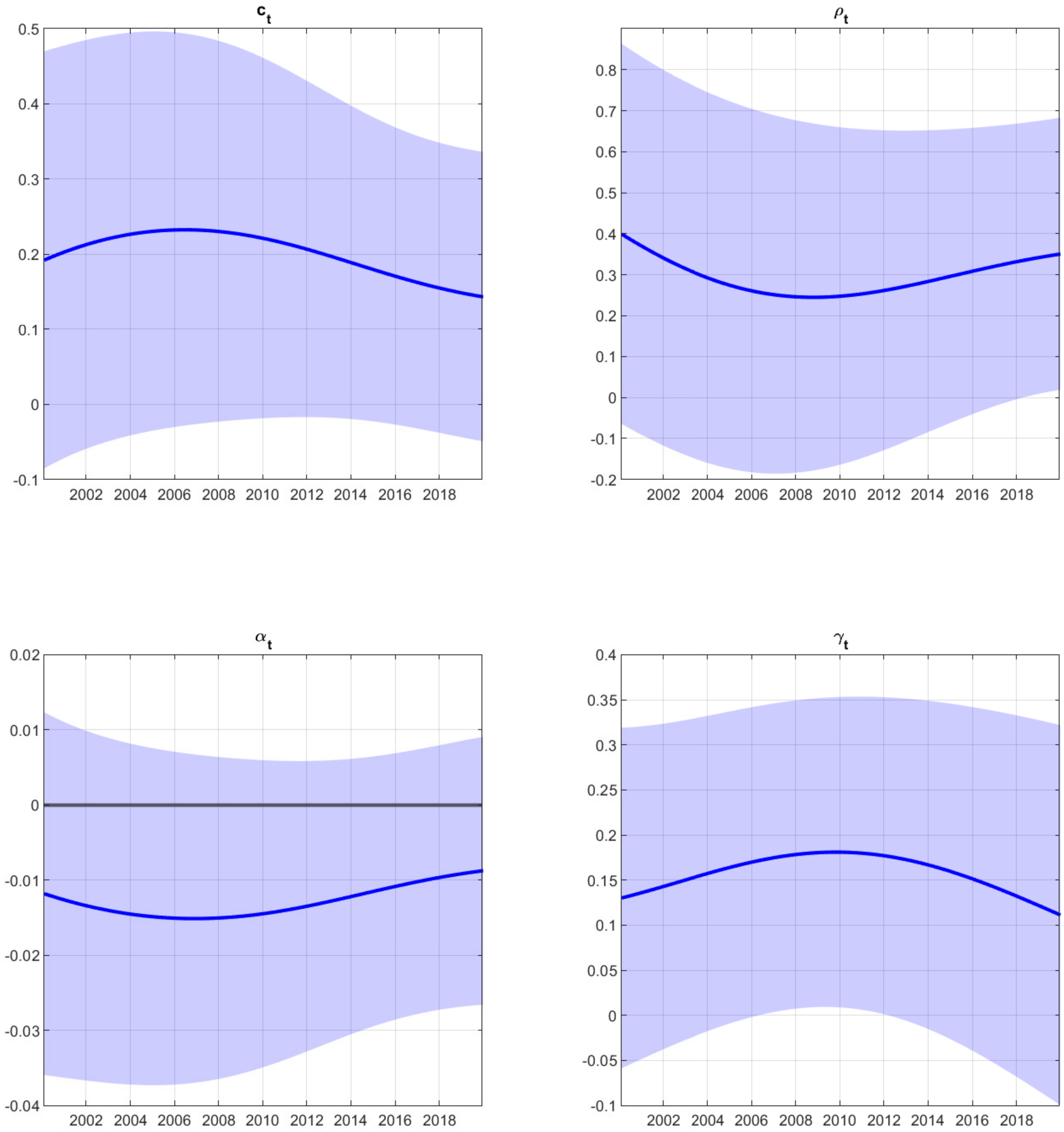

Figure 2: Empirical results for hybrid Phillips curve from the aggregated Eurozone data. Shaded areas show the 95% pointwise confidence intervals, which are computed as $(\hat{\beta}_t - 1.96\,\mathrm{std}(\hat{\beta}_t), \hat{\beta}_t + 1.96\,\mathrm{std}(\hat{\beta}_t))$, where $\mathrm{std}(\hat{\beta}_t)$ are obtained as Theorem 3(ii) in Giraitis et al. (2021).

5. Conclusion

Large heterogeneous panel data models are becoming increasingly popular in empirical applications, but the parameters are typically assumed to be constant over time and regressors are treated as exogenous. However, the vast literature on panels, structural change and parameter instability has highlighted the importance of considering both time variation of parameters and endogeneity. In this paper, the existing literature of large heterogenous panels is extended to the setting when the heterogenous coefficients are changing over time and the regressors are endogenous. We propose time-varying mean group IV estimators, taking a non-parametric approach in order to remain as agnostic as possible regarding the type of parameter evolution.

We derive theoretical properties for the proposed time-varying mean group estimator in the IV context. We show the uniform consistency and derive the (pointwise) asymptotic distribution of the proposed estimator. It is shown that the pointwise asymptotic expansion of the estimator takes a simple form. Thus, the estimator remains valid even in the correlated random coefficient model. Since the asymptotic results rely on the knowledge of the unknown degree of the smoothness of time-varying parameters, we also propose a data-driven bandwidth selection procedure.

Next, we evaluate the finite sample properties of the estimator in a Monte Carlo study. The results show that the performances are satisfactory in terms of both bias and coverage rates. The data-driven bandwidth selection procedure generally performs better than the rule-of-thumb approach particularly when the Gaussian kernel is used, as it delivers lower biases and higher coverage rates in most of the cases. Finally, we provide an empirical application to illustrate in practice the use of the time-varying mean group estimator. We estimate the panel version of time-varying hybrid Phillips curves for 19 Eurozone countries over the period 2000M1–2019M12. We find that Eurozone Phillips curve is alive after 2004, but gradually flattening after 2010, with inflation expectations more relevant than past inflation over the whole sample period. The time series estimation based on aggregated Eurozone data delivers similar point estimates, but with much wider bandwidth. This shows that our panel methods are more efficient than time-series approach.

Acknowledgments

We would like to thank the editor, the associate editor, three anonymous referees and seminar participants at Capital University of Economics and Business, Bocconi University, Dongbei University of Finance and Economics, 2021 North American Summer Meeting of the Econometric Society, 2021 IAAE Annual Conference, 26th International Panel Data Conference and 7th Continuing Education in Macroeconometrics workshop for their useful comments. Bai and Marcellino thank MIUR–PRIN Bando 2017 – prot. 2017TA7TYC for financial support for this research. Bai would also acknowledge financial support from ARC Discovery Project DP210100476.

References

Baltagi, B H and Feng, Q and Kao, C. (2019). Structural changes in heterogeneous panels with endogenous regressors. *Journal of Applied Econometrics,* 34(6), 883–892.

Chen, B. (2015). Modeling and testing smooth structural changes with endogenous regressors. *Journal of Econometrics,* 185(1), 196–215.

Chen, B and Hong, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica,* 80(3), 1157–1183.

Chen, B and Huang, L. (2018). Nonparametric testing for smooth structural changes in panel data models. *Journal of Econometrics,* 202(2), 245–267.

Castagnetti, C and Rossi, E and Trapani, L. (2019). A two-stage estimator for heterogeneous panel models with common factors. *Econometrics and Statistics,* 11, 63–82.

Dendramis, Y and Giraitis, L and Kapetanios, G. (2021). Estimation of time-varying covariance matrices for large datasets. *Econometric Theory,* 37(6), 1100–1134.

Giraitis, L and Kapetanios, G and Yates, T. (2014). Inference on stochastic time-varying coefficient models. *Journal of Econometrics,* 179(1), 46–65.

Giraitis, L and Kapetanios, G and Yates, T. (2018). Inference on multivariate heteroscedastic time varying random coefficient models. *Journal of Time Series Analysis,* 39(2), 129–149.

Giraitis, L and Kapetanios, G and Marcellino, M. (2021). Time-varying instrumental variable estimation. *Journal of Econometrics,* 224(2), 394–415.

Gali, J and Gertler, M. (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics,* 44(2), 195–222.

Haque, N Ul and Pesaran, M H and Sharma, S. (1999). Neglected heterogeneity and dynamics in cross-country savings regressions. *IMF working papers,* 1999(128). International Monetary Fund.

Horváth, L and Trapani, L. (2016). Statistical inference in a random coefficient panel model. *Journal of Econometrics,* 193(1), 54–75.

Hsiao, C and Li, Q and Liang, Z and Xie, W. (2019). Panel data estimation for correlated random coefficients models. *Econometrics,* 7(1), 7.

Hsiao, C and Pesaran, M H and Tahmiscioglu, A K (1999). Bayes Estimation of Short-Run Coefficients in Dynamic Panel Data Models. In *Analysis of Panel Data and Limited Dependent Variable Models,* (pp. 268-296). Cambridge University Press.

Hsiao, C and Pesaran, M H. (2008). Random coefficient models. In *The econometrics of panel data,* (pp. 185-213). Springer, Berlin, Heidelberg.

Jiang, P and Kurozumi, E. (2023). A new test for common breaks in heterogeneous panel data models. *Econometrics and Statistics, forthcoming*.

Li, D and Chen, J and Gao, J. (2011). Non-parametric time-varying coefficient panel data models with fixed effects. *The Econometrics Journal,* 14(3), 387–408.

Del Negro, M and Lenza, M and Primiceri, G E and Tambalotti, A. (2020) What's up with the Phillips Curve? In *Brookings Papers on Economic Activity,* (Spring 2020), pp. 301-357.

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

Robinson P M. (1991). Time-varying nonlinear regression. In *Economic Structural Change,* pp. 179-190. Springer.

Sun, Y and Raymond, J.C and Li, D. (2009). Semiparametric estimation of fixed-effects panel data varying coefficient models. In *Advances in Econometrics,* 25 (2009), pp. 185-213. Emerald Group Publishing Limited.

Stock, J H and Watson, M W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics,* 14(1), 11–30.

Appendix A: Statement and proofs of lemmas

NOTATION: $\|A\|_{sp} = \sqrt{\lambda_{\max}(A'A)}$ is the spectrum norm of matrix A, where $\lambda_{\max}(\cdot)$ is the maximum eigenvalue of $\cdot$. $\|\cdot\|$ is the Euclidean norm. $|\cdot|$ denotes the associated norm when $\cdot$ is one dimensional. We use $\longrightarrow$ to denote the ordinary limit; $\xrightarrow{p}$ and $\xrightarrow{d}$ to denote convergence in distribution and in probability, respectively. All variables are assumed to be defined on a common probability space $(\Omega, \mathcal{F}, \mathcal{P})$.

In the proofs, we shall write $b_{jt} = b_{j,t}(H)$ and

$$K_t = \sum_{j=1}^{T} b_{jt}, \quad K_{2,t} = \sum_{j=1}^{T} b_{jt}^2.$$

We will use repeatedly the following property of the weights (see Giraitis et al., 2014): for $t = [\tau T]$ $(0 < \tau < 1)$, as $H \to \infty$,

$$K_t = O(H), \quad K_{2,t} = O(H).$$

Lemma 1. *Define*

$$r_{T,H,\gamma} = (\frac{H}{T})^{\gamma} + \sqrt{\frac{\log T}{H}}, \quad \bar{r}_{T,H,\gamma,\alpha} = \sqrt{\frac{\log T}{H}} + (\frac{H}{T})^{\gamma} \log^{1/\alpha} T.$$

Consider the model defined in (1)-(2) *and the following sums for each i*

$$S^{i}_{z\psi,t} = \frac{1}{K_t} \sum_{j=1}^{T} b_{jt} \hat{\Psi}'_{ij} z_{ij} x'_{ij},$$

$$\Delta^{i}_{z,t} = \frac{1}{K_t} \sum_{j=1}^{T} b_{jt} \hat{\Psi}'_{ij} z_{ij} x'_{ij} (\beta_{0,j} - \beta_{0,t}),$$

$$S^{i}_{zu,t} = \frac{1}{K_t} \sum_{j=1}^{T} b_{jt} \hat{\Psi}'_{ij} z_{ij} u_{ij},$$

$$S^{i}_{ze,t} = \frac{1}{K_t} \sum_{j=1}^{T} b_{jt} \hat{\Psi}'_{ij} z_{ij} x'_{ij} e_{ij}.$$

Then under Assumptions 2.1–2.5, we have

$$\max_{t=1,2,\cdots,T} \left\| (S^{i}_{z\psi,t})^{-1} \right\|_{sp} = O_p(1), \tag{A.1}$$

$$\left\| \Delta^{i}_{z,t} \right\| = O_p\Big((\frac{H}{T})^{\gamma_2} \Big), \quad \max_{t=1,2,\cdots,T} \left\| \Delta^{i}_{z,t} \right\| = O_p\Big((\log T)^{2/\alpha} (\frac{H}{T})^{\gamma_2} \Big), \tag{A.2}$$

$$\left\| S^{i}_{zu,t} \right\| = O_p\Big(\frac{1}{\sqrt{H}} \Big), \quad \max_{t=1,2,\cdots,T} \left\| S^{i}_{zu,t} \right\| = O_p\Big(\bar{r}_{T,L,\gamma_1,\alpha} + \Big(\log T \Big)^{1/\alpha} r_{T,H,\gamma_2} \Big), \tag{A.3}$$

$$\left\| S^{i}_{ze,t} \right\| = O_p(1), \quad \max_{t=1,2,\cdots,T} \left\| S^{i}_{ze,t} \right\| = O_p\Bigg(\Big(\log T \Big)^{1/\alpha} \Big(\log^{1/\alpha} T + (\frac{H}{T})^{\gamma_3} \Big) \Bigg). \tag{A.4}$$

Proof. *Proof of* (A.1). This has been shown in (56) in Giraitis et al. (2021).

Proof of (A.2). This has been shown in (57), (75) in Giraitis et al. (2021), with the exception that we have a different smoothness parameter γ_2 (which is set to 1 in Giraitis et al. (2021)). The extension is straightforward. Take the uniform rate for an example. We could follow the proofs of (57) in Giraitis et al. (2021) to bound

$$\max_{t=1,2,\cdots,T} \left\| \Delta^{i}_{z,t} \right\| \leqslant \Big(\max_{j=1,2,\cdots,T} \left\| \hat{\Psi}_{ij} \right\| \Big) s_T,$$

where

$$s_T := \max_{t=1,2,\cdots,T} \frac{1}{K_t} \sum_{j=1}^{T} b_{jt} \left\| z_{ij} x'_{ij} (\beta_{0,j} - \beta_{0,t}) \right\|.$$

By Assumption 3, it follows from Theorem 1(i) and (92) in Giraitis et al. (2021) that

$$\max_{j=1,2,\cdots,T} \left\| \hat{\Psi}_{ij} \right\| = O_p\Big((\log T)^{1/\alpha} \Big).$$

To bound s_T, we would have

$$s_T := \max_{j=1,2,\cdots,T} (\left\| \Psi_{ij} \right\| + 1)\, s^*_T, \quad s^*_T = \max_{t=1,2,\cdots,T} \frac{1}{K_t} \sum_{j=1}^{T} b_{jt} (\left\| z_{ij} \right\| + \left\| v_{ij} \right\|) \left\| \beta_{0,j} - \beta_{0,t} \right\|.$$

Notice that the only difference compared to (57) in Giraitis et al. (2021) is that we now have a different smoothness parameter γ_2, so $s^*_T = O_p\Big((\frac{H}{T})^{\gamma_2} \Big)$. Together with the fact that $\max_{j=1,2,\cdots,T} \left\| \Psi_{ij} \right\| = O_p\Big((\log T)^{1/\alpha} \Big)$, we obtain the uniform rate. Pointwise rate follows accordingly.

Proof of (A.3). The uniform rate is shown in (58) in Giraitis et al. (2021), with a different smoothness parameter γ_1 (which is set to 1/2 in Giraitis et al. (2021)). The pointwise rate follows similarly by realizing the fact that

$$S^i_{zu,t} = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}'_{ij} - \Psi'_{ij} + \Psi'_{ij})z_{ij}u_{ij} = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}'_{ij} - \Psi'_{ij})z_{ij}u_{ij} + \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\Psi'_{ij}z_{ij}u_{ij}$$
$$= S^i_{zu,t,1} + S^i_{zu,t,2}.$$

By Theorem 1 in Giraitis et al. (2021), $\max_{1\leqslant j\leqslant T}\left\|\hat{\Psi}'_{ij} - \Psi'_{ij}\right\| = O_p(\bar{r}_{T,L,\gamma_1,\alpha})$. Together with (48) in Dendramis et al. (2021), we have

$$\begin{aligned}\left\|S^i_{zu,t,1}\right\| &\leqslant \Big(\max_{1\leqslant j\leqslant T}\left\|\hat{\Psi}'_{ij} - \Psi'_{ij}\right\|\Big)\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\left\|z_{ij}u_{ij}\right\|\Big)\\ &= O_p(\bar{r}_{T,L,\gamma_1,\alpha})O_p(H^{-1/2}) = o_p(1).\end{aligned}$$

For $S^i_{zu,t,2}$, we have

$$\begin{aligned}\left\|S^i_{zu,t,2}\right\| &= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi'_{ij} - \Psi'_{it} + \Psi'_{it})z_{ij}u_{ij}\\ &= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi'_{ij} - \Psi'_{it})z_{ij}u_{ij} + \Psi'_{it}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}z_{ij}u_{ij}\Big)\\ &= O_p\Big((\frac{H}{T})^{\gamma_1}\Big) + O_p(H^{-1/2}) = O_p(H^{-1/2}),\end{aligned}$$

which again follows from (48) in Dendramis et al. (2021). Then, the pointwise rate follows immediately from the triangular inequality.

Proof of (A.4). Write

$$\begin{aligned}S^i_{ze,t} &= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}(e_{ij} - e_{it} + e_{it})\\ &= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}(e_{ij} - e_{it}) + \Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}\Big)e_{it}\\ &= S^i_{ze,t,1} + S^i_{ze,t,2}.\end{aligned}$$

For $S^i_{ze,t,1}$, this has the similar structure as (A.2), except that $\{e_{it}\}_t$ is now a smooth random process with a different smoothness parameter γ_3. Then, it follows again from (57), (75) in Giraitis et al. (2021) that

$$\left\|S^i_{ze,t,1}\right\| = O_p\Big((\frac{H}{T})^{\gamma_3}\Big),\quad \max_{t=1,2,\cdots,T}\left\|S^i_{ze,t,1}\right\| = O_p\Big((\frac{H}{T})^{\gamma_3}\log^{2/\alpha}T\Big).$$

For $S^{i}_{ze,t,2}$, notice that, by Assumption 2.3(iii), $E\|e_{it}\| = O(1)$ and thus $\|e_{it}\| = O_p(1)$. Together with (A.1), we have

$$\left\|S^{i}_{ze,t,2}\right\| = O_p(1), \quad \max_{t=1,2,\cdots,T}\left\|S^{i}_{ze,t,2}\right\| = O_p(\log^{1/\alpha} T).$$

Then, by triangular inequality, we have

$$\begin{aligned}
&\left\|S^{i}_{ze,t}\right\| \leqslant \left\|S^{i}_{ze,t,1}\right\| + \left\|S^{i}_{ze,t,2}\right\| = O_p(1),\\
&\max_{t=1,2,\cdots,T}\left\|S^{i}_{ze,t}\right\| \leqslant \max_{t=1,2,\cdots,T}\left\|S^{i}_{ze,t,1}\right\| + \max_{t=1,2,\cdots,T}\left\|S^{i}_{ze,t,2}\right\| = O_p\Big(\Big(\log T\Big)^{1/\alpha}\Big(\log^{1/\alpha} T + (\frac{H}{T})^{\gamma_3}\Big)\Big).
\end{aligned}$$

□

Appendix B: Proof of Theorem 1

Let us first rewrite the TVP-IV-MG estimator,

$$\begin{aligned}
\hat{\beta}^{IV}_{MG,t} &= \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}^{IV}_{i,t}\\
&= \frac{1}{N}\sum_{i=1}^{N}\Big[\Big(\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}\Big)^{-1}\Big(\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}y_{ij}\Big)\Big]\\
&= \frac{1}{N}\sum_{i=1}^{N}\Big[\Big(\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}\Big)^{-1}\Big(\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}(x'_{ij}(\beta_{0,j} - \beta_{0,t} + \beta_{0,t} + e_{ij}) + u_{ij})\Big)\Big]\\
&= \beta_{0,t} + \Delta_{z,t} + S_{zu,t} + S_{ze,t},
\end{aligned}$$

where

$$\begin{aligned}
\Delta_{z,t} &= \frac{1}{N}\sum_{i=1}^{N}\Big[(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij})^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}(\beta_{0,j} - \beta_{0,t}))\Big],\\
S_{zu,t} &= \frac{1}{N}\sum_{i=1}^{N}\Big[(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij})^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}u_{ij})\Big],\\
S_{ze,t} &= \frac{1}{N}\sum_{i=1}^{N}\Big[(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij})^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}e_{ij})\Big].
\end{aligned}$$

We will show that

$$\left\|\Delta_{z,t}\right\| = O_p\Big((\frac{H}{T})^{\gamma_2}\Big), \quad \max_{t=1,2,\cdots,T}\left\|\Delta_{z,t}\right\| = O_p\Big(\big(\log T\big)^{2/\alpha}(\frac{H}{T})^{\gamma_2}\Big), \tag{B.1}$$

$$\left\|S_{zu,t}\right\| = O_p\Big(\frac{1}{\sqrt{NH}}\Big), \quad \max_{t=1,2,\cdots,T}\left\|S_{zu,t}\right\| = O_p\bigg(\frac{1}{\sqrt{N}}\Big(\bar{r}_{T,L,\gamma_1,\alpha} + \Big(\log T\Big)^{1/\alpha} r_{T,H,\gamma_2}\Big)\bigg), \tag{B.2}$$

$$\left\|S_{ze,t}\right\| = O_p\Big(\frac{1}{\sqrt{N}}\Big), \quad \max_{t=1,2,\cdots,T}\left\|S_{ze,t}\right\| = O_p\bigg(\frac{\big(\log T\big)^{1/\alpha}}{\sqrt{N}}\Big(\log^{1/\alpha} T + (\frac{H}{T})^{\gamma_3}\Big)\bigg). \tag{B.3}$$

These together establish (i) and (ii).

Proof of (B.1). Observe that

$$\begin{aligned}
\left\|\Delta_{z,t}\right\| &\leqslant \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}\right\|_{sp}^{-1}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}(\beta_{0,j}-\beta_{0,t})\right\| \\
&= O_p\Big((\frac{H}{T})^{\gamma_2}\Big),
\end{aligned}$$

where the second line follows from (A.1) and (A.2) and the fact that by Assumption 2.4(i) all terms are *i.i.d.* across i. Again, by uniform rates derived in (A.1) and (A.2), we have

$$\begin{aligned}
\max_{t=1,2,\cdots,T}\left\|\Delta_{z,t}\right\| &\leqslant \frac{1}{N}\sum_{i=1}^{N}\max_{t=1,2,\cdots,T}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}\right\|_{sp}^{-1}\max_{t=1,2,\cdots,T}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}(\beta_{0,j}-\beta_{0,t})\right\| \\
&= O_p\Big(\big(\log T\big)^{2/\alpha}(\frac{H}{T})^{\gamma_2}\Big).
\end{aligned}$$

Proof of (B.2). Let

$$Z_i^{zu} = (\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij})^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}u_{ij}).$$

Clearly, by (A.1) and (A.3):

$$\left\|Z_i^{zu}\right\| \leqslant \left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}\right\|_{sp}^{-1}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}u_{ij}\right\| = O_p\Big(\frac{1}{\sqrt{H}}\Big).$$

Thus, we have

$$E\left\|\frac{1}{N}\sum_{i=1}^{N} Z_i^{zu}\right\|^2 = \frac{1}{HN^2}\sum_{i=1}^{N} HE\left\|Z_i^{zu}\right\|^2 = O\Big(\frac{1}{NH}\Big),$$

which implies that

$$\left\|S_{zu,t}\right\| = O_p\Big(\frac{1}{\sqrt{NH}}\Big).$$

To derive the uniform rate, we follow a similar reasoning as above, except for the fact that we have to use uniform rates derived in (A.1) and (A.3) to obtain

$$\max_{t=1,2,\cdots,T}\left\|S_{zu,t}\right\| = O_p\Big(\frac{1}{\sqrt{N}}\Big(\bar{r}_{T,L,\gamma_1,\alpha} + \Big(\log T\Big)^{1/\alpha} r_{T,H,\gamma_2}\Big)\Big).$$

Proof of (B.3). Let

$$Z_i^{ze} = (\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij})^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}e_{ij}).$$

Clearly, by (A.1) and (A.4):

$$\left\|Z_i^{ze}\right\| \leqslant \left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}\right\|_{sp}^{-1}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}e_{ij}\right\| = O_p(1).$$

Thus, we have

$$E\left\|\frac{1}{N}\sum_{i=1}^{N} Z_i^{ze}\right\|^2 = \frac{1}{N^2}\sum_{i=1}^{N} E\left\|Z_i^{ze}\right\|^2 = O\Big(\frac{1}{N}\Big),$$

which implies that

$$\left\|S_{ze,t}\right\| = O_p\Big(\frac{1}{\sqrt{N}}\Big).$$

To derive the uniform rate, we follow again a similar reasoning as above, except for the fact that we have to use uniform rates derived in (A.1) and (A.4) to obtain

$$\max_{t=1,2,\cdots,T}\left\|S_{ze,t}\right\| = O_p\Big(\frac{\big(\log T\big)^{1/\alpha}}{\sqrt{N}}\Big(\log^{1/\alpha} T + (\frac{H}{T})^{\gamma_3}\Big)\Big).$$

Now, we sum them up. By (B.1), (B.2) and (B.3), we first obtain the uniform consistency rate

$$\begin{aligned}\max_{t=1,2,\cdots,T}\left\|\hat{\beta}_{MG,t}^{IV} - \beta_{0,t}\right\| &\leqslant \max_{t=1,2,\cdots,T}\left\|\Delta_{z,t}\right\| + \max_{t=1,2,\cdots,T}\left\|S_{zu,t}\right\| + \max_{t=1,2,\cdots,T}\left\|S_{ze,t}\right\| \\ &= O_p\Big(\frac{\bar{r}_{T,L,\gamma_1,\alpha} + \big(\log T\big)^{1/\alpha} r_{T,H,\gamma_2}}{\sqrt{N}} + \frac{\big(\log T\big)^{1/\alpha}}{\sqrt{N}}(\frac{H}{T})^{\gamma_3} + \frac{\log^{2/\alpha} T}{\sqrt{N}}\Big).\end{aligned}$$

Then, we obtain the expansion of the estimator

$$\sqrt{N}(\hat{\beta}_{MG,t}^{IV} - \beta_{0,t}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big[(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij})^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}e_{ij})\Big] + O_p\Big(\sqrt{N}(\frac{H}{T})^{\gamma_2}\Big) + O_p\Big(\frac{1}{\sqrt{H}}\Big).$$

Because we assume that $(\frac{H}{T})^{\gamma_2} = o(N^{-1/2})$ as $(N,T) \longrightarrow \infty$, the last two terms above are $o_p(1)$ and the dominating term is the first one. Recall from the derivation of (A.4), we have that

$$\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}e_{ij} = \Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}\Big)\, e_{it} + o_p(1).$$

Then, by continuing from above, we have

$$\sqrt{N}\,(\hat{\beta}^{IV}_{MG,t} - \beta_{0,t}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_{it} + o_p(1).$$

Since (e_{it}) is $i.i.d.$ across i, we can apply CLT for $i.i.d.$ sequence to obtain

$$\sqrt{N}\,(\Sigma_{e,t})^{-1/2}\,(\hat{\beta}^{IV}_{MG,t} - \beta_{0,t}) \xrightarrow{d} N(0, I_k),$$

where $\Sigma_{e,t}$ is given by

$$\Sigma_{e,t} = \lim_{N\to\infty} \operatorname{Var}\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_{it}\Big).$$

We now show that $\Sigma_{e,t}$ can be consistently estimated by

$$\hat{\Sigma}_{e,t} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\beta}^{IV}_{it} - \hat{\beta}^{IV}_{MG,t})(\hat{\beta}^{IV}_{it} - \hat{\beta}^{IV}_{MG,t})'. \tag{B.4}$$

Because

$$\begin{aligned}\hat{\beta}^{IV}_{it} - \beta_{0,t} &= \Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}\Big)^{-1}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_{ij}z_{ij}x'_{ij}e_{ij}\Big) + O_p\Big(\frac{1}{\sqrt{H}}\Big) + O_p\Big((\frac{H}{T})^{\gamma_2}\Big)\\ &= e_{it} + o_p(1)\end{aligned}$$

and

$$\hat{\beta}^{IV}_{MG,t} - \beta_{0,t} = \frac{1}{N}\sum_{i=1}^{N} e_{it} + o_p(1),$$

we have

$$\hat{\beta}^{IV}_{it} - \hat{\beta}^{IV}_{MG,t} = e_{it} - \frac{1}{N}\sum_{i=1}^{N} e_{it} + o_p(1).$$

Then, as $(N,T) \longrightarrow \infty$,

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\beta}^{IV}_{i,t} - \hat{\beta}^{IV}_{MG,t})(\hat{\beta}^{IV}_{i,t} - \hat{\beta}^{IV}_{MG,t})' \xrightarrow{p} \Sigma_{e,t},$$

which implies that (B.4) is a consistent estimator for $\Sigma_{e,t}$.

$\square$