KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE
OF
ECONOMIC RESEARCH

Discussion Paper No.906

"Panel Data Analysis with Heterogeneous Dynamics"

Ryo Okui and Takahide Yanagi

November 2014

KYOTO UNIVERSITY

KYOTO, JAPAN

Panel Data Analysis with Heterogeneous Dynamics*†

Ryo Okui‡ and Takahide Yanagi§

November 25, 2014

Abstract

This paper proposes the analysis of panel data whose dynamic structure is heterogeneous across individuals. Our aim is to estimate the cross-sectional distributions and/or some distributional features of the heterogeneous mean and autocovariances. We do not assume any specific model for the dynamics. Our proposed method is easy to implement. We first compute the sample mean and autocovariances for each individual and then estimate the parameter of interest based on the empirical distributions of the estimated mean and autocovariances. The asymptotic properties of the proposed estimators are investigated using double asymptotics under which both the cross-sectional sample size (N) and the length of the time series (T) tend to infinity. We prove the functional central limit theorem for the empirical process of the proposed distribution estimator. By using the functional delta method, we also derive the asymptotic distributions of the estimators for various parameters of interest. We show that the distribution estimator exhibits a bias whose order is proportional to $1/\sqrt{T}$. Conversely, when the parameter of interest can be written as the expectation of a smooth function of the heterogeneous mean and/or autocovariances, the bias is of order $1/T$ and can be corrected by the jackknife method. The results of Monte Carlo simulations show that our asymptotic results are informative regarding the finite-sample properties of the estimators. They also demonstrate that the proposed jackknife bias correction is successful.

Keywords: Panel data; heterogeneity; functional central limit theorem; autocovariance; jackknife; long panel.

JEL Classification: C13; C14; C23.

---

*This paper was previously circulated under the title "Dynamic Panel Data Analysis when the Dynamics are Heterogeneous."

†The authors would like to thank Hidehiko Ichimura, Kengo Kato, Katsumi Shimotsu, Hiroyuki Kasahara, Kazuhiko Hayakawa, Yoon-Jae Whang, Simon Lee, Yuya Sasaki, Taisuke Otsu, Hiroaki Kaido, Roger Koenker, Yoosoon Chang, Hashem Pesaran, Yasunori Fujikoshi, seminar participants at the University of Tokyo, Seoul National University, and Hiroshima University, and attendees at the Kansai Econometrics Meeting held in Kyoto, the Institute of Advanced Studies Workshop on Advances in Microeconometrics held in Hong Kong, the 2014 Econometric Society Australasian Meeting, the 20th International Panel Data Conference held in Tokyo, the Mini Conference in Microeconometrics held in Hakone, and the Summer Workshop on Economic Theory 2014 for helpful comments and discussion. Okui acknowledges financial support from the Japan Society for the Promotion of Science (JSPS) under KAKENHI Grants Nos. 22330067, 25780151, and 25285067. Yanagi recognizes financial support from Grant-in-Aid for JSPS Fellows No. 252035. All remaining errors are of course ours.

‡Institute of Economic Research, Kyoto University, Yoshida-Hommachi, Sakyo, Kyoto, Kyoto, 606-8501, Japan. Email: okui@kier.kyoto-u.ac.jp

§Graduate School of Economics, Kyoto University, Yoshida-Hommachi, Sakyo, Kyoto, Kyoto, 606-8501, Japan; Research Fellow of Japan Society for the Promotion of Science. Email: yanagi.takahide.87w@st.kyoto-u.ac.jp

1 Introduction

This paper considers the analysis of panel data whose dynamic structure is stationary across time but heterogeneous across individuals. We propose methods for estimating the distributional features of the mean and autocovariances that are heterogeneous across individuals using panel data. Our estimation procedure is simple to implement. We first estimate the mean or autocovariances for each individual. We then estimate the distribution and other distributional quantities using the empirical distribution of the estimated mean or autocovariances. When the parameter of interest can be written as the expected value of a smooth function of the heterogeneous mean or autocovariances, the jackknife method reduces the bias of the estimator.

Understanding the dynamic nature of an economic variable that is potentially heterogeneous when using panel data is an important research consideration in economics. For example, there is considerable study using panel data on income dynamics (see, e.g., Lillard and Willis, 1978, Meghir and Pistaferri, 2004, Guvenen, 2007, and Browning, Ejrnæs, and Alvarez, 2010, among many others). In particular, Browning et al. (2010) show that income dynamics exhibit considerable heterogeneity in that an income shock may have a persistent effect on the future income profiles of some individuals, whereas for others, the effect may disappear quite quickly.

The contribution of this paper is to propose easy-to-implement methods to analyze panel data whose dynamics are heterogeneous without assuming any specific model. To study the heterogeneous dynamic structure, we investigate the cross-sectional distributions of the mean and autocovariances that are heterogeneous across individuals. Investigating these quantities does not depend on a particular model structure. While the literature on dynamic panel data analysis is already voluminous, many studies assume some specific model for the dynamics (such as the autoregressive (AR) model) and the homogeneity in the dynamics, allowing heterogeneity only in the mean of the process.[1] While several analyses consider either heterogeneous dynamics or model-free analysis (see the section "Related literature" below), we are unaware of any specific study that proposes methods to analyze heterogeneous dynamics using panel data without specifying some particular model. This paper builds on the literature by proposing model-free analysis for a heterogeneous dynamic structure.

The distributions of the heterogeneous mean and autocovariances are informative in various ways. First, the mean and the autocovariances are perhaps the most basic descriptive statistics for dynamics. Indeed, a typical first step in analyzing time-series data is to examine the mean and the autocovariance (or autocorrelation) properties of the data. We believe that the distributions of the heterogeneous mean and autocovariances would also be useful descriptive statistics for understanding the dynamics in panel data analysis. Second, we can use the mean and autocovariances to investigate whether different groups possess dissimilar dynamic structures without relying on some particular model. For example, consider the situation in which we would like to investigate whether males and females face different income dynamics, but we

[1] See, e.g., Baltagi (2008) and Arellano (2003) for excellent reviews of the more important existing contributions on dynamic panel data analysis.

are also aware of the fact that income dynamics are heterogeneous across individuals. In such a case, we can estimate the distributions of the autocovariances for males and females separately and compare them to see if they indeed differ.

Our approach is to estimate the mean and autocovariances for each individual and use the empirical distributions of the estimated mean and autocovariances to estimate the cross-sectional distributions of the heterogeneous mean and autocovariances and other quantities of interest, such as the quantile function. The asymptotic properties of the empirical distributions are derived based on double asymptotics under which both the number of cross-sectional observations, N, and the length of the time series, T, tend to infinity. By using empirical process theory (see, e.g., van der Vaart and Wellner, 1996), we show that the empirical distributions converge weakly to Gaussian processes. However, the condition $N/T \to 0$ is required for this weak convergence because of the bias caused by the estimation error in the estimated mean and autocovariances for each individual. We also derive the asymptotic distributions of the estimators for other distributional characteristics, including quantiles, using the functional delta method.

When the parameter of interest can be written as the expected value of a smooth function of the heterogeneous mean or autocovariances, the condition on the relative magnitudes of N and T can be relaxed. This class of parameters includes the mean, the variance, and other moments of the heterogeneous mean and autocovariances. In this case, the bias becomes of order $O(1/T)$, and the condition $N/T^2 \to 0$ is sufficient for asymptotically unbiased estimation. Moreover, we can analytically evaluate the bias, and jackknife bias correction is available. This bias has two sources. The first is the incidental parameter problem originally discussed in Neyman and Scott (1948) and now well known in the econometrics literature. This type of bias does not affect the estimated mean, but does influence the estimated autocovariances. When we estimate the autocovariance for each individual, we also need to estimate the mean for each individual. Because there are N individual-specific mean parameters to be estimated, this creates incidental parameter bias. The second source of bias arises when the smooth function is nonlinear. This bias affects both the mean and the autocovariances. However, this source of bias does not appear when the parameter of interest is the mean of the heterogeneous mean or autocovariances because the corresponding function is linear. We propose using the half-panel jackknife in Dhaene and Jochmans (2014) to correct the bias.

We also conduct Monte Carlo simulations to investigate the finite-sample properties of the proposed procedures. The results of the Monte Carlo simulations show that the asymptotic analyses in this paper are informative regarding the finite-sample properties of the proposed estimators. They show that the estimators based on the estimated autocovariances have severe bias when T is small compared with N, but the bias decreases as T increases. They also show that the proposed jackknife bias correction decreases this bias. The half-panel jackknife also reduces the bias allocated with the incidental parameter problem and the nonlinearity of the smooth function, even when T is relatively small.

Related literature: This paper most closely relates to the literature on heterogeneous panel AR models. In these models, we capture the heterogeneity in the dynamics by allowing the AR coefficients to be individual specific. The panel AR(1) models with individual-specific AR coefficients are analyzed by, for example, Pesaran and Smith (1995), Hsiao, Pesaran, and Tahmiscioglu (1999), and Pesaran, Shin, and Smith (1999). These analyses are extended to nonstationary panel data by Phillips and Moon (1999), while Pesaran (2006) considers models with a multifactor error structure. The present analysis differs in two ways from these earlier studies. First, we do not assume any specific model to describe the dynamics, while the abovementioned studies consider an AR or linear-process specification. Second, our aim is to estimate the entire distribution of the mean or autocovariances, which are heterogeneous across individuals. In contrast, Pesaran and Smith (1995) and others focus on the estimation of the means of the AR coefficients.

Elsewhere, Mavroeidis, Sasaki, and Welch (2014) consider the identification and estimation of the distribution of the AR coefficients in heterogeneous panel AR models. The advantage of their approach is that T can be fixed, and thus it is applicable to short panels. While we consider the case in which $T \to \infty$, our method is much simpler to implement. We simply need to estimate the mean and autocovariances for each individual and compute the empirical distributions of the estimated mean and autocovariances. By contrast, the estimation method in Mavroeidis et al. (2014) requires the maximization of a kernel-weighting function that is written as an integration over multiple variables. We also emphasize that our method does not depend on model specification. In addition, we note that identification of the distributions of the heterogeneous mean and autocovariances is trivial in our setting because we consider the setting $T \to \infty$. Alternatively, the identification analysis in Mavroeidis et al. (2014) is mathematically involved because they consider fixed T.

Several studies propose model-free methods to investigate the dynamic structure using panel data. For example, Okui (2010, 2011, 2014) considers the estimation of autocovariances using long panel data and assumes that the autocovariance structure is homogeneous across individuals. By contrast, our paper considers a heterogeneous structure. However, we note that it is easy to show that Okui's autocovariance estimator is equivalent to the estimator of the mean of the heterogeneous autocovariances. In other work, Lee, Okui, and Shintani (2013) consider infinite-order panel AR models. Given we can represent a stationary time series by an infinite-order AR process under mild conditions, their approach is essentially model-free. However, they assume that the dynamics are homogeneous.

A different line of research investigates the properties of the estimators for model-based analysis when the assumed model is possibly misspecified. For instance, Okui (2008) examines the probability limits of various estimators for panel AR(1) models when the true dynamics do not follow an AR(1) process and assumes homogeneity in the dynamics, while the mean is allowed to be heterogeneous. Lee (2012) discusses the fixed effects estimator for panel AR models when the lag order is misspecified and also considers the case where the dynamics are homogeneous. Lastly, Galvao and Kato (2014) investigate the asymptotic properties of the fixed effects estimator in general regression models and allow the data-generating process to be generally heterogeneous. However, the purpose of the current study is to propose new methods to analyze panel data with heterogeneous dynamics, not to examine the properties of existing estimators.

The literature on deconvolution techniques examines the identification and estimation of the distribution of individual effects (see, e.g., Horowitz and Markatou, 1996, Székely and Rao, 2000, and Bonhomme and Robin, 2010). In the context of the present analysis, we may employ these deconvolution techniques to identify and estimate the distribution of the individual-specific mean with fixed T. That T can be fixed is an advantage of these techniques. However, our focus is on the distribution of not only the mean, but also the heterogeneous autocovariance. Moreover, we propose methods that are easily implemented under the requirement that $T \to \infty$. On the other hand, the deconvolution techniques involve the computation of the characteristic function, and the rate of convergence is remarkably slow.

While not directly connected, this paper is also somewhat related to the recent literature on random coefficient models. For example, Arellano and Bonhomme (2012) consider linear regression models with random coefficients in panel data analysis and discuss the identification and estimation of the distribution of random coefficients using deconvolution techniques. Note that Chamberlain (1992) and Graham and Powell (2012) consider a model similar to that of Arellano and Bonhomme (2012), but their focus is on the means of the random coefficients. Fernández-Val and Lee (2013) study moment restriction models with random coefficients and their generalized methods of moment estimation. Their analysis on the smooth function of individual effects is closely related to our analysis on the smooth function of means and autocovariances in terms of technique. Finally, Evdokimov (2009) considers a nonparametric panel regression model with individual effects entering the unspecified structural function, but also relies on deconvolution techniques.

Organization of the paper: The remainder of the paper is organized as follows. Section 2 explains the setting. Section 3 introduces the proposed procedures. In Section 4, we derive the asymptotic properties of the distribution estimators. Section 5 considers the estimation of the expected value of a smooth function of the heterogeneous mean or autocovariances, the inference methods, and the jackknife bias correction. Section 6 presents some extensions based on the proposed procedures. Section 7 presents the results of the Monte Carlo simulations. Section 8 concludes the paper. All technical proofs are presented in the Technical appendix.

2 Settings

We observe panel data $\{\{y_{it}\}_{t=1}^{T}\}_{i=1}^{N}$, where y_{it} is a scalar random variable, i represents a cross-sectional unit, and t indicates a time period. The number of cross-sectional observations is N

and the length of the time series is T. We consider situations in which both N and T are large. We assume that $\{y_{it}\}_{t=1}^T$ is independent across individuals.

The law of $\{y_{it}\}_{t=1}^T$ is assumed to be stationary, but its dynamic structure may be heterogeneous. To be specific, we consider the following data-generating process to model the heterogeneous dynamic structure. The unobserved individual effect, α_i, is independently drawn from a distribution common to all individuals. The time series $\{y_{it}\}_{t=1}^T$ for individual i is then drawn from some distribution $L(\{y_{it}\}_{t=1}^T; \alpha_i)$. The dynamic structure of y_{it} is heterogeneous because α_i varies across individuals. However, note that introducing the parameter α_i is a somewhat abstract way to represent heterogeneity in the dynamics across individuals. We do not directly assume anything about the distribution of α_i, because α_i does not explicitly appear in our analysis. For notational simplicity, we denote "$\cdot|\alpha_i$" by "$\cdot|i$"; that is, "conditional on α_i" becomes "conditional on i" below.

Our aim is to develop statistical tools to analyze the cross-sectional distributions of the heterogeneous mean and autocovariances of y_{it}. The mean for unit i is $\mu_i := E(y_{it}|i)$. Note that μ_i is a random variable whose realization differs across individuals. This is because μ_i depends on α_i, which differs among individuals. As we assume stationarity, μ_i is constant over time. The distribution of μ_i represents heterogeneity in the mean of y_{it} across individuals. Let $\gamma_{k,i}$ be the k-th conditional autocovariance of y_{it} given α_i:

$$\gamma_{k,i} := E\left((y_{it} - \mu_i)(y_{i,t-k} - \mu_i)|i\right).$$

In other words, $\gamma_{k,i}$ represents the k-th autocovariance of y_{it} for individual i. Note that $\gamma_{0,i}$ is the variance for individual i. Similarly to the case of μ_i, $\gamma_{k,i}$ is a random variable and its realization may be different among individuals. To understand the possibly heterogeneous dynamics of y_{it}, we aim to estimate quantities that characterize the distributions of μ_i and/or $\gamma_{k,i}$, such as the distribution function, the quantile function, and the moments.

Our setting is very general and includes many situations.

Example 1. The panel AR(1) model with heterogeneous coefficients considered by Pesaran and Smith (1995) and others is a special case of our setting. This model is

$$y_{it} = c_i + \phi_i y_{i,t-1} + \epsilon_{it},$$

where c_i and ϕ_i are the individual-specific intercept and slope coefficients, respectively, and ϵ_{it} follows a strong white noise process with variance σ^2. In this case, $\alpha_i = (c_i, \phi_i)$, $\mu_i = c_i/(1-\phi_i)$, and $\gamma_{k,i} = \sigma^2 \phi_i^k/(1-\phi_i^2)$.

Example 2. Another example is the case in which y_{it} is generated by a linear process with heterogeneous coefficients:

$$y_{it} = c_i + \sum_{j=0}^{\infty} \theta_{j,i} \epsilon_{i,t-j},$$

where c_i and $\{\theta_{j,i}\}_{j=0}^\infty$ are heterogeneous coefficients and ϵ_{it} follows a strong white noise process with variance σ^2. In this example, $\alpha_i = (c_i, \{\theta_{j,i}\}_{j=0}^\infty)$, $\mu_i = c_i$, and $\gamma_{k,i} = \sigma^2 \sum_{j=k}^\infty \theta_{j,i}\theta_{j-k,i}$.

Example 3. Our setting also includes cases in which the true data-generating process follows some nonlinear process. Suppose that y_{it} is generated by

$$y_{it} = m(\alpha_i, \epsilon_{it}),$$

where $m(\cdot, \cdot)$ is some function and ϵ_{it} is stationary over time and independent across individuals. In this case, $\mu_i = E(m(\alpha_i, \epsilon_{it})|\alpha_i)$ and $\gamma_{k,i}$ is the k-th order autocovariance of $w_{it} = y_{it} - \mu_i$ given α_i.

Our focus is on estimating the heterogeneous mean and autocovariance structure; we do not aim to recover the underlying structural form of the data-generating process. For example, even when y_{it} is generated by $y_{it} = m(\alpha_i, \epsilon_{it})$ as in the third example, we estimate not the function $m(\cdot, \cdot)$ but rather the heterogeneous mean and autocovariance structure only. We understand that addressing several important economic questions requires knowledge of the structural function of the dynamics. Nonetheless, we can estimate relatively easily the distribution of the heterogeneous mean and autocovariance without imposing strong assumptions. Moreover, the heterogeneous mean and autocovariance structure can provide valuable information, even if our ultimate goal is to identify the structural function.

3 Procedures

In this section, we present the statistical procedures used to estimate the distribution functions and other distributional characteristics of the heterogeneous mean and autocovariances of y_{it}. The proposed procedures are simple: we estimate the mean and autocovariances for each individual and then use their empirical distributions to estimate our parameter of interest. The following sections provide the theoretical justification for the proposed statistical procedures.

We first estimate the mean and autocovariances for each individual: μ_i and $\gamma_{k,i}$. We estimate these using the sample average and sample autocovariances:

$$\hat{\mu}_i := \bar{y}_i := \frac{1}{T}\sum_{t=1}^{T} y_{it},$$

and

$$\hat{\gamma}_{k,i} := \frac{1}{T-k}\sum_{t=k+1}^{T} (y_{it} - \bar{y}_i)(y_{i,t-k} - \bar{y}_i).$$

We then compute the empirical distributions of $\{\hat{\mu}_i\}_{i=1}^N$ and $\{\hat{\gamma}_{k,i}\}_{i=1}^N$:

$$\mathbb{F}_N^{\hat{\mu}}(a) := \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(\hat{\mu}_i \leq a), \tag{1}$$

and

$$\mathbb{F}_N^{\hat{\gamma}_k}(a) := \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(\hat{\gamma}_{k,i} \leq a), \tag{2}$$

where $\mathbf{1}(\cdot)$ is the indicator function and $a \in \mathbb{R}$. These empirical distributions are interesting in their own right because they are estimators of the cross-sectional distribution functions of μ_i and $\gamma_{k,i}$, respectively. Let F_0^{μ} and $F_0^{\gamma_k}$ denote the distribution functions of μ_i and $\gamma_{k,i}$, respectively, so that $F_0^{\mu}(a) := \Pr(\mu_i \leq a)$ and $F_0^{\gamma_k}(a) := \Pr(\gamma_{k,i} \leq a)$. In Section 4, we show the consistency of $\mathbb{F}_N^{\hat{\mu}}$ and $\mathbb{F}_N^{\hat{\gamma}_k}$ for F_0^{μ} and $F_0^{\gamma_k}$, respectively, and derive the asymptotic distributions of $\mathbb{F}_N^{\hat{\mu}}$ and $\mathbb{F}_N^{\hat{\gamma}_k}$ under the condition $N/T \to 0$.

Remark 1. The condition $N/T \to 0$ implies that the length of the time series T is large relative to the number of cross-sectional observations N. Accordingly, the analysis based on the distribution function would be more suitable for macroeconomic data than for microeconomic data. Macroeconomic panel data, such as multi-country panels or state-level panels, may include a sufficiently long period compared with the cross-sectional sample size.

We can estimate other distributional quantities based on the empirical distributions of $\hat{\gamma}_{k,i}$ or $\hat{\mu}_i$. For example, consider the estimation of quantiles of $\gamma_{k,i}$. Let q_τ be the τ-th quantile of $\gamma_{k,i}$: $q_\tau := \inf\{a : F_0^{\gamma_k}(a) \geq \tau\}$. This is estimated by the τ-th quantile of $\hat{\gamma}_{k,i}$ so that $\hat{q}_\tau := \inf\{a : \mathbb{F}_N^{\hat{\gamma}_k}(a) \geq \tau\}$. Using the functional delta method, we derive the asymptotic distribution of the quantile estimator when $N, T \to \infty$ with $N/T \to 0$.

We can also test parametric specifications of the distribution of the heterogeneous mean or autocovariances based on the empirical distribution. Moreover, we can examine the difference of the heterogeneous dynamic structures across distinct groups based on the empirical distributions. The tests are based on Kolmogorov–Smirnov statistics based on the empirical distributions. We develop these tests in Section 6.

When the parameter of interest is the expectation of a *smooth* function of μ_i or $\gamma_{k,i}$, the condition on the relative magnitudes of N and T can be relaxed. Suppose that we are interested in $G_\mu := E(g(\mu_i))$, where $g(\cdot)$ is a known function. We estimate G_μ by

$$\hat{G}_\mu := \frac{1}{N} \sum_{i=1}^{N} g(\hat{\mu}_i). \tag{3}$$

When our parameter of interest is $G_{\gamma_k} := E(g(\gamma_{k,i}))$, it is estimated by

$$\hat{G}_{\gamma_k} := \frac{1}{N} \sum_{i=1}^{N} g(\hat{\gamma}_{k,i}). \tag{4}$$

Suppose that $g(\cdot)$ is twice continuously differentiable with a bounded second derivative. For example, the mean of $\gamma_{k,i}$ satisfies this condition because, for the mean, g is the identity function. The theoretical results in Section 5 show that this estimator is consistent as $N, T \to \infty$ and that $\sqrt{N}(\hat{G}_a - G_a)$ for $a = \mu$ or γ_k is asymptotically normal with mean zero when $N/T^2 \to 0$.

Remark 2. This result is important because the condition $N/T^2 \to 0$ may be justified, even in the case of microeconomic data, as long as T is moderately large. By contrast, condition $N/T \to 0$ is quite strong in the analysis of microeconomic data because the number of cross-sectional units N is typically larger than the length of the time series T.

The estimation of the variance can also be justified under the weaker condition $N/T^2 \to 0$. Suppose that the parameter of interest is $var(\gamma_{k,i}) = E(\gamma_{k,i}^2) - (E(\gamma_{k,i}))^2$. Thus, it is estimated by

$$\frac{1}{N}\sum_{i=1}^{N}\hat{\gamma}_{k,i}^2 - \left(\frac{1}{N}\sum_{i=1}^{N}\hat{\gamma}_{k,i}\right)^2.$$

Neither of the estimators of $E(\gamma_{k,i}^2)$ and $E(\gamma_{k,i})$ suffer asymptotic bias when $N/T^2 \to 0$. Because the variance is a continuous function of these two moments, it can also be estimated without asymptotic bias when $N/T^2 \to 0$.

We can also estimate the expected value of a smooth function of a vector of the mean and autocovariances. Suppose that we would like to estimate $H := E(h(\theta_i))$, where $h : \mathbb{R}^l \mapsto \mathbb{R}$ is some known smooth function and θ_i is an l-dimensional vector of μ_i and/or $\gamma_{k,i}$s. Let $\hat{\theta}_i$ be the vector of estimators corresponding to the elements of θ_i. This parameter is estimated by

$$\hat{H} := \frac{1}{N}\sum_{i=1}^{N} h(\hat{\theta}_i). \tag{5}$$

For example, if we are interested in estimating $H = E(\mu_i\gamma_{0,i})$, it can be estimated by

$$\hat{H} = \frac{1}{N}\sum_{i=1}^{N}\hat{\mu}_i\hat{\gamma}_{0,i}.$$

The covariance between μ_i and $\gamma_{0,i}$ is thus estimated by

$$\frac{1}{N}\sum_{i=1}^{N}\hat{\mu}_i\hat{\gamma}_{0,i} - \left(\frac{1}{N}\sum_{i=1}^{N}\hat{\mu}_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}\hat{\gamma}_{0,i}\right).$$

The half-panel jackknife (HPJ) proposed by Dhaene and Jochmans (2014) can further reduce the bias in $\hat{G}_\mu$ or $\hat{G}_{\gamma_k}$. The estimator exhibits the bias of order $O(1/T)$ and the HPJ bias correction can delete the bias of this order. It thus allows us to relax the condition on the ratio of N and T. The bias correction is easy to implement. Suppose that T is even.[2] We divide the panel data into two subpanels: $\{\{y_{it}\}_{t=1}^{T/2}\}_{i=1}^N$ and $\{\{y_{it}\}_{t=T/2+1}^{T}\}_{i=1}^N$. The first subpanel, $\{\{y_{it}\}_{t=1}^{T/2}\}_{i=1}^N$, consists of observations from the first half of the overall time period, and the second subpanel, $\{\{y_{it}\}_{t=T/2+1}^{T}\}_{i=1}^N$, consists of those from the second half. Let $G = G_\mu$ or G_{γ_k} and $\hat{G}$ be the estimator of G. Let $\hat{G}^{(1)}$ and $\hat{G}^{(2)}$ be the estimators of G computed using $\{\{y_{it}\}_{t=1}^{T/2}\}_{i=1}^N$ and $\{\{y_{it}\}_{t=T/2+1}^{T}\}_{i=1}^N$, respectively. Let $\bar{G} := (\hat{G}^{(1)} + \hat{G}^{(2)})/2$. The HPJ estimator of G is:

$$\hat{G}^H := \hat{G} - (\bar{G} - \hat{G}) = 2\hat{G} - \bar{G}. \tag{6}$$

[2] If T is odd, we define $\bar{G} = (\hat{G}^{(1,1)} + \hat{G}^{(2,1)} + \hat{G}^{(1,2)} + \hat{G}^{(2,2)})/4$ as in Dhaene and Jochmans (2014, p. 9), where $\hat{G}^{(1,1)}$, $\hat{G}^{(2,1)}$, $\hat{G}^{(1,2)}$, and $\hat{G}^{(2,2)}$ are the estimators of G computed using $\{\{y_{it}\}_{t=1}^{\lceil T/2\rceil}\}_{i=1}^N$, $\{\{y_{it}\}_{t=\lceil T/2\rceil+1}^{T}\}_{i=1}^N$, $\{\{y_{it}\}_{t=1}^{\lfloor T/2\rfloor}\}_{i=1}^N$, and $\{\{y_{it}\}_{t=\lfloor T/2\rfloor+1}^{T}\}_{i=1}^N$, respectively. Here, $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$ are the ceiling and floor functions, respectively. We note that the asymptotic properties of the half-panel jackknife estimator for odd T are the same as those for even T. We thus focus on even T in this paper without loss of generality.

The HPJ estimates the bias in $\hat{G}$ by $\bar{G} - \hat{G}$, and $\hat{G}^H$ corrects the bias in $\hat{G}$ by subtracting the HPJ bias estimate. The bias-corrected estimator $\hat{G}^H$ does not exhibit the bias of order $O(1/T)$ and is asymptotically unbiased even when N/T^2 does not converge to zero. The jackknife bias correction may also be applied to alleviate the bias $\hat{H}$.

When correcting the bias of the variance or covariance estimator, we recommend that the jackknife bias correction is applied for estimation of each expected value, not the variance or covariance estimator itself. For example, to correct the bias for the estimator of $cov(\mu_i, \gamma_{0,i})$, our recommendation is to correct the biases in the estimators of $E(\mu_i \gamma_{0,i})$ and $E(\gamma_{0,i})$ (note that $E(\mu_i)$ can be estimated without bias) and then combine the bias-corrected estimators.

For statistical inferences on parameter G_μ, G_{γ_k}, or H, we suggest the cross-sectional bootstrap. The cross-sectional bootstrap is used to approximate the distribution of the HPJ estimator (or $\hat{G}_\mu$, $\hat{G}_{\gamma_k}$, or $\hat{H}$ when T is sufficiently large). In the cross-sectional bootstrap, we regard the time series from an individual as the unit of observation and approximate the distribution of statistics by that under the empirical distribution of z_i, where $z_i := (y_{i1}, \dots, y_{iT})$. The algorithm is as follows:

1. Randomly draw $z_1^*, \dots, z_N^*$ from $\{z_1, \dots, z_N\}$ with replacement.
2. Compute the statistics of interest, say S, using $z_1^*, \dots, z_N^*$.
3. Repeat 1 and 2 B times. Let $S^*(b)$ be the statistics computed with the b-th bootstrap sample.
4. Compute the distributional quantities of interest for S based on the empirical distribution of $S^*(b)$.

For example, suppose that we are interested in constructing a 95% confidence interval for parameter $G_\mu = E(g(\mu_i))$. We obtain the bootstrap approximation of the distribution of $S = \hat{G}_\mu^H - G_\mu$. Let $\hat{G}_\mu^{H*}(b)$ be the HPJ estimate of G_μ obtained with the b-th bootstrap sample. We then compute the 2.5% and 97.5% quantiles, denoted as $q_{0.025}^*$ and $q_{0.975}^*$, of the empirical distribution of $S^*(b) = \hat{G}_\mu^{H*}(b) - \hat{G}_\mu^H$. The cross-sectional bootstrap 95% confidence interval for G_μ is

$$[\hat{G}_\mu^H - q_{0.975}^*, \hat{G}_\mu^H - q_{0.025}^*].$$

4 Asymptotic analysis for the distribution estimators

This section presents the asymptotic properties of the distribution estimators (1) and (2). We first show the uniform consistency of the empirical distribution of the estimated mean or autocovariance. We then derive the functional central limit theorem for the empirical distributions. We also show that the functional delta method can be applied in this case. All the asymptotic analyses presented in the following sections are under double asymptotics ($N, T \to \infty$). The asymptotic analyses are based on empirical process techniques (see, e.g., van der Vaart and Wellner, 1996).

The following representation is useful for our theoretical analysis. Let $w_{it} := y_{it} - E(y_{it}|i) = y_{it} - \mu_i$. By construction, y_{it} is decomposed as

$$y_{it} = \mu_i + w_{it}.$$

The random variable w_{it} is the unobservable idiosyncratic component that varies over both i and t. Note that, by definition, $E(w_{it}|i) = 0$ for any i and t. Note also that $\gamma_{k,i} = E(w_{it}w_{i,t-k}|i)$.

4.1 Uniform consistency

In this section, we show that the empirical distributions of $\hat{\mu}_i$ and $\hat{\gamma}_{k,i}$ are uniformly consistent for the true distributions of μ_i and $\gamma_{k,i}$,

Because we use empirical process techniques, it is convenient to rewrite the empirical distributions as empirical processes indexed by a class of indicator functions. Let $\mathbb{P}_N^{\hat{\mu}}$ be the empirical measure of $\hat{\mu}_i$:

$$\mathbb{P}_N^{\hat{\mu}} := \frac{1}{N}\sum_{i=1}^{N} \delta_{\hat{\mu}_i},$$

where $\delta_{\hat{\mu}_i}$ is the probability distribution degenerated at $\hat{\mu}_i$. Let $\mathcal{F}$ be the following class of indicator functions:

$$\mathcal{F} := \{\mathbf{1}_{(-\infty,a]} : a \in \mathbb{R}\},$$

where $\mathbf{1}_{(-\infty,a]}(x) := \mathbf{1}(x \leq a)$. We define the probability measure of μ_i as P_0^{μ}. In this notation, the empirical distribution function, $\mathbb{F}_N^{\hat{\mu}}$, is an empirical process indexed by $\mathcal{F}$. For example, $\mathbb{P}_N^{\hat{\mu}} f = \mathbb{F}_N^{\hat{\mu}}(a)$ for $f = \mathbf{1}_{(-\infty,a]}$. Similarly, for $f = \mathbf{1}_{(-\infty,a]}$, $P_0^{\mu} f = F_0^{\mu}(a) = \Pr(\mu_i \leq a)$. The empirical measure of $\hat{\gamma}_{k,i}$, $\mathbb{P}_N^{\hat{\gamma}_k}$ and the probability measure of $\gamma_{k,i}$, $P_0^{\gamma_k}$ are analogously defined.

Our objective in this section is to show that the class $\mathcal{F}$ is P_0-Glivenko–Cantelli for $P_0 = P_0^{\mu}$ or $P_0^{\gamma_k}$, in the sense that

$$\sup_{f \in \mathcal{F}} |\mathbb{P}_N f - P_0 f| \xrightarrow{as} 0, \tag{7}$$

where $\mathbb{P}_N$ is the empirical distribution corresponding to P_0, and $\xrightarrow{as}$ is the almost sure convergence. This is equivalent to the uniform consistency of the empirical distribution function. Note that (7) cannot be directly shown by the usual Glivenko–Cantelli theorem, e.g., Theorem 19.1 in van der Vaart (1998). This is because the true distributions of $\hat{\mu}_i$ and $\hat{\gamma}_{k,i}$ change as T increases. Nonetheless, our proof of (7) follows similar steps to those of the usual Glivenko–Cantelli theorem.

We use the following assumption throughout the paper, which summarizes the conditions imposed in Section 2.

Assumption 1. *The sample space of* α_i *is some Polish space and* y_{it} *is a scalar real random variable.* $\{(\{y_{it}\}_{t=1}^T, \alpha_i)\}_{i=1}^N$ *is i.i.d. across* i*.* $\{y_i\}_{t=1}^T$ *is strictly stationary given* α_i*.*

The following conditions are used to show the consistency of $\mathbb{P}_N^{\hat{\mu}}$.

Assumption 2. $\sum_{k=-\infty}^{\infty} E|\gamma_{k,i}| < \infty$.

Assumption 3. *The random vector* $(\mu_i, \bar{y}_i)$ *is continuously distributed and its joint density is bounded.*

Assumption 2 indicates that the dynamics of w_{it} is a short memory process. We do not here consider the case in which the process has a long memory property. Assumption 3 states that μ_i and $\bar{y}_i$ are continuous random variables. This assumption is restrictive in the sense that it does not allow the case in which the distribution of μ_i is discrete or there is no heterogeneity in the mean (i.e., μ_i is homogeneous so that $\mu_i = \mu$ for some constant μ for any i). It should not be very difficult to relax this assumption, but then we would need to employ a different proof technique.

For the consistency of $\mathbb{P}_N^{\hat{\gamma}_k}$, we need a different set of assumptions.

Assumption 4. *For each* i*,* $\{y_{it}\}_{t=1}^{\infty}$ *is strictly stationary and* α*-mixing given* α_i *with mixing coefficients* $\{\alpha(m|i)\}_{m=0}^{\infty}$*. There exists a sequence* $\{\alpha(m)\}_{m=0}^{\infty}$ *such that for any* i *and* m*,* $\alpha(m|i) \le \alpha(m)$ *and* $\sum_{m=0}^{\infty}(m+1)^3\alpha(m)^{\delta/(4+\delta)} < \infty$ *for some* $\delta > 0$*.*

Assumption 5. $E|w_{it}|^{4+\delta} < \infty$ *for some* $\delta > 0$*.*

Assumption 6. *The random vector* $(\gamma_{k,i}, \hat{\gamma}_{k,i})$ *is continuously distributed and its joint density is bounded.*

Assumption 4 is a mixing condition and restricts the degree of persistency of y_{it}. Assumption 5 requires that w_{it} exhibits some moment higher than 4th order. Assumptions 4 and 5 are satisfied, for example, when y_{it} follows a heterogeneous stationary panel AR(1) model with Gaussian innovations. Assumption 6 is similar to Assumption 3 and is restrictive in the sense that $\gamma_{k,i}$ must be continuously distributed.

The following theorem establishes the uniform consistency of our distribution estimators.

Theorem 1. *Suppose that Assumptions 1, 2, and 3 hold. When* $N, T \to \infty$*, the class* $\mathcal{F}$ *is* P_0^{μ}*-Glivenko–Cantelli in the sense that*

$$\sup_{f\in\mathcal{F}} \left|\mathbb{P}_N^{\hat{\mu}} f - P_0^{\mu} f\right| \xrightarrow{as} 0.$$

Suppose that Assumptions 1, 4, 5, and 6 hold. When $N, T \to \infty$*, the class* $\mathcal{F}$ *is* $P_0^{\gamma_k}$*-Glivenko–Cantelli in the sense that*

$$\sup_{f\in\mathcal{F}} \left|\mathbb{P}_N^{\hat{\gamma}_k} f - P_0^{\gamma_k} f\right| \xrightarrow{as} 0.$$

4.2 Functional central limit theorem

We present the functional central limit theorems for the empirical distributions of $\hat{\mu}_i$ and $\hat{\gamma}_{k,i}$. Our objective is to derive the asymptotic properties of

$$\sqrt{N}(\mathbb{P}_N^{\hat{\mu}} f - P_0^{\mu} f), \quad \text{and} \quad \sqrt{N}(\mathbb{P}_N^{\hat{\gamma}_k} f - P_0^{\gamma_k} f),$$

where $f \in \mathcal{F}$. This is equivalent to investigating the limiting distributions of $\sqrt{N}(\mathbb{F}_N^{\hat{\mu}}(a) - F_0^{\mu}(a))$ and $\sqrt{N}(\mathbb{F}_N^{\hat{\gamma}_k}(a) - F_0^{\gamma_k}(a))$ for every $a \in \mathbb{R}$. This result is interesting in its own right because it provides the asymptotic distributions of the empirical distributions. It is also important because the asymptotic distribution of other quantities of interest can be obtained by the functional delta method based on this result.

The functional central limit theorems for $\mathbb{P}_N^{\hat{\mu}}$ and $\mathbb{P}_N^{\hat{\gamma}_k}$ hold under the same set of assumptions for the uniform consistency. However, it requires a stronger condition on the relative magnitude of N and T. Let $\ell^{\infty}(\mathcal{F})$ be the collection of all bounded real functions on $\mathcal{F}$.

Theorem 2. *Suppose that Assumptions 1, 2, and 3 hold. When $N, T \to \infty$ with $N/T \to 0$, we have*

$$\sqrt{N}(\mathbb{P}_N^{\hat{\mu}} - P_0^{\mu}) \rightsquigarrow \mathbb{G}_{P_0^{\mu}} \quad \text{in} \quad \ell^{\infty}(\mathcal{F}),$$

where $\mathbb{G}_{P_0^{\mu}}$ is a Gaussian process with zero mean and covariance function

$$E(\mathbb{G}_{P_0^{\mu}}(f_i)\mathbb{G}_{P_0^{\mu}}(f_j)) = F_0^{\mu}(a_i \wedge a_j) - F_0^{\mu}(a_i)F_0^{\mu}(a_j),$$

for $f_i = 1_{(-\infty, a_i]}$ and $f_j = 1_{(-\infty, a_j]}$.

Suppose that Assumptions 1, 4, 5, and 6 hold. When $N, T \to \infty$ with $N/T \to 0$, we have

$$\sqrt{N}(\mathbb{P}_N^{\hat{\gamma}_k} - P_0^{\gamma_k}) \rightsquigarrow \mathbb{G}_{P_0^{\gamma_k}} \quad \text{in} \quad \ell^{\infty}(\mathcal{F}),$$

where $\mathbb{G}_{P_0}^{\gamma_k}$ is a Gaussian process with zero mean and covariance function

$$E(\mathbb{G}_{P_0^{\gamma_k}}(f_i)\mathbb{G}_{P_0^{\gamma_k}}(f_j)) = F_0^{\gamma_k}(a_i \wedge a_j) - F_0^{\gamma_k}(a_i)F_0^{\gamma_k}(a_j),$$

for $f_i = \mathbf{1}_{(-\infty, a_i]}$ and $f_j = \mathbf{1}_{(-\infty, a_j]}$.

This theorem shows that the asymptotic laws of the empirical processes are Gaussian. This limiting process is then the same as that for the empirical process constructed using the true μ_i or $\gamma_{k,i}$. However, this result requires that $N/T \to 0$. Put differently, the condition $N/T \to 0$ allows us to ignore the estimation error in $\hat{\mu}_i$ or $\hat{\gamma}_{k,i}$ asymptotically.

Here, we provide a brief summary of the proof and explain the reason why the condition $N/T \to 0$ is required. The same discussion can be applied to both $\mathbb{P}_N^{\hat{\mu}}$ and $\mathbb{P}_N^{\hat{\gamma}_k}$. In the following discussion, we let $\mathbb{P}_N$ be either $\mathbb{P}_N^{\hat{\mu}}$ or $\mathbb{P}_N^{\hat{\gamma}_k}$ and P_0 be the corresponding true distribution.

The key to understanding the mechanism behind the requirement that $N/T \to 0$ is to recognize that $E(\mathbb{P}_N f) \neq P_0 f$. That is, $\mathbb{P}_N f$ is not an unbiased estimator for $P_0 f$. For this reason, the existing results for the empirical process cannot be directly applied to derive the asymptotic distribution. Let P_T be the (true) probability measure of $\hat{\mu}_i$ or $\hat{\gamma}_{k,i}$ so that $P_T f = \Pr(\hat{\mu}_i \leq a)$ or $P_T f = \Pr(\hat{\gamma}_{k,i} \leq a)$, which depends on T. Note that $E(\mathbb{P}_N f) = P_T f$. Let

$$\mathbb{G}_{N,P_T} := \sqrt{N}(\mathbb{P}_N - P_T).$$

We decompose the process in the following way:

$$\sqrt{N}(\mathbb{P}_N f - P_0 f) = \mathbb{G}_{N,P_T} f \tag{8}$$

$$+ \sqrt{N}(P_T f - P_0 f). \tag{9}$$

We analyze the asymptotic behavior of the terms in (8) and (9) separately.

For $\mathbb{G}_{N,P_T}$ in (8), we can directly apply the uniform central limit theorem for the empirical process based on triangular arrays (van der Vaart and Wellner, 1996, Lemma 2.8.7). Note that $E(\mathbb{G}_{N,P_T} f) = 0$. Using Lemma 2.8.7 in van der Vaart and Wellner (1996), we show that

$$\mathbb{G}_{N,P_T} \rightsquigarrow \mathbb{G}_{P_0} \quad \text{in} \quad \ell^\infty(\mathcal{F}).$$

This part of the proof is standard.

The condition $N/T \to 0$ is needed to eliminate the effect of the bias term in the empirical process: $\sqrt{N}(P_T - P_0)$ in (9). In the proof of the theorem, we show that

$$\sup_{f \in \mathcal{F}} \left| \sqrt{N}(P_T f - P_0 f) \right| = O\left(\frac{\sqrt{N}}{\sqrt{T}} \right).$$

This term converges to zero when T is of a higher order than N. This result arises from the fact that the rate of convergence of $\hat{\mu}_i$ to μ_i or $\hat{\gamma}_{k,i}$ to $\gamma_{k,i}$ is $1/\sqrt{T}$. Hence, the difference between the distributions of $\hat{\mu}_i$ and μ_i or $\hat{\gamma}_{k,i}$ and $\gamma_{k,i}$ is of order $1/\sqrt{T}$. This is the reason why the difference between P_T and P_0 is also of order $O(1/\sqrt{T})$.

4.3 Functional delta method

The asymptotic distribution of an estimator that is a function of the empirical distribution may be derived using the functional delta method. Suppose that we are interested in the asymptotics of $\phi(\mathbb{P}_N)$ for $\phi : D(\mathcal{F}) \mapsto \mathbb{R}$, where $\mathbb{P}_N = \mathbb{P}_N^{\hat{\mu}}$ or $\mathbb{P}_N^{\hat{\gamma}_k}$ and $D(\mathcal{F})$ is the collection of all cadlag real functions on $\mathcal{F}$. For example, the quantile function of $\gamma_{k,i}$, $\phi(P_0^{\gamma_k}) = q_\tau = \inf\{t : F_0^{\gamma_k}(t) \geq \tau\}$ for $\tau \in (0,1)$, may be estimated by:

$$\phi(\mathbb{P}_N^{\hat{\gamma}_k}) = \hat{q}_\tau = (\mathbb{F}_N^{\hat{\gamma}_k})^{-1}(\tau) = \inf\{t : \mathbb{F}_N^{\hat{\gamma}_k}(t) \geq \tau\}.$$

The derivation of the asymptotic distribution of $\phi(\mathbb{P}_N)$ is an application of the functional delta method (see, e.g., van der Vaart and Wellner, 1996, Theorem 3.9.4) and Theorem 2. We summarize this result in the following corollary.

Corollary 1. *Let $\mathbb{E}$ be a normed linear space. Let $\phi : D(\mathcal{F}) \subset \ell^\infty(\mathcal{F}) \mapsto \mathbb{E}$ be Hadamard differentiable at P_0^μ. Denote its derivative by $\phi'_{P_0^\mu}$. Under Assumptions 1, 2, and 3, when $N, T \to \infty$ with $N/T \to 0$, we have*

$$\sqrt{N}(\phi(\mathbb{P}_N^{\hat{\mu}}) - \phi(P_0^\mu)) \rightsquigarrow \phi'_{P_0^\mu}(\mathbb{G}_{P_0^\mu}).$$

Similarly, suppose that ϕ has the Hadamard derivative, $\phi'_{P_0^{\gamma_k}}$, at $P_0^{\gamma_k}$. Under Assumptions 1, 4, 5, and 6, when $N, T \to \infty$ and $N/T \to 0$, we have

$$\sqrt{N}(\phi(\mathbb{P}_N^{\hat{\gamma}_k}) - \phi(P_0^{\gamma_k})) \rightsquigarrow \phi'_{P_0^{\gamma_k}}(\mathbb{G}_{P_0^{\gamma_k}}).$$

Proof. This is immediate by the functional delta method and Theorem 2. □

This result can be used, for example, to derive the asymptotic distribution of $\hat{q}_\tau$. The form of $\phi'_{P_0^{\gamma_k}}$ for $\hat{q}_\tau$ is available in Example 20.5 in van der Vaart (1998) and indicates that as $N, T \to \infty$ with $N/T \to 0$,

$$\sqrt{N}(\hat{q}_\tau - q_\tau) \rightsquigarrow N\left(0, \frac{\tau(1-\tau)}{(f_0^{\gamma_k}(q_\tau))^2}\right),$$

where $f_0^{\gamma_k}$ is the density function of $\gamma_{k,i}$.

5 Expected value of a smooth function of the heterogeneous mean and/or autocovariances

In this section, we consider the estimation of the expected value of a smooth function of the heterogeneous mean and/or autocovariances. A close inspection of the asymptotic expansion of the estimator reveals that a milder condition on the relative magnitude of N and T is sufficient for the asymptotically unbiased estimation in this case. Furthermore, half-panel jackknife bias correction can reduce the asymptotic bias in the estimator and further relax the condition on the ratio of N to T.

5.1 Function of the mean

We first analyze the asymptotic property of $\hat{G}_\mu = N^{-1}\sum_{i=1}^N g(\hat{\mu}_i)$ in (3). Recall that the parameter of interest is $G_\mu = E(g(\mu_i))$. We consider the case in which $g(\cdot)$ is sufficiently smooth. We derive the asymptotic distribution of $\hat{G}_\mu$ under the condition $N/T^2 \to 0$.

We make the following assumption on $g(\cdot)$.

Assumption 7. *The function $g(\cdot)$ is twice differentiable. $E(g(\mu_i)^2) < \infty$ and $E(g'(\mu_i)^4) < \infty$. $\sup_a |g''(a)| < M$ for some $M < \infty$.*

Assumption 7 states that the function $g(\cdot)$ is sufficiently smooth. This assumption is satisfied, for example, when the parameter of interest is the mean (i.e., $g(a) = a$) or when it is the p-th order moment (i.e., $g(a) = a^p$). However, this assumption is not satisfied when the distribution function is estimated (i.e., $g(a) = \mathbf{1}(a \leq c)$ for some $c \in \mathbb{R}$) or when a quantile is estimated. The existence of the first derivative is crucial for relaxing the condition $N/T \to 0$. The existence of the second derivative is useful for evaluating the order of the asymptotic bias.

The asymptotic property of $\hat{G}_\mu$ is given in the following theorem.

Theorem 3. *Suppose that Assumptions 1, 4, 5, and 7 hold. When $N, T \to \infty$, we have*

$$\hat{G}_\mu - G_\mu \xrightarrow{p} 0,$$

and

$$\sqrt{N}\left(\hat{G}_\mu - G_\mu - \frac{1}{2}E\left(\bar{w}_i^2 g''(\tilde{\mu}_i)\right)\right) \rightsquigarrow N(0, var(g(\mu_i))),$$

where $\tilde{\mu}_i$ is between μ_i and $\bar{y}_i$. In addition, when $N/T^2 \to 0$, we have

$$\sqrt{N}\left(\hat{G}_\mu - G_\mu\right) \rightsquigarrow N(0, var(g(\mu_i))).$$

This theorem states that $\hat{G}_\mu$ is consistent for G_μ and that the asymptotic distribution of $\hat{G}_\mu$ is normal and centered at zero when $N/T^2 \to 0$. Note that we use the mixing and moment conditions that have been used for $\mathbb{P}_N^{\hat{\gamma}_k}$ here. The remarkable result is that the asymptotically unbiased estimation holds under $N/T^2 \to 0$, which is a markedly weaker condition than $N/T \to 0$. This result is because of the smoothness of $g(\cdot)$ and the fact that $\hat{\mu}_i$ is unbiased for μ_i. In fact, when we are interested in $E(\mu_i)$ (i.e., when $g(a) = a$), no conditions on the relative magnitude of N and T are needed to achieve an asymptotically unbiased estimation (indeed T can be fixed for the estimation of $E(\mu_i)$). However, if $g(\cdot)$ is nonlinear, $N/T^2 \to 0$ is needed to remove the asymptotic bias.

In order to obtain a better understanding of the results in the theorem, we observe the following expansion:

$$\begin{aligned}
&\sqrt{N}\left(\hat{G}_\mu - G_\mu\right) \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^N (g(\mu_i) - E(g(\mu_i))) + \frac{1}{\sqrt{N}}\sum_{i=1}^N (\hat{\mu}_i - \mu_i) g'(\mu_i) + \frac{1}{\sqrt{N}}\sum_{i=1}^N (\hat{\mu}_i - \mu_i)^2 g''(\tilde{\mu}_i) \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^N (g(\mu_i) - E(g(\mu_i))) + \frac{1}{\sqrt{N}}\sum_{i=1}^N \bar{w}_i g'(\mu_i) + \frac{1}{\sqrt{N}}\sum_{i=1}^N (\bar{w}_i)^2 g''(\tilde{\mu}_i). \qquad (10)
\end{aligned}$$

The second term in (10) has a mean of zero and is of order $O_p(1/\sqrt{T})$. The fact that the second term has a mean of zero is the key reason that a milder condition, $N/T^2 \to 0$, is sufficient for the asymptotically unbiased estimation of G_μ. This result relies on the assumption that $g(\cdot)$ is smooth. When $g(\cdot)$ is not smooth, this expansion cannot be executed and we cannot exploit the fact that $\hat{\mu}_i$ is unbiased for μ_i. The third term corresponds to the bias caused by the nonlinearity of $g(\cdot)$. When $g(\cdot)$ is linear, this term does not appear and the parameter can be estimated without any restriction on the relative magnitude between N and T. The nonlinearity bias is of order $O_p(\sqrt{N}/T)$. The condition $N/T^2 \to 0$ is used to eliminate the effect of this bias.

We also note that the joint asymptotic distribution of the estimators of $E(g_1(\mu_i))$ and $E(g_2(\mu_i))$ for different $g_1(\cdot)$ and $g_2(\cdot)$ can be easily derived. This is because the estimator is asymptotically linear. This observation is important, for example, when we are interested in the variance of μ_i. The variance of μ_i is a function of $E(\mu_i^2)$ and $E(\mu_i)$. To derive the asymptotic distribution of the variance estimator, we need the joint asymptotic distribution of $N^{-1}\sum_{i=1}^N \hat{\mu}_i^2$ and $N^{-1}\sum_{i=1}^N \hat{\mu}_i$. The fact that the estimator is asymptotically linear enables us to derive it easily.

5.2 Function of the autocovariances

We next consider the asymptotic properties of $\hat{G}_{\gamma_k} = N^{-1}\sum_{i=1}^N g(\hat{\gamma}_{k,i})$ in (4), which is the estimator for $G_{\gamma_k} = E(g(\gamma_{k,i}))$, and obtain results similar to those for $\hat{G}_\mu$. However, $\hat{G}_{\gamma_k}$ suffers from an additional source of bias, namely incidental parameter bias.

We make the following additional assumptions to study the asymptotic properties of $\hat{G}_{\gamma_k}$.

Assumption 8. *For each i, $\{y_{it}\}_{t=1}^\infty$ is strictly stationary and α-mixing given α_i with mixing coefficients $\{\alpha(m|i)\}_{m=0}^\infty$. There exists a sequence $\{\alpha(m)\}_{m=0}^\infty$ such that for any i and m, $\alpha(m|i) \le \alpha(m)$ and $\sum_{m=0}^\infty (m+1)^3 \alpha(m)^{\delta/(8+\delta)} < \infty$ for some $\delta > 0$.*

Assumption 9. *$E|w_{it}|^{8+\delta} < \infty$ for some $\delta > 0$.*

Assumption 10. *The function $g(\cdot)$ is twice differentiable. $E(g(\gamma_{k,i})^2) < \infty$, $E((g'(\gamma_{k,i}))^4) < \infty$. $\sup_a |g''(a)| < M$ for some $M < \infty$.*

Assumption 8 is a stronger version of Assumption 4 and imposes restrictions on the persistency of w_{it}. Assumption 9 is a stronger version of Assumption 5 and states that w_{it} has some moment of higher order than 8. Assumption 10 is similar to Assumption 7 and states that function $g(\cdot)$ is sufficiently smooth.

The asymptotic property of $\hat{G}_{\gamma_k}$ is given in the following theorem.

Theorem 4. *Suppose that Assumptions 1, 8, 9, and 10 are satisfied. When $N, T \to \infty$, it holds that*

$$\hat{G}_{\gamma_k} - G_{\gamma_k} \xrightarrow{p} 0.$$

Moreover, when additionally $N/T^2 \to 0$ holds, we have

$$\sqrt{N}(\hat{G}_{\gamma_k} - G_{\gamma_k}) \rightsquigarrow N(0, var(g(\gamma_{k,i})).$$

This theorem states that $\hat{G}_{\gamma_k}$ is consistent for G_{γ_k} and that the asymptotic distribution of $\hat{G}_{\gamma_k}$ is normal and is centered at zero when $N/T^2 \to 0$. Similarly to Theorem 3, this theorem merely requires that $N/T^2 \to 0$ because of the smoothness of $g(\cdot)$ and the fact that the leading term in the expansion of $\hat{\gamma}_{k,i}$ has a mean of zero. However, contrary to Theorem 3, even if our parameter of interest is $E(\gamma_{k,i})$ so that $g(\cdot)$ is linear, we cannot relax the condition $N/T^2 \to 0$. This is because $\hat{\gamma}_{k,i}$ is not unbiased for $\gamma_{k,i}$.

The results of the theorem can be better understood by examining the asymptotic expansion of $\hat{\gamma}_{k,i}$ and $\hat{G}_{\gamma_k}$. The autocovariance estimator, $\hat{\gamma}_{k,i}$, is expanded as follows:

$$
\begin{aligned}
\hat{\gamma}_{k,i} &= \frac{1}{T-k}\sum_{t=k+1}^{T}(y_{it}-\bar{y}_i)(y_{i,t-k}-\bar{y}_i) \\
&= \gamma_{k,i} + \frac{1}{T-k}\sum_{t=k+1}^{T}(w_{it}w_{i,t-k}-\gamma_{k,i}) - (\bar{w}_i)^2 + O_p\left(\frac{1}{T^2}\right).
\end{aligned}
$$

It is important to observe that the second term in the second line has a mean of zero although it is of order $T^{-1/2}$. The third term, $(\bar{w}_i)^2$, is the estimation error in $\bar{y}_i$ $(=\hat{\mu}_i)$. This term is of order $O(1/T)$ and is the cause of the incidental parameter bias (Neyman and Scott, 1948; Nickell, 1981). By the Taylor expansion of $\hat{G}_{\gamma_k}$, we have

$$
\begin{aligned}
\sqrt{N}(\hat{G}_{\gamma_k} - G_{\gamma_k}) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(g(\gamma_{k,i}) - E(g(\gamma_{k,i}))) \\
&\quad + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\hat{\gamma}_{k,i}-\gamma_{k,i})g'(\gamma_{k,i}) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\hat{\gamma}_{k,i}-\gamma_{k,i})^2 g''(\tilde{\gamma}_{k,i}) \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(g(\gamma_{k,i}) - E(g(\gamma_{k,i}))) \qquad (11) \\
&\quad + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T} w_{it}w_{i,t-k} - \gamma_{k,i}\right) g'(\gamma_{k,i}) \qquad (12) \\
&\quad - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\bar{w}_i)^2 g'(\gamma_{k,i}) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\hat{\gamma}_{k,i}-\gamma_{k,i})^2 g''(\tilde{\gamma}_{k,i}) + O_p\left(\frac{\sqrt{N}}{T^2}\right), \qquad (13)
\end{aligned}
$$

where the second equality is obtained by plugging the expansion for $\hat{\gamma}_{k,i}$, and $\tilde{\gamma}_{k,i}$ is between $\hat{\gamma}_{k,i}$ and $\gamma_{k,i}$.

Contrary to $\hat{G}_{\mu}$, $\hat{G}_{\gamma_k}$ exhibits the incidental parameter bias that corresponds to the first term in (13). This bias is caused by the estimation of μ_i by $\bar{y}_i$ and is of order $O_p(\sqrt{N}/T)$, but does not appear in the expansion of $\hat{G}_{\mu}$. Because of this term, the condition $N/T^2 \to 0$ is needed even when $g(\cdot)$ is linear. The other terms are similar to those in the expansion of $\hat{G}_{\mu}$. The term on the right-hand side of (11) yields the asymptotic normality of $\hat{G}$. The term in (12) has a mean of zero and is of order $O_p(1/\sqrt{T})$. That this term has a mean of zero is crucial for the condition $N/T^2 \to 0$ to be sufficient for the asymptotic unbiasedness of $\hat{G}_{\gamma_k}$. The second term in (13) is the nonlinearity bias term that also appears in $\hat{G}_{\mu}$. This is also of order $O_p(\sqrt{N}/T)$.

As in the case of $\hat{G}_{\mu}$, as the estimator is asymptotically linear, it is easy to derive the joint asymptotic distribution of the estimators of, say, $E(g_1(\gamma_{k,i}))$ and $E(g_2(\gamma_{k,i}))$ for different $g_1(\cdot)$ and $g_2(\cdot)$. Similarly, it is also easy to derive the joint asymptotic distribution of $\hat{G}_{\gamma_k}$ and $\hat{G}_{\mu}$.

5.3 Function of a vector of the mean and autocovariances

We now discuss the asymptotic properties of $\hat{H} = N^{-1}\sum_{i=1}^{N} h(\hat{\theta}_i)$ in (5), which is the estimator of $H = E(h(\theta_i))$. Recall that $h : \mathbb{R}^l \mapsto \mathbb{R}$ is some known smooth function, θ_i is an l-dimensional random vector of μ_i and/or $\gamma_{k,i}$s, and $\hat{\theta}_i$ is the estimator of θ_i with $\hat{\mu}_i$ for μ_i and $\hat{\gamma}_{k,i}$ for $\gamma_{k,i}$. The asymptotic results and the mechanism behind them are similar to those of $\hat{G}_{\gamma_k}$, and the asymptotically unbiased estimation is achieved when $N/T^2 \to 0$.

We make the following assumptions to develop the asymptotic properties of $\hat{H}$. They impose conditions on the smoothness of $h(\cdot)$ and the existence of the moments, and are similar to Assumptions 7 and 10.

Assumption 11. *The function $h(\cdot)$ is twice differentiable and $\sup_a \left| \frac{\partial^2}{\partial z_{j_1} \partial z_{j_2}} h(z)|_{z=a} \right| < M$ for some $M < \infty$ and any $j_1, j_2 = 1, \ldots, l$.*

Assumption 12. *$E(h^2(\theta_i)) < \infty$ and $E((\frac{\partial}{\partial z_j} h(z)|_{z=\theta_i})^4) < \infty$ for any $j = 1, \ldots, l$.*

The following theorem demonstrates the asymptotic properties of $\hat{H}$.

Theorem 5. *Suppose that Assumptions 1, 8, 9, 11, and 12 hold. When $N, T \to \infty$, it holds that*

$$\hat{H} - H \xrightarrow{p} 0.$$

Moreover, when $N/T^2 \to 0$ holds additionally, we have

$$\sqrt{N}(\hat{H} - H) \rightsquigarrow N(0, var(h(\theta_i))).$$

The theorem states that $\hat{H}$ is consistent when both N and T tend to infinity, and is asymptotically normal with mean zero when $N/T^2 \to 0$. The condition $N/T^2 \to 0$ is needed because of the incidental parameter bias in $\hat{\gamma}_{k,i}$ and the nonlinearity bias. The proof is very similar to that of Theorem 4.

We remark that it is easy to derive the joint asymptotic distribution for the case in which $h(\cdot)$ is multivalued because the estimator is asymptotically linear. Similarly, deriving the joint asymptotic distribution of $\hat{H}$, $\hat{G}_{\gamma_k}$, and $\hat{G}_\mu$ is also possible. For example, when we are interested in the asymptotic distribution of the estimator of $cov(\mu_i, \gamma_{0,i})$, we need to derive the joint asymptotic distribution of $\hat{H} = N^{-1} \sum_{i=1}^N \hat{\mu}_i \hat{\gamma}_{0,i}$, $\hat{G}_\mu = N^{-1} \sum_{i=1}^N \hat{\mu}_i$, and $\hat{G}_{\gamma_k} = N^{-1} \sum_{i=1}^N \hat{\gamma}_{0,i}$. This is possible because of the asymptotic linearity.

5.4 Jackknife bias correction

Here, we provide a theoretical justification of the half-panel jackknife (HPJ) bias-corrected estimator (6), which is based on the bias-correction method proposed by Dhaene and Jochmans (2014). It results that the bias of order $O(1/T)$ in $\hat{G}_\mu$ and $\hat{G}_{\gamma_k}$ is eliminated by the HPJ procedure. Recall the definitions: $G = G_\mu$ or G_{γ_k} and $\hat{G}$ is the corresponding estimator of G; $\hat{G}^{(1)}$ and $\hat{G}^{(2)}$ are the estimators of G computed using $\{\{y_{it}\}_{t=1}^{T/2}\}_{i=1}^N$ and $\{\{y_{it}\}_{t=T/2+1}^{T}\}_{i=1}^N$, respectively, with even T; the HPJ estimator of G is $\hat{G}^H = 2\hat{G} - \bar{G}$, where $\bar{G} = (\hat{G}^{(1)} + \hat{G}^{(2)})/2$.

We make the following additional assumptions to study the asymptotic property of the HPJ estimator of G_μ.

Assumption 13. *The function $g(\cdot)$ is thrice differentiable. $E(g(\mu_i)^2) < \infty$, $E((g'(\mu_i))^4) < \infty$, $E((g''(\mu_i))^4) < \infty$. $\sup_a |g'''(a)| < M$ for some $M < \infty$.*

For the HPJ estimator of G_{γ_k}, we use the following assumptions.

Assumption 14. *$E|w_{it}|^{16+\delta} < \infty$ for some $\delta > 0$.*

Assumption 15. *The function $g(\cdot)$ is thrice differentiable. $E(g(\gamma_{k,i})^2) < \infty$, $E((g'(\gamma_{k,i}))^4) < \infty$, $E((g''(\gamma_{k,i}))^4) < \infty$. $\sup_a |g'''(a)| < M$ for some $M < \infty$. $\lim_{T\to\infty} T^{-1}(\sum_{t=k+1}^{T}(w_{it}w_{i,t-k} - \gamma_{k,i}))^2 g''(\gamma_{k,i})$ exists almost surely.*

Assumption 14 provides a condition on the existence of moments of w_{it}. It is stronger than Assumption 9. A stronger moment condition is called for because the asymptotic expansion needs to be executed for a higher order to derive the asymptotic properties of $\hat{G}^H$. Assumptions 13 and 15 require that $g(\cdot)$ is thrice differentiable, contrary to Assumptions 7 and 10. This condition is also needed to conduct a higher-order asymptotic expansion.

The following theorem shows the asymptotic normality of the HPJ estimator.

Theorem 6. *Suppose that Assumptions 1, 8, 9, and 13 are satisfied. Then, as $N, T \to \infty$ with $N/T^2 \to r$ for some $r \in [0, \infty)$, it holds that*

$$\sqrt{N}(\hat{G}^H_\mu - G_\mu) \rightsquigarrow N(0, var(g(\mu_i))).$$

Suppose that Assumptions 1, 8, 14, and 15 are satisfied. Then, as $N, T \to \infty$ with $N/T^2 \to r$ for some $r \in [0, \infty)$, it holds that

$$\sqrt{N}(\hat{G}^H_{\gamma_k} - G_{\gamma_k}) \rightsquigarrow N(0, var(g(\gamma_{k_i}))).$$

The remarkable result is that the HPJ estimator is asymptotically unbiased even when $N/T^2 \to 0$ is violated. Moreover, this bias correction does not inflate the asymptotic variance. To see how the HPJ works, we observe that

$$\hat{G} = G + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + \frac{B}{T} + O_p\left(\frac{1}{T^2}\right),$$

where B is a constant. Similarly, we have

$$\hat{G}^{(j)} = G + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + \frac{2B}{T} + O_p\left(\frac{1}{T^2}\right),$$

for $j = 1, 2$. Therefore, it holds that

$$\hat{G}^H = G + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T^2}\right).$$

Thus, the HPJ reduces the order of the bias from $O(1/T)$ to $O(1/T^2)$.

The HPJ bias correction for $\hat{H}$ can also be similarly developed and reduces the bias of order $O(1/T)$. The theoretical justification of the HPJ estimator can be done along the same lines as the proof of Theorem 6 under a similar set of assumptions.

Remark 3. One may also consider applying the half-series jackknife (Quenouille, 1949, 1956) to each $\hat{\gamma}_{k,i}$, but we argue that this jackknife is not suitable in the current context. In the method based on the half-series jackknife, we bias-correct each $\hat{\gamma}_{k,i}$ using the jackknife. Suppose T is even. Let $\hat{\gamma}_{k,i}^{(1)}$ be the estimator of $\gamma_{k,i}$ using $\{y_{it}\}_{t=1}^{T/2}$, and $\hat{\gamma}_{k,i}^{(2)}$ be that based on $\{y_{it}\}_{t=T/2+1}^{T}$. Let $\bar{\gamma}_{k,i} = (\hat{\gamma}_{k,i}^{(1)} + \hat{\gamma}_{k,i}^{(2)})/2$. The half-series jackknife bias-corrected estimator of $\gamma_{k,i}$ is $\hat{\gamma}_{k,i}^{H} = \hat{\gamma}_{k,i} - (\bar{\gamma}_{k,i} - \hat{\gamma}_{k,i}) = 2\hat{\gamma}_{k,i} - \bar{\gamma}_{k,i}$. We then estimate G using $\{\hat{\gamma}_{k,i}^{H}\}_{i=1}^{N}$. The half-series jackknife method can reduce the bias of order $1/T$ in $\hat{\gamma}_{k,i}$ and, therefore, the incidental parameter bias in $\hat{G}$. However, it does not reduce the bias caused by the nonlinearity of $g(\cdot)$. Indeed, our Monte Carlo simulations show that the half-series jackknife may not work as well as the half-panel jackknife. Therefore, we do not pursue theoretical investigation of the half-series jackknife in this paper. We note that when $g(\cdot)$ is linear, the half-panel jackknife and the half-series jackknife are numerically equivalent.

Remark 4. We may also consider a higher-order jackknife bias correction. This is discussed in Dhaene and Jochmans (2014). The HPJ bias correction can eliminate bias up to the order of $O(1/T)$. The higher-order jackknife bias correction is expected to eliminate bias of higher order. Here, we consider the third-order jackknife. Suppose that T is a multiple of six.[3] The panel data are divided into three subpanels: $\{\{y_{it}\}_{i=1}^{N}\}_{t=1}^{T/3}$, $\{\{y_{it}\}_{i=1}^{N}\}_{t=T/3+1}^{2T/3}$, and $\{\{y_{it}\}_{i=1}^{N}\}_{t=2T/3+1}^{T}$. Let $\hat{G}^{(3,1)}$, $\hat{G}^{(3,2)}$, and $\hat{G}^{(3,3)}$ be the estimates of G computed from each of these three subpanels. The third-order jackknife estimator is

$$\hat{G}^{J3} = 3\hat{G} - \frac{3}{2}\left(\hat{G}^{(1)} + \hat{G}^{(2)}\right) + \frac{1}{3}\left(\hat{G}^{(3,1)} + \hat{G}^{(3,2)} + \hat{G}^{(3,3)}\right).$$

However, we do not examine its theoretical property in this paper. Our Monte Carlo results indicate that the higher-order jackknife can eliminate the bias effectively in some cases, but in other cases, we observe that it inflates the bias substantially. This result may be related to the caution noted by Dhaene and Jochmans (2014): a higher-order jackknife may inflate the bias by an order higher than that to be corrected. We also find that the variance inflation may be substantial in certain cases.

5.5 Cross-sectional bootstrap

In this section, we present the theorems that justify the use of the cross-sectional bootstrap. The first theorem is concerned with $\hat{G}_{\mu}$ and $\hat{G}_{\gamma_k}$, and the second theorem discusses the case of $\hat{G}_{\mu}^{H}$ and $\hat{G}_{\gamma_k}^{H}$.

We require the following additional assumptions. The following assumption is used to satisfy Lyapunov's conditions for $\hat{G}_{\mu}^{*}$ that is the estimator of G_{μ} obtained with the bootstrap sample.

Assumption 16. *$E(g^6(\mu_i)) < \infty$ and $E((g^2(\mu_i)g'(\mu_i))^4) < \infty$.*

[3] See Dhaene and Jochmans (2014) for the treatment of the case in which T is not a multiple of 6. Note that the asymptotic properties of the third-order jackknife estimator do not depend on whether or not T is a multiple of 6.

The following assumption is for $\hat{G}^*_{\gamma_k}$

Assumption 17. *$E(g^6(\gamma_{k,i})) < \infty$ and $E((g^2(\gamma_{k,i})g'(\gamma_{k,i}))^4) < \infty$.*

The following theorem states that the bootstrap distribution converges to the asymptotic distribution of $\hat{G}_\mu$ or $\hat{G}_{\gamma_k}$ when T is sufficiently large, but it fails to capture the bias term.

Theorem 7. *Suppose that Assumptions 1, 4, 5, 7, and 16 are satisfied. As $N, T \to \infty$, we have*

$$\sup_{x\in\mathbb{R}} \left| \Pr\left(\sqrt{N}(\hat{G}^*_\mu - \hat{G}_\mu) \le x \,\middle|\, \{\{y_{it}\}_{t=1}^T\}_{i=1}^N\right) - \Pr\left(Z_\mu \le x\right) \right| \xrightarrow{p} 0, \tag{14}$$

where $Z_\mu \sim N(0, var(g(\mu_i)))$.

Suppose that Assumptions 1, 8, 9, 10, and 17 are satisfied. As $N, T \to \infty$, we have

$$\sup_{x\in\mathbb{R}} \left| \Pr\left(\sqrt{N}(\hat{G}^*_{\gamma_k} - \hat{G}_{\gamma_k}) \le x \,\middle|\, \{\{y_{it}\}_{t=1}^T\}_{i=1}^N\right) - \Pr\left(Z_{\gamma_k} \le x\right) \right| \xrightarrow{p} 0, \tag{15}$$

where $Z_{\gamma_k} \sim N(0, var(g(\gamma_{k,i})))$.

It is important to note that the bootstrap does not capture the bias properties of $\hat{G}_\mu$ and $\hat{G}_{\gamma_k}$. The bootstrap distribution is asymptotically centered at zero. Thus, even if $\hat{G}_\mu$ or $\hat{G}_{\gamma_k}$ suffers from the bias as seen in Section 5.2, the bootstrap distribution cannot capture the bias. This implies that when T is small, we must be cautious about the use of the bootstrap to make statistical inferences. Galvao and Kato (2014), Gonçalves and Kaffo (2014), and Kaffo (2014) also observe that the bootstrap fails to approximate the bias in dynamic panel data settings for different estimators.

We can also show that the bootstrap can approximate the asymptotic distribution of the HPJ estimator.

Theorem 8. *Suppose that Assumptions 1, 8, 9, 13, and 16 are satisfied. As $N, T \to \infty$, we have*

$$\sup_{x\in\mathbb{R}} \left| \Pr\left(\sqrt{N}(\hat{G}^{H*}_\mu - \hat{G}^H_\mu) \le x \,\middle|\, \{\{y_{it}\}_{t=1}^T\}_{i=1}^N\right) - \Pr\left(Z_\mu \le x\right) \right| \xrightarrow{p} 0.$$

Suppose that Assumptions 1, 8, 14, 15, and 17 are satisfied. As $N, T \to \infty$, we have

$$\sup_{x\in\mathbb{R}} \left| \Pr\left(\sqrt{N}(\hat{G}^{H*}_{\gamma_k} - \hat{G}^H_{\gamma_k}) \le x \,\middle|\, \{\{y_{it}\}_{t=1}^T\}_{i=1}^N\right) - \Pr\left(Z_{\gamma_k} \le x\right) \right| \xrightarrow{p} 0.$$

The proof is analogous to the proof of Theorem 7, and is thus omitted.

The theorem indicates that the cross-sectional bootstrap can approximate the asymptotic distribution of the HPJ estimator correctly under the condition that N/T^2 does not diverge. Because the HPJ estimator does not suffer from bias as long as N/T^2 does not diverge, the bootstrap approximation would be more comfortably used for the HPJ estimator.

6 Extensions

In this section, we present two extensions based on the proposed procedure. The first is a test for parametric specifications on the distribution of the heterogeneous mean or autocovariance. The second is a test for whether the distributions of the mean or autocovariance are the same across different groups.

6.1 Testing parametric specifications

This subsection develops a testing procedure for hypotheses on parametric specifications of the distribution of the heterogeneous mean or autocovariance. The test is based on one-sample Kolmogorov–Smirnov (KS) statistics based on the empirical distributions of $\hat{\mu}_i$ and $\hat{\gamma}_{k,i}$. We derive their asymptotic null distributions. The results indicate that they are equivalent to those of the usual one-sample KS statistics and thus the critical values can be computed easily.

It is not uncommon to impose a parametric specification to model heterogeneous dynamics, and it is important to have a test for such a parametric specification. For example, Browning et al. (2010) develops a parametric model of heterogeneous income dynamics. Hsiao et al. (1999) consider random coefficients panel AR(1) models and impose parametric assumptions to implement a Bayesian procedure. Our test may be used to examine the validity of these parametric specifications.

We consider the following hypotheses:

$$H_0^\mu : P_0^\mu = Q^\mu \quad \text{v.s.} \quad H_1^\mu : P_0^\mu \neq Q^\mu,$$

and

$$H_0^{\gamma_k} : P_0^{\gamma_k} = Q^{\gamma_k} \quad \text{v.s.} \quad H_1^{\gamma_k} : P_0^{\gamma_k} \neq Q^{\gamma_k},$$

where Q^μ and Q^{γ_k} are known continuous probability distributions. The hypotheses are concerned with whether the distributions P_0^μ or $P_0^{\gamma_k}$ are the same as Q^μ or Q^{γ_k}, respectively. We note that Q^μ and Q^{γ_k} cannot be discrete probability distributions. This is because our asymptotic analyses are based on Assumptions 3 and 6.

We consider tests based on one-sample KS statistics (Kolmogorov, 1933; Smirnov, 1944):

$$KS_1^\mu := \sqrt{N} \left\| \mathbb{P}_N^{\hat{\mu}} - Q^\mu \right\|_\infty = \sqrt{N} \sup_{f \in \mathcal{F}} \left| \mathbb{P}_N^{\hat{\mu}} f - Q^\mu f \right|,$$
$$KS_1^{\gamma_k} := \sqrt{N} \left\| \mathbb{P}_N^{\hat{\gamma}_k} - Q^{\gamma_k} \right\|_\infty = \sqrt{N} \sup_{f \in \mathcal{F}} \left| \mathbb{P}_N^{\hat{\gamma}_k} f - Q^{\gamma_k} f \right|,$$

where $\|\cdot\|_\infty$ is the uniform norm. The test statistics measure the distances between the empirical distributions and the null distributions. We note that KS_1^μ and $KS_1^{\gamma_k}$ are different from the usual one-sample KS statistics in the sense that they are based on the empirical distributions of the estimates $\hat{\mu}_i$ and $\hat{\gamma}_{k,i}$, respectively.

We derive the asymptotic distributions of KS_1^μ and $KS_1^{\gamma_k}$ under H_0^μ and $H_0^{\gamma_k}$, respectively, utilizing Theorem 2. The following theorem presents the asymptotic null distributions.

Theorem 9. *Suppose that Assumptions 1, 2, and 3 hold for the case of KS_1^{μ}, and Assumptions 1, 4, 5, and 6 hold for the case of $KS_1^{\gamma_k}$. When $N, T \to \infty$ with $N/T \to 0$, it holds that KS_1^{μ} converges in distribution to $\|\mathbb{G}_{Q^{\mu}}\|_{\infty}$ under H_0^{μ}. Similarly, when $N, T \to \infty$ with $N/T \to 0$, it holds that $KS_1^{\gamma_k}$ converges in distribution to $\|\mathbb{G}_{Q^{\gamma_k}}\|_{\infty}$ under $H_0^{\gamma_k}$.*

This theorem shows that the asymptotic null distributions of KS_1^{μ} and $KS_1^{\gamma_k}$ are the uniform norms of the Gaussian processes. The asymptotic null distributions in the theorem are identical to those of the usual one-sample KS statistics developed in Kolmogorov (1933) and Smirnov (1944) so that they are equivalent to those of the one-sample KS statistics based on the true μ_i and $\gamma_{k,i}$. This is because the estimation errors in $\hat{\mu}_i$ or $\hat{\gamma}_{k,i}$ can be ignored asymptotically under the condition $N/T \to 0$.

Note that the asymptotic distributions do not depend on Q^{μ} or Q^{γ_k}, and critical values can be computed readily. As shown by Kolmogorov (1933) and Smirnov (1944) (for easy reference, see, e.g., Theorem 6.10 in Shao, 2003 or Section 2.1.5 in Serfling, 2002),

$$\Pr(\|\mathbb{G}_{Q^{\mu}}\|_{\infty} \leq a) = \Pr(\|\mathbb{G}_{Q^{\gamma_k}}\|_{\infty} \leq a) = 1 - 2\sum_{j=1}^{\infty}(-1)^{j-1}\exp\left(-2j^2a^2\right), \tag{16}$$

for any continuous distributions Q^{μ} and Q^{γ_k}, with $a > 0$. The far right-hand side of (16) does not depend on Q^{μ} or Q^{γ_k}. Moreover, the critical values are readily available in many statistical software packages and the implementation of our tests is easy.

6.2 Testing the difference in degrees of heterogeneity

Next, we develop tests to examine whether the distributions of the heterogeneous mean or autocovariances differ across distinct groups. The test statistics are two-sample KS statistics based on our empirical distribution estimators. We develop the asymptotic null distributions of the test statistics.

In many applications, it would be interesting to see whether distinct groups possess different heterogeneous dynamic structures. For example, when studying income dynamics, one would be interested in whether the distribution of individual average incomes differs between males and females. One may also be interested in whether the degrees of heterogeneity of income dynamics depend on racial group. We develop test procedures for such hypotheses without any parametric specification. Suppose that we have two panel data sets for two different groups: $\{\{y_{it,(1)}\}_{t=1}^{T_1}\}_{i=1}^{N_1}$ and $\{\{y_{it,(2)}\}_{t=1}^{T_2}\}_{i=1}^{N_2}$. We allow the situation in which $T_1 \neq T_2$ and/or $N_1 \neq N_2$. We define $y_{i,(1)} := \{y_{it,(1)}\}_{t=1}^{T_1}$ for $i = 1, \dots, N_1$ and $y_{i,(2)} := \{y_{it,(2)}\}_{t=1}^{T_2}$ for $i = 1, \dots, N_2$.

We introduce the following assumption on the data sets.

Assumption 18. *Each of $\{\{y_{it,(1)}\}_{t=1}^{T_1}\}_{i=1}^{N_1}$ and $\{\{y_{it,(2)}\}_{t=1}^{T_2}\}_{i=1}^{N_2}$ satisfies Assumptions 1, 2, and 3 for the case of the mean, and Assumptions 1, 4, 5, and 6 for the case of the autocovariances. $(y_{1,(1)}, \dots, y_{N_1,(1)})$ and $(y_{1,(2)}, \dots, y_{N_2,(2)})$ are independent.*

We need the assumptions introduced in the previous sections and require the independence assumption. This assumption implies that our test cannot be used to test the equivalence of the distributions of two variables from the same individuals. Our test is intended to be used to compare the distributions of the same variable from different groups.

We estimate the distribution of the mean or autocovariances for each group. Let $\mu_{i,(a)}$ be the heterogeneous mean of $y_{i,(a)}$ for group $a = 1, 2$. We estimate $\mu_{i,(a)}$ by the sample mean $\hat{\mu}_{i,(a)} := \bar{y}_{i,(a)} := T_a^{-1} \sum_{t=1}^{T_a} y_{it,(a)}$ for $a = 1, 2$. We denote the probability distribution of $\mu_{i,(a)}$ by $P_{0,(a)}^{\mu}$ and the empirical distribution of $\hat{\mu}_{i,(a)}$ by $\mathbb{P}_{N_a,(a)}^{\hat{\mu}}$ for $a = 1, 2$. Similarly, let $\gamma_{k,i,(a)}$ be the k-th individual autocovariance of $y_{i,(a)}$ for $a = 1, 2$. We estimate $\gamma_{k,i,(a)}$ by the sample k-th autocovariance $\hat{\gamma}_{k,i,(a)} := T_a^{-1} \sum_{t=k+1}^{T_a} (y_{it,(a)} - \bar{y}_{i,(a)})(y_{i,t-k,(a)} - \bar{y}_{i,(a)})$ for $a = 1, 2$. We write the distribution of $\gamma_{k,i,(a)}$ by $P_{0,(a)}^{\gamma_k}$ and the empirical distribution of $\hat{\gamma}_{k,i,(a)}$ by $\mathbb{P}_{N_a,(a)}^{\hat{\gamma}_k}$ for $a = 1, 2$.

We focus on the following hypotheses to examine the difference in the degrees of heterogeneity between the two groups:

$$H_0^{\mu} : P_{0,(1)}^{\mu} = P_{0,(2)}^{\mu} \quad \text{v.s.} \quad H_1^{\mu} : P_{0,(1)}^{\mu} \neq P_{0,(2)}^{\mu},$$

and

$$H_0^{\gamma_k} : P_{0,(1)}^{\gamma_k} = P_{0,(2)}^{\gamma_k} \quad \text{v.s.} \quad H_1^{\gamma_k} : P_{0,(1)}^{\gamma_k} \neq P_{0,(2)}^{\gamma_k}.$$

Under the null hypothesis H_0^{μ} ($H_0^{\gamma_k}$), the distribution of the heterogeneous mean (autocovariances) is identical across the two groups.

We investigate the hypotheses using the following two-sample KS statistics based on our empirical distribution estimators:

$$KS_2^{\mu} := \sqrt{\frac{N_1 N_2}{N_1 + N_2}} \left\| \mathbb{P}_{N_1,(1)}^{\hat{\mu}} - \mathbb{P}_{N_2,(2)}^{\hat{\mu}} \right\|_{\infty} = \sqrt{\frac{N_1 N_2}{N_1 + N_2}} \sup_{f \in \mathcal{F}} \left| \mathbb{P}_{N_1,(1)}^{\hat{\mu}} f - \mathbb{P}_{N_2,(2)}^{\hat{\mu}} f \right|,$$
$$KS_2^{\gamma_k} := \sqrt{\frac{N_1 N_2}{N_1 + N_2}} \left\| \mathbb{P}_{N_1,(1)}^{\hat{\gamma}_k} - \mathbb{P}_{N_2,(2)}^{\hat{\gamma}_k} \right\|_{\infty} = \sqrt{\frac{N_1 N_2}{N_1 + N_2}} \sup_{f \in \mathcal{F}} \left| \mathbb{P}_{N_1,(1)}^{\hat{\gamma}_k} f - \mathbb{P}_{N_2,(2)}^{\hat{\gamma}_k} f \right|.$$

The test statistics measure the distances between the empirical distributions of the two groups. KS_2^{μ} and $KS_2^{\gamma_k}$ are different from the usual two-sample KS statistics in the sense that KS_2^{μ} and $KS_2^{\gamma_k}$ are based on the empirical distributions of the estimates $\hat{\mu}_{i,(a)}$ and $\hat{\gamma}_{k,i,(a)}$, respectively, for $a = 1, 2$.

The asymptotic null distributions of KS_2^{μ} and $KS_2^{\gamma_k}$ are derived using Theorem 2.

Theorem 10. *Suppose that Assumption 18 is satisfied. When $N_1, T_1 \to \infty$ with $N_1/T_1 \to 0$ and $N_2, T_2 \to \infty$ with $N_2/T_2 \to 0$, and $N_1/(N_1 + N_2) \to \lambda$ for some $\lambda \in (0, 1)$, it holds that KS_2^{μ} converges in distribution to $\|\mathbb{G}_{P_{0,(1)}^{\mu}}\|_{\infty}$ under H_0^{μ} and that $KS_2^{\gamma_k}$ converges in distribution to $\|\mathbb{G}_{P_{0,(1)}^{\gamma_k}}\|_{\infty}$ under $H_0^{\gamma_k}$.*

This theorem shows that the asymptotic null distributions of KS_2^{μ} and $KS_2^{\gamma_k}$ are the uniform norms of the Gaussian processes. The conditions $N_1/T_1 \to 0$ and $N_2/T_2 \to 0$ are required in order to use the result of Theorem 2. The condition $N_1/(N_1 + N_2) \to \lambda$ implies that N_1 is not much greater or less than N_2 and guarantees the existence of the asymptotic null distributions.

The asymptotic null distributions in the theorem are the same as those in Theorem 9 when we set $Q^{\mu} = P^{\mu}_{0,(1)}$ or $Q^{\gamma_k} = P^{\gamma_k}_{0,(1)}$. Hence, the asymptotic null distributions can be evaluated easily by (16) and the critical values of our test are readily available.

Remark 5. When the true distributions of $\hat{\mu}_{i,(1)}$ and $\hat{\mu}_{i,(2)}$ (or $\hat{\gamma}_{k,i,(1)}$ and $\hat{\gamma}_{k,i,(2)}$) are the same, i.e., when $P^{\hat{\mu}}_{T_1,(1)} = P^{\hat{\mu}}_{T_2,(2)}$ (or $P^{\hat{\gamma}_k}_{T_1,(1)} = P^{\hat{\gamma}_k}_{T_2,(2)}$), neither the condition $N_1/T_1 \to 0$ nor the condition $N_2/T_2 \to 0$ is needed to establish Theorem 10. This is clear from the proof of Theorem 10. In particular, when $T_1 = T_2$ and the mean and dynamic structures across the two groups are completely identical under the null hypothesis, we can test the null hypotheses H_0^{μ} or $H_0^{\gamma_k}$ without restricting the relative order of N_a and T_a for $a = 1, 2$. Note that we still need the condition $N_1/(N_1 + N_2) \to \lambda \in (0, 1)$.

7 Monte Carlo simulations

This section presents the results of the Monte Carlo simulations. We investigate the finite-sample performance of the proposed methods in the simulations. We also evaluate the performance of the proposed bias-correction method. The simulations are conducted with R 3.1.1 for Mac OS X 10.9.5. The number of replications in the simulations is 5000.

7.1 Designs

The data-generating process is the following random coefficients panel ARMA(1,1) process:

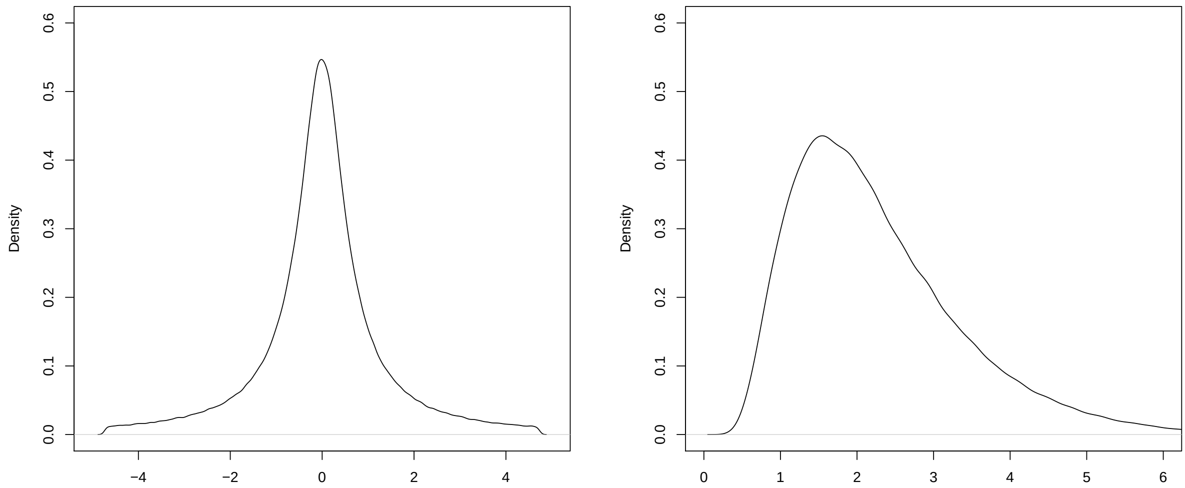
$$y_{it} = \eta_i + \phi_i y_{i,t-1} + \epsilon_{it} + \theta_i \epsilon_{i,t-1},$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$, where $\epsilon_{it} \sim i.i.d.N(0, 1)$. We consider two specifications of the distribution of the random coefficients $(\eta_i, \phi_i, \theta_i)$. In the first specification (design A), $\eta_i \sim i.i.d.N(0, 1)$, $\phi_i \sim i.i.d.U[-0.9, 0.9]$, and $\theta_i = 0$ and η_i, ϕ_i, and θ_i are independent. In the second specification (design B), $\eta_i = \phi_i + \xi_i$ where $\phi_i \sim 0.4 + 0.5\ i.i.d.Beta(5, 2)$, $\xi_i \sim i.i.d.N(0, 0.25)$, and $\theta_i \sim i.i.d.U[-0.2, 0.3]$ and ϕ_i, ξ_i, and θ_i are independent. The second specification is motivated by the empirical results of Browning et al. (2010). However, our specification is simpler and the process here is less persistent. Moreover, the mean part is different. We note that the specification for individual-specific means (η_i) does not affect the estimation of the autocovariances because individual-specific means are eliminated when we estimate these autocovariances. We generate the initial observations from the stationary distribution given $(\eta_i, \phi_i, \theta_i)$:

$$\begin{pmatrix} y_{i0} \\ \epsilon_{i0} \end{pmatrix} \sim N\left(\frac{\eta_i}{1-\phi_i}, \begin{pmatrix} \dfrac{1+\theta_i^2+\phi_i\theta_i}{1-\phi_i^2} & 1 \\ 1 & 1 \end{pmatrix} \right).$$

We set $N = 100$ and 1000, and $T = 24$ and 48.

We estimate the distributions of the mean ($\mu_i = \eta_i/(1-\phi_i)$), the variance ($\gamma_{0,i}$), and the first-order autocovariance ($\gamma_{1,i}$). In particular, we consider the estimation of the mean (Mean),



(a) Design A

(b) Design B

Figure 1: The density function of $\gamma_{1,i}$

variance (Var), and 25%, 50%, and 75% quantiles (25%Q, 50%Q, 75%Q) of the distributions of these quantities. We also estimate the covariances between these quantities. By way of illustration, the densities of $\gamma_{1,i}$ for designs A and B are plotted in Figure 1, parts (a) and (b), respectively.

We consider the following four estimators. The first estimator is a naive estimator based on the empirical distribution of the estimated means or autocovariances. We denote this "NE". The second estimator is the half-panel jackknife estimator (HPJ). Note that for the quantiles, the HPJ estimator is not theoretically justified because they are not expected values of smooth functions. The third estimator is the third-order jackknife estimator (TOJ) discussed in Remark 4. The fourth estimator is based on the half-series jackknife autocovariance estimator (HSJ) considered in Remark 3.

Importantly, as noted in Section 3, the estimation of the variance is done by separately estimating the uncentered second and first moments. Likewise, for the HPJ and TOJ estimators, we do not bias-correct the variance estimator directly. Rather, we separately bias-correct the estimators of the uncentered second and first moments and then compute the variance estimate by combining these bias-corrected estimates. Similarly, when we estimate the covariances with the split-panel bias correction, we bias-correct the cross-moment estimate and the mean estimates separately. We then combine these bias-corrected estimates to form the covariance estimates.

7.2 Results

Tables 1–4 and 5–8 summarize the results of the Monte Carlo simulations with designs A and B, respectively. They present the bias and the standard deviation (std) of each estimator. In

μ_i				NE		HPJ		TOJ		HSJ	
	N	T	true	bias	std	bias	std	bias	std	bias	std
Mean	100	24	-0.001	0.002	0.232	0.002	0.232	0.002	0.232	0.002	0.232
	100	48	-0.001	-0.003	0.234	-0.003	0.234	-0.003	0.234	-0.003	0.234
	1000	24	-0.001	0.002	0.074	0.002	0.074	0.002	0.074	0.002	0.074
	1000	48	-0.001	0.001	0.073	0.001	0.073	0.001	0.073	0.001	0.073
Var	100	24	5.280	0.230	2.433	0.056	2.400	0.025	2.397	0.230	2.433
	100	48	5.280	0.116	2.343	-0.017	2.313	-0.033	2.311	0.116	2.343
	1000	24	5.280	0.154	0.750	0.031	0.747	-0.000	0.746	0.154	0.750
	1000	48	5.280	0.087	0.738	0.002	0.735	-0.013	0.734	0.087	0.738
25%Q	100	24	-0.723	-0.009	0.168	0.006	0.191	0.007	0.231	-0.009	0.168
	100	48	-0.723	-0.006	0.170	0.001	0.187	-0.001	0.218	-0.006	0.170
	1000	24	-0.723	-0.014	0.054	0.002	0.061	0.001	0.076	-0.014	0.054
	1000	48	-0.723	-0.005	0.053	0.003	0.059	0.003	0.070	-0.005	0.053
50%Q	100	24	0.000	0.002	0.128	0.001	0.146	0.000	0.178	0.002	0.128
	100	48	0.000	0.001	0.125	0.002	0.137	0.003	0.160	0.001	0.125
	1000	24	0.000	0.001	0.040	0.001	0.046	0.001	0.057	0.001	0.040
	1000	48	0.000	0.001	0.040	0.001	0.044	0.001	0.053	0.001	0.040
75%Q	100	24	0.722	0.013	0.174	-0.003	0.197	-0.003	0.239	0.013	0.174
	100	48	0.722	0.007	0.171	-0.001	0.187	-0.001	0.216	0.007	0.171
	1000	24	0.722	0.016	0.054	0.001	0.062	0.001	0.076	0.016	0.054
	1000	48	0.722	0.008	0.054	0.000	0.061	0.001	0.072	0.008	0.054

Table 1: Monte Carlo simulation results: distribution of μ_i with design A

$\gamma_{0,i}$				NE		HPJ		TOJ		HSJ	
	N	T	true	bias	std	bias	std	bias	std	bias	std
Mean	100	24	1.636	-0.180	0.116	-0.062	0.131	-0.030	0.149	-0.062	0.131
	100	48	1.636	-0.102	0.102	-0.022	0.116	-0.007	0.130	-0.022	0.116
	1000	24	1.636	-0.179	0.036	-0.061	0.041	-0.030	0.047	-0.061	0.041
	1000	48	1.636	-0.100	0.033	-0.021	0.037	-0.006	0.042	-0.021	0.037
Var	100	24	0.775	0.516	0.972	0.075	0.793	0.001	0.944	0.882	1.099
	100	48	0.775	0.289	0.570	0.039	0.542	0.024	0.731	0.598	0.811
	1000	24	0.775	0.517	0.306	0.096	0.261	0.033	0.314	0.893	0.355
	1000	48	0.775	0.288	0.182	0.047	0.175	0.029	0.237	0.591	0.256
25%Q	100	24	1.053	-0.170	0.047	-0.002	0.075	1.770	0.155	-0.129	0.050
	100	48	1.053	-0.070	0.040	0.029	0.060	1.936	0.128	-0.048	0.041
	1000	24	1.053	-0.173	0.015	-0.004	0.024	1.762	0.049	-0.133	0.016
	1000	48	1.053	-0.073	0.012	0.027	0.019	1.932	0.040	-0.050	0.013
50%Q	100	24	1.254	-0.094	0.063	0.034	0.097	1.562	0.166	-0.032	0.069
	100	48	1.254	-0.030	0.058	0.034	0.085	1.460	0.145	0.004	0.062
	1000	24	1.254	-0.096	0.020	0.032	0.030	1.557	0.052	-0.034	0.021
	1000	48	1.254	-0.032	0.018	0.032	0.026	1.457	0.047	0.002	0.019
75%Q	100	24	1.838	-0.224	0.128	-0.126	0.194	1.009	0.313	-0.104	0.148
	100	48	1.838	-0.153	0.137	-0.077	0.201	0.928	0.323	-0.080	0.151
	1000	24	1.838	-0.226	0.041	-0.132	0.062	1.002	0.100	-0.109	0.047
	1000	48	1.838	-0.151	0.044	-0.075	0.065	0.932	0.104	-0.079	0.048

Table 2: Monte Carlo simulation results: distribution of $\gamma_{0,i}$ with design A

$\gamma_{1,i}$				NE		HPJ		TOJ		HSJ	
	N	T	true	bias	std	bias	std	bias	std	bias	std
Mean	100	24	-0.001	-0.183	0.138	-0.062	0.158	-0.029	0.177	-0.062	0.158
	100	48	-0.001	-0.098	0.134	-0.016	0.150	-0.001	0.164	-0.016	0.150
	1000	24	-0.001	-0.181	0.044	-0.060	0.050	-0.027	0.056	-0.060	0.050
	1000	48	-0.001	-0.100	0.043	-0.019	0.048	-0.003	0.052	-0.019	0.048
Var	100	24	1.816	0.114	1.130	-0.094	1.007	-0.081	1.248	0.726	1.347
	100	48	1.816	0.061	0.726	-0.024	0.732	0.001	0.956	0.539	1.018
	1000	24	1.816	0.115	0.354	-0.065	0.326	-0.043	0.410	0.737	0.430
	1000	48	1.816	0.061	0.233	-0.006	0.239	0.016	0.314	0.534	0.325
25%Q	100	24	-0.565	0.053	0.132	0.033	0.178	-0.797	0.382	0.071	0.133
	100	48	-0.565	0.035	0.137	0.014	0.173	-0.985	0.399	0.044	0.137
	1000	24	-0.565	0.054	0.041	0.031	0.055	-0.794	0.118	0.071	0.041
	1000	48	-0.565	0.031	0.043	0.009	0.055	-0.992	0.124	0.041	0.043
50%Q	100	24	0.000	-0.041	0.069	-0.005	0.099	-0.787	0.214	-0.009	0.074
	100	48	0.000	-0.020	0.080	0.000	0.103	-1.018	0.248	-0.002	0.082
	1000	24	0.000	-0.040	0.021	-0.005	0.031	-0.783	0.067	-0.009	0.023
	1000	48	0.000	-0.020	0.025	-0.001	0.033	-1.016	0.079	-0.003	0.026
75%Q	100	24	0.564	-0.209	0.096	-0.070	0.149	-1.493	0.287	-0.120	0.117
	100	48	0.564	-0.111	0.116	-0.015	0.161	-1.494	0.322	-0.056	0.130
	1000	24	0.564	-0.207	0.030	-0.068	0.047	-1.488	0.090	-0.118	0.037
	1000	48	0.564	-0.110	0.037	-0.013	0.053	-1.491	0.103	-0.056	0.042

Table 3: Monte Carlo simulation results: distribution of $\gamma_{1,i}$ with design A

Cov				NE		HPJ		TOJ		HSJ	
	N	T	true	bias	std	bias	std	bias	std	bias	std
$\mu_i \& \gamma_{0,i}$	100	24	-0.001	0.003	0.311	0.003	0.505	0.002	0.685	0.003	0.508
	100	48	-0.001	0.004	0.377	0.005	0.540	0.007	0.686	0.005	0.545
	1000	24	-0.001	0.004	0.101	0.005	0.164	0.006	0.222	0.005	0.164
	1000	48	-0.001	0.000	0.118	-0.000	0.172	-0.000	0.218	0.000	0.172
$\mu_i \& \gamma_{1,i}$	100	24	-0.002	0.003	0.445	0.003	0.638	0.003	0.806	0.003	0.643
	100	48	-0.002	-0.000	0.540	0.002	0.697	0.004	0.828	0.001	0.704
	1000	24	-0.002	0.005	0.141	0.007	0.205	0.008	0.260	0.007	0.205
	1000	48	-0.002	0.000	0.169	-0.000	0.221	-0.000	0.263	0.000	0.222
$\gamma_{0,i} \& \gamma_{1,i}$	100	24	-0.001	-0.814	1.051	-0.277	0.894	-0.104	1.072	-0.350	1.227
	100	48	-0.001	-0.416	0.652	-0.043	0.641	0.033	0.842	-0.029	0.925
	1000	24	-0.001	-0.802	0.331	-0.263	0.295	-0.082	0.361	-0.327	0.397
	1000	48	-0.001	-0.418	0.204	-0.047	0.200	0.030	0.266	-0.037	0.284

Table 4: Monte Carlo simulation results: covariances with design A

μ_i				NE		HPJ		TOJ		HSJ	
	N	T	true	bias	std	bias	std	bias	std	bias	std
Mean	100	24	3.592	0.006	0.300	0.006	0.300	0.006	0.300	0.006	0.300
	100	48	3.592	0.007	0.295	0.007	0.295	0.007	0.295	0.007	0.295
	1000	24	3.592	0.014	0.095	0.014	0.095	0.014	0.095	0.014	0.095
	1000	48	3.592	0.006	0.094	0.006	0.094	0.006	0.094	0.006	0.094
Var	100	24	8.275	0.948	1.709	0.337	1.685	0.139	1.687	0.948	1.709
	100	48	8.275	0.544	1.657	0.075	1.633	-0.016	1.637	0.544	1.657
	1000	24	8.275	0.965	0.542	0.438	0.540	0.240	0.540	0.965	0.542
	1000	48	8.275	0.516	0.527	0.125	0.524	0.034	0.525	0.516	0.527
25%Q	100	24	1.620	-0.064	0.302	-0.013	0.361	0.005	0.464	-0.064	0.302
	100	48	1.620	-0.027	0.287	0.010	0.335	0.018	0.421	-0.027	0.287
	1000	24	1.620	-0.083	0.096	-0.031	0.117	-0.011	0.153	-0.083	0.096
	1000	48	1.620	-0.045	0.093	-0.007	0.111	0.001	0.142	-0.045	0.093
50%Q	100	24	3.165	0.007	0.333	0.003	0.391	0.008	0.490	0.007	0.333
	100	48	3.165	0.010	0.325	0.009	0.375	0.012	0.462	0.010	0.325
	1000	24	3.165	0.010	0.106	0.003	0.126	0.004	0.159	0.010	0.106
	1000	48	3.165	0.005	0.105	0.004	0.123	0.005	0.153	0.005	0.105
75%Q	100	24	5.130	0.077	0.467	0.018	0.547	-0.003	0.679	0.077	0.467
	100	48	5.130	0.037	0.461	-0.003	0.525	-0.014	0.636	0.037	0.461
	1000	24	5.130	0.106	0.149	0.047	0.176	0.028	0.222	0.106	0.149
	1000	48	5.130	0.053	0.145	0.013	0.168	0.006	0.208	0.053	0.145

Table 5: Monte Carlo simulation results: distribution of μ_i with design B

$\gamma_{0,i}$				NE		HPJ		TOJ		HSJ	
	N	T	true	bias	std	bias	std	bias	std	bias	std
Mean	100	24	2.843	-0.852	0.130	-0.334	0.204	-0.135	0.278	-0.334	0.204
	100	48	2.843	-0.480	0.134	-0.100	0.186	-0.009	0.240	-0.100	0.186
	1000	24	2.843	-0.852	0.042	-0.335	0.065	-0.137	0.088	-0.335	0.065
	1000	48	2.843	-0.480	0.042	-0.098	0.058	-0.007	0.075	-0.098	0.058
Var	100	24	1.179	0.549	0.675	0.449	1.076	0.464	1.496	3.042	2.006
	100	48	1.179	0.602	0.643	0.368	0.938	0.281	1.326	2.270	1.619
	1000	24	1.179	0.549	0.214	0.486	0.346	0.534	0.488	3.034	0.642
	1000	48	1.179	0.607	0.204	0.410	0.309	0.353	0.442	2.287	0.521
25%Q	100	24	2.033	-0.885	0.081	-0.493	0.143	0.898	0.216	-0.752	0.101
	100	48	2.033	-0.543	0.088	-0.200	0.147	1.101	0.213	-0.430	0.102
	1000	24	2.033	-0.893	0.026	-0.503	0.046	0.880	0.070	-0.764	0.032
	1000	48	2.033	-0.552	0.028	-0.210	0.047	1.083	0.067	-0.440	0.032
50%Q	100	24	2.608	-0.963	0.115	-0.487	0.199	0.798	0.290	-0.693	0.152
	100	48	2.608	-0.585	0.122	-0.201	0.197	0.954	0.281	-0.371	0.148
	1000	24	2.608	-0.968	0.036	-0.493	0.062	0.788	0.092	-0.699	0.048
	1000	48	2.608	-0.589	0.039	-0.206	0.064	0.946	0.092	-0.376	0.047
75%Q	100	24	3.411	-0.985	0.198	-0.391	0.342	0.937	0.501	-0.396	0.290
	100	48	3.411	-0.570	0.204	-0.144	0.331	0.976	0.473	-0.141	0.262
	1000	24	3.411	-0.983	0.063	-0.390	0.110	0.946	0.159	-0.396	0.093
	1000	48	3.411	-0.567	0.065	-0.141	0.105	0.988	0.150	-0.138	0.084

Table 6: Monte Carlo simulation results: distribution of $\gamma_{0,i}$ with design B

$\gamma_{1,i}$				NE		HPJ		TOJ		HSJ	
	N	T	true	bias	std	bias	std	bias	std	bias	std
Mean	100	24	2.266	-0.879	0.125	-0.333	0.203	-0.124	0.280	-0.333	0.203
	100	48	2.266	-0.491	0.133	-0.095	0.188	-0.003	0.242	-0.095	0.188
	1000	24	2.266	-0.879	0.040	-0.335	0.065	-0.126	0.089	-0.335	0.065
	1000	48	2.266	-0.490	0.041	-0.094	0.059	-0.003	0.076	-0.094	0.059
Var	100	24	1.229	0.369	0.640	0.353	1.031	0.425	1.433	2.965	2.001
	100	48	1.229	0.515	0.633	0.345	0.937	0.282	1.326	2.265	1.646
	1000	24	1.229	0.368	0.204	0.388	0.334	0.494	0.470	2.954	0.641
	1000	48	1.229	0.520	0.201	0.388	0.310	0.356	0.444	2.282	0.531
25%Q	100	24	1.444	-0.864	0.074	-0.504	0.133	-0.216	0.193	-0.735	0.094
	100	48	1.444	-0.533	0.087	-0.200	0.146	-0.046	0.207	-0.419	0.102
	1000	24	1.444	-0.872	0.023	-0.516	0.041	-0.236	0.060	-0.746	0.029
	1000	48	1.444	-0.542	0.027	-0.210	0.046	-0.064	0.066	-0.430	0.032
50%Q	100	24	2.040	-1.000	0.109	-0.509	0.191	-0.449	0.272	-0.719	0.150
	100	48	2.040	-0.603	0.119	-0.203	0.195	-0.239	0.276	-0.383	0.146
	1000	24	2.040	-1.004	0.034	-0.514	0.060	-0.458	0.086	-0.724	0.048
	1000	48	2.040	-0.607	0.039	-0.207	0.064	-0.246	0.090	-0.388	0.047
75%Q	100	24	2.857	-1.068	0.191	-0.404	0.337	-0.422	0.487	-0.440	0.290
	100	48	2.857	-0.611	0.201	-0.144	0.328	-0.258	0.464	-0.162	0.267
	1000	24	2.857	-1.067	0.061	-0.403	0.107	-0.414	0.152	-0.437	0.093
	1000	48	2.857	-0.608	0.064	-0.142	0.105	-0.245	0.149	-0.159	0.085

Table 7: Monte Carlo simulation results: distribution of $\gamma_{1,i}$ with design B

Cov				NE		HPJ		TOJ		HSJ	
	N	T	true	bias	std	bias	std	bias	std	bias	std
$\mu_i\&\gamma_{0,i}$	100	24	1.331	-0.731	0.489	-0.403	0.787	-0.205	1.093	-0.389	0.790
	100	48	1.331	-0.446	0.509	-0.135	0.727	-0.008	0.956	-0.121	0.732
	1000	24	1.331	-0.731	0.156	-0.397	0.251	-0.201	0.348	-0.392	0.251
	1000	48	1.331	-0.440	0.162	-0.118	0.236	0.007	0.310	-0.116	0.235
$\mu_i\&\gamma_{1,i}$	100	24	1.365	-0.753	0.472	-0.418	0.782	-0.211	1.094	-0.404	0.786
	100	48	1.365	-0.456	0.504	-0.136	0.731	-0.004	0.963	-0.122	0.736
	1000	24	1.365	-0.754	0.149	-0.413	0.249	-0.208	0.347	-0.409	0.248
	1000	48	1.365	-0.451	0.161	-0.119	0.237	0.010	0.313	-0.117	0.237
$\gamma_{0,i}\&\gamma_{1,i}$	100	24	1.203	0.439	0.656	0.405	1.052	0.448	1.462	2.974	2.000
	100	48	1.203	0.549	0.637	0.358	0.937	0.283	1.324	2.254	1.631
	1000	24	1.203	0.438	0.208	0.441	0.340	0.517	0.478	2.964	0.640
	1000	48	1.203	0.555	0.202	0.401	0.310	0.355	0.442	2.272	0.525

Table 8: Monte Carlo simulation results: covariances with design B

the column labeled "true," the true value of the corresponding quantity is presented. We note that all estimates of the mean of μ_i are numerically equivalent by construction, and that the estimates of the mean of each quantity by HPJ and HSJ are numerically equivalent.

The simulation results demonstrate that our asymptotic analyses are informative regarding the finite-sample behavior of the estimators and the importance of bias correction. When both N and T are small, NE exhibits severe biases for many parameters of interest. In particular, large biases are observed in the estimation of $var(\mu_i)$, all quantities of $\gamma_{0,i}$ and $\gamma_{1,i}$, $cov(\mu_i, \gamma_{0,i})$, and $cov(\mu_i, \gamma_{1,i})$ with design B. With design A, the magnitudes of the biases are relatively moderate. However, the estimation of $var(\mu_i)$, $var(\gamma_{0,i})$, and $cov(\gamma_{0,i}, \gamma_{1,i})$ with design A involves large biases. As T increases while holding N fixed, the biases of NE decrease, which is expected from our asymptotic analyses. Nonetheless, a significant portion of the bias remains even with large T with design B. Worse, the biases are often large compared with the standard deviations. This result suggests the importance of developing the bias-correction method. The standard deviations of NE do not decrease as T becomes large with N fixed. However, the standard deviations decrease as N becomes large. This result can also be expected, as our asymptotic results show that the variances are of order $O(1/N)$.

HPJ successfully reduces the bias in most cases, and works especially well when the biases of NE are large. In particular, HPJ succeeds markedly in correcting the biases of the estimation of $var(\gamma_{0,i})$ and $cov(\gamma_{0,i}, \gamma_{1,i})$ with design A and those of $var(\mu_i)$ and $E(\gamma_{0,i})$ with design B, even when both N and T are small. Interestingly, HPJ also eliminates the biases in the quantile estimates in many cases, despite our theoretical justification that HPJ does not apply to the estimation of the distribution function or quantiles. This result indicates that HPJ may in fact be useful, even when the parameter of interest is not the expected value of a smooth function. However, we need to develop alternative asymptotic analyses to show this formally. As expected by our asymptotic results, the biases in HPJ tend to decrease as T increases. When both N and T are large, the biases in HPJ are satisfactorily small in many cases.

While HPJ slightly increases the finite-sample standard deviations in some cases, the inflation of the standard deviations would be acceptable. The biases are more serious than the standard deviations in many cases. Although expected, when NE is almost unbiased (e.g., in the estimation of $cov(\mu_i, \gamma_{0,i})$ in design A), HPJ slightly increases the mean squared errors of estimates in some cases. Otherwise, the bias reduction of HPJ sufficiently compensates for the inflation of the standard deviations except for $var(\gamma_{0,i})$ and $var(\gamma_{1,i})$ in design B. When both N and T are large, the standard deviations of NE and HPJ are similar. This is expected given our asymptotic result that NE and HPJ possess the same asymptotic variance. For these reasons, we stress that HPJ is more reliable than NE.

In some cases, TOJ reduces the bias more successfully than HPJ, but in other cases, TOJ inflates the bias substantially. We observe the large bias of TOJ in estimating the quantiles of $\gamma_{0,i}$ and $\gamma_{1,i}$ with design A and $var(\gamma_{0,i})$, the quantiles of $\gamma_{0,i}$, $var(\gamma_{1,i})$, and $cov(\gamma_{0,i}, \gamma_{1,i})$ with design B. In addition, TOJ often increases the standard deviation considerably. The examples include the estimation of $var(\gamma_{0,i})$, $var(\gamma_{1,i})$, and $cov(\gamma_{0,i}, \gamma_{1,i})$ with design A and $var(\gamma_{0,i})$ with design B. This result corresponds with the note in Remark 4: the higher-order jackknife may inflate the higher-order bias and the small-sample standard deviation. The inflation of the bias or the standard deviation is critical, especially when the biases of NE and HPJ are not large. We thus recommend HPJ rather than the higher-order jackknife as a precaution.

HSJ does not reduce the bias except in the case of the mean. This is because HSJ fails to eliminate the bias caused by the nonlinearity of smooth functions, as discussed in Remark 3. Worse, HSJ substantially increases the biases in some cases. For example, they are observed for the estimation of $var(\gamma_{1,i})$ in design A and for the estimation of $var(\gamma_{0,i})$, $var(\gamma_{1,i})$, and $cov(\gamma_{0,i}, \gamma_{1,i})$ in design B. Because of these Monte Carlo results, we do not pursue a theoretical investigation of HSJ in this paper.

Our preferred procedure is HPJ, given the results of these Monte Carlo experiments. NE is often considerably biased, whereas HPJ can alleviate the bias problem without significant variance inflation. TOJ may be used for the estimation of the mean of a quantity, but in other cases it may inflate both the bias and the variance. The performance of TOJ appears to be highly situation dependent, and we hesitate to recommend its use for the moment. HSJ is not recommended.

8 Conclusion

This paper proposes methods to analyze the heterogeneous dynamic structure using panel data. Our proposed methods do not require model specification and are easy to implement. We first compute the sample mean or the sample autocovariances of each individual. We then use these to estimate the parameters of interest, such as the distribution function, the quantile function, and the other moments of the heterogeneous mean and/or autocovariances. We show that the estimator for the distribution function of the heterogeneous mean or autocovariances exhibits a bias of order $O(1/\sqrt{T})$. When the parameter of interest can be written as the expected value of a smooth function of the heterogeneous mean or autocovariances, the bias of the estimator becomes of order $O(1/T)$ and can be reduced by the half-panel jackknife bias-correction method. We also present extensions based on the proposed procedures to test parametric specifications on the distribution of the heterogeneous mean or autocovariances and to test the difference of the heterogeneous dynamic structures across distinct groups. The results of Monte Carlo simulations show that our asymptotic analyses are informative regarding the finite-sample properties of the proposed estimators. Based on the simulation results, we recommend the half-panel jackknife estimator. We believe that our proposed methods can be used to address several important questions regarding the dynamics of economic variables.

Future work: Several future research topics are possible. First, methods for prediction could be considered based on the proposed analysis. Given that our proposed analysis estimates the distributions of the heterogeneous mean and autocovariances we could, in principle, use them to construct a best linear predictor of future values of y_{it}.

Second, while this paper develops the analysis for stationary panel data, it could be used to extend our analysis to cover nonstationary panel data. Two types of nonstationarity are relevant to our analysis. The first is the effect of initial distributions. In this paper, we assume that the initial values are drawn from the stationary distributions for simplicity. As we consider the case in which T is large, the effect of initial values would be negligible in large samples. However, the effect in a finite sample remains untested. The second type of nonstationarity is a stochastic trend. For example, in the literature on income dynamics, there is debate over whether the income process exhibits a unit root (see, e.g., Browning et al., 2010). As autocovariances are not well defined in the presence of a unit root, we require a different procedure to handle unit root cases.

Third, whereas this paper focuses only on balanced panel data, an analysis based on unbalanced panel data would be useful. We believe that, at least in terms of implementation, this extension is not too difficult. This is because there is no difficulty in estimating the mean and autocovariances for each individual, even when the panel is unbalanced, and there is no change in the procedure after obtaining the individual mean and autocovariance estimates. However, theoretical investigation of the properties of the procedure using an unbalanced panel may not be straightforward.

A Technical appendix

This appendix presents the proofs of the theorems and technical lemmas used to prove the theorems. Section A.1 contains the proofs of the theorems in the main text. The technical lemmas are given in Section A.2.

A.1 Proofs of the theorems

This section contains the proofs of the theorems in the main text. We repeatedly cite the results in van der Vaart and Wellner (1996), subsequently abbreviated as VW.

A.1.1 Proof of Theorem 1

The proof for $\mathbb{P}_N^{\hat{\mu}}$ and that for $\mathbb{P}_N^{\hat{\gamma}_k}$ are basically the same, so we present that for $\mathbb{P}_N^{\hat{\gamma}_k}$ only. Let $\mathbb{P}_N = \mathbb{P}_N^{\hat{\gamma}_k}$, $P_T = P_T^{\hat{\gamma}_k}$, and $P_0 = P_0^{\gamma_k}$.

By the triangle inequality,

$$\sup_{f \in \mathcal{F}} |\mathbb{P}_N f - P_0 f| \leq \sup_{f \in \mathcal{F}} |\mathbb{P}_N f - P_T f| + \sup_{f \in \mathcal{F}} |P_T f - P_0 f| .$$

For the second term on the right-hand side, Lemma 7 (for the case of $\mathbb{P}_N^{\hat{\mu}}$, Lemma 6) implies that $\hat{\gamma}_{k,i}$ converges to $\gamma_{k,i}$ in mean square convergence and thus also implies that $\hat{\gamma}_{k,i}$ converges to $\gamma_{k,i}$ in distribution. By Lemma 2.11 in van der Vaart (1998), it holds that

$$\sup_{f \in \mathcal{F}} |P_T f - P_0 f| \to 0,$$

since $\gamma_{k,i}$ is continuously distributed by Assumption 6 (for the case of $\mathbb{P}_N^{\hat{\mu}}$, Assumption 3).

We then show that the first term converges to 0 almost surely. Note that, for $f = \mathbf{1}_{(-\infty,a]}$, $\mathbb{P}_N f = \mathbb{F}_N^{\hat{\gamma}_{k,i}}(a)$ and $E(\mathbb{F}_N^{\hat{\gamma}_{k,i}}(a)) = \Pr(\hat{\gamma}_{k,i} \leq a) = P_T f$. We first fix a monotone sequence $T = T(N)$ such that $T \to \infty$ as $N \to \infty$, which makes our sample triangular arrays. We use the strong law of large numbers for triangular arrays (see, e.g., Hu, Móricz, and Taylor, 1989, Theorem 2). This is possible because under Assumption 1, $\mathbf{1}(\hat{\gamma}_{k,i} \leq a)$ for any a is i.i.d. across individuals, the condition (1.5) in Hu et al. (1989) is clearly satisfied, and the condition (1.6) in Hu et al. (1989) is also satisfied when we set $X = 2$ in the condition (1.6). Thus, we have $\mathbb{F}_N^{\hat{\gamma}_{k,i}}(a) - \Pr(\hat{\gamma}_{k,i} \leq a) \xrightarrow{as} 0$ and $\mathbb{F}_N^{\hat{\gamma}_{k,i}}(a-) - \Pr(\hat{\gamma}_{k,i} < a) \xrightarrow{as} 0$ for every $a \in \mathbb{R}$, as $N, T \to \infty$. Given a fixed $\varepsilon > 0$, there exists a partition $-\infty = a_0 < a_1 < \cdots < a_k = \infty$ such that $\Pr(\gamma_{k,i} < a_i) - \Pr(\gamma_{k,i} \leq a_{i-1}) < \varepsilon/3$ for every i. We have shown that $\sup_{f \in \mathcal{F}} |P_T f - P_0 f| \to 0$, and this implies that for sufficiently large N, T, $\sup_{f \in \mathcal{F}} |P_T f - P_0 f| < \varepsilon/3$. Therefore, we have $\Pr(\hat{\gamma}_{k,i} < a_i) - \Pr(\hat{\gamma}_{k,i} \leq a_{i-1}) < \varepsilon$ for every i. The rest of the proof is the same as the proof of Theorem 19.1 in van der Vaart (1998). For $a_{i-1} \leq a < a_i$,

$$\mathbb{F}_N^{\hat{\gamma}_{k,i}}(a) - \Pr(\hat{\gamma}_{k,i} \leq a) \leq \mathbb{F}_N^{\hat{\gamma}_{k,i}}(a_i-) - \Pr(\hat{\gamma}_{k,i} < a_i) + \varepsilon,$$
$$\mathbb{F}_N^{\hat{\gamma}_{k,i}}(a) - \Pr(\hat{\gamma}_{k,i} \leq a) \geq \mathbb{F}_N^{\hat{\gamma}_{k,i}}(a_{i-1}-) - \Pr(\hat{\gamma}_{k,i} < a_{i-1}) - \varepsilon.$$

Accordingly, we have $\limsup_{N,T\to\infty}(\sup_{f\in\mathcal{F}} |\mathbb{P}_N f - P_T f|) \leq \varepsilon$ almost surely. This is true for every $\varepsilon > 0$, and thus we get

$$\sup_{f \in \mathcal{F}} |\mathbb{P}_N f - P_T f| \xrightarrow{as} 0. \tag{17}$$

We note that (17) holds for all monotonic diagonal paths $N \to \infty, T(N) \to \infty$. As stated in REMARKS (a) in Phillips and Moon (1999), (17) thus holds under doubly asymptotics $N, T \to \infty$. Consequently, we obtain the desired result by the continuous mapping theorem.

□

A.1.2 Proof of Theorem 2

The proof for $\mathbb{P}_N^{\hat{\mu}}$ and that for $\mathbb{P}_N^{\hat{\gamma}_k}$ are basically the same, so we present that for $\mathbb{P}_N^{\hat{\gamma}_k}$ only. Let $\mathbb{P}_N = \mathbb{P}_N^{\hat{\gamma}_k}$, $P_T = P_T^{\hat{\gamma}_k}$, and $P_0 = P_0^{\gamma_k}$. The proof is based on the decomposition in (8) and (9). To study the asymptotic behavior of (8), we use Lemma 2.8.7 in VW. We first fix a monotone sequence $T = T(N)$ such that $T \to \infty$ as $N \to \infty$, which makes our sample triangular arrays. By Theorem 2.8.3, Example 2.5.4, and Example 2.3.4 in VW, the class $\mathcal{F}$ is Donsker and pre-Gaussian uniformly in $\{P_T\}$. Thus, we need to check the conditions (2.8.5) and (2.8.6) in VW. The condition (2.8.6) in VW is immediately satisfied by setting the envelope function $F = 1$ (constant).

We check the condition (2.8.5) in VW. Let ρ_{P_T} and ρ_{P_0} be the variance semimetrics with respect to P_T and P_0, respectively. Then,

$$\begin{aligned}
&\sup_{f,g\in\mathcal{F}} |\rho_{P_T}(f,g) - \rho_{P_0}(f,g)| \\
= &\sup_{f,g\in\mathcal{F}} |\sqrt{P_T((f-g)-P_T(f-g))^2} - \sqrt{P_0((f-g)-P_0(f-g))^2}| \\
= &\sup_{a,a'\in\mathbb{R}} |\sqrt{P_T(\mathbf{1}_{(-\infty,a]} - \mathbf{1}_{(-\infty,a']} - P_T(\mathbf{1}_{(-\infty,a]} - \mathbf{1}_{(-\infty,a']}))^2} \\
&-\sqrt{P_0(\mathbf{1}_{(-\infty,a]} - \mathbf{1}_{(-\infty,a']} - P_0(\mathbf{1}_{(-\infty,a]} - \mathbf{1}_{(-\infty,a']}))^2}| \\
\leq &\sup_{a,a'\in\mathbb{R}} |P_T(\mathbf{1}_{(-\infty,a]} - \mathbf{1}_{(-\infty,a']} - P_T(\mathbf{1}_{(-\infty,a]} - \mathbf{1}_{(-\infty,a']}))^2 \\
&-P_0(\mathbf{1}_{(-\infty,a]} - \mathbf{1}_{(-\infty,a']} - P_0(\mathbf{1}_{(-\infty,a]} - \mathbf{1}_{(-\infty,a']}))^2|^{1/2},
\end{aligned}$$

where the first inequality follows from the triangle inequality. Without loss of generality, we assume $a > a'$. Then, by simple algebra,

$$\begin{aligned}
&\sup_{f,g\in\mathcal{F}} |\rho_{P_T}(f,g) - \rho_{P_0}(f,g)| \\
\leq &\sup_{a,a'\in\mathbb{R}} \big|(P_T\mathbf{1}_{(-\infty,a]} - P_0\mathbf{1}_{(-\infty,a]}) - (P_T\mathbf{1}_{(-\infty,a]}\mathbf{1}_{(-\infty,a']} - P_0\mathbf{1}_{(-\infty,a]}\mathbf{1}_{(-\infty,a']}) \\
&+ (P_T\mathbf{1}_{(-\infty,a']} - P_0\mathbf{1}_{(-\infty,a']}) - ((P_T\mathbf{1}_{(-\infty,a]})^2 - (P_0\mathbf{1}_{(-\infty,a]})^2) \\
&- ((P_T\mathbf{1}_{(-\infty,a']})^2 - (P_0\mathbf{1}_{(-\infty,a]'})^2) + 2(P_T\mathbf{1}_{(-\infty,a]}P_T\mathbf{1}_{(-\infty,a']} - P_T\mathbf{1}_{(-\infty,a]}P_0\mathbf{1}_{(-\infty,a']}) \\
&+2(P_T\mathbf{1}_{(-\infty,a]}P_0\mathbf{1}_{(-\infty,a']} - P_0\mathbf{1}_{(-\infty,a]}P_0\mathbf{1}_{(-\infty,a']})\big|^{1/2} \\
\leq &11\sup_{a\in\mathbb{R}} \big|P_T\mathbf{1}_{(-\infty,a]} - P_0\mathbf{1}_{(-\infty,a]}\big|^{1/2} \\
\to &0,
\end{aligned}$$

where the last conclusion follows from Lemma 2.11 in van der Vaart (1998), and $\hat{\gamma}_{k,i} \xrightarrow{p} \gamma_{k,i}$, which follows from Lemma 7 (for the case of $\hat{\mu}_i$, Lemma 6). Therefore, condition (2.8.5) in VW is satisfied.

Therefore, by Lemma 2.8.7 in VW, we have shown that

$$\mathbb{G}_{N,P_T} \rightsquigarrow \mathbb{G}_{P_0} \quad \text{in} \quad \ell^\infty(\mathcal{F}). \tag{18}$$

We note that (18) holds for all monotonic diagonal paths $N\to\infty, T(N)\to\infty$. As stated in REMARKS (a) in Phillips and Moon (1999), (18) thus holds under doubly asymptotic $N, T \to \infty$.

Next, we study the asymptotic behavior of (9):

$$\sqrt{N}(P_T f - P_0 f).$$

Because the nonstochastic function sequence $P_T f - P_0 f$ is uniformly bounded in $f \in \mathcal{F}$, we should consider the convergence rate of

$$\sup_{f\in\mathcal{F}} \Big|P_T f - P_0 f\Big|.$$

Lemmas 7 and 8 (for the case of $\mathbb{P}_N^{\hat{\mu}}$, Lemmas 6 and 8) imply that

$$\sup_{f \in \mathcal{F}} \left| P_T f - P_0 f \right| = O\left(\frac{1}{\sqrt{T}}\right).$$

Therefore, given $N/T \to 0$, the desired result holds by Slutsky's theorem.

□

A.1.3 Proof of Theorem 3

By Taylor's theorem, we have the decomposition:

$$\begin{aligned}\sqrt{N}(\hat{G}_\mu - G_\mu) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(g(\mu_i) - E(g(\mu_i))) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left((\bar{y}_i - \mu_i)g'(\mu_i)\right) \\ &+ \frac{1}{2\sqrt{N}}\sum_{i=1}^{N}\left((\bar{y}_i - \mu_i)^2 g''(\tilde{\mu}_i)\right).\end{aligned}$$

The first term on the right-hand side converges in distribution to $N(0, var(g(\mu_i)))$ by Assumptions 1 and 7.

The second term on the right-hand side is

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left((\bar{y}_i - \mu_i)g'(\mu_i)\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\bar{w}_i g'(\mu_i),$$

and the expectation is

$$E\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\bar{w}_i g'(\mu_i)\right) = 0,$$

by the law of iterated expectations. The variance is

$$var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\bar{w}_i g'(\mu_i)\right) = E\left(\left(\bar{w}_i g'(\mu_i)\right)^2\right) \leq \sqrt{E(\bar{w}_i^4)}\sqrt{E(g'(\mu_i)^4)} = O(T^{-1}),$$

where the first inequality follows from the Cauchy–Schwarz inequality and the last equality follows from Lemma 4 and Assumption 7.

For the third term on the right-hand side,

$$E\left(\frac{1}{2\sqrt{N}}\sum_{i=1}^{N}\left((\bar{y}_i - \mu_i)^2 g''(\tilde{\mu}_i) - E(\bar{w}_i^2 g''(\tilde{\mu}_i))\right)\right) = 0,$$

and

$$\begin{aligned}& var\left(\frac{1}{2\sqrt{N}}\sum_{i=1}^{N}\left((\bar{y}_i - \mu_i)^2 g''(\tilde{\mu}_i) - E(\bar{w}_i^2 g''(\tilde{\mu}_i))\right)\right) \\ \leq\ & var(\bar{w}_i^2 g''(\tilde{\mu}_i)) \leq M \cdot var(\bar{w}_i^2) = O(T^{-2}),\end{aligned}$$

where the first inequality follows from the i.i.d. assumption, the second inequality follows from Assumption 7, and the last equality follows from Lemma 4.

Therefore, the first claim of the theorem is obtained by Markov's inequality and Slutsky's theorem. We also have the second claim of the theorem because $|E(\bar{w}_i^2 g''(\tilde{\mu}_i))| = O(T^{-1})$, which follows from Lemma 1 and Assumption 7.

$\square$

A.1.4 Proof of Theorem 4

We concentrate on proving the asymptotic normality of $\sqrt{N}(\hat{G}_{\gamma_k} - G_{\gamma_k})$, because it is clear that $\hat{G}_{\gamma_k}$ is consistent for G_{γ_k} by the following proof and the law of large numbers. By Taylor's theorem, we have

$$\begin{aligned}
\sqrt{N}(\hat{G}_{\gamma_k} - G_{\gamma_k}) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(g(\hat{\gamma}_{k,i}) - E(g(\gamma_{k,i}))) \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(g(\gamma_{k,i}) - E(g(\gamma_{k,i}))) \quad (19)\\
&\quad + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\hat{\gamma}_{k,i} - \gamma_{k,i})g'(\gamma_{k,i}) \quad (20)\\
&\quad + \frac{1}{2\sqrt{N}}\sum_{i=1}^{N}(\hat{\gamma}_{k,i} - \gamma_{k,i})^2 g''(\tilde{\gamma}_{k,i}), \quad (21)
\end{aligned}$$

where $\tilde{\gamma}_{k,i}$ is located between $\gamma_{k,i}$ and $\hat{\gamma}_{k,i}$. We examine each term in this expansion.

For (19), under Assumptions 1 and 10,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(g(\gamma_{k,i}) - E(g(\gamma_{k,i}))) \rightsquigarrow N(0, var(g(\gamma_{k,i}))),$$

by the central limit theorem.

For (20), we use the expansion for $\hat{\gamma}_{k,i}$. We have the following:

$$\begin{aligned}
\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\hat{\gamma}_{k,i} - \gamma_{k,i})g'(\gamma_{k,i}) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T} w_{it}w_{i,t-k} - \gamma_{k,i}\right)g'(\gamma_{k,i}) \quad (22)\\
&\quad - \frac{1}{\sqrt{N}}\frac{T+k}{T-k}\sum_{i=1}^{N}(\bar{w}_i)^2 g'(\gamma_{k,i}) \quad (23)\\
&\quad + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T-k}\sum_{t=1}^{k} w_{it}\bar{w}_i g'(\gamma_{k,i}) \quad (24)\\
&\quad + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T-k}\sum_{t=T-k+1}^{T} w_{it}\bar{w}_i g'(\gamma_{k,i}). \quad (25)
\end{aligned}$$

For (22), its expectation is

$$E\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T} w_{it}w_{i,t-k} - \gamma_{k,i}\right)g'(\gamma_{k,i})\right) = 0,$$

by the law of iterated expectations. The variance of (22) is

$$
\begin{aligned}
& var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T} w_{it}w_{i,t-k}-\gamma_{k,i}\right)g'(\gamma_{k,i})\right) \\
= \; & E\left(\left(\left(\frac{1}{T-k}\sum_{t=k+1}^{T} w_{it}w_{i,t-k}-\gamma_{k,i}\right)g'(\gamma_{k,i})\right)^2\right) \\
\leq \; & \sqrt{E\left(\left(\frac{1}{T-k}\sum_{t=k+1}^{T} w_{it}w_{i,t-k}-\gamma_{k,i}\right)^4\right)}\sqrt{E(g'(\gamma_{k,i})^4)},
\end{aligned}
$$

where the first equality follows from the i.i.d. assumption and the first inequality follows from the Cauchy–Schwarz inequality. We have

$$
E\left(\left(\frac{1}{T-k}\sum_{t=k+1}^{T}(w_{it}w_{i,t-k}-\gamma_{k,i})\right)^4\right) = O\left(\frac{1}{T^2}\right),
$$

by Lemma 5. This result and Assumption 10 imply that

$$
var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T} w_{it}w_{i,t-k}-\gamma_{k,i}\right)g'(\gamma_{k,i})\right) = O(T^{-1}).
$$

Therefore, by Markov's inequality, term (22) is of order $O_p(T^{-1/2})$.

We next examine (23). We observe that:

$$
\begin{aligned}
E\left|\frac{1}{\sqrt{N}}\frac{T+k}{T-k}\sum_{i=1}^{N}(\bar{w}_i)^2 g'(\gamma_{k,i})\right| &\leq \sqrt{N}\frac{T+k}{T-k}E\left|(\bar{w}_i)^2 g'(\gamma_{k,i})\right| \\
&\leq \sqrt{N}\frac{T+k}{T-k}\sqrt{E((\bar{w}_i)^4)}\sqrt{E((g'(\gamma_{k,i}))^2)} \\
&= O\left(\frac{\sqrt{N}}{T}\right),
\end{aligned}
$$

where the second inequality is the Cauchy–Schwarz inequality and the equality follows from Lemma 4 and Assumption 10. Thus, the term in (23) is $O_p(\sqrt{N}/T)$ by Markov's inequality.

For (24), we have by the Cauchy–Schwarz inequality that

$$
\begin{aligned}
E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T-k}\sum_{t=1}^{k} w_{it}\bar{w}_i g'(\gamma_{k,i})\right| &\leq \frac{\sqrt{N}}{T-k}E\left|\sum_{t=1}^{k} w_{it}\bar{w}_i g'(\gamma_{k,i})\right| \\
&\leq \frac{\sqrt{N}}{T-k}\sqrt{E\left[\left(\sum_{t=1}^{k} w_{it}\bar{w}_i\right)^2\right]}\sqrt{E[(g'(\gamma_{k,i}))^2]}.
\end{aligned}
$$

It holds that

$$
E\left(\left(\sum_{t=1}^{k} w_{it}\bar{w}_i\right)^2\right) \leq \left(E\left(\left(\sum_{t=1}^{k} w_{it}\right)^4\right)\right)^{1/2}\left(E\left((\bar{w}_i)^4\right)\right)^{1/2},
$$

by the Cauchy–Schwarz inequality. Lemma 4 implies that $E\left((\bar{w}_i)^4\right) = O(T^{-2})$ under Assumptions 8 and 9. It is easy to see that $E((\sum_{t=1}^k w_{it})^4) = O(k^4)$ by Assumption 9. Thus, it follows that

$$E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{1}{T-k}\sum_{t=1}^k w_{it}\bar{w}_i g'(\gamma_{k,i})\right| = O\left(\frac{\sqrt{kN}}{\sqrt{T}(T-k)}\right),$$

and the fourth term (24) is of order $O_p\left(\sqrt{kN}/(\sqrt{T}(T-k))\right)$ by Markov's inequality. The same argument can be applied to (25), and gives

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{1}{T-k}\sum_{t=T-k+1}^T w_{it}\bar{w}_i g'(\gamma_{k,i}) = O_p\left(\frac{\sqrt{kN}}{\sqrt{T}(T-k)}\right).$$

For (21),

$$\begin{aligned}
E\left|\frac{1}{2\sqrt{N}}\sum_{i=1}^N (\hat{\gamma}_{k,i}-\gamma_{k,i})^2 g''(\tilde{\gamma}_{k,i})\right| &\leq \frac{\sqrt{N}}{2} E\left|(\hat{\gamma}_{k,i}-\gamma_{k,i})^2 g''(\tilde{\gamma}_{k,i})\right| \\
&\leq \frac{\sqrt{N}}{2} M \cdot E\left((\hat{\gamma}_{k,i}-\gamma_{k,i})^2\right) \\
&= O\left(\frac{\sqrt{N}}{T}\right),
\end{aligned}$$

where the second inequality follows from Assumption 10 and the last equality follows from Lemma 7. By Markov's inequality, (21) is of order $O_p(\sqrt{N}/T)$.

Consequently, we obtain the desired result using Slutsky's theorem.

□

A.1.5 Proof of Theorem 5

We show only the asymptotic normality of $\sqrt{N}(\hat{H}-H)$, because the consistency of $\hat{H}$ is clear by the following proof and the law of large numbers. Let $\hat{\theta}_i = (\hat{\theta}_{i,1}, \ldots, \hat{\theta}_{i,l})$ and $\theta_i = (\theta_{i,1}, \ldots, \theta_{i,l})$. By Taylor's theorem, we have

$$\begin{aligned}
&\sqrt{N}(\hat{H}-H) \\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^N \left(h(\hat{\theta}_i) - E\left(h(\theta_i)\right)\right) \\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^N \left(h(\theta_i) - E\left(h(\theta_i)\right)\right) && (26)\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^N\sum_{j=1}^l (\hat{\theta}_{i,j}-\theta_{i,j})\frac{\partial}{\partial z_j}h(z)\Big|_{z=\theta_i} && (27)\\
&+\frac{1}{2\sqrt{N}}\sum_{i=1}^N \sum_{\sum_{s=1}^l j_s = 2} (\hat{\theta}_{i,1}-\theta_{i,1})^{j_1}\cdots(\hat{\theta}_{i,l}-\theta_{i,l})^{j_l}\frac{\partial^2}{\partial z_1^{j_1}\cdots\partial z_l^{j_l}}h(z)\Big|_{z=\tilde{\theta}_i}, && (28)
\end{aligned}$$

where $\tilde{\theta}_i$ is located between θ_i and $\hat{\theta}_i$.

For (26), under Assumption 12, it holds that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(h(\theta_i) - E(h(\theta_i))) \rightsquigarrow N(0, var(h(\theta_i))),$$

by the central limit theorem.

For (27), we have that for any $j = 1, \dots, l$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\hat{\theta}_{i,j} - \theta_{i,j})\frac{\partial}{\partial z_j}h(z)\Big|_{z=\theta_i} = O_p\left(\frac{\sqrt{N}}{T}\right),$$

which follows from the proof similar to that for Theorems 3 and 4 under Assumptions 1, 8, 9, 11, and 12.

For (28), we observe that

$$\begin{aligned}
&E\left|\frac{1}{2\sqrt{N}}\sum_{i=1}^{N}\sum_{\sum_{s=1}^{l} j_s=2}(\hat{\theta}_{i,1} - \theta_{i,1})^{j_1}\cdots(\hat{\theta}_{i,l} - \theta_{i,l})^{j_l}\frac{\partial^2}{\partial z_1^{j_1}\cdots\partial z_l^{j_l}}h(z)\Big|_{z=\tilde{\theta}_i}\right| \\
&\leq \frac{\sqrt{N}}{2}M\sum_{\sum_{s=1}^{l} j_s=2}E\left|(\hat{\theta}_{i,1} - \theta_{i,1})^{j_1}\cdots(\hat{\theta}_{i,l} - \theta_{i,l})^{j_l}\right|,
\end{aligned}$$

by Assumption 11 and the triangular inequality. Note that for any $k_1, k_2 = 1, \dots, l$,

$$\begin{aligned}
E|(\hat{\theta}_{i,k_1} - \theta_{i,k_1})(\hat{\theta}_{i,k_2} - \theta_{i,k_2})| &\leq \sqrt{E\left((\hat{\theta}_{i,k_1} - \theta_{i,k_1})^2\right)}\sqrt{E\left((\hat{\theta}_{i,k_2} - \theta_{i,k_2})^2\right)} \\
&= O(T^{-1}),
\end{aligned}$$

where the inequality follows from the Cauchy–Schwarz inequality and the equality follows from Lemmas 6 and 7 under Assumptions 1, 8, and 9. Hence, it holds that

$$E\left|\frac{1}{2\sqrt{N}}\sum_{i=1}^{N}\sum_{\sum_{s=1}^{l} j_s=2}(\hat{\theta}_{i,1} - \theta_{i,1})^{j_1}\cdots(\hat{\theta}_{i,l} - \theta_{i,l})^{j_l}\frac{\partial^2}{\partial z_1^{j_1}\cdots\partial z_l^{j_l}}h(z)\Big|_{z=\tilde{\theta}_i}\right| = O\left(\frac{\sqrt{N}}{T}\right).$$

Therefore, (28) is $O_p(\sqrt{N}/T)$ by Markov's inequality.

Consequently, we obtain the desired result using Slutsky's theorem.

□

A.1.6 Proof of Theorem 6

We first consider $\hat{G}_\mu$. The Taylor expansion gives

$$\begin{aligned}
\sqrt{N}(\hat{G}_\mu - G_\mu) =& \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(g(\mu_i) - E(g(\mu_i))) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\bar{w}_i g'(\mu_i) \\
&+ \frac{1}{2\sqrt{N}}\sum_{i=1}^{N}(\bar{w}_i)^2 g''(\mu_i) + \frac{1}{6\sqrt{N}}\sum_{i=1}^{N}(\bar{w}_i)^3 g'''(\tilde{\mu}_i),
\end{aligned}$$

where $\tilde{\mu}_i$ is between μ_i and $\bar{y}_i$. As in the proof of Theorem 3, we have $\sum_{i=1}^{N} \bar{w}_i g'(\mu_i)/\sqrt{N}$ is $O_p(1/\sqrt{T})$. We show that

$$\frac{1}{2\sqrt{N}} \sum_{i=1}^{N} (\bar{w}_i)^2 g''(\mu_i) = \frac{\sqrt{N}}{T} B + o_p\left(\frac{\sqrt{N}}{T}\right), \tag{29}$$

$$\frac{1}{6\sqrt{N}} \sum_{i=1}^{N} (\bar{w}_i)^3 g'''(\tilde{\mu}_i) = o_p\left(\frac{\sqrt{N}}{T}\right), \tag{30}$$

for some constant B. When (29) and (30) hold, the asymptotic normality of the HPJ estimator is established following the argument in Dhaene and Jochmans (2014) when $N/T^2 \to r$, where $r \in [0, \infty)$ is some constant. By (29), (30), and simple algebra, we have

$$\sqrt{N}(\hat{G}_\mu - G_\mu) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (g(\mu_i) - E(g(\mu_i))) + \frac{\sqrt{N}}{T} B + o_p\left(\frac{\sqrt{N}}{T}\right),$$

$$\sqrt{N}(\bar{G}_\mu - G_\mu) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (g(\mu_i) - E(g(\mu_i))) - 2\frac{\sqrt{N}}{T} B + o_p\left(\frac{\sqrt{N}}{T}\right).$$

That is, the bias of order $1/T$ of $\bar{G}_\mu$ is twice as large as that of $\hat{G}_\mu$. By the central limit theorem, we have

$$\sqrt{N} \begin{pmatrix} \hat{G}_\mu - G_\mu - T^{-1} B \\ \bar{G}_\mu - G_\mu - 2T^{-1} B_1 \end{pmatrix} \rightsquigarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} var(g(\mu_i)) & var(g(\mu_i)) \\ var(g(\mu_i)) & var(g(\mu_i)) \end{pmatrix} \right),$$

as $N, T \to \infty$ and $N/T^2 \to r$. Consequently, when $N, T \to \infty$ and $N/T^2 \to r$, we have

$$\sqrt{N}(\hat{G}_\mu^H - G_\mu) \rightsquigarrow N(0, var(g(\mu_i))),$$

by the continuous mapping theorem.

We first prove (29). We note that

$$E\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} ((\bar{w}_i)^2 g''(\mu_i))\right) = \frac{\sqrt{N}}{T} E(V_{T,i} g''(\mu_i)),$$

where $V_{T,i} := TE((\bar{w}_i)^2 | i) = \sum_{j=-T}^{T} \gamma_{j,i}(T - |j|)/T$. The variance is

$$var\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} ((\bar{w}_i)^2 g''(\mu_i))\right) = var((\bar{w}_i)^2 g''(\mu_i)).$$

We have

$$\begin{aligned} var((\bar{w}_i)^2 g''(\mu_i)) &\le E((\bar{w}_i)^4 (g''(\mu_i))^2) \\ &\le \sqrt{E((\bar{w}_i)^8)} \sqrt{E((g''(\mu_i))^4)} \\ &\le \frac{1}{T^2} C \sqrt{E((g''(\mu_i))^4)} = O\left(\frac{1}{T^2}\right), \end{aligned}$$

where the second inequality is the Cauchy–Schwarz inequality, the third inequality follows from Assumptions 8 and 9 and Lemma 4, and the last equality follows from Assumption 13. Therefore, (29) holds with $B = \lim_{T \to \infty} E(V_{T,i} g''(\mu_i))/2$.

Next, we prove (30). We have

$$\begin{aligned}E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\bar{w}_i)^3 g'''(\tilde{\mu}_i)\right| \leq & ME\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left|(\bar{w}_i)^3\right|\right)\\ \leq & M\sqrt{N}E(|\bar{w}_i|^3)\\ \leq & M\sqrt{N}\left(E\left((\bar{w}_i)^4\right)\right)^{\frac{3}{4}},\end{aligned}$$

where the first inequality follows from Assumption 13 and the triangle inequality, the second inequality follows from the i.i.d. assumption, and the third inequality is the Lyapunov's inequality. Thus, we only need to evaluate the order of $E((\bar{w}_i)^4)$. Lemma 4 implies $E((\bar{w}_i)^4) = O(1/T^2)$. (30) therefore holds because

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\bar{w}_i)^3 g'''(\tilde{\mu}_i) = O_p\left(\frac{\sqrt{N}}{T\sqrt{T}}\right) = o_p\left(\frac{\sqrt{N}}{T}\right),$$

by the Markov inequality. Thus, the asymptotic normality of $\hat{G}_{\mu}^{H}$ is proved.

Next, we consider $\hat{G}_{\gamma_k}$. The Taylor expansion gives

$$\begin{aligned}\sqrt{N}(\hat{G}_{\gamma_k} - G_{\gamma_k}) = & \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(g(\gamma_{k,i}) - E(g(\gamma_{k,i}))\right)\\ & - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\bar{w}_i)^2 g'(\gamma_{k,i}) + \frac{1}{2\sqrt{N}}\sum_{i=1}^{N}(\hat{\gamma}_{k,i} - \gamma_{k,i})^2 g''(\gamma_{k,i})\\ & + \frac{1}{3!\sqrt{N}}\sum_{i=1}^{N}(\hat{\gamma}_{k,i} - \gamma_{k,i})^3 g'''(\tilde{\gamma}_{k,i}) + O_p\left(\frac{\sqrt{N}}{T^2}\right).\end{aligned} \tag{31}$$

The second and third terms on the right-hand side of (31) are of order $O_p(\sqrt{N}/T)$. To establish the asymptotic normality of the HPJ estimator, we focus on showing the following results:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\bar{w}_i)^2 g'(\gamma_{k,i}) = \frac{\sqrt{N}}{T}B_1 + o_p\left(\frac{\sqrt{N}}{T}\right), \tag{32}$$

$$\frac{1}{2\sqrt{N}}\sum_{i=1}^{N}(\hat{\gamma}_{k,i} - \gamma_{k,i})^2 g''(\gamma_{k,i}) = \frac{\sqrt{N}}{T}B_2 + o_p\left(\frac{\sqrt{N}}{T}\right), \tag{33}$$

$$\frac{1}{3!\sqrt{N}}\sum_{i=1}^{N}(\hat{\gamma}_{k,i} - \gamma_{k,i})^3 g'''(\tilde{\gamma}_{k,i}) = o_p\left(\frac{\sqrt{N}}{T}\right), \tag{34}$$

where B_1 and B_2 are constants. When (32), (33), and (34) hold, we can show the asymptotic normality of the HPJ estimator following an argument similar to that for G_{μ} that is based on Dhaene and Jochmans (2014).

Therefore, we focus on showing (32), (33), and (34) separately.

The incidental parameter bias; Term (32) We first note by the proof of Theorem 4 that

$$E\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left((\bar{w}_i)^2 g'(\gamma_{k,i})\right)\right) = \frac{\sqrt{N}}{T}E(V_{T,i}g'(\gamma_{k,i})) = O\left(\frac{\sqrt{N}}{T}\right),$$

where $V_{T,i} := TE((\bar{w}_i)^2|i) = \sum_{j=-T}^{T} \gamma_{j,i}(T-|j|)/T$. The variance is

$$var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left((\bar{w}_i)^2 g'(\gamma_{k,i})\right)\right) = var((\bar{w}_i)^2 g'(\gamma_{k,i})),$$

by the i.i.d. assumption. We have

$$\begin{aligned}
var((\bar{w}_i)^2 g'(\gamma_{k,i})) &\leq E((\bar{w}_i)^4 (g'(\gamma_{k,i}))^2) \\
&\leq \sqrt{E((\bar{w}_i)^8)}\sqrt{E((g'(\gamma_{k,i}))^4)} \\
&\leq \frac{1}{T^2} C\sqrt{E((g'(\gamma_{k,i}))^4)} = O\left(\frac{1}{T^2}\right),
\end{aligned}$$

where the second inequality is the Cauchy–Schwarz inequality, the third inequality follows from Assumptions 8 and 14 and Lemma 4, and the last equality follows from Assumption 15. We have thus shown that (32) holds with $B_1 = \lim_{T\to\infty} E(V_{T,i} g'(\gamma_{k,i}))$.

The bias caused by the nonlinearity of g; Term (33) We shall compute the expectation of

$$A := \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T} w_{it}w_{i,t-k} - \gamma_{k,i}\right)^2 g''(\gamma_{k,i}),$$

and will show that

$$E\left((A - E(A))^2\right) = var(A) = o(N/T^2).$$

Under this condition, (33) is established by Loéve's c_r inequality and the proof of Theorem 4.

We examine the order of $E(A)$. We observe that

$$E(A) = \frac{\sqrt{N}}{(T-k)^2} E\left(\left(\sum_{t=k+1}^{T}(w_{it}w_{i,t-k} - \gamma_{k,i})\right)^2 g''(\gamma_{k,i})\right).$$

By the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
&E\left(\left(\sum_{t=k+1}^{T}(w_{it}w_{i,t-k} - \gamma_{k,i})\right)^2 g''(\gamma_{k,i})\right) \\
&\leq \sqrt{E\left(\left(\sum_{t=k+1}^{T}(w_{it}w_{i,t-k} - \gamma_{k,i})\right)^4\right)}\sqrt{E((g''(\gamma_{k,i}))^2)}.
\end{aligned}$$

Thus, Assumption 15 and Lemma 5 imply that

$$|E(A)| = O\left(\frac{\sqrt{N}}{T}\right).$$

Further, $(T/\sqrt{N})E(A)$ converges by the dominated convergence theorem under Assumption 15. We set $B_2 = \lim_{T\to\infty}(T/\sqrt{N})E(A)$ (note that $(T/\sqrt{N})E(A)$ does not depend on N).

We next examine $E((A - E(A))^2) = var(A)$ and show that it is of order $O(1/T^2)$. We first note that

$$var(A) = var\left(\left(\frac{1}{T-k}\sum_{t=k+1}^{T} w_{it}w_{i,t-k} - \gamma_{k,i}\right)^2 g''(\gamma_{k,i})\right),$$

by the i.i.d. assumption. It then holds that

$$\begin{aligned}
&var\left(\left(\frac{1}{T-k}\sum_{t=k+1}^{T} w_{it}w_{i,t-k} - \gamma_{k,i}\right)^2 g''(\gamma_{k,i})\right) \\
&\leq E\left(\left(\frac{1}{T-k}\sum_{t=k+1}^{T} w_{it}w_{i,t-k} - \gamma_{k,i}\right)^4 (g''(\gamma_{k,i}))^2\right) \\
&\leq \sqrt{E\left(\left(\frac{1}{T-k}\sum_{t=k+1}^{T} w_{it}w_{i,t-k} - \gamma_{k,i}\right)^8\right)}\sqrt{E((g''(\gamma_{k,i}))^4)} \\
&= O\left(\frac{1}{T^2}\right) = o\left(\frac{N}{T^2}\right),
\end{aligned}$$

where the second inequality is the Cauchy–Schwarz inequality and the third equality follows from Assumption 15 and Lemma 5.

It is therefore shown that (33) holds with $B_2 = \lim_{T\to\infty}(T/\sqrt{N})E(A)$ by Markov's inequality.

The third-order term; Term (34) We have

$$\begin{aligned}
E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\hat{\gamma}_{k,i} - \gamma_{k,i})^3 g'''(\tilde{\gamma}_{k,i})\right| \leq & ME\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left|(\hat{\gamma}_{k,i} - \gamma_{k,i})^3\right|\right) \\
\leq & M\sqrt{N}E(|\hat{\gamma}_{k,i} - \gamma_{k,i}|^3) \\
\leq & M\sqrt{N}\left(E\left((\hat{\gamma}_{k,i} - \gamma_{k,i})^4\right)\right)^{\frac{3}{4}},
\end{aligned} \tag{35}$$

where the first inequality follows from Assumption 15 and the triangle inequality, the second inequality follows from the i.i.d. assumption, and the third inequality is the Lyapunov's inequality. Thus, we only need to evaluate the order of $E((\hat{\gamma}_{k,i} - \gamma_{k,i})^4)$. We can write:

$$\begin{aligned}
E((\hat{\gamma}_{k,i} - \gamma_{k,i})^4) = & E\left(\left(\frac{1}{T-k}\sum_{t=k+1}^{T}(w_{it}w_{i,t-k} - \gamma_{k,i}) - \frac{T+k}{T-k}(\bar{w}_i)^2\right.\right. \\
& \left.\left. + \frac{1}{T-k}\sum_{t=1}^{k} w_{it}\bar{w}_i + \frac{1}{T-k}\sum_{t=T-k+1}^{T} w_{it}\bar{w}_i\right)^4\right).
\end{aligned} \tag{36}$$

Thanks to Loéve's c_r inequality, we only need to examine the fourth-order moment of each term in parentheses on the right-hand side of (36).

For the first term, we have

$$E\left(\left(\frac{1}{T-k}\sum_{t=k+1}^{T}(w_{it}w_{i,t-k} - \gamma_{k,i})\right)^4\right) = O\left(\frac{1}{T^2}\right),$$

by Lemma 5.

For the second term in (36), we first note that $(T+k)/(T-k) = O(1)$. We observe that

$$E(((\bar{w}_i)^2)^4) = E((\bar{w}_i)^8) = O\left(\frac{1}{T^4}\right),$$

by Lemma 4. We thus have that

$$E\left(\left(\frac{T+k}{T-k}(\bar{w}_i)^2\right)^4\right) = O\left(\frac{1}{T^4}\right).$$

For the third term, we first observe that by the Cauchy–Schwarz inequality,

$$\begin{aligned} E\left(\left(\frac{1}{T-k}\sum_{t=1}^{k} w_{it}\bar{w}_i\right)^4\right) &= E\left((\bar{w}_i)^4\left(\frac{1}{T-k}\sum_{t=1}^{k} w_{it}\right)^4\right) \\ &\leq \left(E\left((\bar{w}_i)^8\right)\right)^{1/2}\left(E\left(\left(\frac{1}{T-k}\sum_{t=1}^{k} w_{it}\right)^8\right)\right)^{1/2}. \end{aligned}$$

It is shown in the discussion on the second term that $E\left((\bar{w}_i)^8\right)$ is of order $1/T^4$. Moreover, because k is fixed, it is easy to see that

$$E\left(\left(\frac{1}{T-k}\sum_{t=1}^{k} w_{it}\right)^8\right) = O\left(\frac{1}{T^8}\right).$$

Therefore, it holds that

$$E\left(\left(\frac{1}{T-k}\sum_{t=1}^{k} w_{it}\bar{w}_i\right)^4\right) = O\left(\frac{1}{T^6}\right).$$

The same argument can be used to show that

$$E\left(\left(\frac{1}{T-k}\sum_{t=T-k+1}^{T} w_{it}\bar{w}_i\right)^4\right) = O\left(\frac{1}{T^6}\right).$$

Thus, we have shown that

$$E((\hat{\gamma}_{k,i} - \gamma_{k,i})^4) = O\left(\frac{1}{T^2}\right). \tag{37}$$

Consequently, we have

$$E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\hat{\gamma}_{k,i} - \gamma_{k,i})^3 g'''(\tilde{\gamma}_{k,i})\right| = O\left(\frac{\sqrt{N}}{T\sqrt{T}}\right) = o\left(\frac{\sqrt{N}}{T}\right). \tag{38}$$

Therefore, we get the desired result (34) by Markov's inequality.

□

A.1.7 Proof of Theorem 7

As the proofs for $\hat{G}_{\mu}^{*}$ and $\hat{G}_{\gamma_k}^{*}$ are identical, we discuss the case for $\hat{G}_{\gamma_k}^{*}$ only. We first show that the moments of $\hat{G}_{\gamma_k}^{*}$ under the bootstrap distribution satisfy Lyapunov's conditions. It is then proved that $\sqrt{N}(\hat{G}_{\gamma_k}^{*} - \hat{G}_{\gamma_k})$ converges in distribution to Z_{γ_k} almost surely under a subsequence of any subsequences of the original sequence. This implies that the bootstrap distribution of $\sqrt{N}(\hat{G}_{\gamma_k}^{*} - \hat{G}_{\gamma_k})$ converges almost surely under a subsequence of any subsequences. It then implies that it converges in probability in the original sequence.

We first examine the moments of $\hat{G}_{\gamma_k}^{*}$. By definition, $\hat{G}_{\gamma_k}^{*}$ is the sample average of $g(\hat{\gamma}_{k,i}^{*})$, where $\hat{\gamma}_{k,i}^{*}$ is the sample autocovariance of the time series z_i^{*}. It is easy to see that conditionally on the data, the mean and the variance of $g(\hat{\gamma}_{k,i}^{*})$ are

$$E(g(\hat{\gamma}_{k,i}^{*})|\{\{y_{it}\}_{t=1}^{T}\}_{i=1}^{N}) = \frac{1}{N}\sum_{i=1}^{N} g(\hat{\gamma}_{k,i}) = \hat{G}_{\gamma_k},$$

$$var(g(\hat{\gamma}_{k,i}^{*})|\{\{y_{it}\}_{t=1}^{T}\}_{i=1}^{N}) = \frac{1}{N}\sum_{i=1}^{N}(g(\hat{\gamma}_{k,i}) - \hat{G}_{\gamma_k})^2 = \frac{1}{N}\sum_{i=1}^{N} g(\hat{\gamma}_{k,i})^2 - (\hat{G}_{\gamma_k})^2.$$

The conditional variance converges to

$$E(g(\gamma_{k,i})^2) - (G_{\gamma_k})^2 = var(g(\gamma_{k,i})),$$

in probability by Theorem 4 and the continuous mapping theorem. We then consider the third-order moment. We note that

$$E\left(\frac{1}{\sqrt{N}}|g(\hat{\gamma}_{k,i}^{*}) - \hat{G}_{\gamma_k}|^3|\{\{y_{it}\}_{t=1}^{T}\}_{i=1}^{N}\right) = \frac{1}{N^{3/2}}\sum_{i=1}^{N}|g(\hat{\gamma}_{k,i}) - \hat{G}_{\gamma_k}|^3.$$

We note that

$$|g(\hat{\gamma}_{k,i}) - \hat{G}_{\gamma_k}|^3 \leq 4|g(\hat{\gamma}_{k,i})|^3 + 4|\hat{G}_{\gamma_k}|^3.$$

As $\hat{G}_{\gamma_k}$ converges, we have

$$\frac{1}{N^{1/2}}|\hat{G}_{\gamma_k}|^3 \xrightarrow{p} 0.$$

Let $h(\cdot)$ be a twice-differentiable function such that $h(a) \geq 0$, $h(a) = |a|^3$ for $|a| \geq 1$, and $|h(a) - |a|^3| < 1$ for $|a| < 1$. It follows that

$$\frac{1}{N^{3/2}}\sum_{i=1}^{N}|g(\hat{\gamma}_{k,i})|^3 \leq \frac{1}{N^{3/2}}\sum_{i=1}^{N} h(g(\hat{\gamma}_{k,i})) + \frac{1}{N^{3/2}}\sum_{i=1}^{N}|h(g(\hat{\gamma}_{k,i})) - |g(\hat{\gamma}_{k,i})|^3| \xrightarrow{p} 0,$$

as $N^{-1}\sum_{i=1}^{N} h(g(\hat{\gamma}_{k,i})) = O_p(1)$ by Theorem 4 under the condition of this theorem and the definition of $h(\cdot)$ implies that $\sum_{i=1}^{N}|h(g(\hat{\gamma}_{k,i})) - |g(\hat{\gamma}_{k,i})|^3|/N^{3/2} \leq 1/N^{1/2}$. Thus, we have

$$\frac{1}{N^{3/2}}\sum_{i=1}^{N}|g(\hat{\gamma}_{k,i}) - \hat{G}_{\gamma_k}|^3 \xrightarrow{p} 0.$$

We argue that for any subsequence of the original sequence, there exists a further subsequence under which $\sqrt{N}(\hat{G}^*_{\gamma_k} - \hat{G}_{\gamma_k})$ converges in distribution conditionally on $\{\{y_{it}\}_{t=1}^T\}_{i=1}^N$ almost surely. We have shown that the first, second, and third moments of $g(\hat{\gamma}^*_{k,i})$ satisfy Lyapunov's conditions in probability. For any subsequence of the original sequence, there thus exists a further subsequence under which these moment conditions are satisfied almost surely. Thus, under a subsequence of any subsequences, $\sqrt{N}(\hat{G}^*_{\gamma_k} - \hat{G}_{\gamma_k})$ converges in distribution to Z conditionally almost surely. This implies that for any subsequence, there exists a further subsequence under which

$$\sup_{x\in\mathbb{R}} \left| \Pr\left(\sqrt{N}(\hat{G}^*_{\gamma_k} - \hat{G}_{\gamma_k}) \leq x \,\middle|\, \{\{y_{it}\}_{t=1}^T\}_{i=1}^N \right) - \Pr(Z \leq x) \right|,$$

converges to 0 almost surely. It thus holds that, for the original sequence,

$$\sup_{x\in\mathbb{R}} \left| \Pr\left(\sqrt{N}(\hat{G}^*_{\gamma_k} - \hat{G}_{\gamma_k}) \leq x \,\middle|\, \{\{y_{it}\}_{t=1}^T\}_{i=1}^N \right) - \Pr(Z \leq x) \right| \xrightarrow{p} 0.$$

□

A.1.8 Proof of Theorem 9

We give only the proof for $KS_1^{\gamma_k}$ because that for KS_1^{μ} is the same. The proof is almost identical to the proof of Corollary 19.21 in van der Vaart (1998). We first note that, under $H_0^{\gamma_k}$, $\sqrt{N}(\mathbb{P}_N^{\hat{\gamma}_k} - Q^{\gamma_k}) \rightsquigarrow \mathbb{G}_{Q^{\gamma_k}}$ in $\ell^\infty(\mathcal{F})$ given $N, T \to \infty$ with $N/T \to 0$ by Theorem 2. Therefore, because the norm $\|\cdot\|_\infty$ on $D[-\infty,\infty]$, where $D[-\infty,\infty]$ is the class of all cadlag functions from $[-\infty,\infty]$ into $\mathbb{R}$, is continuous with respect to the uniform norm, we have $KS_1^{\gamma_k} \rightsquigarrow \|\mathbb{G}_{Q^{\gamma_k}}\|_\infty$ under $H_0^{\gamma_k}$ by the continuous mapping theorem.

□

A.1.9 Proof of Theorem 10

We present only the proof for $KS_2^{\gamma_k}$ because that for KS_2^{μ} is the same. We first observe that

$$KS_2^{\gamma_k} = \left\| \sqrt{\frac{N_1 N_2}{N_1+N_2}} (\mathbb{P}^{\hat{\gamma}_k}_{N_1,(1)} - P^{\gamma_k}_{0,(1)}) - \sqrt{\frac{N_1 N_2}{N_1+N_2}} (\mathbb{P}^{\hat{\gamma}_k}_{N_2,(2)} - P^{\gamma_k}_{0,(2)}) + \sqrt{\frac{N_1 N_2}{N_1+N_2}} (P^{\gamma_k}_{0,(1)} - P^{\gamma_k}_{0,(2)}) \right\|_\infty .$$

We note that, under Assumption 18, $\sqrt{N_1}(\mathbb{P}^{\hat{\gamma}_k}_{N_1,(1)} - P^{\gamma_k}_{0,(1)})$ and $\sqrt{N_2}(\mathbb{P}^{\hat{\gamma}_k}_{N_2,(2)} - P^{\gamma_k}_{0,(2)})$ jointly converge in distribution to independent Brownian processes $\mathbb{G}_{P^{\gamma_k}_{0,(1)}}$ and $\mathbb{G}_{P^{\gamma_k}_{0,(2)}}$ given $N_1, T_1 \to \infty$ with $N_1/T_1 \to 0$ and $N_2, T_2 \to \infty$ with $N_2/T_2 \to 0$ by Theorem 2. Therefore, under $H_0^{\gamma_k}: P^{\gamma_k}_{0,(1)} = P^{\gamma_k}_{0,(2)}$, $KS_2^{\gamma_k}$ converges in distribution to

$$\left\| \sqrt{1-\lambda}\,\mathbb{G}_{P^{\gamma_k}_{0,(1)}} - \sqrt{\lambda}\,\mathbb{G}_{P^{\gamma_k}_{0,(2)}} \right\|_\infty ,$$

by the continuous mapping theorem given $N_1/(N_1+N_2) \to \lambda \in (0,1)$. It is easy to see that the distribution of the limit random variable $\sqrt{1-\lambda}\,\mathbb{G}_{P^{\gamma_k}_{0,(1)}} - \sqrt{\lambda}\,\mathbb{G}_{P^{\gamma_k}_{0,(2)}}$ is identical to that of $\mathbb{G}_{P^{\gamma_k}_{0,(1)}}$ under $H_0^{\gamma_k}$. Thus, we have the desired result.

□

A.2 Technical lemmas

Lemma 1 (Galvao and Kato (2014) based on Davydov (1968)). *Let $\{\xi_t\}_{t=1}^{\infty}$ denote a stationary process taking values in $\mathbb{R}$ and let $\alpha(m)$ denote its α-mixing coefficients. Suppose that $E(|\xi_1|^q) < \infty$ and $\sum_{m=1}^{\infty} \alpha(m)^{1-2/q} < \infty$ for some $q > 2$. Then, we have*

$$var\left(\sum_{t=1}^{T} \xi_t\right) \leq CT$$

with $C = 12(E(|\xi_1|^q))^{2/q} \sum_{m=0}^{\infty} \alpha(m)^{1-2/q}$.

Proof. The proof is available in Galvao and Kato (2014) (the discussion after Theorem C.1). □

Lemma 2 (Yokoyama (1980)). *Let $\{\xi_t\}_{t=1}^{\infty}$ denote a strictly stationary α-mixing process taking values in $\mathbb{R}$, and let $\alpha(m)$ denote its α-mixing coefficients. Suppose that $E(\xi_t) = 0$ and for some constants $\delta > 0$ and $r > 2$, $E(|\xi_1|^{r+\delta}) < \infty$. If $\sum_{m=0}^{\infty}(m+1)^{r/2-1}\alpha(m)^{\delta/(r+\delta)} < \infty$, then there exists a constant C independent of T such that*

$$E\left(\left|\sum_{t=1}^{T} \xi_t\right|^r\right) \leq CT^{r/2}.$$

Lemma 3. *Suppose that Assumptions 1 and 4 hold. Then, $\{w_{it}w_{i,t-k}\}_{t=k+1}^{\infty}$ for a fixed k given α_i is strictly stationary and α-mixing and its mixing coefficients $\{\alpha_k(m|i)\}_{m=0}^{\infty}$ possess the following properties: there exists a sequence $\{\alpha_k(m)\}_{m=0}^{\infty}$ such that for any i and m, $\alpha_k(m|i) \leq \alpha_k(m)$ and $\sum_{m=0}^{\infty}(m+1)^3\alpha_k(m)^{\delta/(r+\delta)} < \infty$ for some $\delta > 0$ and $r = 4$. The result holds with $r = 8$ if Assumption 4 is replaced by Assumption 8.*

Proof. The proof is similar to the proof of Theorem 14.1 in Davidson (1994). It is easy to see that for any i and any $0 \leq m < k$, $\alpha_k(m|i) \leq 1$, and that for any i and any $m \geq k$, $\alpha_k(m|i) \leq \alpha(m-k|i) \leq \alpha(m-k)$ by the definition of α-mixing coefficients and Assumption 4 or 8. Thus, we have $\sum_{m=0}^{\infty}(m+1)^3\alpha_k(m)^{\delta/(r+\delta)} \leq \sum_{m=0}^{k-1}(m+1)^3 + \sum_{m=k}^{\infty}(m+1)^3\alpha(m-k)^{\delta/(r+\delta)} < \infty$ for $r = 4$ and 8 under Assumptions 4 and 8, respectively. Thus, the lemma holds. □

Lemma 4. *Suppose that Assumptions 1, 4, and 5 hold. Then, it holds that $E((\bar{w}_i)^r) \leq CT^{-r/2}$ for $r = 2, 4$ and some constant $C < \infty$. If Assumptions 8 and 9 hold additionally, $E((\bar{w}_i)^8) \leq CT^{-4}$ holds as well.*

Proof. We first consider the case with $r = 2$. Given $E(\bar{w}_i|i) = 0$, Lemma 1 states that

$$E\left((\bar{w}_i)^2|i\right) \leq C_i/T,$$

where $C_i = 12(E(|w_{it}|^{(4+\delta)/2}|i))^{4/(4+\delta)} \sum_{m=0}^{\infty} \alpha(m|i)^{\delta/(4+\delta)}$. Assumption 4 implies that

$$C_i \leq 12(E(|w_{it}|^{(4+\delta)/2}|i))^{4/(4+\delta)} \sum_{m=0}^{\infty} \alpha(m)^{\delta/(4+\delta)}.$$

Thus, we have

$$
\begin{aligned}
E\left((\bar{w}_i)^2\right) &= E\left(E\left((\bar{w}_i)^2|i\right)\right) \\
&\leq 12E\left((E(|w_{it}|^{(4+\delta)/2}|i))^{4/(4+\delta)}\right)\sum_{m=0}^{\infty}\alpha(m)^{\delta/(4+\delta)}/T \\
&\leq 12\left(E\left(E(|w_{it}|^{(4+\delta)/2}|i)\right)\right)^{4/(4+\delta)}\sum_{m=0}^{\infty}\alpha(m)^{\delta/(4+\delta)}/T \\
&= 12\left(E\left(|w_{it}|^{(4+\delta)/2}\right)\right)^{4/(4+\delta)}\sum_{m=0}^{\infty}\alpha(m)^{\delta/(4+\delta)}/T \\
&= O\left(\frac{1}{T}\right),
\end{aligned}
$$

where the second inequality is Jensen's inequality and the last equality follows from Assumptions 4 and 5. Hence, the desired result holds for $r = 2$.

Next, we consider the case with $r = 4, 8$. We use Lemma 2. From the proof of Lemma 2 available in Yokoyama (1980), we have

$$
E\left(\left|\sum_{t=1}^{T} w_{it}\right|^r \middle| i\right) \leq K_{r,i}\left(E\left(|w_{it}|^{r+\delta}|i\right)\right)^{r/(r+\delta)} T^{r/2}
$$

for some $\delta > 0$, where $K_{r,i}$ is a polynomial of $A_q(\alpha|i)$ for $q \leq r$ and $A_q(\alpha|i) := \sum_{m=0}^{\infty}(m+1)^{q/2-1}\alpha(m|i)^{\delta/(q+\delta)}$. Note that $A_q(\alpha|i) < \infty$ for $q \leq r$ if $A_r(\alpha|i) < \infty$. By Assumption 4 or 8, there exists a constant $K_r < \infty$ such that $K_{r,i} < K_r$ for all i. Thus, we have

$$
\begin{aligned}
E\left((\bar{w}_i)^r\right) = E\left(E\left((\bar{w}_i)^r|i\right)\right) &\leq K_r E\left(\left(E\left(|w_{it}|^{r+\delta}|i\right)\right)^{r/(r+\delta)}\right) T^{-r/2} \\
&\leq K_r\left(E\left(E\left(|w_{it}|^{r+\delta}|i\right)\right)\right)^{r/(r+\delta)} T^{-r/2} \\
&= K_r\left(E\left(|w_{it}|^{r+\delta}\right)\right)^{r/(r+\delta)} T^{-r/2} \\
&= O(T^{-r/2}),
\end{aligned}
$$

where the second inequality is Jensen's inequality and the last equality follows from Assumption 5 or 9. The proof for $r = 4, 8$ is complete. □

Lemma 5. *Suppose that Assumptions 1, 4, and 5 hold. Then, it holds that* $E((\sum_{t=k+1}^{T}(w_{it}w_{i,t-k} - \gamma_{k,i}))^r) \leq CT^{r/2}$ *for some constant* C *and* $r = 2$. *The result holds for* $r = 4$ *if Assumptions 4 and 5 are replaced by Assumptions 8 and 9. Furthermore, if Assumption 14 holds additionally, the result holds for* $r = 8$ *as well.*

Proof. In view of Lemma 3, the lemma follows along the same line as that of Lemma 4. □

Lemma 6. *Under Assumptions 1 and 2, we have*

$$
E\left((\bar{y}_i - \mu_i)^2\right) = O(T^{-1}).
$$

Proof. Note that $\bar{y}_i = \mu_i + \bar{w}_i$ where $\bar{w}_i := T^{-1}\sum_{t=1}^{T} w_{it}$. Therefore, we have

$$E\left((\bar{y}_i - \mu_i)^2\right) = E\left(\bar{w}_i^2\right) = \frac{1}{T}E(V_{T,i}) = O(T^{-1}),$$

where $V_{T,i} := TE((\bar{w}_i)^2|i) = \sum_{j=-T}^{T} \gamma_{j,i}(T-|j|)/T$, the second equality follows from Assumption 1, and the last equality follows from Assumption 2.

□

Lemma 7. *Under Assumptions 1, 4, and 5, we have*

$$E((\hat{\gamma}_{k,i} - \gamma_{k,i})^2) = O(T^{-1}).$$

Proof. Given the estimator $\hat{\gamma}_{k,i}$ has the following decomposition:

$$\begin{aligned}\hat{\gamma}_{k,i} &= \frac{1}{T-k}\sum_{t=k+1}^{T} w_{it}w_{i,t-k} - \frac{T+k}{T-k}(\bar{w}_i)^2 \\ &\quad + \frac{1}{T-k}\sum_{t=1}^{k} w_{it}\bar{w}_i + \frac{1}{T-k}\sum_{t=T-k+1}^{T} w_{it}\bar{w}_i,\end{aligned}$$

we can write:

$$\begin{aligned}E((\hat{\gamma}_{k,i} - \gamma_{k,i})^2) =& E\left(\left(\frac{1}{T-k}\sum_{t=k+1}^{T}(w_{it}w_{i,t-k} - \gamma_{k,i}) - \frac{T+k}{T-k}(\bar{w}_i)^2\right.\right. \\ &\left.\left. + \frac{1}{T-k}\sum_{t=1}^{k} w_{it}\bar{w}_i + \frac{1}{T-k}\sum_{t=T-k+1}^{T} w_{it}\bar{w}_i\right)^2\right).\end{aligned} \tag{39}$$

Owing to Loéve's c_r inequality, we just need to examine the second-order moment of each term in parentheses on the right-hand side of (39).

For the first term, we have

$$E\left(\left(\frac{1}{T-k}\sum_{t=k+1}^{T}(w_{it}w_{i,t-k} - \gamma_{k,i})\right)^2\right) = O\left(\frac{1}{T}\right),$$

by Lemma 5.

For the second term in (39), we first note that $(T+k)/(T-k) = O(1)$. We observe that

$$E(((\bar{w}_i)^2)^2) = E((\bar{w}_i)^4) = O\left(\frac{1}{T^2}\right),$$

by Lemma 4. We thus have that

$$E\left(\left(\frac{T+k}{T-k}(\bar{w}_i)^2\right)^2\right) = O\left(\frac{1}{T^2}\right).$$

For the third term, we first observe that by the Cauchy–Schwarz inequality,

$$\begin{aligned}E\left(\left(\frac{1}{T-k}\sum_{t=1}^{k} w_{it}\bar{w}_i\right)^2\right) &= E\left((\bar{w}_i)^2\left(\frac{1}{T-k}\sum_{t=1}^{k} w_{it}\right)^2\right) \\ &\le \left(E\left((\bar{w}_i)^4\right)\right)^{1/2}\left(E\left(\left(\frac{1}{T-k}\sum_{t=1}^{k} w_{it}\right)^4\right)\right)^{1/2}.\end{aligned}$$

It is shown in the discussion on the second term that $E\left((\bar{w}_i)^4\right)$ is of order $1/T^2$. Moreover, because k is fixed, it is easy to see that

$$E\left(\left(\frac{1}{T-k}\sum_{t=1}^{k} w_{it}\right)^4\right) = O\left(\frac{1}{T^4}\right).$$

Therefore, it holds that

$$E\left(\left(\frac{1}{T-k}\sum_{t=1}^{k} w_{it}\bar{w}_i\right)^2\right) = O\left(\frac{1}{T^3}\right).$$

The same argument can be used to show that

$$E\left(\left(\frac{1}{T-k}\sum_{t=T-k+1}^{T} w_{it}\bar{w}_i\right)^2\right) = O\left(\frac{1}{T^3}\right).$$

We have thus shown that all of the terms are $O(T^{-1})$ and the statement of the lemma holds. □

Lemma 8. *Let a_T and b_T be continuous random variables indexed by T with bounded joint density. Suppose that as $T \to \infty$,*

$$E(|a_T - b_T|^p) = O(T^c), \tag{40}$$

for some integer p and real number $c < 0$. It then holds that

$$\sup_x |\Pr(a_T < x) - \Pr(b_T < x)| = O(T^{2c/(2+p)}). \tag{41}$$

In particular, if $c = -1$ and $p = 2$, then $2c/(2+p) = -1/2$ and $\sup_x |\Pr(a_T < x) - \Pr(b_T < x)| = O(T^{-1/2})$.

Proof. We have

$$\Pr(a_T < x) = \Pr(a_T < x, b_T < x) + \Pr(a_T < x, b_T \geq x).$$

We take some $\epsilon > 0$. Then, we have

$$\begin{aligned}
&\Pr(a_T < x, b_T \geq x)\\
=&\Pr(a_T < x, b_T \geq x, |a_T - b_T| > \epsilon) + \Pr(a_T < x, b_T \geq x, |a_T - b_T| \leq \epsilon).
\end{aligned}$$

For the first probability on the right-hand side, we have

$$\sup_x \Pr(a_T < x, b_T \geq x, |a_T - b_T| > \epsilon) \leq \Pr(|a_T - b_T| > \epsilon) \leq \frac{E(|a_T - b_T|^p)}{\epsilon^p},$$

by Markov's inequality. For the second probability, we have

$$\begin{aligned}
\sup_x \Pr(a_T < x, b_T \geq x, |a_T - b_T| \leq \epsilon) \leq& \sup_x \Pr(x - \epsilon \leq a_T < x, x \leq b_T \leq x + \epsilon)\\
\leq& \epsilon^2 \sup_x \sup_{x-\epsilon \leq a < x, x \leq b \leq x+\epsilon} f_{a_T,b_T}(a, b)\\
\leq& \epsilon^2 C,
\end{aligned}$$

for some $C > 0$, where f_{a_T, b_T} is the joint density of a_T and b_T. Therefore, we have

$$\sup_x |\Pr(a_T < x) - \Pr(a_T < x, b_T < x)| \leq \frac{E(|a_T - b_T|^p)}{\epsilon^p} + \epsilon^2 C.$$

We now take $\epsilon = T^d$. Then, we have

$$\sup_x |\Pr(a_T < x) - \Pr(a_T < x, b_T < x)| = O\left(\frac{T^c}{T^{dp}} + T^{2d}\right).$$

We note that the above order is minimized by setting $d = c/(2+p)$. Thus, we have

$$\sup_x |\Pr(a_T < x) - \Pr(a_T < x, b_T < x)| = O\left(T^{2c/(2+p)}\right).$$

Similarly, we have

$$\sup_x |\Pr(b_T < x) - \Pr(a_T < x, b_T < x)| = O\left(T^{2c/(2+p)}\right).$$

Therefore, we have

$$\begin{aligned}
&\sup_x |\Pr(a_T < x) - \Pr(b_T < x)| \\
=&\sup_x |(\Pr(a_T < x) - \Pr(a_T < x, b_T < x)) - (\Pr(b_T < x) - \Pr(a_T < x, b_T < x))| \\
\leq&\sup_x |\Pr(a_T < x) - \Pr(a_T < x, b_T < x)| + \sup_x |\Pr(b_T < x) - \Pr(a_T < x, b_T < x)| \\
=&O\left(T^{2c/(2+p)}\right).
\end{aligned}$$

□

References

M. Arellano. *Panel Data Econometrics.* Oxford University Press, 2003.

M. Arellano and S. Bonhomme. Identifying distributional characteristics in random coefficients panel data models. *Review of Economic Studies*, 79:987–1020, 2012.

B. H. Baltagi. *Econometric Analysis of Panel Data.* Wiley, 4 edition, 2008.

S. Bonhomme and J.-M. Robin. Generalized non-parametric deconvolution with an application to earnings dynamics. *Review of Economic Studies*, 77:491–533, 2010.

M. Browning, M. Ejrnæs, and J. Alvarez. Modelling income processes with lots of heterogeneity. *Rview of Economic Studies*, 77:1353–1381, 2010.

G. Chamberlain. Efficiency bounds for semiparametric regression. *Econometrica*, 60(3):567–596, 1992.

J. Davidson. *Stochastic Limit Theory: An Introduction for Econometricians.* Oxford university press, 1994.

Y. A. Davydov. Convergence of distributions generated by stationary stochastic processes. *Theory of Probability & Its Applications*, 13(4):691–696, 1968.

G. Dhaene and K. Jochmans. Split-panel jackknife estimation of fixed effects models. mimeo, 2014.

K. Evdokimov. Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. mimeo, 2009.

I. Fernández-Val and J. Lee. Panel data models with nonadditive unobserved heterogeneity: Estimation and inference. *Quantitative Economics*, 4:453–481, 2013.

A. F. Galvao and K. Kato. Estimation and inference for linear panel data models under misspecification when both n and T are large. *Journal of Business & Economic Statistics*, 32 (2):285–309, 2014.

S. Gonçalves and M. Kaffo. Bootstrap inference for linear dynamic panel data models with individual fixed effects. mimeo, 2014.

B. S. Graham and J. L. Powell. Identification and estimation of average partial effects in "irregular" correlated random coefficient panel data models. *Econometrica*, 80(5):2105–2152, 2012.

F. Guvenen. Learning your earning: Are labor income shocks really very persistent? *American Economic Review*, 97(3):687–712, 2007.

J. L. Horowitz and M. Markatou. Semiparametric estimation of regression models for panel data. *Review of Economic Studies*, 63(1):145–168, 1996.

C. Hsiao, M. H. Pesaran, and A. K. Tahmiscioglu. Bayes estimation of short-run coefficients in dynamic panel data models ", in analysis of panels and limited dependent variables models. In K. L. C. Hsiao, L.F. Lee and M. Pesaran, editors, *Analysis of Panels and Limited Dependent Variables Models*, pages 268–296. Cambridge University press, 1999.

T.-C. Hu, F. Móricz, and R. Taylor. Strong laws of large numbers for arrays of rowwise independent random variables. *Acta Mathematica Hungarica*, 54(1):153–162, 1989.

M. Kaffo. Bootstrap inference for nonlinear dynamic panel data models with individual fixed effects. mimeo, 2014.

A. N. Kolmogorov. *Sulla determinazione empirica di una legge di distribuzione*. 1933.

Y. Lee. Bias in dynamic panel models under time series misspecification. *Journal of Econometrics*, 169:54–60, 2012.

Y.-J. Lee, R. Okui, and M. Shintani. Asymptotic inference for dynamic panel estimators of infinite order autoregressive processes. mimeo, 2013.

L. A. Lillard and R. J. Willis. Dynamic aspects of earning mobility. *Econometrica*, 46(5): 985–1012, 1978.

S. Mavroeidis, Y. Sasaki, and I. Welch. Estimation of heterogenous autoregressive parameters using short panel data. mimeo, 2014.

C. Meghir and L. Pistaferri. Income variance dynamics and heterogeneity. *Econometrics*, 72 (1):1–32, 2004.

J. Neyman and E. L. Scott. Consistent estimates based on partially consistent observations. *Econometrica*, 16:1–32, 1948.

S. Nickell. Biases in dynamic models with fixed effects. *Econometrica*, pages 1417–1426, 1981.

R. Okui. Panel AR(1) estimators under misspecification. *Economics Letters*, 101:210–213, 2008.

R. Okui. Asymptotically unbiased estimation of autocovariances and autocorrelations with long panel data. *Econometric Theory*, 26:1263–1304, 2010.

R. Okui. Asymptotically unbiased estimation of autocovariances and autocorrelations for panel data with incidental trends. *Economics Letters*, 112:49–52, 2011.

R. Okui. Asymptotically unbiased estimation of autocovariances and autocorrelations with panel data in the presence of individual and time effects. *Journal of Time Series Econometrics*, 6 (2):129–181, 2014.

M. H. Pesaran. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1002, 2006.

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

M. H. Pesaran, Y. Shin, and R. P. Smith. Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association*, 94(446):621–634, 1999.

P. C. B. Phillips and H. R. Moon. Linear regression limit theory for nonstationary panel data. *Econometrica*, 67(5):1057–1111, 1999.

M. H. Quenouille. Approximate tests of correlation in time-series 3. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 45, pages 483–484. Cambridge Univ Press, 1949.

M. H. Quenouille. Notes on bias in estimation. *Biometrika*, pages 353–360, 1956.

R. J. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Statistics. Wiley, 2002.

J. Shao. *Mathematical Statistics*. Springer, 2003.

N. V. Smirnov. Approximate laws of distribution of random variables from empirical data. *Uspekhi Matematicheskikh Nauk*, (10):179–206, 1944.

G. J. Székely and C. R. Rao. Identifiability of distributions of independent random variables by linear combinations and moments. *Sankhyä*, 62:193–202, 2000.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Process*. Springer, 1996.

R. Yokoyama. Moment bounds for stationary mixing sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 52(1):45–57, 1980.