Estimation and inference in large heterogeneous panels with stochastic time-varying coefficients *,**

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

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

Abstract

In this paper, we consider kernel-based non-parametric estimation and inferential theory for large heterogeneous panel data models with stochastic time-varying coefficients. We propose mean group and pooled estimators, derive asymptotic distributions and show the uniform consistency and asymptotic normality of path coefficients. Then, we extend the procedures to the case with possibly endogenous regressors and propose a time-varying version of the Hausman exogeneity test. The finite sample performance of the proposed estimators is investigated through a Monte Carlo study and an empirical application on multi-country Phillips curve with time-varying parameters.

Keywords: Non-Parametric Methods, Large Heterogeneous Panels, Time-Varying Parameters, Mean Group and Pooled Estimators, Hausman Exogeneity Test

JEL classification: C14, C26, C51

*This version: February 2022

**We would like to thank 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 and 26th International Panel Data Conference for their useful comments. Bai and Marcellino thank MIUR–PRIN Bando 2017 – prot. 2017TA7TYC for financial support for this research.

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. Surveys of the literature on large heterogeneous panels are provided by Hsiao and Pesaran (2008) and Chapter 28 of Pesaran (2015). 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, (Ul Haque et al., 1999)).

Various methods have been proposed to identify and handle structural change in econometric models. As parameter instability is pervasive (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. Evolutions of parameters can be either discrete and abrupt, such as in Markov Switching models (e.g., Hamilton, 1989), or continuous and smooth. Continuous and smooth time variation can be driven by observed variables, as in smooth transition models (Teräsvirta, 1994), or by unobserved shocks, as in random coefficient models (e.g., Nyblom, 1989). In these models, parameters typically evolve as random walk or autoregressive processes and are mostly estimated by the Kalman Filter in classical context or by Bayesian Markov Chain Monte Carlo (MCMC) methods.

Yet another strand 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 the realization of persistent stochastic processes, as in Giraitis et al. (2014, 2018). 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. Such estimates are nonparametric and can have computational advantages over MCMC and other simulation-based methods.[1] 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. Liu et al. (2018) and Liu et al. (2020) develop methods to estimate time-varying coefficients in large panel data models with cross-sectional dependence, but focus on the case of exogenous regressors.

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

[1] See, for instance, Kapetanios et al. (2019), for a comparison between kernel-based methods and simulation-based methods in a vector autoregression context.

well as instruments and endogenous variables may vary over time. Recently, some papers have attempted to develop estimation methods in the IV framework that account for the possible presence of parameter instability. Hall et al. (2012) develop inferential theory for linear GMM estimator with endogenous regressors in the structural change context. Chen (2015) extends the framework of deterministic smooth evolution of parameters to the IV case. Giraitis et al. (2020) 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 sample path of stochastic processes. There is also limited but growing attention in the panel data literature to models allowing for both endogenous regressors and parameter instability. Baltagi et al. (2016) and Baltagi et al. (2019) develop an estimation procedure for large heterogeneous panels with cross-sectional dependence in the structural change context by extending the work of Pesaran (2006) and Harding and Lamarche (2011).

This paper makes the following contributions to the literature. First, we introduce a new class of large heterogeneous panels in which parameters are not only heterogeneous across cross sections but also vary stochastically over time. The model extends: (1) standard random coefficient panel data model as in Hsiao and Pesaran (2008); (2) time-varying homogenous panel data model with exogenous regressors as in Chen and Huang (2018) and Gao et al. (2020). The extensions are significant. Structural change leads to time-varying mean coefficients in standard random coefficient panel data models. The commonly used assumption in the nonparametric modeling of time-varying parameters literature implies that 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). In ongoing work (Bai et al. (2022)), we develop inference procedures for the degree of smoothness of time variation in time-varying parameter models. We find that the commonly used assumption in Chen and Hong (2012), Chen (2015), Chen and Huang (2018) and Gao et al. (2020) is very often rejected by the data. More details are provided in Remark 3.

We propose non-parametric kernel-based mean group and pooled estimators, derive their asymptotic distributions and show the uniform consistency and asymptotic normality of the time-varying mean coefficients. Then, we extend the work of Giraitis et al. (2020) to large heterogeneous panels. We show that both kernel-based mean group and pooled estimators can be extended to settings with possibly endogenous regressors. We also derive the properties of time-varying IV mean group and pooled estimators and show their uniform consistency and asymptotic normality under similar conditions to the case of time-varying least square estimators. We further propose a pointwise time-varying version of the Hausman exogeneity test in a large heterogeneous panels context, which compares time-varying OLS and IV estimators, possibly also allowing for changes in the endogeneity status of regressors over time. The finite sample performance of proposed estimators and the time-varying Hausman test is evaluated in an extensive Monte Carlo study. For the estimators, we evaluate the biases of point estimates and coverage probabilities of the time-varying mean coefficients under the scenarios of both exogenous and endogenous regressors.

We also compute both the size and power of the time-varying Hausman test. The results are encouraging, and can be also used to provide guidelines on the choice of the kernel bandwidth parameters.

Finally, we provide an empirical application to explore in practice the use of our proposed estimators. We estimate panel versions of time-varying hybrid Phillips curves with 19 Eurozone countries over the period 2000M1–2019M12. We find trade-off between unemployment and inflation is time-varying, but the coefficients are small and only significant roughly around the period of year 2005 and 2014–2016. Endogeneity issues may arise not only because inflation expectation is not observed, but also the fact that inflation is measured with error. In general, IV delivers much larger estimates than OLS for persistent parameters. Backward-looking is a dominating feature for Eurozone inflation, except for the period around year 2015.

The remainder of this paper is organized as follows. Section 2 describes our framework and the time-varying least square estimators, and derives the related theoretical results. Section 3 extends the work to the case of possibly endogenous regressors, proposes time-varying IV estimators, derives their theoretical properties and introduces the time-varying Hausman test. In Section 4 we evaluate our proposed estimators and pointwise Hausman exogeneity test in an extensive Monte Carlo study, under the scenarios of both exogenous and endogenous regressors. Section 5 presents the empirical application related to multi-country Phillips curves. Section 6 summarizes our main results and concludes the paper. The proofs of all results are presented in the appendices.

NOTATION: The letter C stands for a generic finite positive constant, $\|A\|_{sp} = \sqrt{\lambda_{\max}(A'A)}$ is the spectrum norm of matrix A, where $\lambda_{\max}(\cdot)$ is the maximum eigenvalue of $\cdot$. $\|\cdot\|_p$ denotes the L^p norm, $\|\cdot\|$ is the Euclidean norm. $|\cdot|_p$ and $|\cdot|$ denote the associated norm when $\cdot$ is one dimensional. The operator $\xrightarrow{p}$ denotes convergence in probability, and $\xrightarrow{d}$ denotes convergence in distribution. $(N, T) \to \infty$ denotes joint convergence of N and T. All variables are assumed to be defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$.

2. Theoretical considerations

In this section, we present our model and set out the proposed estimators and their properties. We consider the following model:

$$y_{it} = x_{it}'\beta_{it} + u_{it}, \quad i = 1, 2, ..., N, \;\; t = 1, 2, ..., T, \tag{1}$$

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., $x_{1,it} = 1$), and u_{it} is the disturbance term. We make the following assumptions on this model.

Assumption 2.1. *Elements in x_{it}, u_{it} have the properties:*

(i) There exists $\theta > 4$ such that $E\left|x_{\ell,it}\right|^{\theta} < \infty$ and $E\left|u_{\ell,it}\right|^{\theta} < \infty$, uniformly over ℓ, i, t;

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

$$\alpha_k^j \leqslant c_j(\phi_j)^k, \quad k \geqslant 1 \tag{2}$$

for some $0 < \phi_j < 1$ *and* $c_j > 0$*, where* $j = \{x, u\}$.

Assumption 2.2. *The coefficients* β_{it} *follow the random coefficient model:*

$$\beta_{it} = \beta_{0,t} + e_{it}, \quad i = 1, 2, \cdots, N, \quad t = 1, 2, \cdots, T, \tag{3}$$

where $(\beta_{0,t})_t = (E(\beta_{it}))_t$ *are the sequences of time-varying mean coefficients of the processes* $(\beta_{it})_t$*, which are uniformly bounded in t.*

(i) Let $e_i = (e_{i1}, e_{i2}, \cdots, e_{iT})'$, $u_i = (u_{i1}, u_{i2}, \cdots, u_{iT})'$, $x_i = (x'_{i1}, x'_{i2}, \cdots, x'_{iT})'$, $(x_i, u_i, e_i)'$ *are independently distributed over i;*

(ii) $\forall(\ell,i,t)$, $E[e_{\ell,it}|x_{\ell,it}] = 0$*;*

(iii) $\forall(\ell,i,t)$, $E[u_{\ell,it}|x_{\ell,it}, e_{\ell,it}] = 0$.

Assumption 2.3. $\forall(\ell,i,t)$*, elements in* $\beta_{0,t} = (\beta_{\ell,0,t})$, $e_{it} = (e_{\ell,it})$ *satisfy the following smoothness condition:*

$$\left|\beta_{\ell,0,t} - \beta_{\ell,0,s}\right| \leqslant C\left(\frac{|t-s|}{T}\right)^{\gamma_1}, \tag{4}$$

$$\left|e_{\ell,it} - e_{\ell,is}\right| \leqslant \left(\frac{|t-s|}{T}\right)^{\gamma_2} r_{\ell,i,ts} \quad t, s = 1, 2, \cdots, T \tag{5}$$

for some $0 < \gamma_1, \gamma_2 \leqslant 1$ *and the distribution of each variable in* $X_u^{(1)} = \{e_{\ell,it}, r_{\ell,i,ts}\}$ *has a thin tail* $\mathcal{E}(\alpha)$*:*

$$\mathbb{P}(\left|X_u^{(1)}\right| \geqslant \omega) \leqslant \exp(-c_1|\omega|^{\alpha}), \quad \omega > 0, \tag{6}$$

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

Assumption 2.4. $\forall(i,t)$*, we have*

$$E\left[\max_{t=1,\cdots,T} \left\|x_{it}x'_{it}\right\|_{sp}\right] < \infty, \quad \Omega_{e,t} = \lim_{N\to\infty} Var\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_{it}\Big) = O(1).$$

Assumption 2.1(i) imposes some moment conditions on regressors 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 and Huang (2018), since we allow (u_{it}) and $(x_{it}u_{it})$ to be serially correlated sequences. In Assumption 2.2(i), as in (3), $\beta_{0,t}$ is a $K \times 1$ vector of time-varying mean coefficients and $(e_{it})_t$ is a

K-dimensional stochastic process for each i. If $\beta_t = \beta$ and $e_{it} = e_i$, $\forall t$, the model simplifies to the standard random coefficient model settings with time-invariant coefficient (see Hsiao and Pesaran, 2008). In the random coefficient autoregressive panel model framework, 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. Assumption 2.2(ii) and (iii) impose exogeneity conditions for x_{it}. Assumption 2.2(ii) rules out correlated random effects. Assumption 2.2(iii) imposes contemporaneous exogeneity condition on x_{it}. x_{it} are neither correlated with error terms nor with idiosyncratic components in β_{it}.

Remark 1. Consider $x_{1,it} = 1$, let $x_{-1,it}$ be the remaining regressors in x_{it} and $\beta_{-1,it}$ be the associated coefficients. Then, the model (1) can be rewritten as $y_{it} = \beta_{1,0,t} + x'_{-1,it}\beta_{-1,it} + e_{1,it} + u_{it}, \quad i = 1, 2, ..., N, \quad t = 1, 2, ..., T$. As shown in Theorem 1, $\beta_{1,0,t}$ can be consistently estimated by our proposed kernel methods. $e_{1,it}$ captures unobserved heterogeneity across both i and t. We just need to assume that it changes smoothly over time. A sample path of $e_{1,it}$ can be estimated indirectly using (7) below, by replacing x_{ij} with 1 and y_{ij} with $y_{ij} - \hat{\beta}_{1,j} + x'_{-1,it}\hat{\beta}_{-1,ij}$.

Assumption 2.3 imposes restrictions on both mean coefficients $\beta_{0,t}$ and stochastic components e_{it}. If $\beta_{\ell,0,t} = g_\ell(t/T)$, where $g_\ell(\cdot) : [0,1] \longrightarrow \mathbb{R}$ is a deterministic function, (4) implies that $g_\ell(\cdot)$ is Hölder continuous with exponent γ_1. This condition is more general than the ones imposed in the literature, such as Chen and Huang (2018) and Gao et al. (2020), in which $g_\ell(\cdot)$ is assumed to have a continuous second order derivative. $\beta_{\ell,0,t}$ can also be sample path of a persistent stochastic process: $\beta_{\ell,0,t} = \beta_{\ell,t}(\omega)$, for $\omega \in \Omega$. The stochastic process $(\beta_{\ell,t})_t$ is assumed to satisfy a similar condition as in (5): $\left|\beta_{\ell,t} - \beta_{\ell,s}\right| \leqslant \left(\frac{|t-s|}{T}\right)^{\gamma_1} r_{\ell,ts}$, where both $\beta_{\ell,t}$ and $r_{\ell,ts}$ have a thin tail distribution $\mathcal{E}(\alpha)$, as defined in (6). The condition implies that $(\beta_{\ell,t})_t$, $(e_{\ell,it})_t$ are persistent stochastic processes with bounded variation. For example, consider an array of random processes (bounded random walk) defined as $\beta_{\ell,t} = \frac{1}{\sqrt{T}}u_{\ell,t}$, where $u_{\ell,t}$ are random walk processes: $u_{\ell,t} - u_{\ell,t-1} \overset{i.i.d.}{\sim} N(0,1)$, satisfying Assumption 2.3. As shown in Lemma 1 of Dendramis et al. (2020), Assumption 2.3 is satisfied if $\omega_{\ell,t} = u_{\ell,t} - u_{\ell,t-1}$ is α-mixing and has a thin tail.[2] Other allowable processes are discussed by Giraitis et al. (2014, 2018).

Remark 2. The conditions in (4) and (5) are the key assumptions to ensure consistency and applicability of central limit theorem (CLT). As in standard random coefficient model literature (Hsiao and Pesaran (2008)), our focus is the time-varying mean coefficients $\beta_{0,t}$, $t = 1, 2, \cdots, T$, which can be either deterministic functions of time or sample paths of persistent stochastic processes. This is also notably different from microeconometrics literature in which the focus is to identify the distributional structure of β_{it}, see, for instance, Graham and Powell (2012) and Li (2021).

[2] Dendramis et al. (2020) derive this under the case $\gamma_1 = 0.5$, but extensions are quite straightforward.

Remark 3. It is important to emphasise that the use of the parameter γ_1 allows for a wide variety of behaviour for $\beta_{0,t}$. As noted above, the existing literature deals with a subset of cases when γ_1 is equal to 1. In ongoing work (Bai et al. (2022)), we develop inference procedures for γ_1 in time-varying parameter regressions models. We find that $\gamma_1 = 1$ is very often rejected by the data, in favour of $\gamma_1 < 1$.

The first part of Assumption 2.4 implies that the second order moment of (x_{it}) is uniformly bounded for each i. This together with second part of Assumption 2.4 guarantees that the asymptotic variance of the proposed estimators are positive definite.

The main objective in this section is to construct estimators for $\beta_{0,t}$ and derive uniform consistency rates and asymptotic distributions for the estimators. The individual specific estimates can be obtained from a time-varying parameter least-square estimator (TVP-OLS):

$$\hat{\beta}_{i,t} = \Big(\sum_{j=1}^{T} b_{j,t}(H)x_{ij}x_{ij}'\Big)^{-1}\Big(\sum_{j=1}^{T} b_{j,t}(H)x_{ij}y_{ij}\Big). \tag{7}$$

The kernel weights $b_{j,t}(H)$ are defined as

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

where the bandwidth parameter H satisfies $H = o(T)$ as $T \to \infty$. $K(x)$ is a non-negative continuous function with either bounded or unbounded support satisfying

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

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

As in the literature of large heterogenous panels, we propose two types of estimators. The first is a mean group estimator (TVP-OLS-MG), given by

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

where $\hat{\beta}_{i,t}$ is defined in (7). The second is a pooled estimator (TVP-OLS-P):

$$\hat{\beta}_{P,t} = \Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{j,t}(H)x_{ij}x_{ij}'\Big)^{-1}\Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{j,t}(H)x_{ij}y_{ij}\Big). \tag{9}$$

Before analyzing the theoretical properties of these estimators, we assume that the bandwidth parameter

H satisfies:

$$c_1 T^{1/(\theta/4-1)+\delta_1} \leqslant H \leqslant c_2 T^{1-\delta_2} \tag{10}$$

for some $\theta > 4$ as in Assumption 2.1, and $c_1, c_2 > 0$, $\delta_1, \delta_2 > 0$ are sufficiently small. In the next theorem, we establish uniform consistent rate, asymptotic distributions for (8) and (9) and how to obtain consistent estimates of asymptotic covariance matrices.

Theorem 1. *Under Assumptions 2.1–2.4 and assuming that the bandwidth parameter H satisfies* (10)*, as* $(N, T) \to \infty$*, we have the following:*

(1) Uniform consistency: $\hat{\beta}_{MG,t}$, $\hat{\beta}_{P,t}$ *have the property*

$$\max_{t=1,2,\cdots,T} \left\| \hat{\beta}_{\int,t} - \beta_{0,t} \right\| = O_p(r_{N,T,H,\gamma_1,\alpha}),$$

where $r_{N,T,H,\gamma_1,\alpha} = (\frac{H}{T})^{\gamma_1} + \sqrt{\frac{\log T}{NH}} + \frac{\log^{1/\alpha} T}{\sqrt{N}}$ *for* $\int = \{MG, P\}$.

(2) Asymptotic normality: Suppose that $(\frac{H}{T})^{\gamma_1} = o(N^{-1/2})$*, for* $t = \lfloor Tr \rfloor$, $0 < r < 1$,

(i) Mean group estimator:

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

where $\Omega_{e,t}$ *is given by*

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

(ii) Pooled estimator: Let $\overline{\Sigma}_{xx,t} = \operatorname{plim}_{(N,T)\to\infty} \frac{1}{NK_t} \sum_{i=1}^{N} \sum_{j=1}^{T} b_{jt} x_{ij} x_{ij}' = O_p(1)$*, then*

$$\sqrt{N}\, \overline{\Sigma}_{xx,t}\ R_t^{-1/2} \left(\hat{\beta}_{P,t} - \beta_{0,t}\right) \xrightarrow{d} N(0, I_k),$$

where R_t *is given by*

$$R_t = \lim_{N\to\infty} Var\Big(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Sigma_{xx,it} e_{it}\Big),$$

and

$$\Sigma_{xx,it} = \operatorname*{plim}_{T\to\infty} \frac{1}{K_t} \sum_{j=1}^{T} b_{jt} x_{ij} x_{ij}'.$$

Even though both $\Omega_{e,t}$ and R_t contain e_{it}, which is not observed, in Appendix B1, we show that they can be consistently estimated by $\hat{\Omega}_{e,t}$ and $\hat{R}_t$:

$$\hat{\Omega}_{e,t} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\beta}_{i,t} - \hat{\beta}_{MG,t}\right)\left(\hat{\beta}_{i,t} - \hat{\beta}_{MG,t}\right)' \tag{11}$$

$$\hat{R}_t = \frac{1}{N}\sum_{i=1}^{N}\left[\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)(\hat{\beta}_{i,t} - \hat{\beta}_{MG,t})(\hat{\beta}_{i,t} - \hat{\beta}_{MG,t})'\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)\right]. \tag{12}$$

Three comments are in order. First, for the uniform consistency rate $r_{N,T,H,\gamma_1,\alpha}$, the first term is related to the degree of smoothness of time-varying mean coefficients ((4) in Assumption 2.3). Panel dimension N is helpful to improve the rate, as N appears in both the second the third term. The third term is due to the heterogeneity in model parameters e_{it}. Second, as in the setting of standard random coefficient model (Hsiao et al. (1998)), both estimators have standard root N rate of convergence. However, the requirement $(\frac{H}{T})^{\gamma_1} = o(N^{-1/2})$ is different from time series setting (see, for instance, Giraitis et al. (2014)), which is $(\frac{H}{T})^{\gamma_1} = o(H^{-1/2})$. Third, asymptotic variances can be consistently estimated nonparametrically, as in (11).

The use of estimates $\hat{\beta}_{MG,t}$ and $\hat{\beta}_{P,t}$ makes it necessary to choose the bandwidth parameter H. Whereas uniform consistency holds under minimal restrictions on H (see (10)), asymptotic normality results require stronger restrictions: $(\frac{H}{T})^{\gamma_1} = o(H^{-1/2})$. This condition implies that $\sqrt{N}(\frac{H}{T})^{\gamma_1} \to 0$ when both $(N,T) \to \infty$. As in Giraitis et al. (2014, 2020), a practical suggestion for H is to set $H = T^{\overline{\alpha}}$, for some $0 < \overline{\alpha} < 1$. Then, the condition simplifies to $\sqrt{N}/T^{\gamma_1(1-\overline{\alpha})} \to 0$. If we assume that β_t follows the bounded random walk process and set $H = T^{0.5}$, the condition becomes $\sqrt{N}/T^{0.25} \to 0$ as $(N,T) \to \infty$. A practically meaningful implication is that T has to diverge at a faster rate than N.

3. Endogenous regressors

Consider again the model proposed in section 2:

$$\begin{aligned} y_{it} &= x_{it}'\beta_{it} + u_{it}, \\ \beta_{it} &= \beta_{0,t} + e_{it}, \quad i = 1, 2, \cdots, N, \quad t = 1, 2, \cdots, T. \end{aligned}$$

In Assumption 2.2(iii), we impose contemporaneous exogeneity condition on x_{it}, which is rather restrictive. This condition is likely to be violated in many empirical applications, due to simultaneity, measurement error, or omitted variables.

In this section, we consider an extension of the proposed TVP-OLS type estimators to the Instrumental variable regression (IVR) context. Let us consider the following model:

$$y_{it} = x_{it}'\beta_{it} + u_{it}, \tag{13}$$

$$x_{it} = \Psi_{it}'z_{it} + v_{it}, \tag{14}$$

$$\beta_{it} = \beta_{0,t} + e_{it}, \quad i = 1, 2, \cdots, N, \quad t = 1, 2, \cdots, T, \tag{15}$$

where $z_{it} = (z_{1,it}, z_{2,it}, \cdots, z_{p,it})'$ is a $p \times 1$ vector of instruments, $\Psi_{it}' = (\psi_{\ell k,it})$ is a $k \times p$ parameter matrix and $v_{it} = (v_{1,it}, v_{2,it}, \cdots, v_{p,it})'$ is a $k \times 1$ vector of error terms. We shall make the following additional assumptions.

Assumption 3.1. *Elements in z_{it}, v_{it} have the properties:*

(i) There exists $\theta > 4$ such that $E\left|z_{\ell,it}\right|^{\theta} < \infty$ and $E\left|v_{\ell,it}\right|^{\theta} < \infty$, uniformly over ℓ, i, t;

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

$$\alpha_k^j \leqslant c_j(\phi_j)^k, \quad k \geqslant 1 \tag{16}$$

for some $0 < \phi_j < 1$ and $c_j > 0$, where $j = \{z, v\}$.

Assumption 3.2. *The coefficients Ψ_{it} follow the random coefficient model:*

$$\Psi_{it} = \Psi_{0,t} + \Upsilon_{it}, \quad i = 1, 2, \cdots, N, \quad t = 1, 2, \cdots, T, \tag{17}$$

where $\left(\Psi_{0,t}\right)_t = \left(E(\Psi_{it})\right)_t$ are the sequences of time-varying mean coefficients of the processes $\left(\Psi_{it}\right)_t$, which are uniformly bounded in t.

(i) Let $\Upsilon_i = (\Upsilon_{i1}, \Upsilon_{i2}, \cdots, \Upsilon_{iT})'$, $v_i = (v_{i1}, v_{i2}, \cdots, v_{iT})'$, $z_i = (z_{i1}', z_{i2}', \cdots, z_{iT}')'$, $(x_i, z_i, u_i, v_i, v_i, \Upsilon_i)$ are independently distributed over i;

(ii) $\forall \ell, k$, $E[u_{\ell,it}|z_{\ell,it}, e_{\ell,it}] = 0$, $E[v_{\ell,it}|z_{\ell,it}, \Upsilon_{\ell k,it}] = 0$, $\forall (i, t)$;

(iii) $\forall \ell, k$, $E[\Upsilon_{\ell k,it}|z_{\ell,it}] = 0$, $E[e_{\ell,it}|x_{\ell,it}, z_{\ell,it}, \Upsilon_{\ell k,it}] = 0$, $\forall (i, t)$.

Assumption 3.3. *$\forall \ell, k, i, t$, elements in $\Psi_{0,t} = (\psi_{\ell k,0,t})$, $\Upsilon_{it} = (\upsilon_{\ell k,it})$ satisfy the following smoothness condition:*

$$\left|\psi_{\ell k,0,t} - \psi_{\ell k,0,s}\right| \leqslant \Big(\frac{|t-s|}{T}\Big)^{\gamma_3}, \quad \left|\upsilon_{\ell k,it} - \upsilon_{\ell k,is}\right| \leqslant \Big(\frac{|t-s|}{T}\Big)^{\gamma_4} q_{\ell k,i,ts}, \quad t, s = 1, 2, \cdots, T \tag{18}$$

for some $0 < \gamma_3, \gamma_4 < 1$ and the distribution of each variable in $X_u^{(2)} = \{\upsilon_{\ell k,t}, q_{\ell k,i,ts}\}$ has a thin tail $\mathcal{E}(\alpha)$:

$$\mathbb{P}(\left|X_u^{(2)}\right| \geqslant \omega) \leqslant \exp(-c_1|\omega|^{\alpha}), \quad \omega > 0,$$

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

Assumption 3.4. $\forall(i, t)$, *we have*

$$E\left[\max_{t=1,\cdots,T} \left\|z_{it}z_{it}'\right\|_{sp}\right] < \infty, \quad E\left[\max_{t=1,\cdots,T} \left\|z_{it}x_{it}'\right\|_{sp}\right] < \infty.$$

Since we introduce the first stage regression in our IV context, restrictions on moments and mixing conditions are imposed for z_{it} and v_{it} as in Assumptions 3.1. In Assumption 3.2, we impose similar conditions for random coefficients Ψ_{it} as for β_{it}. Additional assumptions, as in Assumption 3.2(ii)-(iii), are required for identification[3]. Assumption 3.3 imposes similar smooth conditions on newly introduced time-varying components: $\psi_{0,t}$ and Υ_{it}, with different smoothness parameters γ_3, γ_4 and random component $q_{\ell k,ts}$. Assumption 3.4 parallels Assumptions 2.4 to ensure identification and the asymptotic variance of the proposed estimators are positive definite.

The main objectives in this section are to construct consistent estimates for $\beta_{0,t}$ and to derive asymptotic distributions for the estimators. We aim to generalize the estimators proposed in Chen (2015) and Giraitis et al. (2020) to the panel setting. The individual specific estimator can be obtained as the kernel-based two-stage least square estimator (2SLS):

$$\hat{\beta}_{i,t}^{IV} = \Big(\sum_{j=1}^{T} b_{j,t}(H)\hat{\Psi}_j' z_{ij}z_{ij}'\hat{\Psi}_j\Big)^{-1}\Big(\sum_{j=1}^{T} b_{j,t}(H)\hat{\Psi}_j' z_{ij}y_{ij}\Big), \tag{19}$$

where $b_{j,t}(H)$ is the kernel function with bandwidth H and $\hat{\Psi}_j$ are the consistent estimates of Ψ_j. At the first stage, $\hat{\Psi}_t$ can be easily obtained as either the TVP-OLS-MG estimator

$$\hat{\Psi}_{MG,t} = \frac{1}{N}\sum_{i=1}^{N} \hat{\Psi}_{i,t}, \tag{20}$$

where

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

or the TVP-OLS-P estimator

$$\hat{\Psi}_{P,t} = \Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{j,t}(L)z_{ij}z_{ij}'\Big)^{-1}\Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{j,t}(L)z_{ij}x_{ij}'\Big), \tag{21}$$

where $b_{j,t}(L)$ is the kernel function with bandwidth L, where L can be different from H. We assume that

[3]Note that, unlike the micro panel literature, these assumptions rule out the correlated random coefficient panel data model. In our context, endogeneity is caused by the fact that x_{it} are correlated with u_{it}.

both bandwidth parameters H, L satisfy (10).

Define

$$r_{N,T,H,\gamma,\alpha} = \left(\frac{H}{T}\right)^{\gamma} + \frac{\log^{1/\alpha} T}{\sqrt{N}} + \sqrt{\frac{\log T}{NH}}. \tag{22}$$

In the next lemma, we establish uniform consistency results for the estimates $\hat{\Psi}_{MG,t}$ and $\hat{\Psi}_{P,t}$. Since Assumptions 2.1–2.4 are satisfied for the first stage regression, it follows immediately from Theorem 1(1), with a possibly different bandwidth L and smoothness parameter γ_3.

Lemma 1. *Under Assumptions 3.1–3.4 and assuming that the bandwidth parameters L satisfy* (10)*, as* $(N,T) \to \infty$*, we have*

$$\max_{t=1,2,\cdots,T} \left\| \hat{\Psi}_{\int,t} - \Psi_{0,t} \right\|_{sp} = O_p(r_{N,T,L,\gamma_3,\alpha}).$$

Then, at the second stage, we propose two estimators for $\beta_{0,t}$. As in the previous section, we consider both mean group and pooled estimators. The TVP-IV-MG estimator is defined as

$$\hat{\beta}^{IV}_{MG,t} = \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}^{IV}_{i,t}, \tag{23}$$

where $\hat{\beta}^{IV}_{i,t}$ is given by (19). The TVP-IV-P estimator can be computed as

$$\hat{\beta}^{IV}_{P,t} = \Big(\sum_{i=1}^{N} \sum_{j=1}^{T} b_{j,t}(H) \hat{\Psi}'_j z_{ij} z'_{ij} \hat{\Psi}_j \Big)^{-1} \Big(\sum_{i=1}^{N} \sum_{j=1}^{T} b_{j,t}(H) \hat{\Psi}'_j z_{ij} y_{ij} \Big). \tag{24}$$

In the next theorem, we establish uniform consistency rates and asymptotic distributions for (23) and (24).

Theorem 2. *Under Assumptions 2.1–2.3 (except Assumption 2.2(iii)), Assumptions 3.1–3.4 and assuming that the bandwidth parameters H, L satisfy* (10)*, then as* $(N,T) \to \infty$ *we have the following:*

1. *Uniform consistency:* $\hat{\beta}^{IV}_{MG,t}$, $\hat{\beta}^{IV}_{P,t}$ *have the property*

$$\max_{t=1,2,\cdots,T} \left\| \hat{\beta}_{\int,t} - \beta_{0,t} \right\| = O_p\left(r_{N,T,H,\gamma_1,\alpha} + \frac{r_{N,T,L,\gamma_3,\alpha}}{\sqrt{N}} \right),$$

   where $\int = \{MG, P\}$ *and* $r_{N,T,H,\gamma,\alpha}$ *is defined as in* (22)*;*

2. *Asymptotic normality: Suppose that* $(\frac{H}{T})^{\gamma_1} = o(N^{-1/2})$, $\log^{1/\alpha} T = o(N^{-1/2})$, *for* $t = \lfloor Tr \rfloor$, $0 < r < 1$,

   (i) Mean group estimator:

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

where $\Omega^{IV}_{e,t}$ is given by

$$\Omega^{IV}_{e,t} = \lim_{N\to\infty} Var\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Sigma^{-1}_{\Psi zz\Psi,it}\Sigma_{\Psi zx,it}e_{it}\Big),$$

and

$$\Sigma_{\Psi zz\Psi,it} = \Psi'_{0,t}\Sigma_{zz,it}\Psi_{0,t},\quad \Sigma_{zz,it} = \operatorname*{plim}_{T\to\infty}\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}z_{ij}z'_{ij}$$

$$\Sigma_{\Psi zx,it} = \Psi'_{0,t}\Sigma_{zx,it},\quad \Sigma_{zx,it} = \operatorname*{plim}_{T\to\infty}\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}z_{ij}x'_{ij}.$$

(ii) Pooled estimator: Suppose that $\overline{\Sigma}_{\Psi zz\Psi,t} = \operatorname{plim}_{(N,T)\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\Psi'_j z_{ij}z'_{ij}\Psi_j = O_p(1)$, then

$$\sqrt{N}\,\overline{\Sigma}_{\Psi zz\Psi,t}\left(R^{IV}_{P,t}\right)^{-1/2}\left(\hat{\beta}^{IV}_{P,t} - \beta_{0,t}\right) \xrightarrow{d} N(0, I_k),$$

where $R^{IV}_{P,t}$ is given by

$$R^{IV}_{P,t} = \lim_{N\to\infty} Var\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Sigma_{\Psi zx,it}e_{it}\Big),$$

and $\Sigma_{\Psi zx,it}$ is defined in (ii).

As in (11), the above inference results can be operationalised by replacing Ψ_t with $\hat{\Psi}_t$ and the fact that

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

$$\hat{R}^{IV}_{P,t} = \frac{1}{N}\sum_{i=1}^{N}\left[\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}x'_{ij}\Big)(\hat{\beta}^{IV}_{i,t} - \hat{\beta}^{IV}_{MG,t})(\hat{\beta}^{IV}_{i,t} - \hat{\beta}^{IV}_{MG,t})'\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}x'_{ij}\Big)\right] \xrightarrow{p} R^{IV}_{P,t}.$$

Several comments are in order. First, the uniform rates derived are different from Theorem 1. Since they are two-step estimators, the uniform rate from the first stage estimation is involved. Second, the IV estimators also have standard root N rate of convergence, except for an additional requirement $\log^{1/\alpha} T = o(N^{-1/2})$, which implies that T cannot diverge too fast compared to N. Finally, asymptotic variances can be consistently estimated similarly as in the TVP-OLS case.

As the endogeneity of x_{it} can change over time, we can apply the Hausman test for the null hypothesis of exogeneity for each t. The pointwise null hypothesis is formally stated as: $\mathcal{H}_0 : E(u_{it}v_{it}) = 0, \forall i$, for each t, $t = 1, 2, \cdots, T$. A common formulation of the Hausman test is in terms of the quadratic differences between the (time-varying) OLS and IV estimators. In our case, it is given by

$$\mathcal{H}_t = N\Big(\hat{\beta}^{IV}_{\int,t} - \hat{\beta}_{\int,t}\Big)' \hat{V}_H^{-1}\Big(\hat{\beta}^{IV}_{\int,t} - \hat{\beta}_{\int,t}\Big), \tag{25}$$

where $\hat{V}_H = \text{Avar}\Big(\hat{\beta}^{IV}_{\int,t} - \beta_t\Big) - \text{Avar}\Big(\hat{\beta}_{\int,t} - \beta_t\Big)$, $\int = \{MG, P\}$ and Avar is the asymptotic variance given in Theorems 1 and 2. Under Assumptions in Theorems 1 and 2, it can easily be shown that for each t, $\mathcal{H}_t \xrightarrow{d} \chi^2_k$, under the pointwise null hypothesis, that the regressors are exogenous for each t.

4. Monte Carlo study

In this section, we conduct Monte-Carlo experiments to evaluate the finite sample performance of the time-varying OLS and IV estimators (8), (9), (23), (24) and the pointwise time-varying Hausman test (25). We generate data using the model defined in (13) with one regressor x_{it}:

$$\begin{aligned} y_{it} &= x'_{it}\beta_{it} + u_{it}, \\ x_{it} &= \Psi'_{it} z_{it} + c_1 v_{it}, \quad i = 1, 2, ..., N, \quad t = 1, 2, ..., T, \end{aligned}$$

where we introduce an additional parameter c_1 to control the strength of instruments and we set it to 0.5. We introduce time-varying correlation between u_{it} and v_{it} by specifying them as

$$u_{it} = a(\alpha_{it} + s)e_{1,it} + e_{2,it}, \quad v_{it} = (\alpha_{it} + s)e_{1,it} + e_{3,it},$$

where $e_{j,it}$, $j = 1, 2, 3$ are generated independently from $N(0, 1)$ and $s = 1$. We set $a = 1$ if x_{it} is endogenous, otherwise (if exogenous) we set $a = 0$. We also introduce a time-varying component α_{it} to measure the time-varying correlations between u_{it} and v_{it}.

The time-varying parameters β_{it}, Ψ_{it} and α_{it} are generated according to

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

where elements in $X^{(1)}_{MC} = \{\beta_{0,t}, e_{it}, \psi_{0,t}, \Upsilon_{it}, \alpha_{0,t}, \iota_{it}\}$ are generated from the scaled random walk processes, such that $X_{\ell,t} = \xi_{\ell,t}/\sqrt{t}$, for $\xi_{\ell,t} - \xi_{\ell,t-1} \overset{i.i.d.}{\sim} N(0, 1)$, $t = 1, 2, \cdots, T$. The instrument z_{it} is again generated from $N(0, 1)$ and it is independent from $e_{j,it}$, $j = 1, 2, 3$ and elements in $X^{(1)}_{MC}$. Each experiment was replicated 1,000 times for each (N, T) pair with $N, T = 50, 100, 200, 500$.

To compute both TVP-OLS and TVP-IV type estimators, we use a two-sided normal kernel $K(x) = \exp(-x^2/2)$ with bandwidth set to take values $T^{\overline{\alpha}}$ for $\overline{\alpha} = 0.2, 0.4, 0.5, 0.7$. Lower values of $\overline{\alpha}$ increase the robustness of estimates to parameter changes but decrease efficiency, and in the panel case lower values for $\overline{\alpha}$ also make it more likely the condition $(\frac{H}{T})^{\gamma_1} = o(N^{-1/2})$ holds. Thus, it is interesting to evaluate the impact of the bandwidth on the performance of estimators and of the pointwise Hausman test for exogeneity. The global performance of the estimators 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, computed starting from the second-half of the sample period, $t = [T/2]+1, \cdots, T$. For the time-varying Hausman test, we report both the size and power of the test evaluated at the middle point $t = [T/2]$.

To get a rough idea of the estimates and 95% confidence intervals, we report a single replication of the estimates for $N = 50, T = 200$ and $H = T^{0.5}$, $L = H = T^{0.5}$ in Figures 1 and 2. Evidently, TVP-OLS estimators perform well when x_{it} is exogenous, but they lead to substantial bias and poor coverage rates when x_{it} is endogenous. The performance of TVP-IV estimators is quite satisfactory under the case of endogeneity of x_{it}.

Table 1 reports the average MAD and coverage probability of both TVP-OLS-MG and TVP-OLS-P estimators for the normal kernel with various values of bandwidth H. A number of comments can be made. First, both cross-sectional and time series dimensions are useful to reduce the bias of the estimates. All values become closer to zero as (N, T) increases. Second, regarding the bandwidth parameter $H = T^{\overline{\alpha}}$, smaller values of $\overline{\alpha}$ often yield the lowest values of MAD and higher coverage probabilities. However, results are very similar if $\overline{\alpha}$ is replaced with a value lower than 0.5. If $\overline{\alpha}$ is larger than 0.5, MAD increases and coverage probability decreases. Recall that the bandwidth parameter $H = T^{\overline{\alpha}}$ is required to be such that $H = o(T)$ and asymptotic normality requires $\frac{\sqrt{N}}{T^{(1-\overline{\alpha})\gamma}} \to 0$. This means that we cannot set $\overline{\alpha}$ to be a value close to 1 otherwise these conditions will fail. Finally, TVP-OLS-MG and TVP-OLS-P estimates deliver very similar results.

Tables 2 and 3 present average MADs and coverage probabilities, respectively, in the case of endogeneity. For the bandwidth parameters, we set $H = L$. We report the results from TVP-IV type estimates obtained either from mean group or pooled estimators at the first stage. For comparison purposes, we also include OLS results in the bottom panel of each table. First, TVP-OLS estimates clearly yield much large biases and lower coverage probabilities when x_{it} is endogenous. There is no sign of convergence as (N, T) increases, as expected. Second, TVP-IV estimates have much better performance with significantly lower MADs and higher coverage probabilities. Regarding the bandwidth, it seems that setting $H = T^{0.2} \sim T^{0.4}$ leads to the lowest MAD and $H = T^{0.2}$ has the best overall performance. However, smaller bandwidth is often associated with larger confidence intervals, which implies higher uncertainty around point estimates. In general, $H = T^{0.4}$ and $H = T^{0.5}$ deliver comparable results with $H^{0.2}$. Recall that asymptotic normality again requires $\frac{\sqrt{N}}{T^{(1-\overline{\alpha})\gamma}} \to 0$. If we let $\overline{\alpha}$ increase, we should also let T be much larger than N to make the above condition hold. As shown in Tables 2 and 3, when $\overline{\alpha}$ gets larger, the lowest MADs are obtained when $(N, T) = (500, 500)$ and the best coverage probabilities are always obtained when $(N, T) = (50, 500)$. Third, it does not seem to make a difference whether $\hat{\Psi}_j$ is obtained either by mean group or pooled estimates at the first stage. However, when the bandwidth is small, $H = T^{0.2}$, the pooled estimator seems to be slightly worse than the mean group estimator for MAD at the second stage, but not for coverage probability. As the bandwidth H increases, both mean group and pooled estimators deliver similar results in terms of both MAD and coverage probability.

Table 4 reports the size and power of the time-varying Hausman test with nominal size equal to 5%. The bandwidth parameters are set to $H = L$. First, we find that setting $H = T^{0.2}$ or $H = T^{0.4}$ leads to sizes close to the nominal value. As the bandwidth H increases, size distortion becomes sizable, especially when N is larger than T. Recall that an asymptotically χ^2 distribution of the test statistic requires asymptotic normality of the estimators to hold. As explained in the previous paragraph, if we increase α, we also need very large T (compared to N) to make the condition $\frac{\sqrt{N}}{T^{(1-\bar{\alpha})\gamma}} \to 0$ hold. For instance, consider the case when $\hat{\Psi}_j$ is obtained by pooled estimator and $H = T^{0.4}$; empirical size is 0.033 when $(N, T) = (50, 500)$, which is close to 0.05. However, when $(N, T) = (500, 50)$, empirical size becomes 0.256, which is clearly oversized. Second, in terms of power, because the convergence rate of the estimator is the square root of N, we expect that power increases with N, and this is confirmed by Table 4. The power also increases slightly with bandwidth H, and with larger (N, T). Setting $H = T^{0.5}$ leads to the highest power when $(N, T) = (500, 500)$. In general, test power is very similar across different values of bandwidth H.

5. Empirical application

In this section, we consider an empirical application on modeling inflation dynamics. We estimate a multi-country version of the Phillips curve that links inflation to unemployment and also possibly to inflation expectations. The main goals are to understand whether unemployment is indeed significantly related to inflation, when and if forward looking behavior of inflation dominates backward looking behavior and whether there are changes in these features over time.

We use monthly data for 19 Eurozone countries[4] over the period 2000M1–2019M12. We consider the hybrid Phillips curve, along the lines of Galí and Gertler (1999),

$$\pi_{i,t} = c_{i,t} + \gamma_{i,t}\pi_{i,t-1} + \alpha_{i,t}u_{i,t} + \rho_{i,t}\pi^e_{i,t+1} + \nu_{i,t}. \tag{26}$$

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

$$\pi_{i,t} = c_{i,t} + \gamma_{i,t}\pi_{i,t-1} + \alpha_{i,t}u_{i,t} + \rho_{i,t}\pi_{i,t+1} + \epsilon_{i,t}. \tag{27}$$

It is clear that, since $\pi^e_{i,t+1}$ is not observed, we have to replace it by $\pi_{i,t+1}$, endogeneity arises again due to measurement error. In view of Hansen and Lunde (2014) and Galí and Gambetti (2019), endogeneity may

[4]Data are taken from Eurostat. The countries are Austria, Belgium, Cyprus, Estonia, Finland, France, Germany, Greece, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Portugal, Slovenia, Slovakia and Spain.

also arise due to measurement error of $\pi_{i,t}$ and simultaneity of $\pi_{i,t}$ and $u_{i,t}$. We use an intercept, three lags of inflation and two lags of unemployment as instruments: $(1, \pi_{i,t-2}, \pi_{i,t-3}, \pi_{i,t-4}, u_{i,t-1}, u_{i,t-2})'$[5].

Figure 3 provides plots of estimates for $\rho_{0,t}$, $\alpha_{0,t}$ and $\gamma_{0,t}$. Left panel of Figure 3 provides plots of time-varying estimates for $\rho_{0,t}$, the mean coefficients on inflation expectations. OLS estimates are small (less than 0.2), only significant around 2005–2013 and remain almost stable over time. IV estimates are more volatile, and are significant in shorter periods, around 2004 and 2014. Middle panel of Figure 3 provides plots of time-varying estimates of $\hat{\alpha}_t$, the common part of coefficients on unemployment. Point estimates generally fluctuate below the zero line, which indicates that there is a trade-off between unemployment and inflation, but the coefficients are very small (less than -0.1) over the whole sample period. The estimates from both OLS and IV are roughly similar (except around 2008–2014), but IV estimates are smaller and are less ”significant” than OLS estimates. Both estimates are statistically significant and very similar around 2014–2018. Moreover, IV estimates are significant around 2005, but OLS estimates are significant around 2005-2010. Turning to the persistence parameter estimates, which are shown in the right panel of Figure 3, we see that OLS estimates are small (less than 0.2) and significant from 2016–2013. However, IV estimates are much larger until 2015. Our findings are in line with Hansen and Lunde (2014), who find that OLS estimates are biased towards zero for persistent parameter when time series are measured with error.

Plots of p-values of time-varying pointwise Hausman test from (27) are provided in Figure 4. The null of exogeneity is rejected in most of the period (until 2016). Indeed, we see from Figure 3 that OLS and IV estimates are very different for ρ_t and γ_t, but estimates for α_t from OLS and IV are almost identical after 2016.

As a robustness check, we report estimation results and p-value of time-varying pointwise Hausman test from both backward-looking Phillips curve and forward-looking Phillips curve, obtained either by setting $\rho_{i,t} = 0$ or $\gamma_{i,t} = 0$. The results are similar compared to Figures 3 and 4, but there are also some differences, particularly for forward-looking Phillips curve. First, IV estimates for α_t are also significant around 2005–2008. Second, there are shorter periods in which the null of exogeneity is rejected, round around 2005, 2010–2012 and 2014.

In summary, this simple but economically interesting empirical application highlights the importance of allowing for parameter time variation and usefulness of the IV method. There is a small but varying impact of unemployment on inflation. OLS deliver smaller persistence parameter estimates than IV. Forward-looking feature only dominates backward-looking feature for inflation around year 2015. Endogeneity may not only come from inflation expectation but also measurement error of inflation series. In terms of sources of endogeneity, measurement error is likely to be more important than unobservable of

[5]Model (27) is a dynamic heterogeneous panel and lagged variables are used as instruments. As is well known in the literature (e.g., Pesaran and Smith, 1995), pooling gives inconsistent estimates. Thus, we only report results for mean group estimates.

inflation expectation and simultaneity between inflation and unemployment. Ignoring these features lead to bias and misleading results.

6. 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, we introduce a new class of large heterogeneous panel data models whose parameters are not only heterogeneous but also vary stochastically over time. We propose time-varying mean group and pooled least square and 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 and pooled estimators in both the least square and IV contexts. We show the uniform consistency and derive the asymptotic distributions of the proposed estimators. We also propose a pointwise time-varying Hausman exogeneity test, which compares time-varying least square and IV estimators, possibly also allowing for changes in the endogeneity status of the regressors over time.

Next, we evaluate the finite sample properties of the estimators and size and power of the time-varying Hausman tests in an extensive Monte Carlo study. The results show that least square type estimates perform very well when regressors are exogenous, but have large biases and low coverage probabilities when regressors are endogenous. The IV type estimates have small finite-sample biases and satisfactory coverage probabilities when regressors are endogenous, especially if the bandwidth is chosen to be a value smaller than $T^{0.5}$. The size of the time-varying Hausman test statistic is also reasonable if bandwidth is smaller than $T^{0.5}$. The test has a good power and is not strongly affected by the bandwidth choice.

Finally, we provide an empirical application to illustrate in practice the use of time-varying mean group and pooled estimators. We estimate the panel version of time-varying hybrid Phillips curves for 19 Eurozone countries over the period 2000M1–2019M12. This simple but economically meaningful empirical application highlights the relevance of allowing for both parameter time variation and endogeneity in the panel framework.

References

Bai, Y., Kapetanios, G., Marcellino, M., 2022. Inference on the smoothness of parameter variation in time varying parameter models. Mimeo .

Baltagi, B.H., Feng, Q., Kao, C., 2016. Estimation of heterogeneous panels with structural breaks. Journal of Econometrics 191, 176–195.

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

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

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

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

Dendramis, Y., Giraitis, L., Kapetanios, G., 2020. Estimation of time-varying covariance matrices for large datasets. Econometric Theory, *forthcoming* .

Galí, J., Gambetti, L., 2019. Has the US wage Phillips curve flattened? A semi-structural exploration. Technical Report. National Bureau of Economic Research.

Galí, J., Gertler, M., 1999. Inflation dynamics: A structural econometric approach. Journal of Monetary Economics 2, 195–222.

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

Giraitis, L., Kapetanios, G., Marcellino, M., 2020. Time-varying instrumental variable estimation. Journal of Econometrics, *forthcoming* .

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

Giraitis, L., Kapetanios, G., Yates, T., 2018. Inference on multivariate heteroscedastic stochastic time varying coefficient models. Journal of Time Series Analysis 39, 129–149.

Graham, B.S., Powell, J.L., 2012. Identification and estimation of average partial effects in "irregular" correlated random coefficient panel data models. Econometrica 80, 2105–2152.

Hall, A.R., Han, S., Boldea, O., 2012. Inference regarding multiple structural changes in linear models with endogenous regressors. Journal of Econometrics 170, 281–302.

Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica: Journal of the Econometric Society 57, 357–384.

Hansen, P.R., Lunde, A., 2014. Estimating the persistence and the autocorrelation function of a time series that is measured with error. Econometric Theory 30, 60–93.

Harding, M., Lamarche, C., 2011. Least squares estimation of a panel data model with multifactor error structure and endogenous covariates. Economics Letters 111, 197–199.

Horváth, L., Trapani, L., 2016. Statistical inference in a random coefficient panel model. Journal of econometrics 193, 54–75.

Hsiao, C., Pesaran, M.H., 2008. Random coefficient models, in: Matyas, L., Sevestre, P. (Eds.), The econometrics of panel data, pp. 185–213. doi:10.1007/978-3-540-75892-1_6.

Hsiao, C., Pesaran, M.H., Tahmiscioglu, A.K., et al., 1998. Bayes estimation of short-run coefficients in dynamic panel data models.

Kapetanios, G., Marcellino, M., Venditti, F., 2019. Large time-varying parameter VARs: A nonparametric approach. Journal of Applied Econometrics 34, 1027–1049.

Li, M., 2021. A time-varying endogenous random coefficient model with an application to production functions. *Mimeo* .

Liu, F., Gao, J., Yang, Y., 2018. Nonparametric estimation in panel data models with heterogeneity and time-varyingness. Available at SSRN 3214046 .

Liu, F., Gao, J., Yang, Y., 2020. Time-varying panel data models with an additive factor structure. Available at SSRN 3729869 .

Nyblom, J., 1989. Testing for the constancy of parameters over time. Journal of the American Statistical Association 84, 223–230.

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

Pesaran, M.H., 2015. Time series and panel data econometrics. Oxford University Press. doi:10.1093/acprof:oso/9780198736912.001.0001.

Pesaran, M.H., Smith, R., 1995. Estimating long-run relationships from dynamic heterogeneous panels. Journal of Econometrics 68, 79–113. doi:10.1016/0304-4076(94)01644-F.

Robinson, P.M., 1991. Time-varying nonlinear regression, in: HacklAnders, P., Westlund, H. (Eds.), Economic Structural Change. Springer, pp. 179–190.

Stock, J.H., Watson, M.W., 1996. Evidence on structural instability in macroeconomic time series relations. Journal of Business & Economic Statistics 14, 11–30. doi:10.1080/07350015.1996.10524626.

Teräsvirta, T., 1994. Specification, estimation, and evaluation of smooth transition autoregressive models. Journal of the American Statistical Association 89, 208–218.

Ul Haque, N., Pesaran, M.H., Sharma, S., 1999. Neglected heterogeneity and dynamics in cross-country savings regressions. *IMF working paper*, International Monetary Fund, Washington, D.C. doi:10.5089/9781451855036.001.

Figures and Tables

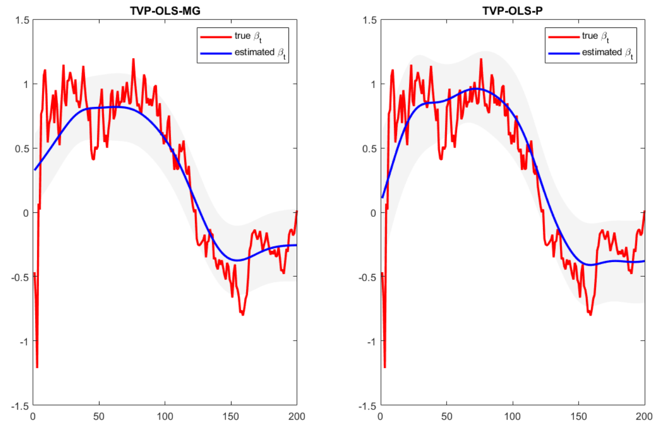

Figure 1: Realization of β_t, TVP-OLS-MG and TVP-OLS-P estimates with a two-sided normal kernel and $H = T^{0.5}$ for $(N, T) = (50, 200)$. The solid red lines show the true realization of β_t. The solid blue lines show the point estimates and the grey shaded areas show the 95% pointwise confidence intervals.

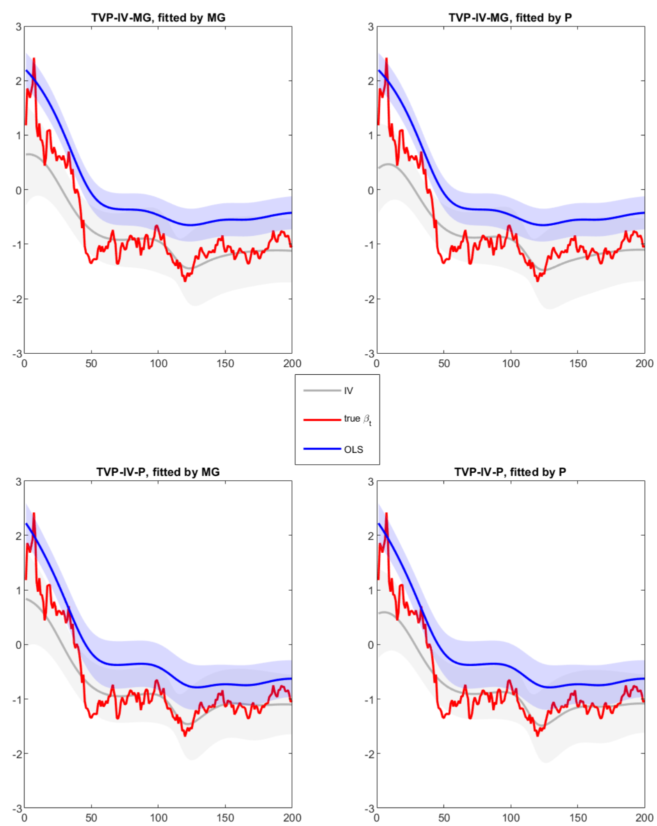

Figure 2: Realization of β_t, TVP-OLS-MG, TVP-IV-MG, TVP-OLS-P and TVP-IV-P estimates with a two-sided normal kernel and $H = T^{0.5}$, $H = L = T^{0.5}$ for $(N, T) = (50, 200)$. The solid red lines show the true realization of β_t. The solid blue lines show the OLS point estimates and the blue shaded areas show the 95% pointwise confidence intervals. The solid grey lines show the IV point estimates and the grey shaded areas show the 95% pointwise confidence intervals.

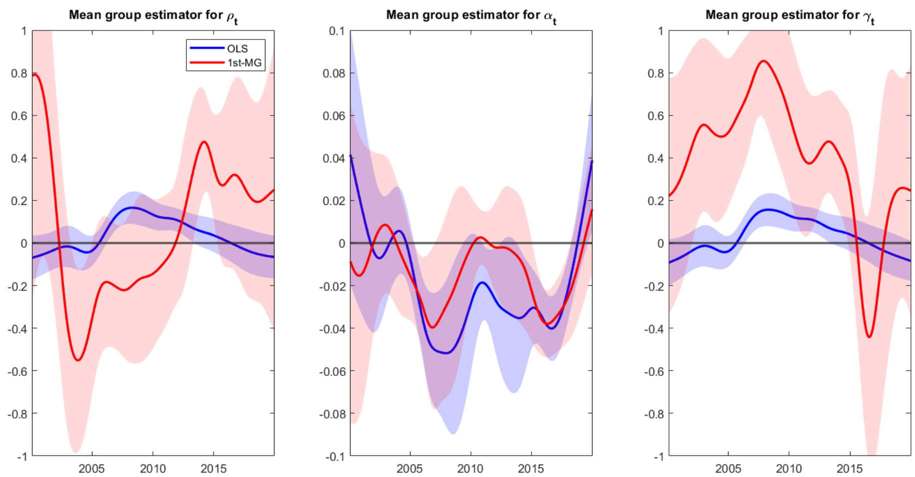

Figure 3: Empirical results for model (27). The solid blue lines show the point TVP-OLS estimates and the blue shaded areas show the 95% pointwise confidence intervals. The red solid lines show the point TVP-IV estimates and the red shaded areas show the 95% pointwise confidence intervals.

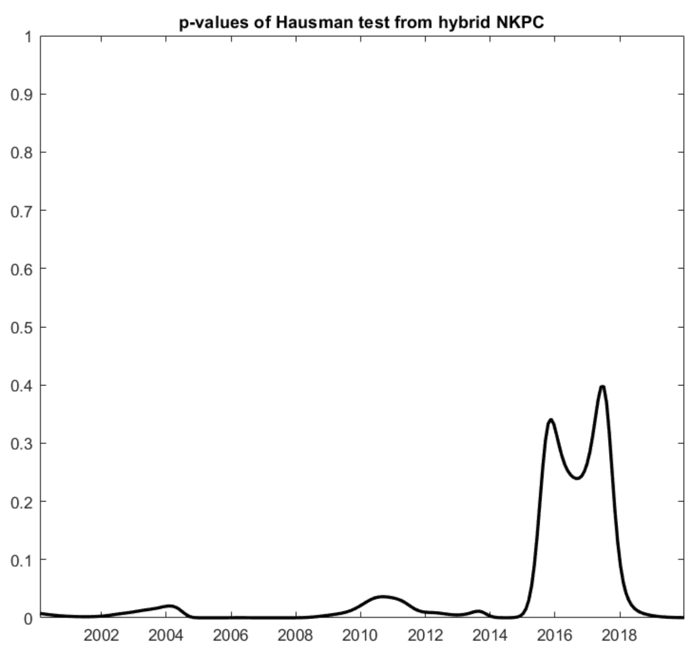

Figure 4: p-values of time-varying Hausman test for model (27).

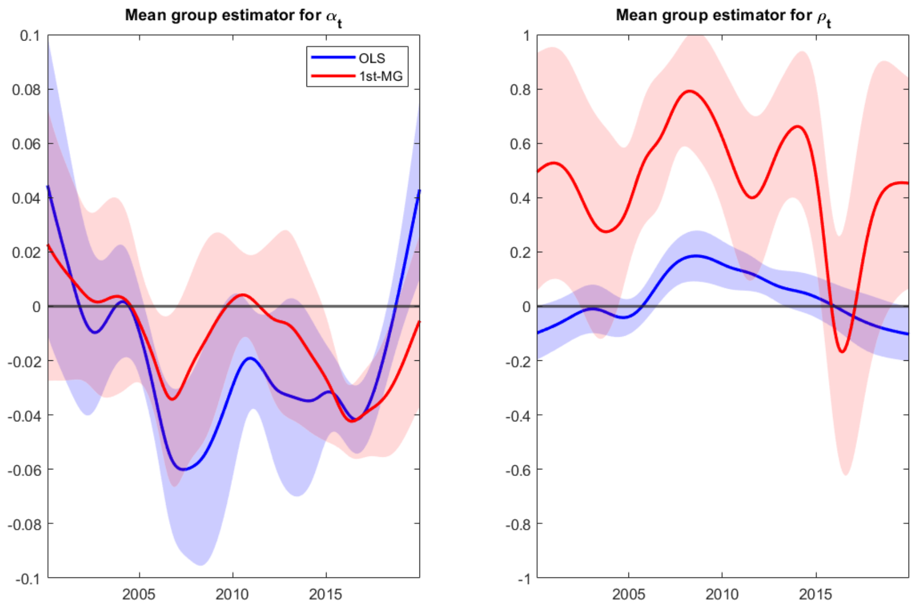

Figure 5: Empirical results for backward-looking Phillips curve. The solid blue lines show the point TVP-OLS estimates and the blue shaded areas show the 95% pointwise confidence intervals. The red solid lines show the point TVP-IV estimates and the red shaded areas show the 95% pointwise confidence intervals.

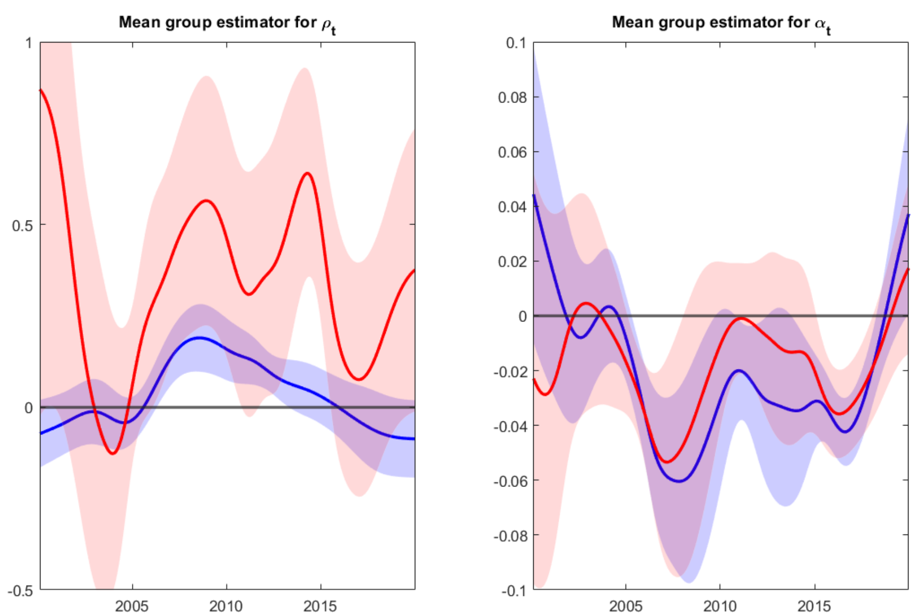

Figure 6: Empirical results for forward-looking Phillips curve. The solid blue lines show the point TVP-OLS estimates and the blue shaded areas show the 95% pointwise confidence intervals. The red solid lines show the point TVP-IV estimates and the red shaded areas show the 95% pointwise confidence intervals.

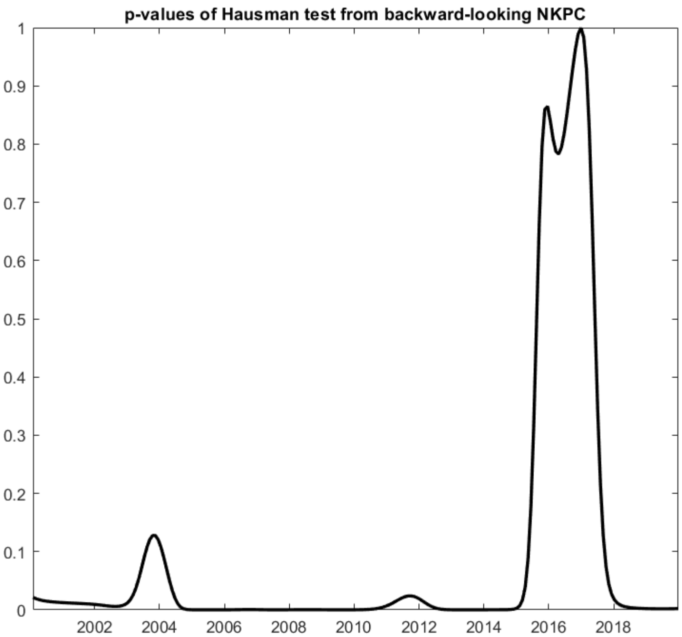

Figure 7: p-values of time-varying Hausman test for backward-looking Phillips curve.

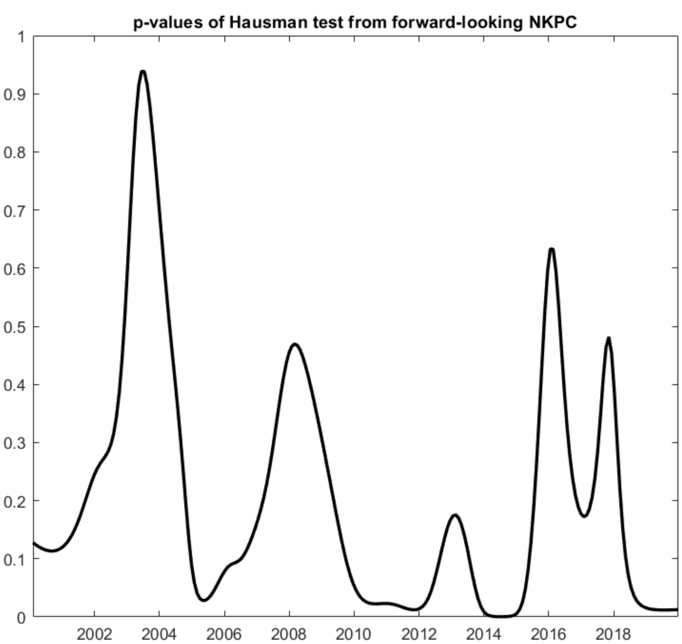

Figure 8: p-values of time-varying Hausman test for forward-looking Phillips curve.

Table 1: Small sample properties of TVP-OLS-MG and TVP-OLS-P estimators in the case of an exogenous regressor: Average MAD and coverage probability

(N, T)	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
MAD	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
TVP-OLS-MG																
50	0.157	0.143	0.131	0.126	0.194	0.165	0.150	0.141	0.217	0.193	0.172	0.158	0.287	0.271	0.248	0.227
100	0.132	0.119	0.104	0.091	0.179	0.146	0.129	0.112	0.204	0.176	0.154	0.132	0.279	0.260	0.238	0.212
200	0.117	0.106	0.086	0.071	0.172	0.136	0.116	0.097	0.199	0.169	0.145	0.120	0.276	0.258	0.232	0.204
500	0.109	0.095	0.073	0.055	0.167	0.129	0.108	0.086	0.196	0.164	0.139	0.112	0.280	0.253	0.232	0.201
TVP-OLS-P																
50	0.161	0.147	0.136	0.131	0.195	0.167	0.151	0.142	0.217	0.193	0.172	0.158	0.287	0.270	0.248	0.227
100	0.134	0.121	0.107	0.095	0.177	0.146	0.129	0.112	0.202	0.176	0.154	0.132	0.277	0.259	0.238	0.212
200	0.116	0.106	0.087	0.074	0.169	0.135	0.116	0.097	0.197	0.168	0.144	0.120	0.274	0.257	0.232	0.204
500	0.105	0.094	0.073	0.056	0.164	0.127	0.107	0.086	0.193	0.162	0.138	0.112	0.278	0.252	0.231	0.201
Coverage probability	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
TVP-OLS-MG																
50	0.875	0.889	0.916	0.917	0.765	0.828	0.865	0.880	0.702	0.756	0.807	0.840	0.530	0.565	0.612	0.664
100	0.813	0.847	0.887	0.923	0.652	0.748	0.795	0.851	0.578	0.652	0.708	0.784	0.410	0.439	0.485	0.548
200	0.720	0.753	0.838	0.898	0.522	0.625	0.695	0.771	0.447	0.515	0.585	0.672	0.305	0.320	0.363	0.428
500	0.552	0.596	0.717	0.838	0.359	0.451	0.526	0.622	0.298	0.359	0.416	0.505	0.190	0.220	0.243	0.284
TVP-OLS-P																
50	0.888	0.894	0.920	0.921	0..783	0.838	0.873	0.884	0.702	0.756	0.807	0.840	0.543	0.573	0.619	0.666
100	0.838	0.860	0.894	0.926	0.673	0.763	0.805	0.856	0.578	0.652	0.708	0.784	0.423	0.446	0.489	0.551
200	0.763	0.783	0.858	0.907	0.552	0.646	0.712	0.779	0.447	0.515	0.585	0.672	0.314	0.325	0.368	0.430
500	0.607	0.637	0.749	0.857	0.385	0.475	0.544	0.635	0.298	0.359	0.416	0.505	0.198	0.224	0.246	0.285

Note: See Section 4 for details on the computations of mean absolute deviation (MAD) and coverage probability.

Table 2: Small sample properties of TVP-OLS and TVP-IV types of estimators in the case of an endogenous regressor: Average MAD

	MAD															
(N, T)	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
IV, MG in first stage	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
TVP-IV-MG																
50	0.332	0.289	0.274	0.264	0.359	0.301	0.277	0.260	0.391	0.333	0.298	0.273	0.516	0.456	0.404	0.367
100	0.249	0.233	0.205	0.196	0.289	0.259	0.221	0.202	0.318	0.298	0.255	0.220	0.434	0.427	0.377	0.332
200	0.213	0.189	0.165	0.148	0.268	0.219	0.189	0.166	0.305	0.264	0.225	0.193	0.435	0.394	0.355	0.303
500	0.177	0.154	0.125	0.106	0.244	0.193	0.161	0.136	0.285	0.240	0.205	0.169	0.428	0.376	0.333	0.303
TVP-IV-P																
50	0.352	0.307	0.293	0.283	0.372	0.314	0.284	0.264	0.401	0.340	0.302	0.276	0.529	0.458	0.408	0.368
100	0.262	0.242	0.217	0.207	0.293	0.264	0.225	0.206	0.321	0.301	0.257	0.222	0.436	0.431	0.376	0.333
200	0.220	0.197	0.173	0.156	0.273	0.224	0.192	0.169	0.307	0.267	0.226	0.194	0.438	0.396	0.355	0.304
500	0.180	0.158	0.129	0.111	0.245	0.196	0.163	0.138	0.285	0.241	0.207	0.170	0.428	0.376	0.333	0.303
IV, P in first stage	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
TVP-IV-MG																
50	0.333	0.291	0.276	0.267	0.357	0.301	0.277	0.260	0.386	0.333	0.296	0.273	0.516	0.455	0.402	0.366
100	0.251	0.233	0.207	0.198	0.288	0.258	0.222	0.202	0.317	0.299	0.256	0.220	0.430	0.427	0.372	0.332
200	0.213	0.188	0.167	0.149	0.267	0.220	0.190	0.166	0.303	0.263	0.224	0.193	0.429	0.394	0.355	0.302
500	0.177	0.154	0.125	0.107	0.241	0.193	0.161	0.136	0.284	0.239	0.204	0.169	0.424	0.376	0.332	0.303
TVP-IV-P																
50	0.337	0.295	0.282	0.272	0.361	0.307	0.280	0.261	0.394	0.336	0.300	0.274	0.517	0.455	0.405	0.366
100	0.251	0.236	0.209	0.200	0.289	0.261	0.222	0.204	0.316	0.300	0.255	0.221	0.430	0.429	0.371	0.332
200	0.212	0.192	0.168	0.151	0.268	0.221	0.190	0.168	0.304	0.264	0.225	0.193	0.431	0.394	0.355	0.303
500	0.175	0.154	0.125	0.108	0.241	0.194	0.161	0.136	0.283	0.240	0.205	0.169	0.424	0.376	0.332	0.303
OLS	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
TVP-OLS-MG																
50	0.590	0.589	0.583	0.586	0.586	0.588	0.581	0.584	0.586	0.586	0.580	0.584	0.595	0.592	0.582	0.585
100	0.582	0.585	0.579	0.578	0.580	0.583	0.577	0.576	0.579	0.581	0.576	0.575	0.588	0.586	0.579	0.572
200	0.583	0.573	0.589	0.578	0.580	0.572	0.587	0.577	0.580	0.571	0.586	0.575	0.581	0.576	0.587	0.571
500	0.589	0.569	0.583	0.578	0.585	0.568	0.581	0.577	0.585	0.567	0.580	0.576	0.593	0.574	0.580	0.573
TVP-OLS-P																
50	0.532	0.528	0.518	0.523	0.537	0.532	0.520	0.524	0.542	0.537	0.524	0.526	0.562	0.556	0.540	0.542
100	0.516	0.517	0.507	0.503	0.525	0.520	0.511	0.506	0.529	0.526	0.515	0.510	0.549	0.545	0.534	0.524
200	0.512	0.500	0.512	0.498	0.521	0.506	0.516	0.502	0.526	0.512	0.521	0.505	0.542	0.532	0.539	0.519
500	0.514	0.495	0.507	0.498	0.523	0.500	0.512	0.502	0.528	0.506	0.516	0.506	0.548	0.528	0.533	0.520

Note: See Section 4 for details on the computations of mean absolute deviation (MAD) and coverage probability. For bandwidth parameters for TVP-IV estimators, we set $H = L$.

Table 3: Small sample properties of TVP-IV-MG and TVP-IV-P estimators in the case of an endogenous regressor: Coverage probability, $t = [T/2]$

	Coverage probability															
(N, T)	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
IV, MG in first stage	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
TVP-IV-MG																
50	0.939	0.953	0.956	0.961	0.892	0.931	0.940	0.950	0.857	0.896	0.915	0.933	0.758	0.779	0.808	0.841
100	0.928	0.941	0.955	0.963	0.845	0.900	0.922	0.940	0.793	0.842	0.874	0.910	0.647	0.676	0.706	0.752
200	0.893	0.910	0.936	0.957	0.758	0.839	0.871	0.912	0.687	0.748	0.792	0.855	0.536	0.562	0.570	0.644
500	0.807	0.835	0.898	0.946	0.616	0.708	0.775	0.846	0.538	0.592	0.657	0.749	0.380	0.406	0.429	0.483
TVP-IV-P																
50	0.940	0.948	0.954	0.959	0.896	0.929	0.940	0.949	0.866	0.895	0.916	0.934	0.763	0.781	0.809	0.842
100	0.930	0.944	0.952	0.961	0.850	0.903	0.922	0.941	0.800	0.842	0.877	0.911	0.657	0.678	0.709	0.754
200	0.899	0.913	0.939	0.957	0.768	0.844	0.876	0.914	0.698	0.756	0.796	0.858	0.545	0.570	0.574	0.645
500	0.823	0.849	0.908	0.948	0.634	0.721	0.786	0.850	0.557	0.602	0.664	0.753	0.391	0.412	0.433	0.485
IV, P in first stage	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
TVP-IV-MG																
50	0.933	0.951	0.954	0.959	0.889	0.930	0.939	0.949	0.854	0.895	0.914	0.933	0.756	0.779	0.808	0.841
100	0.923	0.940	0.952	0.961	0.842	0.898	0.920	0.940	0.789	0.840	0.872	0.910	0.648	0.674	0.708	0.752
200	0.889	0.907	0.934	0.955	0.758	0.838	0.870	0.911	0.686	0.746	0.792	0.855	0.537	0.559	0.569	0.644
500	0.804	0.834	0.895	0.944	0.615	0.708	0.774	0.846	0.536	0.593	0.657	0.748	0.377	0.404	0.429	0.482
TVP-IV-P																
50	0.943	0.952	0.956	0.962	0.899	0.932	0.941	0.951	0.867	0.898	0.916	0.935	0.762	0.782	0.809	0.843
100	0.933	0.946	0.955	0.963	0.853	0.904	0.921	0.942	0.802	0.844	0.877	0.913	0.660	0.677	0.711	0.754
200	0.902	0.918	0.942	0.959	0.771	0.846	0.878	0.915	0.699	0.756	0.798	0.858	0.546	0.567	0.573	0.645
500	0.832	0.853	0.912	0.952	0.636	0.725	0.787	0.852	0.555	0.604	0.667	0.754	0.390	0.411	0.433	0.484
OLS	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
TVP-OLS-MG																
50	0.220	0.203	0.194	0.200	0.216	0.200	0.187	0.189	0.226	0.208	0.195	0.195	0.236	0.225	0.228	0.218
100	0.112	0.088	0.094	0.082	0.130	0.097	0.101	0.085	0.140	0.114	0.116	0.096	0.162	0.150	0.149	0.136
200	0.052	0.045	0.026	0.027	0.077	0.061	0.039	0.035	0.091	0.077	0.055	0.047	0.119	0.107	0.096	0.091
500	0.020	0.015	0.012	0.007	0.040	0.027	0.019	0.012	0.051	0.039	0.029	0.020	0.067	0.067	0.058	0.049
TVP-OLS-P																
50	0.426	0.416	0.428	0.430	0.387	0.393	0.398	0.399	0.372	0.375	0.385	0.390	0.334	0.339	0.358	0.353
100	0.310	0.282	0.302	0.293	0.276	0.261	0.277	0.262	0.268	0.252	0.266	0.255	0.242	0.240	0.249	0.247
200	0.191	0.185	0.166	0.176	0.179	0.177	0.155	0.160	0.179	0.176	0.158	0.163	0.178	0.171	0.162	0.172
500	0.090	0.084	0.067	0.059	0.099	0.090	0.074	0.063	0.104	0.099	0.083	0.074	0.105	0.108	0.101	0.098

Note: See Section 4 for details on the computations of mean absolute deviation (MAD) and coverage probability. For bandwidth parameters for TVP-IV estimators, we set $H = L$.

Table 4: Size and power of time-varying Hausman H test for exogeneity

(N, T)	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
Mean group in first stage																
Size	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
Mean group																
50	0.037	0.036	0.031	0.034	0.062	0.049	0.030	0.033	0.090	0.081	0.043	0.039	0.157	0.155	0.142	0.111
100	0.034	0.031	0.027	0.020	0.088	0.059	0.034	0.027	0.143	0.090	0.073	0.042	0.213	0.239	0.194	0.173
200	0.046	0.034	0.028	0.028	0.156	0.074	0.049	0.038	0.222	0.166	0.109	0.079	0.296	0.343	0.334	0.259
500	0.092	0.063	0.041	0.028	0.256	0.156	0.094	0.064	0.383	0.268	0.190	0.123	0.493	0.470	0.478	0.384
Pooled																
50	0.027	0.034	0.025	0.021	0.055	0.045	0.033	0.020	0.086	0.065	0.048	0.030	0.153	0.169	0.158	0.109
100	0.046	0.034	0.025	0.019	0.088	0.058	0.040	0.026	0.144	0.099	0.083	0.035	0.219	0.270	0.227	0.187
200	0.054	0.040	0.026	0.014	0.155	0.081	0.051	0.034	0.233	0.169	0.121	0.071	0.320	0.358	0.346	0.298
500	0.094	0.059	0.035	0.019	0.279	0.176	0.105	0.065	0.412	0.294	0.225	0.130	0.513	0.485	0.501	0.427
Power	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
Mean group																
50	0.415	0.460	0.421	0.405	0.528	0.516	0.500	0.489	0.550	0.571	0.532	0.516	0.570	0.600	0.579	0.600
100	0.595	0.594	0.580	0.566	0.712	0.653	0.659	0.639	0.730	0.693	0.703	0.668	0.717	0.694	0.707	0.767
200	0.725	0.728	0.741	0.698	0.789	0.789	0.791	0.778	0.803	0.806	0.806	0.790	0.790	0.799	0.822	0.809
500	0.838	0.858	0.850	0.832	0.905	0.891	0.905	0.884	0.896	0.894	0.929	0.904	0.875	0.908	0.889	0.901
Pooled																
50	0.363	0.416	0.394	0.346	0.497	0.485	0.468	0.424	0.534	0.546	0.508	0.463	0.553	0.596	0.575	0.569
100	0.518	0.524	0.525	0.491	0.645	0.595	0.609	0.584	0.677	0.648	0.652	0.622	0.684	0.676	0.689	0.740
200	0.663	0.681	0.670	0.642	0.752	0.737	0.743	0.725	0.760	0.767	0.768	0.743	0.775	0.783	0.796	0.800
500	0.795	0.822	0.810	0.787	0.868	0.859	0.878	0.865	0.866	0.868	0.901	0.883	0.858	0.889	0.874	0.879
Pooled in first stage																
Size	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
Mean group																
50	0.040	0.039	0.035	0.034	0.057	0.050	0.032	0.034	0.095	0.077	0.040	0.040	0.158	0.158	0.141	0.111
100	0.028	0.037	0.026	0.025	0.087	0.059	0.038	0.028	0.138	0.090	0.075	0.040	0.212	0.237	0.197	0.175
200	0.051	0.033	0.036	0.027	0.144	0.077	0.053	0.036	0.221	0.162	0.106	0.076	0.297	0.339	0.337	0.261
500	0.094	0.062	0.045	0.030	0.257	0.148	0.091	0.062	0.372	0.265	0.187	0.116	0.491	0.465	0.481	0.380
Pooled																
50	0.023	0.028	0.019	0.023	0.053	0.044	0.028	0.019	0.088	0.066	0.043	0.029	0.157	0.168	0.160	0.107
100	0.039	0.025	0.021	0.016	0.087	0.055	0.040	0.024	0.135	0.107	0.085	0.036	0.220	0.265	0.222	0.187
200	0.038	0.036	0.022	0.015	0.144	0.078	0.049	0.034	0.232	0.160	0.120	0.071	0.316	0.355	0.345	0.295
500	0.079	0.053	0.031	0.018	0.270	0.169	0.101	0.063	0.409	0.290	0.221	0.128	0.517	0.486	0.500	0.427
Power	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
Mean group																
50	0.427	0.463	0.421	0.405	0.545	0.529	0.509	0.489	0.554	0.572	0.531	0.513	0.566	0.600	0.576	0.597
100	0.595	0.596	0.577	0.570	0.715	0.656	0.650	0.645	0.743	0.693	0.700	0.672	0.708	0.694	0.710	0.764
200	0.727	0.728	0.739	0.699	0.789	0.793	0.795	0.774	0.804	0.808	0.813	0.786	0.792	0.796	0.823	0.808
500	0.848	0.858	0.852	0.828	0.906	0.891	0.908	0.883	0.895	0.894	0.933	0.910	0.879	0.906	0.890	0.900
Pooled																
50	0.369	0.412	0.389	0.348	0.506	0.480	0.469	0.419	0.535	0.547	0.511	0.463	0.543	0.597	0.572	0.574
100	0.533	0.529	0.523	0.493	0.647	0.604	0.602	0.585	0.679	0.656	0.649	0.627	0.685	0.673	0.686	0.736
200	0.661	0.672	0.668	0.644	0.759	0.738	0.742	0.716	0.764	0.769	0.769	0.741	0.776	0.790	0.794	0.801
500	0.806	0.830	0.818	0.790	0.870	0.859	0.879	0.862	0.865	0.876	0.900	0.885	0.864	0.887	0.874	0.881

Note: $\hat{V}_H$ in the Hausman test statistic (25) is computed as in footnote 2. The bandwidth parameters are set to $H = L$.

Appendix A. Statement and proof of Lemmas

In the proof, 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 $H \to \infty$,

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

Moreover, we use H for the bandwidth for simplicity of notation and only introduce L when two bandwidth parameters interact.

Appendix A.1. Auxiliary results

Let (ξ_t) be an univariate strong-mixing sequence with mixing coefficient α_k^{ξ} satisfying

$$\alpha_k^{\xi} \leqslant c_{\xi}\phi^k, \quad k \geqslant 1 \tag{A.1}$$

for some $0 < \phi < 1$ and $c_{\xi} > 0$. Assume that

$$\max_{t=1,2,\cdots,T} E|\xi_t|^{\theta} \leqslant C < \infty \tag{A.2}$$

for some $0 < \theta < 4$. The condition above implies that ξ_t has a fat tail $\mathcal{H}(\theta)$,

$$\mathbb{P}(|\xi_t| \geqslant \omega) \leqslant c|\omega|^{-\theta}, \quad \omega > 0,$$

for some $c > 0$ which does not depend on t. We shall write $(v_t) \in \mathcal{E}(\alpha)$, $\alpha > 0$ to denote that

$$\mathbb{P}(|v_t| \geqslant \omega) \leqslant c_0 \exp\left(-c_1|\omega|^{\alpha}\right), \quad \omega > 0,$$

for some $c_0, c_1 > 0$ do not depend on t.

Lemma A1. *Let* (ξ_t) *be an univariate strong-mixing sequence satisfying* (A.2) *and mixing coefficient satisfying* (A.1). *Consider the sums*

$$S_{T,t} := \frac{1}{\sqrt{K_t}} \sum_{j=1}^{T} b_{jt}(\xi_j - E\xi_j)$$

$$\Delta_{T,t} := \frac{1}{K_t} \sum_{j=1}^{T} b_{jt}(\beta_j - \beta_t)\xi_j,$$

where (β_j) *is either*

$$\left|\beta_j - \beta_t\right| \leqslant \Big(\frac{|j-t|}{T}\Big)^{\gamma} \tag{A.3}$$

for $0 < \gamma < 1$*,* $1 \leqslant t, j \leqslant T$ *or*

$$\left|\beta_j - \beta_t\right| \leqslant \Big(\frac{|j-t|}{T}\Big)^{\gamma} \nu_{jt} \tag{A.4}$$

for $0 < \gamma < 1$*,* $1 \leqslant t, j \leqslant T$ *and* $\beta_t \in \mathcal{E}(\alpha)$*,* $\nu_{jt} \in \mathcal{E}(\alpha)$ *for some* $\alpha > 0$*. Assume that the bandwidth parameter* H *satisfies*

$$cT^{1/(\theta/4-1)} \leqslant H \leqslant T$$

for some $c > 0$ *and* $\delta > 0$*. Then, for any* $\epsilon > 0$

$$\max_{t=1,2,\cdots,T} \left|S_{T,t}\right| = O_p\Big(\log^{1/2} T + (TH)^{1/\theta} H^{\epsilon - 1/2}\Big) \tag{A.5}$$

and

$$\max_{t=1,2,\cdots,T} \left|\Delta_{T,t}\right| = \begin{cases} O_p\Big((H/T)^{\gamma}\Big) & \text{if (A.3) holds} \\ O_p\Big((H/T)^{\gamma} (\log T)^{1/\alpha}\Big) & \text{if (A.4) holds.} \end{cases} \tag{A.6}$$

Moreover, for $g_t \in \mathcal{E}(\alpha)$*, we have*

$$\max_{t=1,2,\cdots,T} |g_t| = O_p(\log^{1/\alpha} T). \tag{A.7}$$

Proof. See Dendramis et al. (2020), for details of the proof. (A.5) is shown in (51), (A.7) is shown in (C.3) and (A.6) is shown in (69) and (70) of that paper. We need to replace $(\frac{H}{T})$ and $(\frac{H}{T})^{1/2}$ with $(\frac{H}{T})^{\gamma}$ because we have a slightly different smoothness condition. □

Appendix A.2. Useful lemmas

Lemma A2. *Consider model* (1) *and the following sums*

$$S_{xx,t} = \frac{1}{K_t} \sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij}$$

$$\Delta_{x,t} = \frac{1}{K_t} \sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij} (\beta_{0,j} - \beta_{0,t})$$

$$S_{xu,t} = \frac{1}{K_t} \sum_{j=1}^{T} b_{jt} x_{ij} u_{ij}$$

$$S_{xe,t} = \frac{1}{K_t} \sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij} e_{ij}$$

Then under Assumptions 2.1-2.4,

$$\left\|S_{xx,t}\right\|_{sp} = O_p(1), \quad \max_{t=1,2,\cdots,T}\left\|S_{xx,t}\right\|_{sp} = O_p(1); \tag{A.8}$$

$$\left\|\Delta_{x,t}\right\| = O_p\Big((H/T)^{\gamma_1}\Big), \quad \max_{t=1,2,\cdots,T}\left\|\Delta_{x,t}\right\| = O_p\Big((H/T)^{\gamma_1}\Big); \tag{A.9}$$

$$\left\|S_{xu,t}\right\| = O_p\Big(\frac{1}{\sqrt{H}}\Big), \quad \max_{t=1,2,\cdots,T}\left\|S_{xu,t}\right\| = O_p\Big(H^{-\frac{1}{2}}\log^{1/2}T\Big); \tag{A.10}$$

$$\left\|S_{xe,t}\right\| = O_p(1), \quad \max_{t=1,2,\cdots,T}\left\|S_{xe,t}\right\| = O_p\Big(\log^{1/\alpha}T(1+(H/T)^{\gamma_2})\Big). \tag{A.11}$$

Proof. *Proof of* (A.8). Notice that, by Markov's inequality, Assumption 2.4 implies that $\max_{t=1,2,\cdots,T}\left\|x_{it}x_{it}'\right\|_{sp} = O_p(1)$, $\forall i$. Therefore,

$$\begin{aligned}\left\|S_{xx,t}\right\|_{sp} &\leqslant \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\left\|x_{ij}x_{ij}'\right\|_{sp} \\ &\leqslant \max_{t=1,2,\cdots,T}\left\|x_{it}x_{it}'\right\|_{sp}\cdot\frac{1}{K_t}\sum_{j=1}^{T} b_{jt} = O_p(1).\end{aligned}$$

For the second part, write

$$\begin{aligned}S_{xx,t} &= \frac{1}{H}\sum_{j=1}^{T} b_{jt}E(x_{ij}x_{ij}') + \frac{1}{H}\sum_{j=1}^{T} b_{jt}(x_{ij}x_{ij}' - E(x_{ij}x_{ij}')) \\ &= S_{xx,it}^{(1)} + S_{xx,it}^{(2)} = S_{xx,it}^{(1)}(1+\tilde{\Delta}_t),\end{aligned} \tag{A.12}$$

where $\tilde{\Delta}_t = (S_{xx,t}^{(1)})^{-1}(S_{xx,t} - S_{xx,t}^{(1)})$. For $S_{xx,t}^{(1)}$, by Assumption 2.4,

$$\max_{t=1,2,\cdots,T}\left\|S_{xx,t}^{(1)}\right\|_{sp} \leqslant \max_{t=1,2,\cdots,T}\left\|\Sigma_{xx,t}\right\|_{sp}\cdot\frac{1}{K_t}\sum_{j=1}^{T} b_{jt} = O_p(1).$$

Next, consider

$$\max_{t=1,2,\cdots,T}\left\|S_{xx,t} - S_{xx,t}^{(1)}\right\|_{sp} = \max_{t=1,2,\cdots,T}\left\|\frac{1}{H}\sum_{j=1}^{T} b_{jt}(x_{ij}x_{ij}' - E(x_{ij}x_{ij}'))\right\|_{sp}.$$

By Assumption 2.1(i)–(ii), the (ℓ,k)-th component $\omega_{\ell k,ij} = x_{\ell,ij}x_{k,ij} - Ex_{\ell,ij}x_{k,ij}$ is strong-mixing. Moreover, let $\theta' = \theta/2$, $E\left|x_{\ell,ij}x_{k,ij}\right|^{\theta'} \leqslant C < \infty$. Together with the assumption of bandwidth in (10), we apply (A.5) to obtain

$$\max_{t=1,2,\cdots,T}\left\|\frac{1}{H}\sum_{j=1}^{T} b_{jt}(x_{ij}x_{ij}' - E(x_{ij}x_{ij}'))\right\|_{sp} = O_p(H^{-\frac{1}{2}}\log^{\frac{1}{2}}T + (TH)^{\frac{1}{\theta}}H^{\varepsilon-1})$$

for any $\varepsilon > 0$. We can always make ε small enough to have $(TH)^{\frac{1}{\theta}} H^{\varepsilon-1} \leqslant H^{-\frac{1}{2}}$. This then implies that

$$\max_{t=1,2,\cdots,T} \left\| \frac{1}{H} \sum_{j=1}^{T} b_{jt}(x_{ij}x'_{ij} - E(x_{ij}x'_{ij})) \right\|_{sp} = O_p(H^{-\frac{1}{2}} \log^{1/2} T) = o_p(1).$$

From (A.12), we have

$$\begin{aligned} \max_{t=1,2,\cdots,T} \left\| S_{xx,t} \right\|_{sp} &\leqslant \max_{t=1,2,\cdots,T} \left\| S^{(1)}_{xx,t} \right\|_{sp} \max_{t=1,2,\cdots,T} \left\| I + \tilde{\Delta}_t \right\|_{sp} \\ &\leqslant \max_{t=1,2,\cdots,T} \left\| S^{(1)}_{xx,t} \right\|_{sp} (1 + \max_{t=1,2,\cdots,T} \left\| \tilde{\Delta}_t \right\|_{sp}) = O_p(1). \end{aligned}$$

Proof of (A.9). Notice that, (4) implies that $\left\| \beta_{0,j} - \beta_{0,t} \right\| = O\Big((H/T)^{\gamma_1}\Big)$. Then,

$$\left\| \Delta_{x,t} \right\| \leqslant \frac{1}{K_t} \sum_{j=1}^{T} b_{jt} \left\| x_{ij}x'_{ij}(\beta_{0,j} - \beta_{0,t}) \right\| \leqslant \frac{1}{K_t} \sum_{j=1}^{T} b_{jt} \left\| x_{ij}x'_{ij} \right\|_{sp} \left\| \beta_{0,j} - \beta_{0,t} \right\| = O_p\Big((H/T)^{\gamma_1}\Big).$$

For the second part, elements in $\Delta_{x,it}$ involve a finite number of linear combinations of sums

$$s_t := \frac{1}{K_t} \sum_{j=1}^{T} b_{jt} \omega_{\ell k,it}(\beta_{m,0,j} - \beta_{m,0,t}),$$

where $\omega_{\ell k,ij} = x_{\ell,ij}x_{k,ij} - Ex_{\ell,ij}x_{k,ij}$ is strong-mixing and $E\left|x_{\ell,ij}x_{k,ij}\right|^{\theta'} \leqslant C < \infty$. By Assumption 2.3, we can apply (A.6) to obtain $\max_{t=1,2,\cdots,T} |s_t| = O_p\Big((H/T)^{\gamma_1}\Big)$, which implies that

$$\max_{t=1,2,\cdots,T} \left\| \Delta_{x,t} \right\| = O_p\Big((H/T)^{\gamma_1}\Big).$$

Proof of (A.10). By the arguments used in Lemma A.4 in Giraitis et al. (2014) or Lemma 6.2 in Giraitis et al. (2018), we could show that

$$\frac{1}{K_{2,t}^{1/2}} \sum_{j=1}^{T} b_{jt} x_{ij} u_{ij} \xrightarrow{d} \mathcal{N},$$

with additional mixing or martingale difference type of assumptions on the process $(x_{it}u_{it})$, where $\mathcal{N}$ denotes Normal distribution. This implies that

$$\frac{K_t}{K_{2,t}^{1/2}} S_{xu,t} = O_p(1) \Longrightarrow S_{xu,t} = O_p\Big(\frac{1}{\sqrt{H}}\Big),$$

since $\frac{K_t}{K_{2,t}^{1/2}} = O(\sqrt{H})$. For the second part, elements in $S_{xu,t}$ involve finite a number of linear combinations of sums

$$\bar{s}_t := \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{\ell,ij}u_{ij}.$$

Because $Ex_{\ell,ij}u_{ij} = 0$, $(x_{\ell,ij}u_{ij})$ is strong-mixing with $E\left|x_{\ell,ij}u_{ij}\right|^{\theta'} \leqslant C$. We can then apply (A.5) to obtain $\max_{t=1,2,\cdots,T}|\bar{s}_t| = O_p(H^{-\frac{1}{2}}\log^{\frac{1}{2}} T + (TH)^{\frac{1}{\theta}}H^{\varepsilon-1}) = O_p(H^{-\frac{1}{2}}\log^{\frac{1}{2}} T)$, by the same reasoning used in the second part of (A.8), which implies that

$$\max_{t=1,2,\cdots,T}\left\|S_{xu,t}\right\| = O_p(H^{-\frac{1}{2}}\log^{\frac{1}{2}} T).$$

Proof of (A.11). Write

$$\begin{aligned}
S_{xe,t} &= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'e_{ij}\\
&= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'(e_{ij}-e_{it}) + \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'e_{it}\\
&= S_{xe_1,t} + S_{xe_2,t}.
\end{aligned}\tag{A.13}$$

From (5), by similar arguments as in the proof of (A.9), we obtain

$$\left\|S_{xe_1,t}\right\| = O_p\Big((\frac{H}{T})^{\gamma_2}\Big) = o_p(1)$$

and

$$\max_{t=1,2,\cdots,T}\left\|S_{xe_1,t}\right\| = O_p\Big((H/T)^{\gamma_2}(\log T)^{1/\alpha}\Big).$$

For $S_{xe_2,t}$, notice that

$$\left\|S_{xe_2,t}\right\| \leqslant \|e_{it}\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\left\|x_{ij}x_{ij}'\right\| \leqslant \|e_{it}\| \max_{t=1,2,\cdots,T}\left\|x_{it}x_{it}'\right\|_{sp}\frac{1}{K_t}\sum_{j=1}^{T} b_{jt} = O_p(1),$$

since $\|e_{it}\| = O_p(1)$ by Assumption 2.3. Moreover, elements in $S_{xe_2,it}$ involve linear combinations of elements in e_{it} multiplying the sums $\frac{1}{H}\sum_{j=1}^{T} b_{jt}x_{\ell,ij}x_{k,ij}$, by (A.7), $\max_{t=1,2,\cdots,T}|e_{it}| = O_p\Big(\log^{1/\alpha} T\Big)$ and by (A.8), $\max_{t=1,2,\cdots,T}\left|\frac{1}{H}\sum_{j=1}^{T} b_{jt}x_{\ell,ij}x_{k,ij}\right| = O_p(1)$. Then, we obtain

$$\max_{t=1,2,\cdots,T}\left\|S_{xe_2,t}\right\| = O_p\Big(\log^{1/\alpha} T\Big).$$

So, by continuing from (A.13), we have

$$\left\|S_{xe,t}\right\| \leqslant \left\|S_{xe_1,t}\right\| + \left\|S_{xe_2,t}\right\| = O_p(1)$$
$$\max_{t=1,2,\cdots,T}\left\|S_{xe,t}\right\| \leqslant \max_{t=1,2,\cdots,T}\left\|S_{xe_1,t}\right\| + \max_{t=1,2,\cdots,T}\left\|S_{xe_2,t}\right\| = O_p\Big(\log^{1/\alpha} T(1+(H/T)^{\gamma_2})\Big).$$

□

Lemma A3. *Define*

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

Consider model (13) *and the following sums*

$$S_{z\psi,t} = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j$$
$$\Delta_{z,t} = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij}(\beta_{0,j}-\beta_{0,t})$$
$$S_{zu,t} = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} u_{ij}$$
$$S_{ze,t} = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij} e_{ij}$$

Then under Assumptions 2.1–2.3 (except Assumption 2.2(iii)), Assumptions 3.1-3.4, and $\log^{1/\alpha} T = o(N^{-1/2})$*,*

$$\left\|S_{z\psi,t}\right\|_{sp} = O_p(1), \quad \max_{t=1,2,\cdots,T}\left\|S_{z\psi,t}\right\|_{sp} = O_p(1); \tag{A.14}$$
$$\left\|\Delta_{z,t}\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big), \quad \max_{t=1,2,\cdots,T}\left\|\Delta_{z,t}\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big); \tag{A.15}$$
$$\left\|S_{zu,t}\right\| = O_p\Big(\frac{1}{\sqrt{H}}\Big), \quad \max_{t=1,2,\cdots,T}\left\|S_{zu,t}\right\| = O_p\Big(\bar{r}_{T,H,\gamma_3} + r_{N,T,L,\gamma_3,\alpha}\Big); \tag{A.16}$$
$$\left\|S_{ze,t}\right\| = O_p(1), \quad \max_{t=1,2,\cdots,T}\left\|S_{ze,t}\right\| = O_p\bigg(\Big(\log T\Big)^{1/\alpha}\Big(1+(\frac{H}{T})^{\min(\gamma_3,\gamma_4)}\Big)\bigg); \tag{A.17}$$

Proof. Proof of (A.14). Write

$$
\begin{aligned}
&\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j \\
=&\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}'_j-\Psi'_{0,j}+\Psi'_{0,j}-\Psi'_{0,t}+\Psi'_{0,t}) z_{ij} z'_{ij}(\hat{\Psi}_j-\Psi_{0,j}+\Psi_{0,j}-\Psi_{0,t}+\Psi_{0,t}) \\
=&\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}'_j-\Psi'_{0,j}) z_{ij} z'_{ij}(\hat{\Psi}_j-\Psi_{0,j})+\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi'_{0,j}-\Psi'_{0,t}) z_{ij} z'_{ij}(\hat{\Psi}_j-\Psi_{0,j}) \\
&\quad+\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}'_j-\Psi'_{0,j}) z_{ij} z'_{ij}(\Psi_{0,j}-\Psi_{0,t})+\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi'_{0,j}-\Psi'_{0,t}) z_{ij} z'_{ij}(\Psi_{0,j}-\Psi_{0,t}) \\
&\qquad+\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi'_{0,j}-\Psi'_{0,t}) z_{ij} z'_{ij}\Big)\Psi_{0,t}+\Psi'_{0,t}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt} z_{ij} z'_{ij}(\Psi_{0,j}-\Psi_{0,t})\Big)+\Psi'_{0,t}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt} z_{ij} z'_{ij}\Big)\Psi_{0,t}
\end{aligned}
$$

By Lemma 1, the condition $\log^{1/\alpha} T = o(N^{-1/2})$ guarantees that $\max_{t=1,2,\cdots,T}\left\|\hat{\Psi}'_t-\Psi'_{0,t}\right\|_{sp} = o_p(1)$. Consider

$$
\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}'_j-\Psi'_{0,j}) z_{ij} z'_{ij}(\hat{\Psi}_j-\Psi_{0,j})\right\|_{sp} \leqslant \left(\max_{t=1,2,\cdots,T}\left\|\hat{\Psi}'_t-\Psi'_{0,t}\right\|_{sp}\right)^2 \frac{1}{K_t}\sum_{j=1}^{T} b_{jt} \max_{t=1,2,\cdots,T}\left\|z_{it} z'_{it}\right\|_{sp} = o_p(1).
$$

Similarly, terms above involving $\hat{\Psi}'_j-\Psi'_{0,j}$ are also $o_P(1)$. Next, consider

$$
\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi'_{0,j}-\Psi'_{0,t}) z_{ij} z'_{ij}(\Psi_{0,j}-\Psi_{0,t})\right\|_{sp} \leqslant \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\left\|\Psi'_{0,j}-\Psi'_{0,t}\right\| \max_{t=1,2,\cdots,T}\left\|z_{it} z'_{it}\right\|_{sp} = O_p\Big(\big(\frac{H}{T}\big)^{\gamma_3}\Big) = o_p(1).
$$

Similarly, terms above involving $\Psi'_{0,j}-\Psi'_{0,t}$ are also $o_p(1)$. Then,

$$
\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j = \Psi'_{0,t}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt} z_{ij} z'_{ij}\Big)\Psi_{0,t}+o_p(1) = \Psi'_{0,t}\Sigma_{zz,i,t}\Psi_{0,t}+o_p(1)
$$

which implies that

$$
\left\|S_{z\psi,t}\right\|_{sp} = O_p(1),
$$

since by Assumptions 3.2 and 3.4, $\Psi'_{0,t}\Sigma_{zz,i,t}\Psi_{0,t}$ is positive definite and $\max_{t=1,2,\cdots,T}\left\|\Sigma_{\Psi zz\Psi,it}\right\|_{sp} < \infty$, $\forall i$. Then, it follows from the expansion above and similar derivation of $S^{(1)}_{xx,t}$ in (A.8) that

$$
\max_{t=1,2,\cdots,T}\left\|S_{z\psi,t}\right\|_{sp} = \max_{t=1,2,\cdots,T}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\Psi'_{0,j} z_{ij} z'_{ij}\Psi_{0,j}\right\|_{sp} + o_p(1) = O_p(1).
$$

Proof of (A.15). Write

$$\begin{aligned}\left\|\Delta_{z,t}\right\| &\leqslant \left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}'_j - \Psi'_{0,j})z_{ij}x'_{ij}(\beta_{0,j}-\beta_{0,t})\right\| + \left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\Psi'_j z_{ij}x'_{ij}(\beta_{0,j}-\beta_{0,t})\right\| \\ &\leqslant \max_{t=1,2,\cdots,T}\left\|\hat{\Psi}'_t - \Psi'_{0,t}\right\|_{sp} \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\left\|z_{ij}x'_{ij}\right\|_{sp}\left\|\beta_{0,j}-\beta_{0,t}\right\| + \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\left\|\Psi'_{0,j}\right\|_{sp}\left\|z_{ij}x'_{ij}\right\|_{sp}\left\|\beta_{0,j}-\beta_{0,t}\right\| \\ &\leqslant \max_{t=1,2,\cdots,T}\left\|\hat{\Psi}'_t - \Psi'_{0,t}\right\|_{sp} \max_{t=1,2,\cdots,T}\left\|z_{it}x'_{it}\right\|_{sp} \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\left\|\beta_{0,j}-\beta_{0,t}\right\| + \max_{t=1,2,\cdots,T}\left\|z_{it}x'_{it}\right\|_{sp} \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\left\|\Psi'_{0,j}\right\|_{sp}\left\|\beta_{0,j}-\beta_{0,t}\right\|\end{aligned}$$

By (4), $\left\|\beta_j - \beta_t\right\| = O\big((\frac{H}{T})^{\gamma_1}\big)$. Together with Lemma 1, Assumptions 3.3 and 3.4, we have

$$\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}x'_{ij}(\beta_{0,j}-\beta_{0,t})\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big).$$

Second part follows similarly from (57) in Giraitis et al. (2020):

$$\max_{t=1,2,\cdots,T}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}x'_{ij}(\beta_j-\beta_t)\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big).$$

Proof of (A.16). (26) in Giraitis et al. (2020) implies that

$$\frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} b_{jt}z_{ij}u_{ij} = O_p(1) \Longrightarrow \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}z_{ij}u_{ij} = O_p\Big(\frac{1}{\sqrt{H}}\Big),$$

since $\frac{K_t}{K_{2,t}^{1/2}} = O(\sqrt{H})$. Write

$$\begin{aligned}S_{zu,t} &= \frac{1}{K_t}\sum_{j=1}^{T} b_{j,t}\hat{\Psi}'_j z_{ij}u_{ij} \\ &= \frac{1}{K_t}\sum_{j=1}^{T} b_{j,t}(\hat{\Psi}'_j - \Psi'_{0,j} + \Psi'_{0,j} - \Psi'_{0,t} + \Psi'_{0,t})z_{ij}u_{ij} \\ &= \frac{1}{K_t}\sum_{j=1}^{T} b_{j,t}(\hat{\Psi}'_j - \Psi'_{0,j})z_{ij}u_{ij} + \frac{1}{K_t}\sum_{j=1}^{T} b_{j,t}(\Psi'_{0,j} - \Psi'_{0,t})z_{ij}u_{ij} + \Psi'_{0,t}\frac{1}{K_t}\sum_{j=1}^{T} b_{j,t}z_{ij}u_{ij} \\ &= S_{zu,t,1} + S_{zu,t,2} + S_{zu,t,3}.\end{aligned}$$

Since $S_{zu,t,1}$ involves $\hat{\Psi}'_j - \Psi'_{0,j}$, $\hat{\Psi}_t$ is uniformly consistent, $S_{zu,t,1}$ is asymptotically negligible. For $S_{zu,t,2}$, we have

$$\left\|S_{zu,t,2}\right\| \leqslant \frac{1}{K_t}\sum_{j=1}^{T} b_{j,t}\left\|\Psi_{0,j}' - \Psi_{0,t}'\right\|_{sp}\left\|z_{ij}u_{ij}\right\| = O_p\Big((\frac{H}{T})^{\gamma_3}\Big) = o_p(1).$$

For $S_{zu,t,3}$, we have

$$\left\|S_{zu,t,3}\right\| \leqslant \left\|\Psi_{0,t}'\right\|_{sp}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{j,t}z_{ij}u_{ij}\right\| = O_p\Big(\frac{1}{\sqrt{H}}\Big).$$

It follows immediately from triangular inequality that

$$\left\|S_{zu,t}\right\| \leqslant \left\|S_{zu,t,1}\right\| + \left\|S_{zu,t,2}\right\| + \left\|S_{zu,t,3}\right\| = O_p\Big(\frac{1}{\sqrt{H}}\Big).$$

Second part follows similarly from (58) in Giraitis et al. (2020):

$$\max_{t=1,2,\cdots,T}\left\|S_{zu,t}\right\| = O_p\Big(\bar{r}_{T,H,\gamma_3} + r_{N,T,L,\gamma_3,\alpha}\Big),$$

where L is the bandwidth parameter used to obtain $\hat{\Psi}_t$ and H is used to obtain $\hat{\beta}_t$.

Proof of (A.17). Write

$$\begin{aligned}
S_{ze,t} &= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}_j' - \Psi_{0,j}' + \Psi_{0,j}' - \Psi_{0,t}' + \Psi_{0,t}')z_{ij}x_{ij}'(e_{ij} - e_{it} + e_{it}) \\
&= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}_j' - \Psi_{0,j}')z_{ij}x_{ij}'(e_{ij} - e_{it}) + \Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}_j' - \Psi_{0,j}')z_{ij}x_{ij}'\Big)e_{it} \\
&\quad + \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi_{0,j}' - \Psi_{0,t}')z_{ij}x_{ij}'(e_{ij} - e_{it}) + \Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi_{0,j}' - \Psi_{0,t}')z_{ij}x_{ij}'\Big)e_{it} \\
&\quad + \Psi_{0,t}'\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}z_{ij}x_{ij}'(e_{ij} - e_{it})\Big) + \Psi_{0,t}'\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}z_{ij}x_{ij}'\Big)e_{it}.
\end{aligned}$$

Consider the first term:

$$\begin{aligned}
\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}_j' - \Psi_{0,j}')z_{ij}x_{ij}'(e_{ij} - e_{it})\right\| &\leqslant \max_{t=1,2,\cdots,T}\left\|\hat{\Psi}_t' - \Psi_{0,t}'\right\|_{sp}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}z_{ij}x_{ij}'(e_{ij} - e_{it})\right\| \\
&\leqslant \max_{t=1,2,\cdots,T}\left\|\hat{\Psi}_t' - \Psi_{0,t}'\right\|_{sp}\max_{t=1,2,\cdots,T}\left\|z_{it}x_{it}'\right\|_{sp}\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\left\|e_{ij} - e_{it}\right\| \\
&= o_p(1).
\end{aligned}$$

Thus, terms involving $\hat{\Psi}_j' - \Psi_{0,j}'$ are also $o_p(1)$. Consider the third term:

$$\begin{aligned}\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi_{0,j}' - \Psi_{0,t}')z_{ij}x_{ij}'(e_{ij} - e_{it})\Big\| &\leqslant \max_{t=1,2,\cdots,T}\big\|z_{it}x_{it}'\big\|_{sp}\ \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\big\|\Psi_{0,j}' - \Psi_{0,t}'\big\|_{sp}\big\|e_{ij} - e_{it}\big\| \\ &= O_p\Big((\frac{H}{T})^{\gamma_3+\gamma_4}\Big) = o_p(1).\end{aligned}$$

Thus, terms involving $\Psi_{0,j}' - \Psi_{0,t}'$ are also $o_p(1)$. The dominating term is the last one. Then, we have

$$\begin{aligned}\big\|S_{ze,t}\big\| &= \Big\|\Psi_{0,t}'\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{j,t}z_{ij}x_{ij}'\Big)e_{it}\Big\| + o_p(1) \\ &\leqslant \big\|\Psi_{0,t}'\big\|_{sp}\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{j,t}z_{ij}x_{ij}'\Big\|_{sp}\ \|e_{it}\| + o_p(1) \\ &= O_p(1).\end{aligned}$$

To derive the uniform rate, it follows the similar reasoning as above, except the fact that we have to use (A.6) and (A.7). Since by Assumption imposed in Lemma 1, $\hat{\Psi}_t$ is uniformly consistent, we have

$$\begin{aligned}\max_{t=1,2,\cdots,T}\big\|S_{ze,t}\big\| \leqslant{}& \max_{t=1,2,\cdots,T}\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi_{0,j}' - \Psi_{0,t}')z_{ij}x_{ij}'(e_{ij} - e_{it})\Big\| + \max_{t=1,2,\cdots,T}\Big\|\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi_{0,j}' - \Psi_{0,t}')z_{ij}x_{ij}'\Big)e_{it}\Big\| \\ &+ \max_{t=1,2,\cdots,T}\Big\|\Psi_{0,t}'\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}z_{ij}x_{ij}'(e_{ij} - e_{it})\Big)\Big\| + \max_{t=1,2,\cdots,T}\Big\|\Psi_{0,t}'\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}z_{ij}x_{ij}'\Big)e_{it}\Big\| \\ ={}& O_p\bigg(\Big(\log T\Big)^{1/\alpha}\Big(1 + (\frac{H}{T})^{\min(\gamma_3,\gamma_4)}\Big)\bigg).\end{aligned}$$

$\square$

Appendix B. Mathematical proofs

Appendix B.1. Proof of Theorem 1

Under (1) and (3), the TVP-OLS-MG estimator defined in (8) can be written as:

$$\begin{aligned}
\hat{\beta}_{MG,t} &= \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_{i,t} \\
&= \frac{1}{N}\sum_{i=1}^{N}\Big[\Big(\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)^{-1}\sum_{j=1}^{T} b_{jt}x_{ij}y_{ij}\Big] \\
&= \frac{1}{N}\sum_{i=1}^{N}\Big[\Big(\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)^{-1}\sum_{j=1}^{T} b_{jt}x_{ij}(x_{ij}'\beta_{ij}+u_{ij})\Big] \\
&= \frac{1}{N}\sum_{i=1}^{N}\Big[\Big(\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)^{-1}\sum_{j=1}^{T} b_{jt}x_{ij}(x_{ij}'(\beta_{0,j}-\beta_{0,t}+\beta_{0,t}+e_{ij})+u_{ij})\Big].
\end{aligned}$$

Then, we have

$$\hat{\beta}_{MG,t}-\beta_{0,t} = \Delta_{x,it}+S_{xu,it}+S_{xe,it}, \tag{B.1}$$

where:

$$\begin{aligned}
\Delta_{x,it} &= \frac{1}{N}\sum_{i=1}^{N}\Big[\big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\big)^{-1}\big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'(\beta_{0,j}-\beta_{0,t})\big)\Big] \\
S_{xu,it} &= \frac{1}{N}\sum_{i=1}^{N}\Big[\big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\big)^{-1}\big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}u_{ij}\big)\Big] \\
S_{xe,it} &= \frac{1}{N}\sum_{i=1}^{N}\Big[\big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\big)^{-1}\big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'e_{ij}\big)\Big].
\end{aligned}$$

We will show that

$$\|\Delta_{x,it}\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big),\quad \max_{t=1,2,\cdots,T}\|\Delta_{x,it}\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big); \tag{B.2}$$

$$\|S_{xu,it}\| = O_p\Big(\frac{1}{\sqrt{NH}}\Big),\quad \max_{t=1,2,\cdots,T}\|S_{xu,it}\| = O_p\Big((NH)^{-1/2}\log^{1/2}T\Big); \tag{B.3}$$

$$\|S_{xe,it}\| = O_p\Big(\frac{1}{\sqrt{N}}\Big),\quad \max_{t=1,2,\cdots,T}\|S_{xe,it}\| = O_p\Big(\frac{1}{\sqrt{N}}\log^{1/\alpha}T(1+(\frac{H}{T})^{\gamma_2})\Big); \tag{B.4}$$

These together establish (1) and (2)(i) in Theorem 1.

Proof of (B.2). Notice that

$$\begin{aligned}\|\Delta_{x,it}\| &\leqslant \frac{1}{N}\sum_{i=1}^{N}\Big\|(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}')^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'(\beta_{0,j}-\beta_{0,t}))\Big\| \\ &\leqslant \frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big\|_{sp}^{-1}\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'(\beta_{0,j}-\beta_{0,t})\Big\| \\ &= O_p\Big((\frac{H}{T})^{\gamma_1}\Big).\end{aligned}$$

Similarly, we have

$$\begin{aligned}\max_{t=1,2,\cdots,T}\|\Delta_{x,it}\| &\leqslant \frac{1}{N}\sum_{i=1}^{N}\max_{t=1,2,\cdots,T}\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big\|_{sp}^{-1}\max_{t=1,2,\cdots,T}\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'(\beta_{0,j}-\beta_{0,t})\Big\| \\ &= O_p\Big((\frac{H}{T})^{\gamma_1}\Big).\end{aligned}$$

The final inequalities above all follow from (A.8) and (A.9) in Lemma A2, and the fact that all terms are *i.i.d.* across i (Assumption 2.2(i)).

Proof of (B.3). Let us define

$$Z_i^{xu} = (\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}')^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}u_{ij}).$$

By (A.8) and (A.10), we have

$$\|Z_i^{xu}\| \leqslant \Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big\|_{sp}^{-1}\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}u_{ij}\Big\| = O_p\Big(\frac{1}{\sqrt{H}}\Big).$$

Then, we have

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

where the first equality follows from the fact that (Z_i^{xu}) is *i.i.d.* over i. This implies that

$$\|S_{xu,it}\| = O\Big(\frac{1}{\sqrt{NH}}\Big).$$

To derive the uniform rate, it follows a similar reasoning as above, except for the fact that we have to use the uniform rate of (A.8) and (A.10) in Lemma A2. This implies that

$$\max_{t=1,2,\cdots,T}\left\|S_{xu,it}\right\| = O_p\Big((NH)^{-1/2}\log^{1/2} T\Big)$$

Proof of (B.4). Let us define

$$Z_i^{xe} = \Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)^{-1}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'e_{ij}\Big).$$

By (A.8) and (A.11), we have

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

Then, we have

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

where the first line follows from the fact that (Z_i^{xe}) is *i.i.d.* over i. This implies that

$$\left\|S_{xe,it}\right\| = O\Big(\frac{1}{\sqrt{N}}\Big).$$

To derive the uniform rate, it follows a similar reasoning as above, except for the fact that we have to use the uniform rate of (A.8) and (A.11) in Lemma A2. This implies that

$$\max_{t=1,2,\cdots,T}\left\|S_{xe,it}\right\| = O_p\Big(\frac{1}{\sqrt{N}}\log^{1/\alpha} T\big(1+(\frac{H}{T})^{\gamma_2}\big)\Big).$$

Now by combining (B.2), (B.3) and (B.4), we have

$$\begin{aligned}\sqrt{N}(\hat{\beta}_{MG,t}-\beta_{0,t}) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big[\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)^{-1}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'e_{ij}\Big)\Big] + O_p\Big(\frac{1}{\sqrt{H}}\Big) + O_p\Big(\sqrt{N}(\frac{H}{T})^{\gamma_1}\Big)\\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_{it} + o_p(1),\end{aligned}$$

since it is assumed that $(\frac{H}{T})^{\gamma_1} = o(N^{-1/2})$ as $(N,T)\to\infty$ and according to (A.11)

$$\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'e_{ij} = \Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)e_{it} + o_p(1).$$

By Assumption 2.2(i), we can apply CLT for *i.i.d.* sequences to obtain

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

where

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

and by Assumption 2.2(iv) it is positive definite.

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

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\beta}_{it}-\hat{\beta}_{MG,t})(\hat{\beta}_{it}-\hat{\beta}_{MG,t})'. \tag{B.5}$$

Because

$$\begin{aligned}\hat{\beta}_{it}-\beta_{0,t}&=(\frac{1}{H}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}')^{-1}(\frac{1}{H}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'e_{ij})+O_p\Big(\frac{1}{\sqrt{H}}\Big)+O_p\Big((\frac{H}{T})^{\gamma_1}\Big)\\&=e_{it}+o_p(1)\end{aligned}$$

and

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

we have

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

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

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

which implies that (B.5) is a consistent estimator for $\Omega_{e,t}$.

By (B.1), we see that

$$\begin{aligned}\max_{t=1,2,\cdots,T}\left\|\hat{\beta}_{MG,t}-\beta_{0,t}\right\|&\leqslant\max_{t=1,2,\cdots,T}\left\|\Delta_{x,it}\right\|+\max_{t=1,2,\cdots,T}\left\|S_{xu,it}\right\|+\max_{t=1,2,\cdots,T}\left\|S_{xe,it}\right\|\\&=O_p\Big((\frac{H}{T})^{\gamma_1}+\frac{\log^{1/\alpha}T}{\sqrt{N}}+\sqrt{\frac{\log T}{NH}}\Big),\end{aligned}$$

which completes the proof of Theorem 1(1) for $\hat{\beta}_{MG,t}$ and Theorem 1(2)(i).

We now explore the pooled estimator (9). Notice that

$$\begin{aligned}
\hat{\beta}_{P,t} - \beta_{0,t} &= \Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}y_{ij} - \beta_{0,t} \\
&= \Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}(x_{ij}'\beta_{ij} + u_{ij}) - \beta_{0,t} \\
&= \Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}(x_{ij}'(\beta_{0,j} - \beta_{0,t} + \beta_{0,t} + e_{ij}) + u_{ij}) - \beta_{0,t} \\
&= \Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}u_{ij} + \Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'e_{ij} \\
&\quad + \Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'(\beta_{0,j} - \beta_{0,t}) \\
&= S_{xx,p,t}^{-1}(S_{xu,p,t} + S_{xe,p,t} + \Delta_{x,p,t}), \qquad \text{(B.6)}
\end{aligned}$$

where

$$\begin{aligned}
S_{xx,p,t} &= \frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}' \\
\Delta_{x,p,t} &= \frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'(\beta_{0,j} - \beta_{0,t}) \\
S_{xu,p,t} &= \frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}u_{ij} \\
S_{xe,p,t} &= \frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'e_{ij}
\end{aligned}$$

We will show that

$$\|S_{xx,p,t}\|_{sp} = O_p(1), \quad \max_{t=1,2,\cdots,T}\|S_{xx,p,t}\|_{sp} = O_p(1); \tag{B.7}$$

$$\|\Delta_{x,p,t}\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big), \quad \max_{t=1,2,\cdots,T}\|\Delta_{x,p,t}\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big); \tag{B.8}$$

$$\|S_{xu,p,t}\| = O_p\Big(\frac{1}{\sqrt{NH}}\Big), \quad \max_{t=1,2,\cdots,T}\|S_{xu,p,t}\| = O_p\Big((NH)^{-1/2}\log^{1/2} T\Big); \tag{B.9}$$

$$\|S_{xe,p,t}\| = O_p\Big(\frac{1}{\sqrt{N}}\Big), \quad \max_{t=1,2,\cdots,T}\|S_{xe,p,t}\| = O_p\Big(\frac{1}{\sqrt{N}}\log^{1/\alpha} T(1 + (\frac{H}{T})^{\gamma_2})\Big); \tag{B.10}$$

These together establish (1) and (2)(ii) in Theorem 1.

Proof of (B.7). This follows immediately from (A.8), since by Assumption 2.2(i), all are *i.i.d.* across i:

$$\Big\|\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big\|_{sp} \leqslant \frac{1}{N}\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big\|_{sp} = O_p(1)$$

$$\max_{t=1,2,\cdots,T}\Big\|\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big\|_{sp} \leqslant \frac{1}{N}\max_{t=1,2,\cdots,T}\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big\|_{sp} = O_p(1)$$

Proof of (B.8). By Assumption 2.2(i) and (A.9), we have

$$\Big\|\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'(\beta_{0,j}-\beta_{0,t})\Big\| \leqslant \frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'(\beta_{0,j}-\beta_{0,t})\Big\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big)$$

$$\max_{t=1,2,\cdots,T}\Big\|\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'(\beta_{0,j}-\beta_{0,t})\Big\| \leqslant \frac{1}{N}\sum_{i=1}^{N}\max_{t=1,2,\cdots,T}\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'(\beta_{0,j}-\beta_{0,t})\Big\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big)$$

Proof of (B.9). Define

$$Z_{i,p,t}^{xu} = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}u_{ij}.$$

By Assumption 2.2(i) and (A.10), we have

$$E\big\|S_{xu,p,t}\big\|^2 = \frac{1}{N^2}\sum_{i=1}^{N} E\big\|Z_{i,p,t}^{xu}\big\|^2 = O\Big(\frac{1}{NH}\Big).$$

This establishes the first part of (B.9). To derive uniform rate, it follows same reasoning as above, except that we need to use uniform rate in (A.10). Then, we have

$$\max_{t=1,2,\cdots,T}\big\|S_{xu,p,t}\big\| = O_p\Big((NH)^{-1/2}\log^{1/2} T\Big).$$

Proof of (B.10). Define

$$Z_{i,p,t}^{xe} = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'e_{ij}.$$

By Assumption 2.2(i) and (A.11), we have

$$E\big\|S_{xe,p,t}\big\|^2 = \frac{1}{N^2}\sum_{i=1}^{N} E\big\|Z_{i,p,t}^{xe}\big\|^2 = O\Big(\frac{1}{N}\Big).$$

This establishes the first part of (B.10). To derive uniform rate, it follows same reasoning as above, except that we need to use uniform rate in (A.11). Then, we have

$$\max_{t=1,2,\cdots,T}\left\|S_{xu,p,t}\right\| = O_p\Big(\frac{1}{\sqrt{N}}\log^{1/\alpha} T(1+(\frac{H}{T})^{\gamma_2})\Big).$$

Now by combining (B.7), (B.8), (B.9), and (B.10), we have

$$\begin{aligned}\sqrt{N}(\hat{\beta}_{P,t}-\beta_{0,t}) &= \Big(\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)^{-1}\Big(\frac{1}{\sqrt{N}K_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'e_{ij}\Big) + O_p\Big(\frac{1}{\sqrt{H}}\Big) + O_p\Big(\sqrt{N}(\frac{H}{T})^{\gamma_1}\Big)\\ &= \left(\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)e_{it}\right) + o_p(1)\end{aligned}$$

Because it is assumed that $(\frac{H}{T})^{\gamma_1} = o(N^{-1/2})$ as $(N,T)\to\infty$, the last two terms in the first line are $o_p(1)$ and the dominating term is the first one. As in the case of Mean group estimator, we can apply CLT for *i.i.d.* sequences to obtain

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)e_{it} \xrightarrow{d} N(0,R_t),$$

where

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

and

$$\Sigma_{xx,it} = \operatorname*{plim}_{T\to\infty}\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}'x_{ij}.$$

Then, since

$$\operatorname*{plim}_{(N,T)\to\infty}\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}' = \overline{\Sigma}_{xx,t} = O(1),$$

it follows from Slutsky's theorem that

$$\sqrt{N}\,\overline{\Sigma}_{xx,t}\ R_t^{-1/2}\left(\hat{\beta}_{P,t}-\beta_t\right) \xrightarrow{d} N(0,I_k).$$

Consider

$$\hat{R}_t = \frac{1}{N}\sum_{i=1}^{N}\left[\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)(\hat{\beta}_{i,t}-\hat{\beta}_{MG,t})(\hat{\beta}_{i,t}-\hat{\beta}_{MG,t})'\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big)\right].$$

We now show that $\hat{R}_t$ is a consistent estimator of R_t. Write

$$\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'(\hat{\beta}_{i,t}-\hat{\beta}_{MG,t}) = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\Big(e_{it}-\frac{1}{N}\sum_{i=1}^{N} e_{it}\Big)$$

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

$$\hat{R}_t \xrightarrow{p} R_t$$

which establishes the claim.

By (B.6), together with (B.7), (B.8), (B.9), and (B.10), we have

$$\begin{aligned}\max_{t=1,2,\cdots,T}\left\|\hat{\beta}_{P,t} - \beta_{0,t}\right\| &\leqslant \max_{t=1,2,\cdots,T}\left\|S_{xx,p,t}\right\|_{sp}^{-1}\Big(\max_{t=1,2,\cdots,T}\left\|\Delta_{x,p,t}\right\| + \max_{t=1,2,\cdots,T}\left\|S_{xu,p,t}\right\| + \max_{t=1,2,\cdots,T}\left\|S_{xe,p,t}\right\|\Big)\\ &= O_p\Big((\frac{H}{T})^{\gamma_1} + \frac{\log^{1/\alpha} T}{\sqrt{N}} + \sqrt{\frac{\log T}{NH}}\Big).\end{aligned}$$

Appendix B.2. Proof of Theorem 2

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

$$\begin{aligned}\hat{\beta}_{MG,t}^{IV} &= \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_{i,t}^{IV}\\ &= \frac{1}{N}\sum_{i=1}^{N}\Big[(\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j)^{-1}(\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} y_{ij})\Big].\end{aligned}$$

A similar expansion to the TVP-OLS-MG case gives

$$\hat{\beta}_{MG,t}^{IV} - \beta_{0,t} = \Delta_{z,t} + S_{zu,t} + S_{ze,t},$$

where

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

We will show that

$$\left\|\Delta_{z,t}\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big), \quad \max_{t=1,2,\cdots,T}\left\|\Delta_{z,t}\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big) \tag{B.11}$$

$$\left\|S_{zu,t}\right\| = O_p\Big(\frac{1}{\sqrt{NH}}\Big), \quad \max_{t=1,2,\cdots,T}\left\|S_{zu,t}\right\| = O_p\left(\frac{1}{\sqrt{N}}\Big((\bar{r}_{T,H,\gamma_3} + r_{N,T,L,\gamma_3,\alpha}\Big)\right) \tag{B.12}$$

$$\left\|S_{ze,t}\right\| = O_p\Big(\frac{1}{\sqrt{N}}\Big), \quad \max_{t=1,2,\cdots,T}\left\|S_{ze,t}\right\| = O_p\left(\frac{(\log T)^{1/\alpha}}{\sqrt{N}}\Big(1 + (\frac{H}{T})^{\min(\gamma_3,\gamma_4)}\Big)\right). \tag{B.13}$$

These together establishes Theorem 2(i) and Theorem 2(ii).

Proof of (B.11). Observe that

$$\begin{aligned}\left\|\Delta_{z,t}\right\| &\leqslant \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j\right\|_{sp}^{-1}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij}(\beta_{0,j} - \beta_{0,t})\right\| \\ &= O_p\Big((\frac{H}{T})^{\gamma_1}\Big),\end{aligned}$$

where the second line follows from (A.14) and (A.15) and the fact that by Assumption 3.2(i) all terms are *i.i.d.* accross i. Again, by uniform rates derived in (A.14) and (A.15), we have

$$\max_{t=1,2,\cdots,T}\left\|\Delta_{z,t}\right\| \leqslant \frac{1}{N}\sum_{i=1}^{N}\max_{t=1,2,\cdots,T}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j\right\|_{sp}^{-1}\max_{t=1,2,\cdots,T}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij}(\beta_{0,j} - \beta_{0,t})\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big).$$

Proof of (B.12). Let

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

Clearly, by (A.14) and (A.16):

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

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, it follows similar reasoning as above, except for the fact that we have to use uniform rates derived in (A.14) and (A.16) to obtain

$$\max_{t=1,2,\cdots,T}\left\|S_{zu,t}\right\| = O_p\left(\frac{1}{\sqrt{N}}\left(\bar{r}_{T,H,\gamma_3} + r_{N,T,L,\gamma_3,\alpha}\right)\right).$$

Proof of (B.13). Let

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

Clearly, by (A.14) and (A.17):

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

Thus, we have

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

which implies that

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

To derive the uniform rate, it follows again similar reasoning as above, except for the fact that we have to use uniform rates derived in (A.14) and (A.17) to obtain

$$\max_{t=1,2,\cdots,T}\left\|S_{ze,t}\right\| = O_p\left(\frac{(\log T)^{1/\alpha}}{\sqrt{N}}\left(1 + \left(\frac{H}{T}\right)^{\min(\gamma_3,\gamma_4)}\right)\right).$$

Now, we sum them up. By (B.11), (B.12) and (B.13), we first obtain the uniform consistency rate

$$\begin{aligned}\max_{t=1,2,\cdots,T}\left\|\hat{\beta}_{MG,t}^{IV} - \beta_{0,t}\right\| &\leqslant \max_{t=1,2,\cdots,T}\|\Delta_{xt}\| + \max_{t=1,2,\cdots,T}\left\|S_{xu,t}\right\| + \max_{t=1,2,\cdots,T}\left\|S_{xe,t}\right\| \\ &= O_p\left(r_{N,T,H,\gamma_1,\alpha} + \frac{r_{N,T,L,\gamma_3,\alpha}}{\sqrt{N}}\right).\end{aligned}$$

Then, we obtain the expansion of the estimator

$$\sqrt{N}(\hat{\beta}_{MG,t}^{IV} - \beta_{0,t}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\left(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j\right)^{-1}\left(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij} e_{ij}\right)\right] + O_p\left(\sqrt{N}\left(\frac{H}{T}\right)^{\gamma_1}\right) + O_p\left(\frac{1}{\sqrt{H}}\right).$$

Because we assume that $(\frac{H}{T})^{\gamma_1} = o(N^{-1/2})$ as $(N,T) \to \infty$, the last two terms above are $o_p(1)$ and the dominating term is the first one. Recall from the derivation of (A.14) and (A.17), we have that

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

where

$$\Sigma_{\Psi zz\Psi,it} = \Psi'_{0,t}\Sigma_{zz,it}\Psi_{0,t}, \quad \Sigma_{zz,it} = \operatorname*{plim}_{T\to\infty}\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H} z_{ij} z'_{ij}$$
$$\Sigma_{\Psi zx,it} = \Psi'_{0,t}\Sigma_{zx,it}, \quad \Sigma_{zx,it} = \operatorname*{plim}_{T\to\infty}\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H} z_{ij} x'_{ij}.$$

Then, we have

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

By Assumptions 3.2 and 3.4, $\Sigma^{-1}_{\Psi zz\Psi,it}$ and $\Sigma_{\Psi zx,it}$ are all positive definite. Since (e_{it}) is *i.i.d.* across i, we can apply CLT for *i.i.d.* sequence to obtain

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

where $\Omega^{IV}_{e,t}$ is given by

$$\Omega^{IV}_{e,t} = \lim_{N\to\infty}\ \mathrm{Var}\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Sigma^{-1}_{\Psi zz\Psi,it}\Sigma_{\Psi zx,it} e_{it}\Big).$$

Consider

$$\hat{\beta}^{IV}_{it} - \beta_{0,t} = \Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j\Big)^{-1}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij} e_{ij}\Big) + o_p(1).$$

Similar expansion as in TVP-OLS case gives

$$\hat{\beta}^{IV}_{it} - \beta_{0,t} = \Sigma^{-1}_{\Psi zz\Psi,it}\Sigma_{\Psi zx,it} e_{it} + o_p(1).$$

Then, it follows similarly as in the proof of TVP-OLS-MG case that

$$\sum_{i=1}^{N}(\hat{\beta}^{IV}_{it} - \hat{\beta}^{IV}_{MG,t})(\hat{\beta}^{IV}_{it} - \hat{\beta}^{IV}_{MG,t})'$$

is a consistent estimator for $\Omega^{IV}_{MG,t}$.

In the next step, we consider the pooled estimator

$$\hat{\beta}^{IV}_{P,t} = \Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j\Big)^{-1}\Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} y_{i,j}\Big).$$

Similarly to the TVP-OLS-P case, we have

$$\hat{\beta}^{IV}_{P,t} - \beta_{0,t} = S^{-1}_{z\Psi,p,t}\Big(\Delta_{z,p,t} + S_{zu,p,t} + S_{ze,p,t}\Big),$$

where

$$S_{z\Psi,p,t} = \frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} z'_{i,j}\hat{\Psi}_j$$

$$\Delta_{z,p,t} = \frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} x'_{i,j}(\beta_{0,j} - \beta_{0,t})$$

$$S_{zu,p,t} = \frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} u_{i,j}$$

$$S_{ze,p,t} = \frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} x'_{i,j} e_{i,j}.$$

We will show that

$$\left\|S_{z\Psi,p,t}\right\|_{sp} = O_p(1), \quad \max_{t=1,2,\cdots,T}\left\|S_{z\Psi,p,t}\right\|_{sp} = O_p(1) \tag{B.14}$$

$$\left\|\Delta_{z,p,t}\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big), \quad \max_{t=1,2,\cdots,T}\left\|\Delta_{z,p,t}\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big) \tag{B.15}$$

$$\left\|S_{zu,p,t}\right\| = O_p\Big(\frac{1}{\sqrt{NH}}\Big), \quad \max_{t=1,2,\cdots,T}\left\|S_{zu,p,t}\right\| = O_p\left(\frac{1}{\sqrt{N}}\Big(\bar{r}_{T,H,\gamma_3} + r_{N,T,L,\gamma_3,\alpha}\Big)\right) \tag{B.16}$$

$$\left\|S_{ze,p,t}\right\| = O_p\Big(\frac{1}{\sqrt{N}}\Big), \quad \max_{t=1,2,\cdots,T}\left\|S_{ze,p,t}\right\| = O_p\left(\frac{(\log T)^{1/\alpha}}{\sqrt{N}}\Big(1 + (\frac{H}{T})^{\min(\gamma_3,\gamma_4)}\Big)\right). \tag{B.17}$$

These together establishes uniform consistency and asymptotic normality for the pooled estimator.

Proof of (B.14). This follows immediately from (A.14) and Assumption 3.2(i):

$$\left\|\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} z'_{i,j}\hat{\Psi}_j\right\|_{sp} \leqslant \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} z'_{i,j}\hat{\Psi}_j\right\|_{sp} = O_p(1)$$

$$\max_{t=1,2,\cdots,T}\left\|\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} z'_{i,j}\hat{\Psi}_j\right\|_{sp} \leqslant \frac{1}{N}\sum_{i=1}^{N}\max_{t=1,2,\cdots,T}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} z'_{i,j}\hat{\Psi}_j\right\|_{sp} = O_p(1)$$

Proof of (B.15). By Assumption 3.2(1) and (A.15), we have

$$\left\|\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} x'_{i,j}(\beta_{0,j}-\beta_{0,t})\right\| \leqslant \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} x'_{i,j}(\beta_{0,j}-\beta_{0,t})\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big)$$

$$\max_{t=1,2,\cdots,T}\left\|\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} x'_{i,j}(\beta_{0,j}-\beta_{0,t})\right\| \leqslant \frac{1}{N}\sum_{i=1}^{N}\max_{t=1,2,\cdots,T}\left\|\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} x'_{i,j}(\beta_{0,j}-\beta_{0,t})\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big).$$

Proof of (B.16). Define

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

By Assumption 3.2(i) and (A.16), we have

$$E\left\|S_{zu,p,t}\right\|^2 = \frac{1}{N^2}\sum_{i=1}^{N} E\left\|Z^{zu}_{i,p,t}\right\|^2 = O\Big(\frac{1}{NH}\Big),$$

which establishes the first part of (B.16). To derive uniform rate, it follows a similar reasoning as above, except that we need to use uniform rate in (A.16). Then, we have

$$\max_{t=1,2,\cdots,T}\left\|S_{zu,p,t}\right\| = O_p\bigg(\frac{1}{\sqrt{N}}\Big(\bar{r}_{T,H,\gamma_3} + r_{N,T,L,\gamma_3,\alpha}\Big)\bigg).$$

Proof of (B.17). Define

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

By Assumption 3.2(1) and (A.17), we have

$$E\left\|S_{ze,p,t}\right\|^2 = \frac{1}{N^2}\sum_{i=1}^{N} E\left\|Z^{ze}_{i,p,t}\right\|^2 = O\Big(\frac{1}{N}\Big),$$

which establishes the first part of (B.17). To derive uniform rate, it follows a similar reasoning as above, except that we need to use uniform rate in (A.17). Then, we have

$$\max_{t=1,2,\cdots,T}\left\|S_{ze,p,t}\right\| = O_p\Big(\frac{(\log T)^{1/\alpha}}{\sqrt{N}}\big(1+(\frac{H}{T})^{\min(\gamma_3,\gamma_4)}\big)\Big).$$

Now we combine the equations. First, by (B.14), (B.15), (B.16), and (B.17), we have

$$\begin{aligned}\max_{t=1,2,\cdots,T}\left\|\hat{\beta}^{IV}_{P,t}-\beta_{0,t}\right\| & \max_{t=1,2,\cdots,T}\left\|S_{z\Psi,p,t}\right\|^{-1}_{sp}\Big(\max_{t=1,2,\cdots,T}\left\|\Delta_{z,p,t}\right\| + \max_{t=1,2,\cdots,T}\left\|S_{zu,p,t}\right\| + \max_{t=1,2,\cdots,T}\left\|S_{ze,p,t}\right\|\Big)\\ &= O_p\Big(r_{N,T,H,\gamma_1,\alpha}+\frac{r_{N,T,L,\gamma_3,\alpha}}{\sqrt{N}}\Big).\end{aligned}$$

Then, we obtain the following expansion for the pooled estimator

$$\sqrt{N}(\hat{\beta}^{IV}_{P,t}-\beta_{0,t}) = \Big(\frac{1}{NH}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}z'_{ij}\hat{\Psi}_j\Big)^{-1}\Big(\frac{1}{\sqrt{N}H}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}x'_{ij}e_{ij}\Big) + O_p\Big(\sqrt{N}(\frac{H}{T})^{\gamma_1}\Big) + O_p\Big(\frac{1}{\sqrt{H}}\Big).$$

Because it is assumed that $(\frac{H}{T})^{\gamma_1} = o(N^{-1/2})$ as $(N,T)\to\infty$, the last two terms are $o_p(1)$ and the dominating term is the first one. Recall from the derivations of (A.14) and (A.17), and the fact that $\overline{\Sigma}_{\Psi zz\Psi,t} = \mathrm{plim}_{(N,T)\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\Psi'_j z_{ij}z'_{ij}\Psi_j = O_p(1)$, we have that

$$\begin{aligned}\frac{1}{NH}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}z'_{ij}\hat{\Psi}_j &= \overline{\Sigma}_{\Psi zz\Psi,t} + o_p(1)\\ \frac{1}{\sqrt{N}H}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}x'_{ij}e_{ij} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Sigma_{\Psi zx,it}e_{it} + o_p(1).\end{aligned}$$

Since both $\overline{\Sigma}_{\Psi zz\Psi,t}$ and $\Sigma_{\Psi zx,it}$ are positive definite and the fact that (e_{it}) are $i.i.d.$ across i, we apply CLT for $i.i.d.$ sequence to obtain:

$$\frac{1}{\sqrt{N}H}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\Sigma_{\Psi zx,it}e_{it} \xrightarrow{d} N\big(0, R^{IV}_{P,t}\big),$$

where $R^{IV}_{P,t}$ is given by

$$R^{IV}_{t} = \lim_{N\to\infty}\mathrm{Var}\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Sigma_{\Psi zx,it}e_{it}\Big).$$

By Slutsky's theorem, we have

$$\sqrt{N}\,\overline{\Sigma}_{\Psi zz\Psi,t}\left(R^{IV}_{P,t}\right)^{-1/2}\big(\hat{\beta}^{IV}_{P,t}-\beta_{0,t}\big) \xrightarrow{d} N(0, I_k).$$

Consider

$$\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}x'_{ij}\Big)\Big(\hat{\beta}^{IV}_{i,t}-\hat{\beta}^{IV}_{MG,t}\Big)=\Psi'_{0,t}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}z_{ij}x'_{ij}\Big)(e_{it}-\frac{1}{N}\sum_{i=1}^{N}e_{it})+o_p(1).$$

Then, we have

$$\frac{1}{N}\sum_{i=1}^{N}\Bigg[\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}x'_{ij}\Big)\Big(\hat{\beta}^{IV}_{i,t}-\hat{\beta}^{IV}_{MG,t}\Big)\Big(\hat{\beta}^{IV}_{i,t}-\hat{\beta}^{IV}_{MG,t}\Big)'\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}x'_{ij}\Big)'\Bigg]=R^{IV}_{p,t}+o_p(1),$$

which shows that above is a consistent estimator for $R^{IV}_{p,t}$.