Panel Data Analysis with Heterogeneous Dynamics∗

Ryo Okui† and Takahide Yanagi‡

March, 2016

Abstract

This paper proposes the analysis of panel data whose dynamic structure is heterogeneous across individuals. Our proposed method is easy to implement and does not rely on any specific model for the dynamics. We first compute the sample mean, autocovariances, and/or autocorrelations for each individual, and then estimate the parameter of interest based on the empirical distributions of the estimated mean, autocovariances, and/or autocorrelations. We illustrate the usefulness of our proposed procedures by applying them to the study of earnings and productivity dynamics and find that both exhibit substantial heterogeneity. We then investigate the asymptotic properties of the proposed estimators using double asymptotics under which both the cross-sectional sample size and the length of the time series tend to infinity. We prove the functional central limit theorem for the proposed distribution estimator. Further, if we can write the parameter of interest as the expectation of a smooth function of the individual mean and/or autocovariances, we can reduce bias using split-panel jackknife bias-correction. We also develop an inference procedure based on the cross-sectional bootstrap. The results of Monte Carlo simulations confirm the usefulness of our procedures in finite samples and show that our asymptotic results are informative regarding the finite-sample properties.

Keywords: Panel data; heterogeneity; autocorrelation structure; functional central limit theorem; jackknife; bootstrap.

JEL Classification: C13; C14; C23.

∗This paper was previously circulated under the title “Dynamic Panel Data Analysis when the Dynamics are Heterogeneous.”

†Department of Econometrics and OR, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, The Netherlands; and Institute of Economic Research, Kyoto University, Yoshida-Hommachi, Sakyo, Kyoto, Kyoto, 606-8501, Japan. Email: okui.ryo.3@gmail.com

‡Graduate School of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo, 186-8601, Japan. Email: t.yanagi@r.hit-u.ac.jp
1 Introduction

This paper develops the analysis of panel data whose dynamic structure is heterogeneous across individuals. We propose methods for estimating the distributional features of the mean, autocovariances, and autocorrelations that are heterogeneous across individuals using panel data. Our proposed procedure is simple to implement. We first estimate the mean, autocovariances, and/or autocorrelations for each individual. We then estimate the distribution and other distributional quantities using the empirical distributions of the estimated mean, autocovariances, and/or autocorrelations. When the parameter of interest can be written as the expected value of a smooth function of the heterogeneous mean and/or autocovariances, the split-panel jackknife reduces the bias of the estimator and the cross-sectional bootstrap is justified for statistical inference.

Understanding the dynamic nature of an economic variable that is potentially heterogeneous is an important research consideration in economics. For example, a considerable amount of study has been done 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 the present paper is to propose easy-to-implement methods to analyze panel data, \( \{y_{it}\}_{i=1}^{T} \), whose dynamics are heterogeneous, without assuming any specific model. To study the heterogeneous dynamic structure, we investigate the cross-sectional distributional properties of the mean, autocovariances, and autocorrelations of \( y_{it} \) 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 voluminous, many studies assume some specific model for the dynamics (such as the autoregressive (AR) model) and 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 panel data methods to analyze heterogeneous dynamics without specifying some particular model. This paper builds on the literature by proposing model-free analysis for heterogeneous dynamics.

The distributional properties of the heterogeneous mean, autocovariances, and autocor-

\(^1\) See, e.g., Baltagi (2008) and Arellano (2003) for excellent reviews of existing contributions on dynamic panel data analysis.
relations are informative in various ways. First, the mean, autocovariances, and autocorrelations are perhaps the most basic descriptive statistics for dynamics. Indeed, a typical first step in analyzing time-series data is to examine these properties of the data. As we demonstrate in this paper, the distributions of the heterogeneous mean, autocovariances, and autocorrelations can also be useful descriptive statistics for understanding heterogeneous dynamics in panel data. For example, understanding the properties of the income shocks individuals face is important in the literature on income dynamics (Meghir and Pistaferri, 2011). In this case, the mean and the variance of the heterogeneous variance respectively measure the average amount and variability of income shocks that people face. Moreover, our analysis provides the entire distribution of the variance of income shocks. We can also examine the correlations of the heterogeneous variance and heterogeneous autocorrelations that inform us whether the magnitude of a income shock faced by people relates to the persistency of the income dynamics. This focus by us contrasts our analysis with many existing analyses in the econometric literature that emphasize only the means of heterogeneous coefficients (such as for AR coefficients). Second, we can use the distributions to investigate whether different groups possess dissimilar dynamic structures without relying on some particular model. For instance, consider the situation in which we would like to investigate whether males and females face different income dynamics, but where we are also aware of the fact that income dynamics are heterogeneous across individuals. In such a case, we may estimate the distributions of the autocorrelations for males and females separately and compare them to see if they indeed differ. Third, we may use the estimated distributions to explore adequate parametric specifications for the dynamics.

We demonstrate the usefulness of the proposed procedures through two empirical applications. Our first application is to examine the heterogeneous dynamic structure of earnings dynamics using data from the Panel Study of Income Dynamics. We find that the dynamics exhibit substantial heterogeneity and that the degree of heterogeneity varies considerably by educational attainment (i.e., across high school and college graduates). Our second application analyzes the productivity dynamics of Chilean firms in the food industry. Our analysis again suggests considerable heterogeneity in the dynamics. We also find that firms with high permanent productivity face relatively large shocks, but that such shocks tend to disappear quickly.

The asymptotic properties of the empirical distributions of the estimated individual mean, autocovariances, and autocorrelations 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 under
a condition on the relative magnitude of $N$ and $T$. This condition is required because the empirical distributions include bias caused by the estimation error in the estimated mean, autocovariances, and autocorrelations for each individual. We also derive the asymptotic distributions of the estimators for other distributional characteristics, including the quantile function, 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 and/or autocovariances, we evaluate the exact order of the bias and provide a sharp condition on the relative magnitude of $N$ and $T$. This class of parameters includes the mean, the variance, and other moments (such as correlations) of the heterogeneous mean, autocovariances, and/or autocorrelations. 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 to reduce the bias. 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 econometric literature. This type of bias does not affect the estimated mean, but does influence the estimated autocovariances. The second source of bias arises when the smooth function is nonlinear. This bias affects both the mean and the autocovariances.

The bias in the estimator for the expected value of a smooth function can be reduced through use of the split-panel jackknife in Dhaene and Jochmans (2015). In particular, we consider the half-panel jackknife (HPJ), which can correct both the incidental parameter bias and the nonlinearity bias. We find the HPJ bias-corrected estimator is asymptotically unbiased even when the condition $N/T^2 \to 0$ is violated and, in addition, does not inflate the asymptotic variance.

We propose the cross-sectional bootstrap to conduct hypothesis testing and to construct confidence intervals. The bootstrap distribution is asymptotically equivalent to the distribution of the estimator under the condition on the relative magnitude of $N$ and $T$, which ensures that the bias of the estimator is asymptotically negligible.

Monte Carlo simulations based on the data generating processes imitating the application results allow us to investigate the finite-sample properties of the proposed procedures. These show that the estimators suffer from severe bias when $T$ is small compared with $N$, but the HPJ bias reduction is successful even when $T$ is relatively small. They also demonstrate the noticeable performance of the cross-sectional bootstrap.

**Related literature:** This paper most closely relates to the literature on heterogeneous panel AR models. In these models, the heterogeneity in the dynamics is captured by allowing the AR coefficients to be individual specific. They 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. We note that the mean group estimator in Pesaran and Smith (1995) is identical to our estimator for the mean of the heterogeneous first-order autocorrelation if the AR(1) model in their paper does not contain exogenous covariates. Hsiao et al. (1999) show that the mean group estimator is asymptotically unbiased under $N/T^2 \to 0$, which is the same rate condition we obtain without the bias-correction. However, we stress that the present paper differs in two ways from these earlier studies. First, we do not assume any specific model to describe the dynamics, while the aforementioned studies consider, for example, an AR specification. Second, our aim is to estimate the entire distributions of the individual mean, autocovariances, and autocorrelations, which are heterogeneous across individuals. In contrast, Pesaran and Smith (1995) and others focus on estimating the means of the AR coefficients.

Elsewhere, Mavroeidis, Sasaki, and Welch (2015) 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 impose $T \to \infty$, our method is much simpler to implement. By contrast, the estimation method in Mavroeidis et al. (2015) 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.

The present paper also relates to the literature on the grouped heterogeneity of panel data structures. Recently, Bonhomme and Manresa (2015) and Su, Shi, and Phillips (2014) developed linear panel data analyses in which the cross-sectional heterogeneity is divided into finite unknown groups. The distinction in their approach is that the speed of $T$ going to infinity can be very slow, although $T \to \infty$ is still required. Their analyses, however, do not allow cases in which all individuals face different dynamics. In addition, their analyses employ linear model specifications, which contrast with our model-free approach.

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. 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, these studies assume that the dynamics are homogeneous, while our paper considers a heterogeneous structure.

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, 2015)
examines the probability limits of various estimators for panel AR(1) models when the true dynamics do not follow an AR(1) process. Lee (2012) discusses the fixed effects estimator for panel AR models with the homogeneous dynamics when the lag order is misspecified. Lastly, Galvao and Kato (2014) investigate the asymptotic properties of the fixed effects estimator in regression models and allow the data-generating process to be generally heterogeneous. We stress that 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.

While not directly connected, this paper also relates somewhat 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. 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) examine moment restriction models with random coefficients using generalized methods of moment estimation. Their analysis on the smooth function of individual effects closely relates to our analysis on the smooth function of means and autocovariances, at least in terms of technique. Finally, Evdokimov (2010) considers a nonparametric panel regression model with individual effects entering the unspecified structural function.

**Organization of the paper:** Section 2 explains the setting. Section 3 introduces the proposed procedures. Section 4 illustrates the proposed procedures using two empirical applications. In Section 5, we derive the asymptotic properties of the distribution estimators. Section 6 considers the estimation of the expected value of a smooth function of the heterogeneous mean and/or autocovariances, the jackknife bias-correction, and the cross-sectional bootstrap. Section 7 presents two Kolmogorov–Smirnov-type tests based on the proposed distribution estimator. Section 8 presents the Monte Carlo simulation results. Section 9 concludes. 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 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 het-
heterogeneous. To be specific, we consider the following data-generating process to model the heterogeneous dynamic structure, in a spirit similar to Galvao and Kato (2014). The unobserved individual effect, $\alpha_i$, 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 $\alpha_i$ varies across individuals. However, note that introducing the parameter $\alpha_i$ is a somewhat abstract way to represent heterogeneity in the dynamics across individuals. We do not directly assume anything about the distribution of $\alpha_i$, because $\alpha_i$ does not explicitly appear in our analysis.

For notational simplicity, we denote “$\cdot|\alpha_i$” by “$\cdot|i$”; that is, “conditional on $\alpha_i$” becomes “conditional on $i$” below.

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

$$
\gamma_{k,i} := E \left( (y_{it} - \mu_i)(y_{it-k} - \mu_i) | i \right),
\rho_{k,i} := \frac{\gamma_{k,i}}{\gamma_{0,i}}.
$$

Note that $\gamma_{0,i}$ is the variance for individual $i$. Similarly to the case of $\mu_i$, $\gamma_{k,i}$ and $\rho_{k,i}$ are random variables and their 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 $\mu_i$, $\gamma_{k,i}$, and/or $\rho_{k,i}$ such as the distribution function, the quantile function, and the moments.

We consider situations in which both $N$ and $T$ are large. For example, in our empirical illustrations, we use panel data with $(N,T) = (481,8)$ and $(465,18)$. A large $N$ allows us to consistently estimate the cross-sectional distributions of $\mu_i$, $\gamma_{k,i}$, and $\rho_{k,i}$. A large $T$ is required to identify and estimate $\mu_i$, $\gamma_{k,i}$, and $\rho_{k,i}$ based on each individual time series. The requirement is popular in the literature on the dynamic panel data model with heterogeneous coefficients. We also note that a large $T$ could also guarantee that the initial values condition (see, e.g., Section 4.3.2 in Hsiao, 2014) is negligible in our analysis. That is, even if the stationary distribution did not generate the initial values, the effect would disappear in the large sample. That said, while empirical researchers should be generally cautious of the initial values condition problem with relatively small $T$, such a problem is not peculiar to our analysis.

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

**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 \( \epsilon_{it} \) follows a strong white noise process with variance \( \sigma^2 \). In this example, \( \alpha_i = (c_i, \{\theta_{j,i}\}_{j=0}^{\infty}), \mu_i = c_i, \gamma_{k,i} = \sigma^2 \sum_{j=k}^{\infty} \theta_{j,i} \theta_{j-k,i}, \) and \( \rho_{k,i} = \sum_{j=k}^{\infty} \theta_{j,i} \theta_{j-k,i}/\sum_{j=0}^{\infty} \theta_{j,i}^2 \).

**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 \( \epsilon_{it} \) is stationary over time and independent across individuals. In this case, \( \mu_i = E(m(\alpha_i, \epsilon_{it})|\alpha_i) \) and \( \gamma_{k,i} \) and \( \rho_{k,i} \) are the \( k \)-th order autocovariance and autocorrelation of \( w_{it} = y_{it} - \mu_i \) given \( \alpha_i \), respectively.

Our focus is on estimating the heterogeneous mean, autocovariance, and autocorrelation structure; we do not aim to recover the underlying structural form of the data-generating process. 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, autocovariance, and autocorrelations without imposing strong assumptions. Moreover, these distributions can provide valuable information, even if our ultimate goal is to identify the structural form.

### 3 Procedures

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

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

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

$$\hat{\rho}_{k,i} := \frac{\hat{\gamma}_{k,i}}{\hat{\gamma}_{0,i}}.$$

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

$$F_{\hat{\mu}}(a) := \frac{1}{N} \sum_{i=1}^N 1(\hat{\mu}_i \leq a), \quad F_{\hat{\gamma}_{k}}(a) := \frac{1}{N} \sum_{i=1}^N 1(\hat{\gamma}_{k,i} \leq a), \quad F_{\hat{\rho}_{k}}(a) := \frac{1}{N} \sum_{i=1}^N 1(\hat{\rho}_{k,i} \leq a),$$

(1)

where $1(\cdot)$ 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 $\mu_i$, $\gamma_{k,i}$, and $\rho_{k,i}$, respectively. Let $F_0^\mu$, $F_0^\gamma_k$, and $F_0^\rho_k$ denote the distribution functions of $\mu_i$, $\gamma_{k,i}$, and $\rho_{k,i}$, respectively, so that $F_0^\mu(a) := \Pr(\mu_i \leq a)$, $F_0^\gamma_k(a) := \Pr(\gamma_{k,i} \leq a)$, and $F_0^\rho_k(a) := \Pr(\rho_{k,i} \leq a)$. In Section 5, we show the consistency of these distribution estimators and derive their asymptotic distributions.

We can estimate other distributional quantities based on the empirical distributions of $\hat{\mu}_i$, $\hat{\gamma}_{k,i}$, or $\hat{\rho}_{k,i}$. For example, consider the estimation of quantiles of $\gamma_{k,i}$. Let $q_\tau$ be the $\tau$-th quantile of $\gamma_{k,i}$: $q_\tau := \inf\{a \in \mathbb{R} : F_0^\gamma_k(a) \geq \tau\}$. This is estimated by the $\tau$-th quantile of $\hat{\gamma}_{k,i}$ so that $\hat{q}_\tau := \inf\{a \in \mathbb{R} : F_{\hat{\gamma}_{k}}(a) \geq \tau\}$. Using the functional delta method, we derive the asymptotic distribution of the quantile estimator.

We can also test parametric specifications of the distribution of the heterogeneous mean, autocovariances, or autocorrelations based on the empirical distribution. It is also possible to examine the difference in 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 7.

We can also estimate a function of the expected value of a smooth function of a vector
of the heterogeneous mean and autocovariances straightforwardly. For example, if we are interested in the variance of the mean, we compute the sample variance of \( \hat{\mu}_t \):

\[
\frac{1}{N} \sum_{i=1}^{N} (\hat{\mu}_t)^2 - \left( \frac{1}{N} \sum_{i=1}^{N} \hat{\mu}_i \right)^2.
\]

The mean of the first-order autocorrelation may be estimated by

\[
\frac{1}{N} \sum_{i=1}^{N} \rho_{1,i}.
\]

The correlation between the mean and variance can be estimated by

\[
\frac{\frac{1}{N} \sum_{i=1}^{N} (\hat{\mu}_t)^2 - \left( \frac{1}{N} \sum_{i=1}^{N} \hat{\mu}_i \right)^2}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{\mu}_i)^2 - \left( \frac{1}{N} \sum_{i=1}^{N} \hat{\mu}_i \right)^2} \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{\gamma}_{0,i})^2 - \left( \frac{1}{N} \sum_{i=1}^{N} \hat{\gamma}_{0,i} \right)^2}}.
\]

In general, when the parameter of interest is \( H := h(E(g(\theta_i))) \), where \( \theta_i \) is a \( l \times 1 \) vector whose elements belong to a subset of \( (\mu_i, \gamma_{0,i}, \gamma_{1,i}, \ldots) \) and \( g : \mathbb{R}^l \rightarrow \mathbb{R}^m \) and \( h : \mathbb{R}^m \rightarrow \mathbb{R}^n \) are smooth vector-valued functions, we estimate \( H \) by

\[
\hat{H} := h \left( \frac{1}{N} \sum_{i=1}^{N} g(\hat{\theta}_i) \right).
\]

The theoretical results in Section 6 show that this estimator is consistent as \( N, T \rightarrow \infty \) and is asymptotically normal with mean zero when \( N/T^2 \rightarrow 0 \).

The half-panel jackknife (HPJ) proposed by Dhaene and Jochmans (2015) can further reduce the bias in \( \hat{H} \). The estimator exhibits bias of order \( O(1/T) \) and the HPJ bias-correction can delete the bias of this order. The bias-correction is easy to implement. Suppose that \( T \) is even. We divide the panel data into two subpanels: \( \{y_{it}\}_{t=1}^{T/2} \) and \( \{y_{it}\}_{t=T/2+1}^{T} \). The first subpanel, \( \{y_{it}\}_{t=1}^{T/2} \), consists of observations from the first half of the overall time period, and the second subpanel, \( \{y_{it}\}_{t=T/2+1}^{T} \), consists of those from the second half. Let \( \hat{H}^{(1)} \) and \( \hat{H}^{(2)} \) be the estimators of \( H \) computed using \( \{y_{it}\}_{t=1}^{T/2} \) and \( \{y_{it}\}_{t=T/2+1}^{T} \), respectively. Let \( \hat{H} := (\hat{H}^{(1)} + \hat{H}^{(2)})/2 \). The HPJ estimator of \( H \) is

\[
\hat{H}^H := \hat{H} - (\hat{H} - \hat{H}) = 2\hat{H} - \hat{H}.
\]
The HPJ estimates the bias in $\hat{H}$ by $\bar{H} - \hat{H}$, and $\hat{H}^H$ corrects the bias in $\hat{H}$ by subtracting the HPJ bias estimate. The bias-corrected estimator $\hat{H}^H$ does not exhibit bias of order $O(1/T)$ and is asymptotically unbiased even when $N/T^2$ does not converge to zero.

For statistical inferences on parameter $H$, we suggest the cross-sectional bootstrap. The cross-sectional bootstrap is used to approximate the distribution of the HPJ estimator (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 $y_i$, where $y_i := (y_{i1}, \ldots, y_{iT})^\top$. The algorithm is as follows:

1. Randomly draw $y^*_1, \ldots, y^*_N$ from $\{y_1, \ldots, y_N\}$ with replacement.
2. Compute the statistics of interest, say $S$, using $y^*_1, \ldots, y^*_N$.
3. Repeat 1 and 2 $B$ times. Let $S^*(b)$ be the statistics computed in the $b$-th bootstrap.
4. Compute distributional quantities of interest for $S$ by the empirical distribution of $S^*(b)$.

For example, suppose that we are interested in constructing a 95% confidence interval for parameter $H$. We obtain the bootstrap approximation of the distribution of $S = \hat{H}^H - H$. Let $\hat{H}^H*(b)$ be the HPJ estimate of $H$ 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 $S^*(b) = \hat{H}^H*(b) - \hat{H}^H$. The cross-sectional bootstrap 95% confidence interval for $H$ is

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

**Remark 1.** Our procedures could allow the presence of covariates. If there are discrete covariates with relatively small support, we split the data $\{y_{it}\}_{i=1}^N$ into “cells” according to the realization of the discrete covariates. We then estimate the distributional properties of $\mu_i, \gamma_{k,i},$ and $\rho_{k,i}$ of $\{y_{it}\}_{i=1}^N$ in each cell. If there are discrete covariates with large support or continuous covariates, we obtain the residuals $\{\hat{v}_{it}\}_{i=1}^N$ from the regression of $y_{it}$ on the covariates to control the effect. We then compute the distributional properties of the heterogeneous mean, autocovariances, and autocorrelations of $\{\hat{v}_{it}\}_{i=1}^N$. Of course, we could develop the statistical properties of these procedures, but this would be beyond the scope of the present paper.

### 4 Empirical applications

This section illustrates our proposed procedures through applying them to the study of earnings and productivity dynamics. In the tables in this section, “ED” refers to the estimator
based on the empirical distribution and “HPJ” to the HPJ estimator. The 95% confidence intervals are computed using the cross-sectional bootstrap.

### 4.1 Earnings dynamics

We study the properties of earnings dynamics without relying on any specific models. Earnings (or income) dynamics are an important research theme in economics (see, e.g., Meghir and Pistaferri, 2011 for a review). We examine the distribution of quantities that describe the earnings dynamics. We find that the earnings dynamics exhibit considerable heterogeneity. Our model-free analysis supports the empirical results in Browning et al. (2010) and Ejrnæs and Browning (2014) in which they find that the determinants of the income process are significantly heterogeneous.

We use a panel data set of white male workers’ income extracted from the Panel Study of Income Dynamics (PSID). The data were originally prepared by Meghir and Pistaferri (2004) and we use a subsample of the version created by Browning et al. (2010). This version of data is also used in Ejrnæs and Browning (2014). We draw a subsample that consists of individuals without any missing values in earnings between 1967 and 1974. The sample size is \( N = 481 \). The length of time series is \( T = 8 \): \( t = 1 \) refers to year 1967 and \( t = 8 \) refers to year 1974.

We investigate the distribution of individual means, variances, and first-order autocorrelations of residual log earnings. We obtain the residuals by regressing log earnings on year and age dummies.\(^4\) The regressions are run for three educational categories (high school dropouts, high school graduates, and college graduates).

Table 1 summarizes the distributional features of the means, variances, and first-order autocorrelations. We also estimate the correlations between these three quantities.

The results show that the earnings dynamics exhibit considerable heterogeneity. They also show that the HPJ estimates can be very different from the ED estimates and the HPJ estimates imply more persistent dynamics than those implied by the ED estimates. Our estimates imply that the variances of income shocks are markedly heterogeneous with a skewed distribution, which is consistent with the finding in Browning et al. (2010). The results also indicate that individuals with high permanent earnings face relatively small shocks (in percentage terms). Conversely, there is no statistical evidence that the persistency of an

---

\(^4\) While this paper does not investigate the influence of eliminating the time effects by year dummies, such control would not affect the formulas of the asymptotic biases of our estimators. This is because the time effects can be estimated at the rate of \( \sqrt{N} \), and a similar observation is developed by Okui (2014) for the autocorrelation estimators under homogeneous dynamics. We also note that eliminating the time effects may alter the asymptotic variances of our estimators, but the cross-sectional bootstrap would be valid even in such situations.
Table 1: The distribution of earnings dynamics

(a) The distribution of $\mu_i$

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Var</th>
<th>Q25</th>
<th>Q50</th>
<th>Q75</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED</td>
<td>0.010</td>
<td>0.131</td>
<td>-0.189</td>
<td>0.016</td>
<td>0.256</td>
</tr>
<tr>
<td>95% CI</td>
<td>(-0.020, 0.042)</td>
<td>(0.095, 0.159)</td>
<td>(-0.229, -0.147)</td>
<td>(-0.033, 0.045)</td>
<td>(0.212, 0.289)</td>
</tr>
<tr>
<td>HPJ</td>
<td>0.122</td>
<td>-0.183</td>
<td>0.006</td>
<td>0.238</td>
<td></td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.087, 0.150)</td>
<td>(-0.238, -0.131)</td>
<td>(-0.059, 0.032)</td>
<td>(0.176, 0.275)</td>
<td></td>
</tr>
</tbody>
</table>

(b) The distribution of $\gamma_{0,i}$

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Var</th>
<th>Q25</th>
<th>Q50</th>
<th>Q75</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED</td>
<td>0.030</td>
<td>0.005</td>
<td>0.000</td>
<td>0.012</td>
<td>0.028</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.024, 0.036)</td>
<td>(0.001, 0.008)</td>
<td>(0.005, 0.007)</td>
<td>(0.010, 0.013)</td>
<td>(0.023, 0.031)</td>
</tr>
<tr>
<td>HPJ</td>
<td>0.039</td>
<td>0.009</td>
<td>0.009</td>
<td>0.018</td>
<td>0.040</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.031, 0.046)</td>
<td>(0.001, 0.008)</td>
<td>(0.008, 0.011)</td>
<td>(0.015, 0.020)</td>
<td>(0.033, 0.046)</td>
</tr>
</tbody>
</table>

(c) The distribution of $\rho_{1,i}$

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Var</th>
<th>Q25</th>
<th>Q50</th>
<th>Q75</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED</td>
<td>0.155</td>
<td>0.115</td>
<td>-0.090</td>
<td>0.168</td>
<td>0.425</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.127, 0.188)</td>
<td>(0.102, 0.127)</td>
<td>(-0.129, -0.046)</td>
<td>(0.123, 0.214)</td>
<td>(0.399, 0.461)</td>
</tr>
<tr>
<td>HPJ</td>
<td>0.503</td>
<td>0.253</td>
<td>0.530</td>
<td>0.739</td>
<td></td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.450, 0.558)</td>
<td>(0.071, 0.122)</td>
<td>(0.171, 0.335)</td>
<td>(0.439, 0.613)</td>
<td>(0.679, 0.801)</td>
</tr>
</tbody>
</table>

(d) Correlation structure

<table>
<thead>
<tr>
<th></th>
<th>$\mu_i$ vs $\gamma_{0,i}$</th>
<th>$\mu_i$ vs $\rho_{1,i}$</th>
<th>$\gamma_{0,i}$ vs $\rho_{1,i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED</td>
<td>-0.270</td>
<td>0.055</td>
<td>0.015</td>
</tr>
<tr>
<td>95% CI</td>
<td>(-0.384, -0.170)</td>
<td>(-0.044, 0.156)</td>
<td>(-0.051, 0.082)</td>
</tr>
<tr>
<td>HPJ</td>
<td>-0.274</td>
<td>0.091</td>
<td>0.016</td>
</tr>
<tr>
<td>95% CI</td>
<td>(-0.440, -0.148)</td>
<td>(-0.080, 0.255)</td>
<td>(-0.107, 0.127)</td>
</tr>
</tbody>
</table>

earnings shock is associated with either the level of permanent earnings or the magnitude of income shocks.

We next examine whether the distribution of the means is normal. This is of interest because it is common to model the distribution of earnings using a lognormal distribution. As an intuitive investigation, Figure 1 provides the QQ plot of the means of earnings against the normal distribution with the same mean and standard deviation. In a formal manner, the Kolmogorov–Smirnov test cannot reject the null of normality (or lognormality of earnings) given a test statistic of 0.130 and a $p$-value of 0.189. These results suggest that the lognormal approximation of earnings may be reasonable.

Finally, we examine whether the distribution of earnings dynamics differs according to distinct educational attainments. In particular, we investigate whether the distribution of the first-order autocorrelations for high school graduates differs from that of college graduates. In our sample, there are 220 high school graduates and 104 college graduates. Table 2 summarizes the distribution of the first-order autocorrelations of high school graduates and
that of college graduates and Figure 2 provides the QQ plot. These indicate that college graduates face more persistent earnings shocks than high school graduates. Moreover, the first-order autocorrelations of earnings for college graduates are relatively more concentrated (i.e., more homogeneous) than those for high school graduates. The value of the two-sample Kolmogorov–Smirnov test is 0.178 with a $p$-value of 0.028. The null hypothesis is then rejected at the 5% significance level and as such there is statistical evidence that high school and college graduates face different earnings dynamics.

### 4.2 Productivity dynamics

Our second application examines the total factor productivity (TFP) dynamics of Chilean firms. We find considerable heterogeneity in the productivity of Chilean firms.

Our analysis is based on plant-level panel data on productivity in the Chilean food industry (Industry 311). This data set is used by Levinsohn and Petrin (2003), Balat, Brambilla, and Sasaki (2015), and others. Detailed information about the data is provided in Balat et al. (2015).\footnote{Yuya Sasaki kindly agrees to share his data set with the authors.} We use the TFP data computed by the method in Levinsohn and Petrin (2003).\footnote{While Balat et al. (2015) also compute TFPs, they use different time periods and a different method. Note also that Levinsohn and Petrin (2003) use a similar data set but their TFP values are slightly different.} We employ a subsample consisting of firms with no missing observations in the entire sample period.
Table 2: The distributions of $\rho_{1,i}$ for high school and college graduates

(a) The distribution of $\rho_{1,i}$ for high school graduates

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Var</th>
<th>Q25</th>
<th>Q50</th>
<th>Q75</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED</td>
<td>0.128</td>
<td>0.123</td>
<td>-0.132</td>
<td>0.128</td>
<td>0.415</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.081, 0.173)</td>
<td>(0.105, 0.140)</td>
<td>(-0.187, -0.076)</td>
<td>(0.066, 0.194)</td>
<td>(0.381, 0.497)</td>
</tr>
<tr>
<td>HPJ</td>
<td>0.472</td>
<td>0.115</td>
<td>0.181</td>
<td>0.486</td>
<td>0.752</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.391, 0.552)</td>
<td>(0.080, 0.148)</td>
<td>(0.064, 0.293)</td>
<td>(0.367, 0.605)</td>
<td>(0.665, 0.908)</td>
</tr>
</tbody>
</table>

(b) The distribution of $\rho_{1,i}$ for college graduates

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Var</th>
<th>Q25</th>
<th>Q50</th>
<th>Q75</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED</td>
<td>0.230</td>
<td>0.101</td>
<td>0.007</td>
<td>0.251</td>
<td>0.472</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.171, 0.297)</td>
<td>(0.078, 0.126)</td>
<td>(-0.120, 0.117)</td>
<td>(0.186, 0.306)</td>
<td>(0.395, 0.541)</td>
</tr>
<tr>
<td>HPJ</td>
<td>0.588</td>
<td>0.072</td>
<td>0.364</td>
<td>0.588</td>
<td>0.781</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.481, 0.701)</td>
<td>(0.022, 0.118)</td>
<td>(0.122, 0.583)</td>
<td>(0.439, 0.703)</td>
<td>(0.607, 0.922)</td>
</tr>
</tbody>
</table>

In this application, $y_{it}$ is the TFP of firm $i$ at time $t$. The cross-sectional sample size is $N = 465$. The length of time series is $T = 18$: $t = 1$ refers to year 1979 and $t = 18$ refers to year 1996. The observations are in 1980 Chilean pesos.

Table 3 summarizes the distributional features of the heterogeneous means, variances, and first-order autocorrelations. We also estimate the correlations between these three quantities. The results show that the productivity dynamics exhibit considerable heterogeneity. The results also indicate that firms with high permanent productivity face relatively large shocks, but those shocks may be less persistent. We also obtain a slightly insignificant (at the 5% level) result that firms facing large shocks may not suffer (or benefit) from these shocks for long. We also find that the HPJ estimates differ markedly from the ED estimates, in particular for the first-order autocorrelation. Figure 3 provides the QQ plot of the productivity means against a normal distribution with the same mean and standard deviation. Because the value for a Kolmogorov–Smirnov test of the null hypothesis of a normal distribution is 0.155 and the $p$-value is very close to zero, we reject the null hypothesis.

4.3 Summary

These empirical examples demonstrate the usefulness of our proposed procedures. Our simple and easy-to-implement method identifies several interesting features of earnings and productivity dynamics, indicating that both exhibit considerable heterogeneity. For the earnings dynamics, we identify new empirical evidence that individuals with high permanent earnings may face relatively small earnings shocks and that high school graduates face earnings dynamics that differ from those for college graduates. For the productivity dynamics, the results indicate that firms with high levels of permanent productivity face relatively large, but less persistent, shocks.
Table 3: The distribution of the productivity dynamics

(a) The distribution of $\mu_i$

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Var</th>
<th>Q25</th>
<th>Q50</th>
<th>Q75</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED</td>
<td>2.96</td>
<td>0.085</td>
<td>2.750</td>
<td>2.863</td>
<td>3.142</td>
</tr>
<tr>
<td>95% CI</td>
<td>(2.93, 2.98)</td>
<td>(0.073, 0.098)</td>
<td>(2.73, 2.77)</td>
<td>(2.85, 2.90)</td>
<td>(3.08, 3.23)</td>
</tr>
<tr>
<td>HPJ</td>
<td>0.080</td>
<td>2.755</td>
<td>2.864</td>
<td>3.159</td>
<td></td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.068, 0.093)</td>
<td>(2.73, 2.78)</td>
<td>(2.85, 2.91)</td>
<td>(3.10, 3.28)</td>
<td></td>
</tr>
</tbody>
</table>

(b) The distribution of $\gamma_{0,i}$

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Var</th>
<th>Q25</th>
<th>Q50</th>
<th>Q75</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED</td>
<td>0.035</td>
<td>0.0015</td>
<td>0.0145</td>
<td>0.0235</td>
<td>0.0374</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.032, 0.038)</td>
<td>(0.0008, 0.0021)</td>
<td>(0.0134, 0.0156)</td>
<td>(0.0217, 0.0254)</td>
<td>(0.0332, 0.0407)</td>
</tr>
<tr>
<td>HPJ</td>
<td>0.041</td>
<td>0.010</td>
<td>0.0194</td>
<td>0.0305</td>
<td>0.0423</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.037, 0.045)</td>
<td>(0.0006, 0.0015)</td>
<td>(0.0178, 0.0214)</td>
<td>(0.0278, 0.0336)</td>
<td>(0.0349, 0.0478)</td>
</tr>
</tbody>
</table>

(c) The distribution of $\rho_{1,i}$

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Var</th>
<th>Q25</th>
<th>Q50</th>
<th>Q75</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED</td>
<td>0.41</td>
<td>0.071</td>
<td>0.227</td>
<td>0.450</td>
<td>0.611</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.384, 0.432)</td>
<td>(0.063, 0.079)</td>
<td>(0.184, 0.282)</td>
<td>(0.423, 0.484)</td>
<td>(0.591, 0.637)</td>
</tr>
<tr>
<td>HPJ</td>
<td>0.57</td>
<td>0.023</td>
<td>0.446</td>
<td>0.620</td>
<td>0.697</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.540, 0.609)</td>
<td>(0.008, 0.039)</td>
<td>(0.364, 0.546)</td>
<td>(0.573, 0.672)</td>
<td>(0.658, 0.746)</td>
</tr>
</tbody>
</table>

(d) Correlation structure

<table>
<thead>
<tr>
<th></th>
<th>$\mu_i$ vs $\gamma_{0,i}$</th>
<th>$\mu_i$ vs $\rho_{1,i}$</th>
<th>$\gamma_{0,i}$ vs $\rho_{1,i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED</td>
<td>0.231</td>
<td>-0.195</td>
<td>-0.041</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.138, 0.328)</td>
<td>(-0.280, -0.120)</td>
<td>(-0.137, 0.048)</td>
</tr>
<tr>
<td>HPJ</td>
<td>0.258</td>
<td>-0.235</td>
<td>-0.126</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.137, 0.385)</td>
<td>(-0.362, -0.116)</td>
<td>(-0.277, 0.012)</td>
</tr>
</tbody>
</table>

5 Asymptotic analysis for the distribution estimators

This section presents the asymptotic properties of the distribution estimators (1). We first show the uniform consistency of the empirical distributions. We then derive the functional central limit theorem for them. 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 $\gamma_{k,i} = E(w_{it}w_{i,t-k}|i)$. 

16
5.1 Uniform consistency

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

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 $\mathbb{P}^\mu_N$ be the empirical measure of $\hat{\mu}_i$:

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

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

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

where $1_{(-\infty,a]}(x) := 1(x \leq a)$. We define the probability measure of $\mu_i$ as $P^\mu_0$. In this notation, the empirical distribution function, $\mathbb{P}^\mu_N$, is an empirical process indexed by $\mathcal{F}$. For example, $\mathbb{P}^\mu_N f = \mathbb{P}^\mu_N (a)$ for $f = 1_{(-\infty,a]}$. Similarly, for $f = 1_{(-\infty,a]}$, $P^\mu_0 f = F^\mu_0 (a) = Pr(\mu_i \leq a)$. The empirical measures of $\hat{\gamma}_{k,i}$ and $\hat{\rho}_{k,i}$, $\mathbb{P}^{\gamma_k}_N$ and $\mathbb{P}^{\rho_k}_N$, and the probability measures of $\gamma_{k,i}$ and $\rho_{k,i}$, $P^{\gamma_k}_0$ and $P^{\rho_k}_0$, are analogously defined.

We use the following assumption throughout the paper, which summarizes the conditions imposed in Section 2.
Assumption 1. The sample space of $\alpha_i$ 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 $i$.

The following assumptions depend on natural numbers $r_m$ and $r_d$, which will be specified in the theorems that use this assumption.

Assumption 2. For each $i$, $\{y_{it}\}_{t=1}^\infty$ is strictly stationary and $\alpha$-mixing given $\alpha_i$ with mixing coefficients $\{\alpha(m|i)\}_{m=0}^\infty$. There exists a natural number $r_m$ and a sequence $\{\alpha(m)\}_{m=0}^\infty$ such that for any $i$ and $m$, $\alpha(m|i) \leq \alpha(m)$ and $\sum_{m=0}^\infty (m+1)^{r_m/2-1} \alpha(m)^{\delta/(r_m+\delta)} < \infty$ for some $\delta > 0$.

Assumption 3. There exists a natural number $r_d$ such that $E|w_{it}|^{r_d+\delta} < \infty$ for some $\delta > 0$.

Assumptions 2 and 3 are mild regularity conditions on the process of $y_{it}$. Assumption 2 is a mixing condition depending on $r_m$ and restricts the degree of persistency of $y_{it}$ across time. Assumption 3 requires that $w_{it}$ exhibits some moment higher than $r_d$-th order. These assumptions are satisfied, for example, when $y_{it}$ follows a heterogeneous stationary panel ARMA model with Gaussian innovations.

Additionally, we need Assumptions 4, 5, and 6 for the consistency of $\hat{P}_N^\mu$, that of $\hat{P}_N^{\hat{\gamma}_i}$, and that of $\hat{P}_N^{\hat{\rho}_i}$, respectively.

Assumption 4. The random vector $(\mu_i, \hat{\mu}_i)$ is continuously distributed and its joint density is bounded.

Assumption 5. The random vector $(\gamma_{k,i}, \hat{\gamma}_{k,i})$ is continuously distributed and its joint density is bounded.

Assumption 6. The random vector $(\rho_{k,i}, \hat{\rho}_{k,i})$ is continuously distributed and its joint density is bounded. There exists a constant $\epsilon > 0$ such that $\gamma_{0,i} > \epsilon$ almost surely.

Assumption 4 states that $\mu_i$ and $\hat{\mu}_i$ are continuous random variables. This assumption is restrictive in the sense that it does not allow the case in which the distribution of $\mu_i$ is discrete or there is no heterogeneity in the mean (i.e., $\mu_i$ is homogeneous so that $\mu_i = \mu$ for some constant $\mu$ for any $i$). The uniform consistency would not hold without this assumption. Assumption 5 is similar to Assumption 4. Assumption 6 is also similar to Assumptions 4 and 5, but it also imposes that the variance $\gamma_{0,i}$ is bounded away from zero.

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

Theorem 1. Suppose that Assumptions 1, 2, 3, and 4 hold for $r_m = 2$ and $r_d = 2$. When $N, T \to \infty$, the class $\mathcal{F}$ is $P_0^\mu$-Glivenko–Cantelli in the sense that

$$\sup_{f \in \mathcal{F}} \left| P_N^\mu f - P_0^\mu f \right| \overset{as}{\to} 0,$$
where $\overset{a.s.}\to$ is the almost sure convergence.

Suppose that Assumptions 1, 2, 3, and 5 hold for $r_m = 4$ and $r_d = 4$. When $N, T \to \infty$, the class $\mathcal{F}$ is $P_0^{\gamma_k}$-Glivenko–Cantelli in the sense that

$$\sup_{f \in \mathcal{F}} \left| \mathbb{P}_N^\gamma f - P_0^{\gamma_k} f \right| \overset{a.s.}\to 0.$$ 

Suppose that Assumptions 1, 2, 3, and 6 hold for $r_m = 4$ and $r_d = 4$. When $N, T \to \infty$, the class $\mathcal{F}$ is $P_0^{\rho_k}$-Glivenko–Cantelli in the sense that

$$\sup_{f \in \mathcal{F}} \left| \mathbb{P}_N^\rho f - P_0^{\rho_k} f \right| \overset{a.s.}\to 0.$$ 

Note that Theorem 1 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 $\hat{\mu}_i$, $\hat{\gamma}_{k,i}$, and $\hat{\rho}_{k,i}$ change as $T$ increases. Nonetheless, our proof of Theorem 1 follows similar steps to those of the usual Glivenko–Cantelli theorem.

5.2 Functional central limit theorem

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

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

where $f \in \mathcal{F}$. 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 can be obtained by the functional delta method based on this result.

The functional central limit theorems for $\mathbb{P}_N^\mu$, $\mathbb{P}_N^\gamma$, and $\mathbb{P}_N^\rho$ hold under the same set of assumptions for the uniform consistency. However, it requires a 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, 3, and 4 hold for $r_m = 2r$ and $r_d = 2r$, where $r \geq 1$ is some natural number. When $N, T \to \infty$ with $N^{2+r}/T^{2r} \to 0$, we have

$$\sqrt{N}(\mathbb{P}_N^\mu - P_0^\mu) \overset{\ell^\infty(\mathcal{F})}{\sim} \mathbb{G}_{P_0^\mu},$$

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, 2, 3, and 5 hold for \( r_m = 2r \) and \( r_d = 2r \), where \( r \geq 2 \) is some natural number. When \( N, T \to \infty \) with \( N^{2+r}/T^{2r} \to 0 \), we have

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

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

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

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

Suppose that Assumptions 1, 2, 3, and 6 hold for \( r_m = 2r \) and \( r_d = 2r \), where \( r \geq 2 \) is some natural number. When \( N, T \to \infty \) with \( N^{2+r}/T^{2r} \to 0 \), we have

\[
\sqrt{N}(\mathbb{P}_N^{\hat{\rho}_k} - P_0^{\rho_k}) \Rightarrow \mathbb{G}^{\rho_k}_{P_0^k} \quad \text{in} \quad \ell^\infty(\mathcal{F}),
\]

where \( \mathbb{G}^{\rho_k}_{P_0^k} \) is a Gaussian process with zero mean and covariance function

\[
E(\mathbb{G}^{\rho_k}_{P_0^k}(f_i)\mathbb{G}^{\rho_k}_{P_0^k}(f_j)) = F_0^{\rho_k}(a_i \wedge a_j) - F_0^{\rho_k}(a_i)F_0^{\rho_k}(a_j),
\]

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

This theorem shows that the asymptotic laws of the empirical processes are Gaussian. This limiting process is then identical to that for the empirical process constructed using the true \( \mu_i, \gamma_{k,i}, \) or \( \rho_{k,i} \). However, this result requires that \( N^{2+r}/T^{2r} \to 0 \). This condition allows us to ignore the estimation error in \( \hat{\mu}_i, \hat{\gamma}_{k,i}, \) and \( \hat{\rho}_{k,i} \) asymptotically. Note that \( N^{2+r}/T^{2r} \to 0 \) can be arbitrarily close to \( N/T^2 \to 0 \) as \( r \) becomes large.

Here, we provide a brief summary of the proof and explain the reason why the condition \( N^{2+r}/T^{2r} \to 0 \) is required. The same discussion can be applied to all of \( \mathbb{P}_N^\mu, \mathbb{P}_N^{\hat{\gamma}_k}, \) and \( \mathbb{P}_N^{\hat{\rho}_k} \). In the following, we let \( \mathbb{P}_N \) be either \( \mathbb{P}_N^\mu, \mathbb{P}_N^{\hat{\gamma}_k}, \) or \( \mathbb{P}_N^{\hat{\rho}_k} \) and \( P_0 \) be the corresponding true distribution.

The key to understanding the mechanism behind the requirement that \( N^{2+r}/T^{2r} \to 0 \) 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 \( \hat{\mu}_i, \hat{\gamma}_{k,i}, \) or \( \hat{\rho}_{k,i} \). Note that \( P_T \) depends on \( T \) and \( P_T \neq P_0 \). Observe 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:

\[
\sqrt{N}(\mathbb{P}_N f - P_0 f) = \mathbb{G}_{N,P_T} f + \sqrt{N}(P_T f - P_0 f). \tag{4}
\]

\[
\sqrt{N}(P_T f - P_0 f). \tag{5}
\]
For $G_{N,P}$ in (4), 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) and obtain

$$G_{N,P} \rightsquigarrow G_{P_0} \quad \text{in} \quad \ell^\infty(F),$$

as $N \to \infty$. This part of the proof is standard.

The condition $N^{2+r}/T^{2r} \to 0$ is needed to eliminate the effect of the bias term in the empirical process: $\sqrt{N}(P_T - P_0)$ in (5). In the proof of the theorem, we show that

$$\sup_{f \in F} \left| \sqrt{N} (P_T f - P_0 f) \right| = O \left( \frac{\sqrt{N}}{T^{r/(2+r)}} \right).$$

This term converges to zero when $T$ is of a higher order than $N$. This result arises from the fact that the $2r$-th order moment of $\hat{\mu}_i - \mu_i$, $\hat{\gamma}_{k,i} - \gamma_{k,i}$, or $\hat{\rho}_{k,i} - \rho_{k,i}$ is of order $O(T^{-r})$. This fact leads to the result that the difference between $P_T$ and $P_0$ is of order $O(T^{-r/(2+r)})$.

### 5.3 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 $\phi(P_N)$ for $\phi : D(F) \to \mathbb{R}$, where $P_N = P_{\hat{\mu}}^N, P_{\hat{\gamma}}^N$, or $P_{\hat{\rho}}^N$ and $D(F)$ is the collection of all cadlag real functions on $F$. For example, the $\tau$-th quantile of $\gamma_{k,i}$, $\phi(P_0^{\gamma_k}) = q_\tau := \inf\{a \in \mathbb{R} : F_0^{\gamma_k}(a) \geq \tau\}$ for $\tau \in (0, 1)$, may be estimated by the empirical $\tau$-th quantile of $\hat{\gamma}_{k,i}$:

$$\phi(P_N^{\hat{\gamma}_k}) = \hat{q}_\tau := (F_N^{\hat{\gamma}_k})^{-1}(\tau) = \inf\{a \in \mathbb{R} : F_N^{\hat{\gamma}_k}(a) \geq \tau\}.$$ 

As a more general case, we can consider the estimation of the process of quantiles. That is, we estimate the quantile process of $F_0^{\gamma_k}$, $(F_0^{\gamma_k})^{-1}$, using the empirical quantile process of $F_N^{\hat{\gamma}_k}$, $(F_N^{\hat{\gamma}_k})^{-1}$.

The derivation of the asymptotic distribution of $\phi(P_N)$ 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 $E$ be a normed linear space. Let $\phi : D(F) \subset \ell^\infty(F) \to E$ be Hadamard differentiable at $P_0^\mu$. Denote its derivative by $\phi'_{P_0^\mu}$. Suppose that Assumptions 1, 2, 3, and 4 hold for $r_m = 2r$ and $r_d = 2r$, where $r \geq 1$ is some natural number. When $N, T \to \infty$ with $N^{2+r}/T^{2r} \to 0$, we have

$$\sqrt{N} \left( \phi(P_N^\mu) - \phi(P_0^\mu) \right) \rightsquigarrow \phi'_{P_0^\mu}(G_{P_0^\mu}).$$
Similarly, suppose that $\phi$ has the Hadamard derivative, $\phi'_{\gamma_k}$, at $P_0^{\gamma_k}$. Suppose that Assumptions 1, 2, 3, and 5 hold for $r_m = 2r$ and $r_d = 2r$, where $r \geq 2$ is some natural number. When $N, T \to \infty$ with $N^{2+r}/T^{2r} \to 0$, we have

$$\sqrt{N}(\phi(P_0^{\gamma_k}) - \phi(P_0^{\gamma_k})) \asymp \phi'_{\gamma_k}(G_{P_0^{\gamma_k}}).$$

Similarly, suppose that $\phi$ has the Hadamard derivative, $\phi'_{\rho_k}$, at $P_0^{\rho_k}$. Suppose that Assumptions 1, 2, 3, and 6 hold for $r_m = 2r$ and $r_d = 2r$, where $r \geq 2$ is some natural number. When $N, T \to \infty$ with $N^{2+r}/T^{2r} \to 0$, we have

$$\sqrt{N}(\phi(P_0^{\rho_k}) - \phi(P_0^{\rho_k})) \asymp \phi'_{\rho_k}(G_{P_0^{\rho_k}}).$$

**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 $\hat{q}_r$. The form of $\phi'_{\gamma_k}$ for $\hat{q}_r$ is available in Example 20.5 in van der Vaart (1998) and indicates that as $N, T \to \infty$ with $N^{2+r}/T^{2r} \to 0$,

$$\sqrt{N}(\hat{q}_r - q_r) \asymp N \left(0, \frac{\tau(1-\tau)}{(f_{\gamma_k}(q_r))^2} \right),$$

where $f_{\gamma_k}(\cdot)$ is the density function of $\gamma_{k,i}$. We can also derive the asymptotic distribution of the empirical quantile process $(F_0^{\gamma_k})^{-1}$. If $f_{\gamma_k}(\cdot)$ is continuous and positive on the interval $[(F_0^{\gamma_k})^{-1}(p) - \varepsilon, (F_0^{\gamma_k})^{-1}(q) + \varepsilon]$ for some $0 < p < q < 1$ and $\varepsilon > 0$, then Corollary 1 means that

$$\sqrt{N}((F_0^{\gamma_k})^{-1} - (F_0^{\gamma_k})^{-1}) \asymp -\frac{G_{P_0^{\gamma_k}} \circ F_0^{\gamma_k}((F_0^{\gamma_k})^{-1})}{f_{\gamma_k}((F_0^{\gamma_k})^{-1})} \quad \text{in} \quad \ell^\infty[p, q],$$

as $N, T \to \infty$ with $N^{2+r}/T^{2r} \to 0$. This process is known to be Gaussian with zero mean and known covariance function (see, e.g., Example 3.9.24 in van der Vaart and Wellner, 1996).

## 6 Function of the expected value of a smooth function of the heterogeneous mean and/or autocovariances

In this section, we consider the estimation of a function of the expected value of a smooth function of the heterogeneous mean and/or autocovariances.\(^7\) We also show that half-panel jackknife bias-correction can reduce the asymptotic bias in the estimator and that it relaxes the condition on the ratio of $N$ to $T$. We then develop the asymptotic justification of the cross-sectional bootstrap inference procedure.

\(^7\)Note that an autocorrelation is a function of the autocovariances.
6.1 Asymptotic results

We derive the asymptotic properties of \( \hat{H} = h(N^{-1} \sum_{i=1}^{N} g(\hat{\theta}_i)) \) in (2), which is the estimator of \( H = h(E(g(\theta_i))) \). Recall that \( h = (h_1, h_2, \ldots, h_n) : \mathbb{R}^m \to \mathbb{R}^n \) and \( g = (g_1, g_2, \ldots, g_m) : \mathbb{R}^l \to \mathbb{R}^m \) are known smooth functions with fixed \( m, n \), and \( l, \theta_i \) is an \( l \)-dimensional random vector whose elements are \( \mu_i \) and/or \( \gamma_{k,i} \)'s, and \( \hat{\theta}_i \) is the estimator of \( \theta_i \) with \( \hat{\mu}_i \) for \( \mu_i \) and \( \hat{\gamma}_{k,i} \) for \( \gamma_{k,i} \). Define \( G := E(g(\theta_i)) \) so that \( H = h(G) \). We show that \( \hat{H} \) is consistent for \( H \) and that the asymptotically unbiased estimation is achieved when \( N/T^2 \to 0 \).

We make the following assumptions to develop the asymptotic properties of \( \hat{H} \).

**Assumption 7.** The function \( h(\cdot) \) is continuously in a neighborhood of \( G \).

**Assumption 8.** The function \( h(\cdot) \) is continuously differentiable in a neighborhood of \( G \). The matrix of the first derivatives \( \nabla h(G) := (\nabla h_1(G)^\top, \nabla h_2(G)^\top, \ldots, \nabla h_n(G)^\top)^\top \) is of full row rank.

**Assumption 9.** The function \( g = (g_1, g_2, \ldots, g_m) : \mathcal{O} \to \mathbb{R}^m \) is twice-continuously differentiable where \( \mathcal{O} \subset \mathbb{R}^l \) is a convex open subset. The covariance matrix \( \Gamma := E((g(\theta_i) - E(g(\theta_i)))(g(\theta_i) - E(g(\theta_i)))^\top) \) exists and is nonsingular. For any \( p = 1, 2, \ldots, m \), the elements of the Hessian matrix of \( g_p(\cdot) \) are bounded functions. For any \( p = 1, 2, \ldots, m \), the function \( g_p(\cdot) \) satisfies \( E(((\partial / (\partial z_j))g_p(z)|_{z=\theta_i})^4) < \infty \) for any \( j = 1, 2, \ldots, m \).

These impose conditions on the smoothness of \( h(\cdot) \) and \( g(\cdot) \) and the existence of the moments. Assumption 7 is used for the continuous mapping theorem for the proof of consistency. Assumption 8 is stronger than Assumption 7 and is used to apply the delta method to derive the asymptotic distribution. Assumption 9 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 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., \( g(a) = 1(a \leq c) \) for some \( c \in \mathbb{R} \)) or when a quantile is estimated. The existence of the first derivative is crucial for analyzing the asymptotic property of \( \hat{H} \) under \( N/T^2 \to 0 \). The existence of the second derivative is useful for evaluating the order of the asymptotic bias. This assumption also guarantees that the asymptotic variance is finite.

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

**Theorem 3.** Let \( r^* = 4 \) if \( \theta_i = \mu_i \) such that \( H = h(E(g(\mu_i))) \) for some \( h \) and \( g \), and \( r^* = 8 \) if \( \theta_i \) contains \( \gamma_{k,i} \) for some \( k \). Suppose also that Assumptions 1, 2, 3, 7, and 9 hold for \( r_m = 4 \) and \( r_d = r^* \). As \( N, T \to \infty \), we have

\[
\hat{H} \overset{p}{\to} H.
\]
Moreover, suppose that Assumption 8 holds in addition. As \( N, T \to \infty \) with \( N/T^2 \to 0 \),

\[
\sqrt{N}(\hat{H} - H) \rightsquigarrow N \left(0, \nabla h(G)\Gamma(\nabla h(G))^\top\right).
\]

The theorem states that \( \hat{H} \) is consistent when both \( N \) and \( T \) tend to infinity, and is asymptotically normal with mean zero when \( N/T^2 \to 0 \). Importantly, contrary to the discussions in Section 5, the distribution of \( \theta_i \) does not need to be continuous and can be discrete. The remarkable result is that the asymptotically unbiased estimation holds under \( N/T^2 \to 0 \). This condition is weaker than that for \( P \) which is \( N^{2+r}/T^{2r} \to 0 \). This result is because of the smoothness of \( g(\cdot) \) and the fact that \( \hat{\theta}_i \) is first-order unbiased for \( \theta_i \). As we shall see below, the condition \( N/T^2 \to 0 \) is needed because of the incidental parameter bias in \( \hat{\gamma}_{k,i} \) and the bias caused by the nonlinearity of \( g(\cdot) \).

In order to obtain a better understanding of the results in the theorem, we first consider the case in which \( \theta_i = \mu_i \) (so that \( l = 1 \)), \( h \) is an identity function (so that \( n = 1 \)), and \( g \) is a scalar function (so that \( m = 1 \)). Denote our parameter of interest as \( G_\mu := E(g(\mu_i)) \) and let \( \hat{G}_\mu := N^{-1} \sum_{i=1}^{N} g(\hat{\mu}_i) \). We observe the following expansion:

\[
\sqrt{N} \left( \hat{G}_\mu - G_\mu \right) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (g(\mu_i) - E(g(\mu_i))) + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \bar{w}_i g'(\mu_i) + \frac{1}{2\sqrt{N}} \sum_{i=1}^{N} (\bar{w}_i)^2 g''(\tilde{\mu}_i), \quad (6)
\]

where \( \tilde{\mu}_i \) is between \( \mu_i \) and \( \hat{\mu}_i \). The second term in (6) 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 \( G_\mu \). 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 \( \mu_i \). 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.

When our parameter of interest involves \( \gamma_{k,i} \) for some \( k \), we encounter an additional source of bias. Let us consider the case in which \( \theta_i = \gamma_{k,i} \) for some \( k \) (so that \( l = 1 \)), \( h \) is an identity function (so that \( n = 1 \)), and \( g \) is a scalar function (so that \( m = 1 \)). Denote our parameter of interest as \( G_{\gamma_k} := E(g(\gamma_{k,i})) \) and let \( \hat{G}_{\gamma_k} := N^{-1} \sum_{i=1}^{N} g(\hat{\gamma}_{k,i}) \). The autocovariance estimator, \( \hat{\gamma}_{k,i} \), is expanded as follows:

\[
\hat{\gamma}_{k,i} = \gamma_{k,i} + \frac{1}{T-k} \sum_{t=k+1}^{T} (w_{it}w_{i,t-k} - \gamma_{k,i}) - (\bar{w}_i)^2 + O_p \left( \frac{1}{T^2} \right).
\]
It is important to observe that the second term 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 Taylor’s theorem, we have

$$\sqrt{N}(\hat{G}_{\gamma_k} - G_{\gamma_k}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( g(\gamma_{k,i}) - E(g(\gamma_{k,i})) \right)$$

$$+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\hat{\gamma}_{k,i} - \gamma_{k,i}) g'(\gamma_{k,i}) + \frac{1}{2\sqrt{N}} \sum_{i=1}^{N} (\hat{\gamma}_{k,i} - \gamma_{k,i})^2 g''(\gamma_{k,i})$$

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

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

$$- \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\bar{w}_i)^2 g'(\gamma_{k,i}) + \frac{1}{2\sqrt{N}} \sum_{i=1}^{N} (\hat{\gamma}_{k,i} - \gamma_{k,i})^2 g''(\gamma_{k,i}) + O_p \left( \frac{\sqrt{N}}{T^{3/2}} \right),$$

where the second equality is obtained by 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 (9). This bias is caused by the estimation of $\mu_i$ 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 (7) yields the asymptotic normality of $\hat{G}$. The term in (8) has a mean of zero and is of order $O_p(1/\sqrt{T})$. The second term in (9) is the nonlinearity bias term that also appears in $\hat{G}_{\mu}$. This is also of order $O_p(\sqrt{N}/T)$.

### 6.2 Jackknife bias-correction

We provide a theoretical justification of the half-panel jackknife (HPJ) bias-corrected estimator (3), which is based on the bias-correction method proposed by Dhaene and Jochmans (2015). It results that the bias of order $O(1/T)$ in $\hat{H}$ is eliminated by the HPJ procedure. Recall the definitions: let $\hat{H}^{(1)}$ and $\hat{H}^{(2)}$ be the estimators of $H$ computed using $\{y_{it}\}_{t=1}^{T/2}$ and $\{y_{it}\}_{t=T/2+1}^{T}$, respectively, with even $T$. The HPJ estimator of $H$ is $\hat{H}^{H} := 2\hat{H} - \bar{H}$, where $\bar{H} := (\hat{H}^{(1)} + \hat{H}^{(2)})/2$.

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

25
Assumption 10. The function $g(\cdot)$ is thrice differentiable. The covariance matrix $\Gamma := E((g(\theta_i) - E(g(\theta_i)))(g(\theta_i) - E(g(\theta_i)))^\top)$ exists and is nonsingular. For any $p = 1, 2, \ldots, m$, the function $g_p(\cdot)$ satisfies $E(((\partial/(\partial z_j))g_p(z)|_{z=\theta_i})^4) < \infty$ for any $j = 1, 2, \ldots, m$, and $E(((\partial^2/(\partial z_{j_1}\partial z_{j_2}))g_p(z)|_{z=\theta_i})^4) < \infty$ for any $j_1, j_2 = 1, 2, \ldots, m$. All of the third-order derivatives of $g(\cdot)$ are bounded.

Assumption 10 requires that $g(\cdot)$ is thrice differentiable, contrary to Assumption 9, and imposes stronger moment conditions. This condition is needed to conduct a higher-order asymptotic expansion of $\hat{H}$.

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

**Theorem 4.** Let $r^* = 8$ if $\theta_i = \mu_i$ such that $H = h(E(g(\mu_i)))$ for some $h$ and $g$, and $r^* = 16$ if $\theta_i$ contains $\gamma_{k,i}$ for some $k$. Suppose that Assumptions 1, 2, 3, 8, and 10 are satisfied for $r_m = 8$ and $r_d = r^*$. Then, as $N, T \to \infty$ with $N/T^2 \to \nu$ for some $\nu \in [0, \infty)$, it holds that

$$\sqrt{N}(\hat{H}^H - H) \rightsquigarrow N \left(0, \nabla h(G)\Gamma(\nabla h(G))^\top \right).$$

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

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

where $B$ is a constant. Similarly, we have

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

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

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

Thus, the HPJ eliminates the bias term of order $O(1/T)$.

Remark 2. One may consider applying the half-series jackknife (HSJ) in Quenouille (1949, 1956) to each $\hat{\gamma}_{k,i}$, but we argue that the HSJ is not suitable in the current context. In the HSJ method, 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 $\{y_{it}\}^T_{t=1}$, and $\hat{\gamma}_{k,i}^{(2)}$ be that based on $\{y_{it}\}^T_{t=T/2+1}$. Let $\hat{\gamma}_{k,i} := (\hat{\gamma}_{k,i}^{(1)} + \hat{\gamma}_{k,i}^{(2)})/2$. The HSJ bias-corrected estimator of $\gamma_{k,i}$ is $\hat{\gamma}_{k,i}^H := \hat{\gamma}_{k,i} - (\hat{\gamma}_{k,i} - \hat{\gamma}_{k,i}) = 2\hat{\gamma}_{k,i} - \hat{\gamma}_{k,i}$. We then estimate $H$ using $\{\hat{\gamma}_{k,i}^H\}^N_{i=1}$. The HSJ can reduce the bias of order $1/T$ in $\hat{\gamma}_{k,i}$, and therefore, the incidental parameter bias in $\hat{H}$. However, it does not reduce the
bias caused by the nonlinearity of \( g(\cdot) \) nor that of \( h(\cdot) \). Indeed, our Monte Carlo simulations show that the HSJ cannot work as well as HPJ. Thus, we do not pursue here theoretical investigation of the HSJ. We note that when both \( h(\cdot) \) and \( g(\cdot) \) are linear, the HPJ and HSJ are numerically equivalent.

**Remark 3.** We may also consider a higher-order jackknife bias-correction. This is discussed in Dhaene and Jochmans (2015). The higher-order jackknife bias-correction is expected to eliminate bias of higher order than \( O(1/T) \). Here, we consider the third-order jackknife (TOJ). Suppose that \( T \) is a multiple of six.\(^8\) The panel data are divided into three subpanels: \( \{ (y_{it})^N_{i=1} \}^T_{t=1} \), \( \{ (y_{it})^N_{i=1} \}^{2T/3}_{t=T/3+1} \), and \( \{ (y_{it})^N_{i=1} \}^{T}_{t=2T/3+1} \). Let \( \hat{H}^{(3,1)} \), \( \hat{H}^{(3,2)} \), and \( \hat{H}^{(3,3)} \) be the estimates of \( H \) computed from each of these three subpanels. Following Dhaene and Jochmans (2015), the TOJ bias-corrected estimator has the following form:

\[
(1 + a_2 + a_3) \hat{H} - a_2 \left( \frac{\hat{H}^{(1)} + \hat{H}^{(2)}}{2} \right) - a_3 \left( \frac{\hat{H}^{(3,1)} + \hat{H}^{(3,2)} + \hat{H}^{(3,3)}}{3} \right),
\]

where \( a_2 \) and \( a_3 \) are constants, which are set so as to eliminate the first-order bias and the second-order bias of the estimator \( \hat{H} \). To determine values of \( a_2 \) and \( a_3 \), we observe the following expansion of \( \hat{H} \) shown in the proof of Theorem 4:

\[
\hat{H} = H + \frac{B}{T} + O_p \left( \frac{1}{T^{3/2}} \right) + o_p \left( \frac{1}{T} \right).
\]

Accordingly, if we assume the third and fourth terms in the right-hand side can be written as \( C/T^{3/2} + o_p(T^{-3/2}) \) for a constant \( C \), we can eliminate the first-order bias and the second-order bias of \( \hat{H} \) by setting

\[
\left( \frac{1 + a_2 + a_3}{T} - \frac{a_2}{T/2} - \frac{a_3}{T/3} \right) B = 0,
\]

\[
\left( \frac{1 + a_2 + a_3}{T^{3/2}} - \frac{a_2}{(T/2)^{3/2}} - \frac{a_3}{(T/3)^{3/2}} \right) C = 0.
\]

Solving these equations, we obtain

\[
a_2 = 1 - \frac{2(2^{3/2} - 2)}{2 \cdot 2^{3/2} - 3^{3/2} - 1} \approx 4.072, \quad a_3 = \frac{2^{3/2} - 2}{2 \cdot 2^{3/2} - 3^{3/2} - 1} \approx -1.536.
\]

Therefore, we have the following TOJ bias-corrected estimator for \( H \):\(^9\)

\[
\hat{H}^{TOJ} := 3.536 \hat{H} - 4.072 \left( \frac{\hat{H}^{(1)} + \hat{H}^{(2)}}{2} \right) + 1.536 \left( \frac{\hat{H}^{(3,1)} + \hat{H}^{(3,2)} + \hat{H}^{(3,3)}}{3} \right).
\]

---

\(^8\)See Dhaene and Jochmans (2015) for the treatment of the case in which \( T \) is not a multiple of 6.

\(^9\)This TOJ estimator is different from the one in Dhaene and Jochmans (2015, page 1018). They consider cases in which there is no bias of order \( T^{-3/2} \) and apply the TOJ to correct the bias of order \( T^{-2} \).
However, we do not examine its theoretical property here. Our Monte Carlo results indicate that the TOJ 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 (2015): a higher-order jackknife may inflate the bias at an order higher than that to be corrected. We also find that the variance inflation may be substantial in certain cases.

6.3 Cross-sectional bootstrap

In this section, we present the justification of the use of the cross-sectional bootstrap. The first theorem is concerned with $\hat{H}$, and the second theorem discusses the case of $\tilde{H}$.

We require the following additional assumptions. The following assumption is used to satisfy Lyapunov’s conditions for $\hat{G}^*$ that represents the estimator of $G$ obtained with the bootstrap sample. Note that $\hat{H} = h(\hat{G})$. Specifically, we can write $\hat{G}^* := N^{-1} \sum_{i=1}^{N} z_{Ni} g(\hat{\theta}_i)$, where $(z_{N1}, z_{N2}, \ldots, z_{NN})$ is independent of $\{\{y_{jt}\}_{t=1}^{N}\}_{j=1}^{N}$ and follows a certain distribution with parameters $N$ and $p = (N^{-1}, N^{-1}, \ldots, N^{-1})$ such that $z_{Ni}$ is the number of times that $\{y_{jt}\}_{t=1}^{T}$ is drawn from $\{\{y_{jt}\}_{t=1}^{N}\}_{j=1}^{N}$ in the bootstrap sample. Let $P_z$ denote the probability measure with respect to $(z_{N1}, z_{N2}, \ldots, z_{NN})$ holding $\{\{y_{jt}\}_{t=1}^{N}\}_{j=1}^{N}$ fixed.

**Assumption 11.** The function $g = (g_1, g_2, \ldots, g_m) : \mathcal{O} \rightarrow \mathbb{R}^m$ is twice-continuously differentiable where $\mathcal{O} \subset \mathbb{R}^l$ is a convex open subset. The covariance matrix of $g(\theta_i)$, $\Gamma$, exists and is non-singular. The elements of the Hessian matrices of $g_p(\cdot)$ for $p = 1, 2, \ldots, m$, $g_{p_1}(\cdot)g_{p_2}(\cdot)$ for $p_1, p_2 = 1, 2, \ldots, m$, and $(g(\cdot)^\top g(\cdot))$ are bounded. For any $p = 1, 2, \ldots, m$, the function $g_p(\cdot)$ satisfies $E(((\partial/\partial z_j)g_p(z)\big|_{z=\theta_j})^4) < \infty$ for any $j = 1, 2, \ldots, l$. For any $p_1, p_2 = 1, 2, \ldots, m$, it holds that $E(((\partial/\partial z_j)g_{p_1}(z)\big|_{z=\theta_j}, g_{p_2}(\theta))2) < \infty$. For any $j = 1, 2, \ldots, l$, $E((\Gamma(\theta_i)^\top g(\theta_i)(\partial/\partial z_j)g_{p_1}(z)\big|_{z=0}, g_{p_2}(\theta))2) < \infty$ is satisfied.

The following theorem states that the bootstrap distribution converges to the asymptotic distribution of $\hat{H}$, but fails to capture the bias term.

**Theorem 5.** Let $r^* = 4$ if $\theta_i = \mu_i$ such that $H = h(E(g(\mu_i)))$ for some $h$ and $g$, and $r^* = 8$ if $\theta_i$ contains $\gamma_{k,i}$ for some $k$. Suppose that Assumptions 1, 2, 3, 8, and 11 hold for $r_m = 4$ and $r_d = r^*$. As $N, T \rightarrow \infty$, we have

$$\sup_{x \in \mathbb{R}} \left| P_z \left( \sqrt{N}(\hat{H} - \bar{H}) \leq x \right) - \Pr \left( N \left( 0, \nabla h(G)\Gamma(\nabla h(G))\top \right) \leq x \right) \right| \xrightarrow{p} 0.$$  

It is important to note that the bootstrap does not capture the bias properties of $\hat{G}$. The bootstrap distribution is asymptotically centered at zero regardless of the relative magnitude.
of \( N \) and \( T \). Thus, even if \( \hat{G} \) suffers from the bias as seen in Section 6.1, 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 (2015), 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. The proof is analogous to the proof of Theorem 5, and is thus omitted.

**Theorem 6.** Let \( r^* = 4 \) if \( \theta_i = \mu_i \) such that \( H = h(E(g(\mu_i))) \) for some \( h \) and \( g \), and \( r^* = 8 \) if \( \theta_i \) contains \( \gamma_{k,i} \) for some \( k \). Suppose that Assumptions 1, 2, 3, 8, and 11 are satisfied for \( r_m = 4 \) and \( r_d = r^* \). As \( N, T \to \infty \), we have

\[
\sup_{x \in \mathbb{R}} \left| P_z \left( \sqrt{N} (\hat{H}^* - \hat{H}) \leq x \right) - \Pr \left( N (0, \nabla h(G) \Gamma(\nabla h(G))^{\top}) \leq x \right) \right| \xrightarrow{p} 0.
\]

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 has a smaller bias, the bootstrap approximation would be more comfortably used for the HPJ estimator.

## 7 KS tests

This section presents two Kolmogorov–Smirnov (KS) tests as applications of the weak convergence results for the distribution estimators. We first consider a test for parametric specifications on the distribution of the heterogeneous mean, autocovariance, or autocorrelation. The second application tests whether the distributions of the mean, autocovariance, or autocorrelation are the same across different groups.

### 7.1 Testing parametric specifications

This subsection develops a testing procedure for hypotheses on parametric specifications of the distribution of \( \mu_i \), \( \gamma_{k,i} \), or \( \rho_{k,i} \). The test is based on one-sample KS statistics based on the empirical distributions of \( \hat{\mu}_i \), \( \hat{\gamma}_{k,i} \), and \( \hat{\rho}_{k,i} \). 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 are readily available.

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:

\[ H_0 : P_0 = Q \text{ v.s. } H_1 : P_0 \neq Q, \]

where \( P_0 \) is \( P_0^{\mu}, P_0^{\gamma_k}, \) or \( P_0^{\rho_k} \) and \( Q \) is a known continuous distribution. The hypothesis is concerned with whether the distribution \( P_0 \) is the same as \( Q \). Note that \( Q \) cannot be discrete probability distributions because our asymptotic analyses are based on Assumption 4, 5, or 6.

Our test is based on one-sample KS statistic (Kolmogorov, 1933; Smirnov, 1944):

\[
KS_1 := \sqrt{N} \left\| \mathbb{P}_N - Q \right\|_\infty = \sqrt{N} \sup_{f \in F} | \mathbb{P}_N f - Q f |,
\]

where \( \left\| \cdot \right\|_\infty \) is the uniform norm and \( \mathbb{P}_N = \mathbb{P}_N^{\mu}, \mathbb{P}_N^{\gamma_k} \), or \( \mathbb{P}_N^{\rho_k} \), depending on \( P_0 \) in the hypothesis. The test statistic measures the distance between the empirical distribution and the null distribution. We note that \( KS_1 \) is different from the usual one-sample KS statistic in the sense that they are based on the empirical distributions of the estimates \( \hat{\mu}_i, \hat{\gamma}_{k,i} \), or \( \hat{\rho}_{k,i} \).

We derive the asymptotic distribution of \( KS_1 \) under \( H_0 \), utilizing Theorem 2. The following theorem presents the asymptotic null distribution.

**Theorem 7.** Suppose that Assumptions 1, 2, 3, and 4 hold with \( r_m = 2r \), \( r_d = 2r \), and a natural number \( r \geq 1 \) for the case that \( \mathbb{P}_N = \mathbb{P}_N^{\mu} \) and \( P_0 = P_0^{\mu} \), Assumptions 1, 2, 3, and 5 hold with \( r_m = 2r \), \( r_d = 2r \), and a natural number \( r \geq 2 \) for the case that \( \mathbb{P}_N = \mathbb{P}_N^{\gamma_k} \) and \( P_0 = P_0^{\gamma_k} \), and Assumptions 1, 2, 3, and 6 hold with \( r_m = 2r \), \( r_d = 2r \), and a natural number \( r \geq 2 \) for the case that \( \mathbb{P}_N = \mathbb{P}_N^{\rho_k} \) and \( P_0 = P_0^{\rho_k} \). When \( N, T \to \infty \) with \( N^{2+r}/T^{2r} \to 0 \), it holds that \( KS_1 \) converges in distribution to \( \left\| G_Q \right\|_\infty \) under \( H_0 \).

This theorem shows that the asymptotic null distribution of \( KS_1 \) is the uniform norm of a Gaussian process. The asymptotic null distribution in the theorem is identical to those of the usual one-sample KS statistics developed in Kolmogorov (1933) and Smirnov (1944) so that it is equivalent to that of the one-sample KS statistics based on the true \( \mu_i, \gamma_{k,i}, \) or \( \rho_{k,i} \). This is because the estimation errors in \( \hat{\mu}_i, \hat{\gamma}_{k,i} \), and \( \hat{\rho}_{k,i} \) can be ignored asymptotically under the condition \( N^{2+r}/T^{2r} \to 0 \).

Note that the asymptotic distributions do not depend on \( Q \), 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),

\[
\Pr(\left\| G_Q \right\|_\infty \leq a) = 1 - 2 \sum_{j=1}^{\infty} (-1)^{j-1} \exp \left(-2j^2a^2\right),
\]

(10)
for any continuous distribution $Q$, with $a > 0$. The right-hand side of (10) does not depend on $Q$. Moreover, the critical values are readily available in many statistical software packages and the implementation of our tests is easy.

### 7.2 Testing the difference in degrees of heterogeneity

We develop a test to examine whether the distributions of $\mu_i$, $\gamma_{k,i}$, and $\rho_{k,i}$ differ across distinct groups. The test statistic is a two-sample KS statistic based on our empirical distribution estimators. We develop the asymptotic null distribution of the test statistic.

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 a testing procedure for such hypotheses without any parametric specification. Suppose that we have two panel data sets for males and females. One may also be interested in whether the distribution of individual average incomes differs between 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 a testing procedure 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 $N_1 \neq N_2$. Define $y_{i,(1)} := \{y_{it,(1)}\}_{t=1}^{T_1}$ for $i = 1, \ldots, N_1$ and $y_{i,(2)} := \{y_{it,(2)}\}_{t=1}^{T_2}$ for $i = 1, \ldots, N_2$.

We estimate the distributions of the mean, autocovariances, or autocorrelations 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 $P_{\hat{\mu}_{0,(a)}}$ and the empirical distribution of $\hat{\mu}_{i,(a)}$ by $\hat{P}_{N_a,(a)}$ for $a = 1, 2$, $P^{\gamma_{k,i}}_{0,(a)}$, $P^{\gamma_{k,i}}_{N_a,(a)}$, $P^{\rho_{k,i}}_{0,(a)}$, and $P^{\rho_{k,i}}_{N_a,(a)}$, for $a = 1, 2$ are defined analogously.

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

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

where $P_{0,(a)} = P^{\mu}_{0,(a)}$, $P^{\gamma_{k,i}}_{0,(a)}$, or $P^{\rho_{k,i}}_{0,(a)}$ for $a = 1, 2$. Under the null hypothesis $H_0$, the distribution of the heterogeneous mean, autocovariance, or autocorrelation is identical across the two groups.

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

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

where $\hat{P}_{N_a,(a)} = \hat{P}^{\mu}_{N_a,(a)}$, $\hat{P}^{\gamma_{k,i}}_{N_a,(a)}$, or $\hat{P}^{\rho_{k,i}}_{N_a,(a)}$, depending on $H_0$. The test statistic measures the distance between the empirical distributions of the two groups. $KS_2$ differs from the usual
two-sample KS statistic in the sense that \( KS_2 \) is based on the empirical distributions of the estimates (e.g., \( \hat{\mu}_{i,(a)} \) if we consider the distribution of the mean).

We introduce the following assumption on the data sets.

**Assumption 12.** 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, 3, and 4 with \( r_m = 2r \), \( r_d = 2r \) for some natural number \( r \geq 1 \) for the case that \( P_{0,(a)} = P_{0,(a)}^\mu \), and Assumptions 1, 2, 3, and 5 with \( r_m = 2r \), \( r_d = 2r \) for some natural number \( r \geq 2 \) for the case that \( P_{0,(a)} = P_{0,(a)}^{\gamma_k} \), and Assumptions 1, 2, 3, and 6 with \( r_m = 2r \), \( r_d = 2r \) for some natural number \( r \geq 2 \) for the case that \( P_{0,(a)} = P_{0,(a)}^{\rho_k} \). \( (y_{1,(1)}, \ldots, y_{N_1,(1)}) \) and \( (y_{1,(2)}, \ldots, y_{N_2,(2)}) \) 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.

The asymptotic null distribution of \( KS_2 \) is derived using Theorem 2.

**Theorem 8.** Suppose that Assumption 12 is satisfied. When \( N_1, T_1 \to \infty \) with \( N_1^{2+r}/T_1^{2r} \to 0 \) and \( N_2, T_2 \to \infty \) with \( N_2^{2+r}/T_2^{2r} \to 0 \), and \( N_1/(N_1 + N_2) \to \lambda \) for some \( \lambda \in (0, 1) \), it holds that \( KS_2 \) converges in distribution to \( \| G_{P_{0,(1)}} \|_\infty \) under \( H_0 \).

This theorem shows that the asymptotic null distribution of \( KS_2 \) is the uniform norm of a Gaussian process. The condition \( N_1^{2+r}/T_1^{2r} \to 0 \) is 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 7 when we set \( Q = P_{0,(1)} \). Hence, the asymptotic null distributions can be evaluated easily by (10) and the critical values of our test are readily available.

**Remark 4.** When the true distributions of \( \hat{\mu}_{i,(1)} \) and \( \hat{\mu}_{i,(2)} \) are the same, i.e., when \( P_{T_1,(1)}^\mu = P_{T_2,(2)}^\mu \), neither the condition \( N_1^{2+r}/T_1^{2r} \to 0 \) nor the condition \( N_2^{2+r}/T_2^{2r} \to 0 \) is needed to establish Theorem 8. This is clear from the proof of Theorem 8. 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 hypothesis \( H_0 \) without restricting the relative order of \( N_a \) and \( T_a \) for \( a = 1, 2 \). Note that we still need the condition \( N_1/(N_1 + N_2) \to \lambda \in (0, 1) \). Of course, the same argument holds for autocovariances and autocorrelations.
8 Monte Carlo simulations

This section presents Monte Carlo simulation results. The simulations designs are motivated by our empirical applications in Section 4. The purpose of the simulations is to demonstrate the finite-sample properties of our proposed procedure and the performance of the proposed bias-correction and cross-sectional bootstrap. We conduct the simulations using R 3.2.1 for Windows 8.1, with 5000 replications.

8.1 Designs

The data are generated by an AR(1) process

\[ y_{it} = (1 - \phi_i) \xi_i + \phi_i y_{i,t-1} + \sqrt{(1 - \phi_i^2)} \sigma_i^2 \epsilon_{it}, \]

for \( i = 1, 2, \ldots, N \) and \( t = 1, 2, \ldots, T \), where \( \epsilon_{it} \sim i.i.d. N(0, 1) \). The random variables \( \xi_i, \phi_i, \) and \( \sigma_i^2 \) are individual-specific variables whose joint distribution is specified below. We generate the initial observations \((y_{i0}, \epsilon_{i0})\) for \( i = 1, 2, \ldots, N \) from the following stationary distribution:

\[ \begin{pmatrix} y_{i0} \\ \epsilon_{i0} \end{pmatrix} \sim N \left( \begin{pmatrix} \xi_i \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_i^2 & 1 \\ 1 & 1 \end{pmatrix} \right). \]

We set \( N = 100 \) and 1000, and \( T = 12, 24, \) and 48.

It is important to note that in this process, we have

\[
\begin{align*}
\mu_i &= \xi_i, & \gamma_{0,i} &= \sigma_i^2, & \rho_{1,i} &= \phi_i, \\
E(\mu_i) &= E(\xi_i), & E(\gamma_{0,i}) &= E(\sigma_i^2), & E(\rho_{1,i}) &= E(\phi_i), \\
var(\mu_i) &= var(\xi_i), & var(\gamma_{0,i}) &= var(\sigma_i^2), & var(\rho_{1,i}) &= var(\phi_i), \\
Cor(\mu_i, \gamma_{0,i}) &= Cor(\xi_i, \sigma_i), & Cor(\mu_i, \rho_{1,i}) &= Cor(\xi_i, \phi_i), & Cor(\gamma_{0,i}, \rho_{1,i}) &= Cor(\sigma_i, \phi_i).
\end{align*}
\]

We consider two designs on \((\xi_i, \phi_i, \sigma_i^2)\) motivated by our empirical estimates.

**Design 1:** Design 1 is motivated by the empirical results for earnings dynamics. We generate \( \phi_i \sim i.i.d. 0.95 Beta(0.8, 0.7) \) independent of \((\xi_i, \sigma_i^2)\). The independence is reflected by the empirical result of no statistical evidence for the correlation of individual first-order autocorrelations and means or variances. We generate \((\xi_i, \sigma_i^2)\) using the truncated normal distribution:

\[ \begin{pmatrix} \xi_i \\ \sigma_i^2 \end{pmatrix} \sim i.i.d. N \left( \begin{pmatrix} m_\xi \\ m_\sigma \end{pmatrix}, \begin{pmatrix} s_\xi^2 & \rho_{\xi,\sigma} s_\xi s_\sigma \\ \rho_{\xi,\sigma} s_\xi s_\sigma & s_\sigma^2 \end{pmatrix} \right), \]
conditional on $|\xi_i - m_{\xi}| < 2.58 s_{\xi}$ and $|\sigma_i^2 - m_{\sigma}| < 2.58 s_{\sigma}$. Motivated by our empirical estimates, we set

$$m_{\xi} = 0.010, \quad m_{\sigma} = 0.185,$$

$$s_{\xi}^2 = 0.122, \quad s_{\sigma}^2 = 0.005, \quad \rho_{\xi,\sigma} = -0.274.$$

These values are the same as the HPJ estimates in our application except for $m_{\sigma}$. This exception and the truncation guarantee that $|\phi_i| < 1$ and $\sigma_i^2 > 0$ for any $i$.

**Design 2:** Design 2 is motivated by the empirical results for productivity dynamics. To reflect the dependence of $(\xi_i, \phi_i, \sigma_i^2)$, they are generated by the following truncated normal distribution:

$$\begin{pmatrix} \xi_i \\ \sigma_i^2 \\ \phi_i \end{pmatrix} \sim \text{i.i.d. } N \begin{pmatrix} m_{\xi} \\ m_{\sigma} \\ m_{\phi} \end{pmatrix},$$

$$\begin{pmatrix} s_{\xi}^2 & \rho_{\xi,\sigma} & \rho_{\xi,\phi} \\ \rho_{\xi,\sigma}^2 & s_{\sigma}^2 & \rho_{\sigma,\phi} \\ \rho_{\xi,\phi}^2 & \rho_{\phi,\sigma} & s_{\phi}^2 \end{pmatrix},$$

conditional on $|\xi_i - m_{\xi}| < 2.58 s_{\xi}$, $|\sigma_i^2 - m_{\sigma}| < 2.58 s_{\sigma}$, $|\phi_i - m_{\phi}| < 2.58 s_{\phi}$. We set

$$m_{\xi} = 2.960, \quad m_{\sigma} = 0.082, \quad m_{\phi} = 0.570,$$

$$s_{\xi}^2 = 0.080, \quad s_{\sigma}^2 = 0.001, \quad s_{\phi}^2 = 0.023,$$

$$\rho_{\xi,\sigma} = -0.235, \quad \rho_{\sigma,\phi} = -0.056, \quad \rho_{\xi,\phi} = 0.258.$$

These values are the same as the HPJ estimates in our application except for $m_{\sigma}$. This exception and the truncation guarantee that $|\phi_i| < 1$ and $\sigma_i^2 > 0$ for any $i$.

**Parameters:** We estimate the distributional properties of the heterogeneous mean $\mu_i$, the heterogeneous variance $\gamma_{0,i}$, and the heterogeneous first-order autocorrelation $\rho_{1,i}$. Specifically, the parameters of interest are the means, the variances, the 25%, 50%, and 75% quantiles, and the correlation coefficients between $\mu_i$, $\gamma_{0,i}$, and $\rho_{1,i}$.

**Estimators:** We consider four estimators. The first is the estimator based on the empirical distribution (ED). The second is the HPJ bias-corrected estimator (HPJ). The third is the TOJ bias-corrected estimator (TOJ) in Remark 3. The fourth is the HSJ bias-corrected estimator (HSJ) in Remark 2.

**8.2 Results**

Tables 4–6 and 7–9 summarize the results of Monte Carlo simulations with designs 1 and 2, respectively. They present the bias and the standard deviation (std) of each estimator and the coverage probability (cp) of the 95% confidence interval based on the cross-sectional
bootstrap for each estimator. In the column labeled “true,” the true value of the corresponding quantity is presented. We note that all estimates of \( E(\mu_i) \) 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 for ED are informative regarding the finite-sample behavior and the importance of bias-correction. First, when \( T \) is small, ED has large biases in some parameters of interest, e.g., \( E(\rho_{1,i}) \), the quantiles of \( \rho_{1,i} \), \( Corr(\mu_i, \rho_{1,i}) \), and \( Corr(\mu_i, \rho_{1,i}) \) with designs 1 and 2. The biases decline when \( T \) increases while holding \( N \) fixed, which is expected by our theoretical investigation. However, a large portion of biases remain even when \( N = 100 \) and \( T = 48 \). Second, the standard deviations of ED decrease as \( N \) becomes large along with the other estimators. This result is expected because our asymptotic results show that the variances are of order \( O(1/N) \). Finally, the coverage probabilities of ED for many parameters differ significantly from 0.95 because of the large biases. These points indicate the importance of developing the bias-correction method.

The HPJ bias-correction method is successful in reducing the biases and in improving the coverage probabilities. The HPJ biases are moderate, even when \( T \) is small. It is remarkable that HPJ works well, especially when the biases in ED are large, e.g., the quantities regarding \( \rho_{1,i} \) in design 2. In addition, the biases in HPJ tend to decrease as \( T \) increases, as expected by our asymptotic results. The coverage probabilities of HPJ are satisfactory for many parameters in many situations. For example, they are considerably close to 0.95 for \( E(\mu_i) \), \( var(\mu_i) \), \( var(\gamma_{0,i}) \), \( E(\rho_{1,i}) \), and \( Corr(\mu_i, \rho_{1,i}) \) with designs 1 and 2, \( var(\rho_{1,i}) \) with design 1, and \( E(\gamma_{0,i}) \) and \( Corr(\mu_i, \gamma_{0,i}) \) with design 2. Interestingly, HPJ also works to estimate quantiles in many cases, although our theoretical justification of 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.

While in some cases the TOJ bias-correction is more successful than the HPJ bias-correction, there are also many cases in which the performance of the TOJ bias-correction is significantly worse, especially when estimating quantiles. In addition, the standard deviations of TOJ tend to be larger than for HPJ, even when both \( N \) and \( T \) are large. For example, they are observed when estimating \( Corr(\mu_i, \gamma_{0,i}) \), \( Corr(\mu_i, \rho_{1,i}) \), and \( Corr(\gamma_{0,i}, \rho_{1,i}) \) with designs 1 and 2. What is worse, the coverage probabilities of TOJ are inadequate, especially when estimating quantiles because of the large biases. This result corresponds with the note in Remark 3: 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 ED and HPJ are not large. We thus recommend HPJ over
the higher-order jackknife as a precaution.

The performance of the HSJ bias-corrected estimator is significantly worse in terms of the bias-correction and the coverage probability. This is because the HSJ bias-correction cannot reduce the bias caused by a nonlinear function, such as the variance, as we discuss in Remark 2. Moreover, HSJ substantially increases biases in some cases. For example, they are observed for $\text{var}(\rho_{1,i})$ with designs 1 and 2, and $\text{Corr}(\gamma_{0,i}, \rho_{1,i})$ with design 2. Based on these results, we do not recommend HSJ bias-correction.

8.3 Summary

Given the results of these Monte Carlo experiments, we conclude that our preferred procedure is HPJ. ED is often considerably biased, whereas HPJ can alleviate the bias problem without significant variance inflation. TOJ may be used for the estimation of smooth functions, 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.

9 Conclusion

This paper proposes methods to analyze heterogeneous dynamic structure using panel data. Our proposed methods are easy to implement without requiring model specification. We first compute the sample mean, the sample autocovariances, and/or the sample autocorrelations 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, autocovariances, and/or autocorrelations. We show that the estimator for the distribution function has a bias of order slower than $1/T$. When the parameter of interest can be written as the expected value of a smooth function of the heterogeneous mean and/or autocovariances, the bias of the estimator becomes of order $1/T$ and can be reduced by the half-panel jackknife bias-correction method. In addition, we develop an inference procedure based on the cross-sectional bootstrap. We also present two extensions based on the proposed procedures: testing parametric specifications on the distributions and testing the differences in heterogeneous dynamic structures across distinct groups. We apply our developed procedures to earnings dynamics in the US and the productivity dynamics of Chilean firms and obtain new empirical evidence that both dynamics exhibit much heterogeneity. The results of Monte Carlo simulations show that our asymptotic analyses are informative regarding the finite-sample properties of the proposed estimators and inference procedures. Based on the simulation results, we recommend the half-panel jackknife estimator.
### Table 4: Monte Carlo simulation results with design 1 ($N = 100, 1000$ and $T = 12$)

<table>
<thead>
<tr>
<th></th>
<th>$N = 100$</th>
<th></th>
<th></th>
<th>$N = 1000$</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>true</td>
<td>bias</td>
<td>std</td>
<td>cp</td>
<td>true</td>
<td>bias</td>
<td>std</td>
<td>cp</td>
</tr>
<tr>
<td>$\mu$ Mean</td>
<td>0.010</td>
<td>-0.000</td>
<td>0.042</td>
<td>0.943</td>
<td>-0.000</td>
<td>0.042</td>
<td>0.943</td>
<td>-0.000</td>
</tr>
<tr>
<td>$\mu$ Var</td>
<td>0.112</td>
<td>0.059</td>
<td>0.025</td>
<td>0.295</td>
<td>0.032</td>
<td>0.025</td>
<td>0.811</td>
<td>0.019</td>
</tr>
<tr>
<td>$\gamma$ 25%Q</td>
<td>-0.222</td>
<td>-0.039</td>
<td>0.056</td>
<td>0.823</td>
<td>-0.021</td>
<td>0.071</td>
<td>0.837</td>
<td>-0.012</td>
</tr>
<tr>
<td>$\gamma$ 50%Q</td>
<td>0.010</td>
<td>0.005</td>
<td>0.051</td>
<td>0.873</td>
<td>0.003</td>
<td>0.064</td>
<td>0.842</td>
<td>0.003</td>
</tr>
<tr>
<td>$\gamma$ 75%Q</td>
<td>0.242</td>
<td>0.044</td>
<td>0.056</td>
<td>0.797</td>
<td>0.024</td>
<td>0.070</td>
<td>0.818</td>
<td>0.013</td>
</tr>
<tr>
<td>$\gamma_0$ Mean</td>
<td>0.185</td>
<td>-0.059</td>
<td>0.010</td>
<td>0.000</td>
<td>-0.032</td>
<td>0.012</td>
<td>0.260</td>
<td>-0.019</td>
</tr>
<tr>
<td>$\gamma_0$ Var</td>
<td>0.005</td>
<td>0.005</td>
<td>0.002</td>
<td>0.316</td>
<td>0.004</td>
<td>0.003</td>
<td>0.799</td>
<td>0.003</td>
</tr>
<tr>
<td>$\gamma_0$ 25%Q</td>
<td>0.138</td>
<td>-0.082</td>
<td>0.008</td>
<td>0.000</td>
<td>-0.057</td>
<td>0.013</td>
<td>0.055</td>
<td>0.075</td>
</tr>
<tr>
<td>$\gamma_0$ 50%Q</td>
<td>0.185</td>
<td>-0.081</td>
<td>0.011</td>
<td>0.000</td>
<td>-0.047</td>
<td>0.017</td>
<td>0.258</td>
<td>0.157</td>
</tr>
<tr>
<td>$\gamma_0$ 75%Q</td>
<td>0.232</td>
<td>-0.060</td>
<td>0.016</td>
<td>0.087</td>
<td>-0.024</td>
<td>0.025</td>
<td>0.700</td>
<td>0.217</td>
</tr>
<tr>
<td>$\rho$ Mean</td>
<td>0.506</td>
<td>-0.243</td>
<td>0.033</td>
<td>0.000</td>
<td>-0.009</td>
<td>0.051</td>
<td>0.938</td>
<td>0.004</td>
</tr>
<tr>
<td>$\rho$ Var</td>
<td>0.090</td>
<td>0.022</td>
<td>0.014</td>
<td>0.647</td>
<td>-0.009</td>
<td>0.025</td>
<td>0.901</td>
<td>-0.047</td>
</tr>
<tr>
<td>$\rho$ 25%Q</td>
<td>0.236</td>
<td>-0.204</td>
<td>0.052</td>
<td>0.071</td>
<td>0.069</td>
<td>0.089</td>
<td>0.806</td>
<td>0.273</td>
</tr>
<tr>
<td>$\rho$ 50%Q</td>
<td>0.523</td>
<td>-0.235</td>
<td>0.047</td>
<td>0.016</td>
<td>-0.006</td>
<td>0.081</td>
<td>0.873</td>
<td>0.366</td>
</tr>
<tr>
<td>$\rho$ 75%Q</td>
<td>0.785</td>
<td>-0.262</td>
<td>0.042</td>
<td>0.001</td>
<td>-0.062</td>
<td>0.076</td>
<td>0.810</td>
<td>0.447</td>
</tr>
<tr>
<td>$\mu$ vs $\gamma_0$</td>
<td>-0.255</td>
<td>0.156</td>
<td>0.101</td>
<td>0.632</td>
<td>0.129</td>
<td>0.142</td>
<td>0.811</td>
<td>0.109</td>
</tr>
<tr>
<td>$\mu$ vs $\rho_1$</td>
<td>-0.001</td>
<td>0.000</td>
<td>0.099</td>
<td>0.932</td>
<td>0.001</td>
<td>0.161</td>
<td>0.929</td>
<td>0.000</td>
</tr>
<tr>
<td>$\gamma_0$ vs $\rho_1$</td>
<td>-0.000</td>
<td>-0.005</td>
<td>0.104</td>
<td>0.912</td>
<td>-0.066</td>
<td>0.170</td>
<td>0.901</td>
<td>-0.147</td>
</tr>
</tbody>
</table>

### Future work:
Several future research topics are possible. First, we may consider estimating the density function of the heterogeneous mean, autocovariances, or autocorrelations. We are currently working on this extension.

Second, we could consider prediction methods based on the proposed analysis. Given that our analysis estimates the distributions of heterogeneous means and autocorrelations, we could use them to construct best linear predictors of future values of $y_{it}$.

Third, it would be interesting to examine other quantities than mean and autocovari-
Table 5: Monte Carlo simulation results with design 1  \((N = 100, 1000\text{ and } T = 24)\)

<table>
<thead>
<tr>
<th></th>
<th>ED</th>
<th>HPJ</th>
<th>TOJ</th>
<th>HSJ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>true bias std cp</td>
<td>true bias std cp</td>
<td>true bias std cp</td>
<td>true bias std cp</td>
</tr>
<tr>
<td>(\mu) Mean</td>
<td>0.010 -0.001 0.039 0.946</td>
<td>-0.001 0.039 0.946</td>
<td>-0.001 0.039 0.946</td>
<td>-0.001 0.039 0.946</td>
</tr>
<tr>
<td>(\mu) Var</td>
<td>0.112 0.038 0.021 0.607</td>
<td>0.017 0.021 0.917</td>
<td>0.007 0.023 0.948</td>
<td>0.038 0.021 0.607</td>
</tr>
<tr>
<td>(\mu) 25%Q</td>
<td>-0.222 -0.024 0.052 0.850</td>
<td>-0.010 0.065 0.840</td>
<td>-0.004 0.103 0.881</td>
<td>-0.024 0.052 0.850</td>
</tr>
<tr>
<td>(\mu) 50%Q</td>
<td>-0.010 0.003 0.048 0.871</td>
<td>0.002 0.059 0.841</td>
<td>0.001 0.093 0.874</td>
<td>0.003 0.048 0.871</td>
</tr>
<tr>
<td>(\mu) 75%Q</td>
<td>-0.242 0.028 0.052 0.846</td>
<td>0.011 0.064 0.830</td>
<td>0.003 0.101 0.874</td>
<td>0.028 0.052 0.846</td>
</tr>
<tr>
<td>(\gamma_0) Mean</td>
<td>0.185 -0.038 0.009 0.026</td>
<td>-0.017 0.011 0.613</td>
<td>-0.007 0.015 0.859</td>
<td>-0.017 0.011 0.613</td>
</tr>
<tr>
<td>(\gamma_0) Var</td>
<td>0.005 0.004 0.002 0.300</td>
<td>0.003 0.002 0.868</td>
<td>0.002 0.004 0.952</td>
<td>0.008 0.004 0.150</td>
</tr>
<tr>
<td>(\gamma_0) 25%Q</td>
<td>-0.138 -0.057 0.009 0.003</td>
<td>-0.032 0.011 0.373</td>
<td>0.134 0.033 0.052</td>
<td>-0.047 0.010 0.022</td>
</tr>
<tr>
<td>(\gamma_0) 50%Q</td>
<td>-0.185 -0.053 0.011 0.016</td>
<td>-0.024 0.016 0.509</td>
<td>0.188 0.035 0.007</td>
<td>-0.037 0.012 0.178</td>
</tr>
<tr>
<td>(\gamma_0) 75%Q</td>
<td>-0.232 -0.036 0.015 0.318</td>
<td>-0.011 0.022 0.787</td>
<td>0.204 0.044 0.028</td>
<td>-0.011 0.017 0.782</td>
</tr>
<tr>
<td>(\rho_1) Mean</td>
<td>0.506 -0.120 0.031 0.023</td>
<td>0.003 0.039 0.942</td>
<td>0.010 0.061 0.942</td>
<td>0.003 0.039 0.942</td>
</tr>
<tr>
<td>(\rho_1) Var</td>
<td>0.090 0.006 0.011 0.926</td>
<td>-0.010 0.016 0.875</td>
<td>-0.008 0.032 0.934</td>
<td>0.302 0.024 0.000</td>
</tr>
<tr>
<td>(\rho_1) 25%Q</td>
<td>-0.236 -0.075 0.049 0.588</td>
<td>0.055 0.074 0.804</td>
<td>0.391 0.152 0.329</td>
<td>-0.011 0.056 0.858</td>
</tr>
<tr>
<td>(\rho_1) 50%Q</td>
<td>0.523 -0.113 0.047 0.366</td>
<td>0.008 0.072 0.854</td>
<td>0.365 0.133 0.289</td>
<td>-0.009 0.055 0.866</td>
</tr>
<tr>
<td>(\rho_1) 75%Q</td>
<td>0.785 -0.148 0.039 0.092</td>
<td>-0.035 0.063 0.831</td>
<td>0.271 0.117 0.410</td>
<td>0.008 0.053 0.861</td>
</tr>
</tbody>
</table>

For example, Arellano, Blundell, and Bonhomme (2015) point out the importance of nonlinearity and conditional skewness in earnings and consumption dynamics. It may be a valuable extension to develop methods to investigate the heterogeneity of quantities that describe nonlinearity and/or skewness.

Fourth, while we develop our analysis for stationary panel data, it would be interesting to consider nonstationary panel data. Two types of nonstationarity are relevant. The first is the effect of initial distributions. In this paper, we assume for simplicity that the initial
values are drawn from the stationary distributions. As we consider the case in which $T$ is large, the effect of the 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 and autocorrelations are not well defined in this situation, we require a different procedure to handle unit root cases.
Table 7: Monte Carlo simulation results with design 2 ($N = 100, 1000$ and $T = 12$)

<table>
<thead>
<tr>
<th></th>
<th>ED</th>
<th>HPJ</th>
<th>TOJ</th>
<th>HSJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$ Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.960</td>
<td>0.000</td>
<td>0.031</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td>0.073</td>
<td>0.024</td>
<td>0.013</td>
<td>0.590</td>
</tr>
<tr>
<td>$\mu$ Var</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.772</td>
<td>-0.022</td>
<td>0.042</td>
<td>0.845</td>
</tr>
<tr>
<td>$\mu$ 25%Q</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.960</td>
<td>-0.001</td>
<td>0.040</td>
<td>0.876</td>
</tr>
<tr>
<td>$\mu$ 50%Q</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.148</td>
<td>0.022</td>
<td>0.043</td>
<td>0.840</td>
</tr>
<tr>
<td>$\mu$ 75%Q</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.082</td>
<td>-0.023</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>$\gamma_0$ Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.001</td>
<td>0.001</td>
<td>0.000</td>
<td>0.461</td>
</tr>
<tr>
<td>$\rho_1$ Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.570</td>
<td>-0.253</td>
<td>0.029</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Finally, while we focus only on balanced panel data, an analysis based on unbalanced panel data would be useful. This extension should not be too difficult because even when the panel is unbalanced, there is no difficulty in estimating the mean and autocovariances for each individual, 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 require some effort.
<table>
<thead>
<tr>
<th></th>
<th>true</th>
<th>ED</th>
<th>HPJ</th>
<th>TOJ</th>
<th>HSJ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N = 100</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>μ Mean</strong></td>
<td>2.960</td>
<td>0.901</td>
<td>0.942</td>
<td>0.942</td>
<td>0.942</td>
</tr>
<tr>
<td><strong>μ Var</strong></td>
<td>0.973</td>
<td>0.013</td>
<td>0.012</td>
<td>0.831</td>
<td>0.936</td>
</tr>
<tr>
<td><strong>μ 25%Q</strong></td>
<td>2.772</td>
<td>-0.011</td>
<td>0.041</td>
<td>0.857</td>
<td>0.857</td>
</tr>
<tr>
<td><strong>μ 50%Q</strong></td>
<td>2.960</td>
<td>0.001</td>
<td>0.038</td>
<td>0.874</td>
<td>0.874</td>
</tr>
<tr>
<td><strong>μ 75%Q</strong></td>
<td>3.148</td>
<td>0.012</td>
<td>0.041</td>
<td>0.857</td>
<td>0.857</td>
</tr>
<tr>
<td>γ₀ Mean</td>
<td>0.082</td>
<td>-0.013</td>
<td>0.004</td>
<td>0.109</td>
<td>0.109</td>
</tr>
<tr>
<td>γ₀ Var</td>
<td>0.001</td>
<td>0.001</td>
<td>0.000</td>
<td>0.109</td>
<td>0.109</td>
</tr>
<tr>
<td>γ₀ 25%Q</td>
<td>0.061</td>
<td>-0.021</td>
<td>0.004</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>γ₀ 50%Q</td>
<td>0.082</td>
<td>-0.020</td>
<td>0.004</td>
<td>0.031</td>
<td>0.031</td>
</tr>
<tr>
<td>γ₀ 75%Q</td>
<td>0.103</td>
<td>-0.014</td>
<td>0.006</td>
<td>0.388</td>
<td>0.388</td>
</tr>
<tr>
<td>ρ₁ Mean</td>
<td>0.570</td>
<td>-0.124</td>
<td>0.022</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>ρ₁ Var</td>
<td>0.021</td>
<td>0.028</td>
<td>0.007</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>ρ₁ 25%Q</td>
<td>0.469</td>
<td>-0.166</td>
<td>0.034</td>
<td>0.013</td>
<td>0.013</td>
</tr>
<tr>
<td>ρ₁ 50%Q</td>
<td>0.570</td>
<td>-0.106</td>
<td>0.029</td>
<td>0.094</td>
<td>0.094</td>
</tr>
<tr>
<td>ρ₁ 75%Q</td>
<td>0.671</td>
<td>-0.063</td>
<td>0.027</td>
<td>0.388</td>
<td>0.388</td>
</tr>
<tr>
<td><strong>N = 1000</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>μ Mean</strong></td>
<td>2.960</td>
<td>0.000</td>
<td>0.009</td>
<td>0.948</td>
<td>0.948</td>
</tr>
<tr>
<td><strong>μ Var</strong></td>
<td>0.073</td>
<td>0.013</td>
<td>0.004</td>
<td>0.035</td>
<td>0.035</td>
</tr>
<tr>
<td><strong>μ 25%Q</strong></td>
<td>2.772</td>
<td>-0.015</td>
<td>0.013</td>
<td>0.770</td>
<td>0.770</td>
</tr>
<tr>
<td><strong>μ 50%Q</strong></td>
<td>2.960</td>
<td>0.000</td>
<td>0.012</td>
<td>0.925</td>
<td>0.925</td>
</tr>
<tr>
<td><strong>μ 75%Q</strong></td>
<td>3.148</td>
<td>0.015</td>
<td>0.013</td>
<td>0.760</td>
<td>0.760</td>
</tr>
<tr>
<td>γ₀ Mean</td>
<td>0.082</td>
<td>-0.013</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>γ₀ Var</td>
<td>0.001</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>γ₀ 25%Q</td>
<td>0.061</td>
<td>-0.021</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>γ₀ 50%Q</td>
<td>0.082</td>
<td>-0.020</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>γ₀ 75%Q</td>
<td>0.103</td>
<td>-0.013</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>ρ₁ Mean</td>
<td>0.570</td>
<td>-0.124</td>
<td>0.007</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>ρ₁ Var</td>
<td>0.021</td>
<td>0.028</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>ρ₁ 25%Q</td>
<td>0.469</td>
<td>-0.168</td>
<td>0.011</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>ρ₁ 50%Q</td>
<td>0.570</td>
<td>-0.106</td>
<td>0.009</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>ρ₁ 75%Q</td>
<td>0.671</td>
<td>-0.061</td>
<td>0.009</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>µ vs γ₀</strong></td>
<td>0.239</td>
<td>-0.072</td>
<td>0.098</td>
<td>0.875</td>
<td>0.875</td>
</tr>
<tr>
<td><strong>µ vs ρ₁</strong></td>
<td>-0.217</td>
<td>0.108</td>
<td>0.099</td>
<td>0.806</td>
<td>0.806</td>
</tr>
<tr>
<td><strong>γ₀ vs ρ₁</strong></td>
<td>-0.112</td>
<td>0.336</td>
<td>0.093</td>
<td>0.065</td>
<td>0.065</td>
</tr>
</tbody>
</table>

### 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.
<table>
<thead>
<tr>
<th></th>
<th>true</th>
<th>ED</th>
<th>HPJ</th>
<th>TOJ</th>
<th>HSJ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>2.960</td>
<td>0.000</td>
<td>0.028</td>
<td>0.945</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Var</strong></td>
<td>0.073</td>
<td>0.007</td>
<td>0.011</td>
<td>0.918</td>
<td>0.001</td>
</tr>
<tr>
<td><strong>25%Q</strong></td>
<td>2.772</td>
<td>-0.006</td>
<td>0.039</td>
<td>0.871</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>50%Q</strong></td>
<td>2.960</td>
<td>0.000</td>
<td>0.036</td>
<td>0.884</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>75%Q</strong></td>
<td>3.148</td>
<td>0.000</td>
<td>0.039</td>
<td>0.866</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>γ0 Mean</strong></td>
<td>0.082</td>
<td>-0.007</td>
<td>0.004</td>
<td>0.494</td>
<td>-0.001</td>
</tr>
<tr>
<td><strong>γ0 Var</strong></td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.567</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>γ0 25%Q</strong></td>
<td>0.061</td>
<td>-0.012</td>
<td>0.004</td>
<td>0.187</td>
<td>-0.004</td>
</tr>
<tr>
<td><strong>γ0 50%Q</strong></td>
<td>0.082</td>
<td>-0.012</td>
<td>0.004</td>
<td>0.268</td>
<td>-0.003</td>
</tr>
<tr>
<td><strong>γ0 75%Q</strong></td>
<td>0.103</td>
<td>-0.007</td>
<td>0.006</td>
<td>0.661</td>
<td>-0.001</td>
</tr>
<tr>
<td><strong>μ vs γ0</strong></td>
<td>0.239</td>
<td>-0.049</td>
<td>0.099</td>
<td>0.904</td>
<td>-0.027</td>
</tr>
<tr>
<td><strong>μ vs p1</strong></td>
<td>-0.217</td>
<td>0.066</td>
<td>0.097</td>
<td>0.892</td>
<td>0.024</td>
</tr>
<tr>
<td><strong>γ0 vs p1</strong></td>
<td>-0.112</td>
<td>0.235</td>
<td>0.101</td>
<td>0.328</td>
<td>0.137</td>
</tr>
</tbody>
</table>

\[ N = 1000 \]

<table>
<thead>
<tr>
<th></th>
<th>true</th>
<th>ED</th>
<th>HPJ</th>
<th>TOJ</th>
<th>HSJ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>2.960</td>
<td>0.000</td>
<td>0.009</td>
<td>0.945</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Var</strong></td>
<td>0.073</td>
<td>0.007</td>
<td>0.003</td>
<td>0.435</td>
<td>0.001</td>
</tr>
<tr>
<td><strong>25%Q</strong></td>
<td>2.772</td>
<td>-0.008</td>
<td>0.013</td>
<td>0.874</td>
<td>-0.001</td>
</tr>
<tr>
<td><strong>50%Q</strong></td>
<td>2.960</td>
<td>0.000</td>
<td>0.011</td>
<td>0.926</td>
<td>0.001</td>
</tr>
<tr>
<td><strong>75%Q</strong></td>
<td>3.148</td>
<td>0.000</td>
<td>0.013</td>
<td>0.858</td>
<td>0.002</td>
</tr>
<tr>
<td><strong>γ0 Mean</strong></td>
<td>0.082</td>
<td>-0.007</td>
<td>0.001</td>
<td>0.000</td>
<td>-0.001</td>
</tr>
<tr>
<td><strong>γ0 Var</strong></td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>γ0 25%Q</strong></td>
<td>0.061</td>
<td>-0.012</td>
<td>0.001</td>
<td>0.000</td>
<td>-0.004</td>
</tr>
<tr>
<td><strong>γ0 50%Q</strong></td>
<td>0.082</td>
<td>-0.012</td>
<td>0.001</td>
<td>0.000</td>
<td>-0.003</td>
</tr>
<tr>
<td><strong>γ0 75%Q</strong></td>
<td>0.103</td>
<td>-0.007</td>
<td>0.002</td>
<td>0.059</td>
<td>-0.000</td>
</tr>
<tr>
<td><strong>p1 Mean</strong></td>
<td>0.570</td>
<td>-0.060</td>
<td>0.006</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td><strong>p1 Var</strong></td>
<td>0.021</td>
<td>0.013</td>
<td>0.001</td>
<td>0.000</td>
<td>-0.003</td>
</tr>
<tr>
<td><strong>p1 25%Q</strong></td>
<td>0.469</td>
<td>-0.081</td>
<td>0.009</td>
<td>0.000</td>
<td>0.006</td>
</tr>
<tr>
<td><strong>p1 50%Q</strong></td>
<td>0.570</td>
<td>-0.049</td>
<td>0.007</td>
<td>0.000</td>
<td>0.006</td>
</tr>
<tr>
<td><strong>p1 75%Q</strong></td>
<td>0.671</td>
<td>-0.027</td>
<td>0.007</td>
<td>0.069</td>
<td>0.006</td>
</tr>
<tr>
<td><strong>μ vs γ0</strong></td>
<td>0.239</td>
<td>-0.049</td>
<td>0.031</td>
<td>0.650</td>
<td>-0.027</td>
</tr>
<tr>
<td><strong>μ vs p1</strong></td>
<td>-0.217</td>
<td>0.066</td>
<td>0.031</td>
<td>0.443</td>
<td>0.023</td>
</tr>
<tr>
<td><strong>γ0 vs p1</strong></td>
<td>-0.112</td>
<td>0.237</td>
<td>0.032</td>
<td>0.000</td>
<td>0.139</td>
</tr>
</tbody>
</table>

### 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. We also denote a generic constant as \( M \) in the proofs.
A.1.1 Proof of Theorem 1

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

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 6 (for the case of $\mathbb{P}_N^\mu$, Corollary 2 and for the case of $\mathbb{P}_N^{\rho}$, Lemma 7) implies that $\hat{\gamma}_{k,i}$ converges to $\gamma_{k,i}$ in mean square convergence and thus also implies that $\gamma_{k,i}$ 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 5 (for the case of $\mathbb{P}_N^\mu$, Assumption 4 and for the case of $\mathbb{P}_N^{\rho}$, Assumption 6).

We then show that the first term converges to 0 almost surely. Note that, for $f = 1_{(-\infty, a]}$, $\mathbb{P}_N f = \mathbb{F}_N^{\hat{\gamma}_{k,i}}(a)$ 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, $1(\hat{\gamma}_{k,i} \leq a)$ 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 $\mathbb{F}_N^{\hat{\gamma}_{k,i}}(a) - \Pr(\hat{\gamma}_{k,i} \leq a) \xrightarrow{a.s.} 0$ and $\mathbb{F}_N^{\hat{\gamma}_{k,i}}(a-) - \Pr(\hat{\gamma}_{k,i} < a) \xrightarrow{a.s.} 0$ for every $a \in \mathbb{R}$, when $T(N) \to \infty$ as $N \to \infty$. Given a fixed $\varepsilon > 0$, there exists a partition $-\infty \equiv a_0 < a_1 < \cdots < a_k = \infty$ such that $\Pr(\gamma_{k,i} < a) - \Pr(\gamma_{k,i} \leq a_{i-1}) \leq \varepsilon / 3$ 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 $N, T$, $\sup_{f \in \mathcal{F}} |P_T f - P_0 f| \leq \varepsilon / 3$. Therefore, we have $\Pr(\gamma_{k,i} < a) - \Pr(\gamma_{k,i} \leq a_{i-1}) < \varepsilon$ for every $i$. The rest of the proof is the same as the proof of Theorem 19.1 in van der Vaart (1998). For $a_{i-1} \leq a < a_i$,
\[
\begin{align*}
\mathbb{F}_N^{\hat{\gamma}_{k,i}}(a) - \Pr(\hat{\gamma}_{k,i} \leq a) &\leq \mathbb{F}_N^{\hat{\gamma}_{k,i}}(a_{i-}) - \Pr(\hat{\gamma}_{k,i} < a_i) + \varepsilon, \\
\mathbb{F}_N^{\hat{\gamma}_{k,i}}(a) - \Pr(\hat{\gamma}_{k,i} \leq a) &\geq \mathbb{F}_N^{\hat{\gamma}_{k,i}}(a_{i-}) - \Pr(\hat{\gamma}_{k,i} < a_{i-1}) - \varepsilon.
\end{align*}
\]

Accordingly, we have $\lim_{N,T \to \infty} \sup_{f \in \mathcal{F}} |\mathbb{P}_N f - P_T f| \leq \varepsilon$ almost surely. This is true for every $\varepsilon > 0$, and thus we get
\[
\sup_{f \in \mathcal{F}} |\mathbb{P}_N f - P_T f| \xrightarrow{a.s.} 0.
\]
We note that (11) holds for all monotonic diagonal paths \( N \rightarrow \infty, T(N) \rightarrow \infty \). As stated in REMARKS (a) in Phillips and Moon (1999), (11) thus holds under double asymptotics \( N, T \rightarrow \infty \). Consequently, we obtain the desired result by the continuous mapping theorem.

\[
\square
\]

### A.1.2 Proof of Theorem 2

The proofs for \( \mathbb{P}_N^{P_A}, \mathbb{P}_N^{P_B} \), and \( \mathbb{P}_N^{P_0} \) are basically the same, so we present that for \( \mathbb{P}_N^{P_0} \) only. Let \( \mathbb{P}_N = \mathbb{P}_N^{P_0}, \mathbb{P}_T = \mathbb{P}_T^{P_0}, \) and \( P_0 = P_0^{P_0} \). The proof is based on the decomposition in (4) and (5).

To study the asymptotic behavior of (4), we use Lemma 2.8.7 in VW. We first fix a monotone sequence \( T = T(N) \) such that \( T(N) \rightarrow \infty \) as \( N \rightarrow \infty \), 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 immediate for the envelope function \( F = 1 \) (constant).

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

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

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

\[
\sup_{f,g \in \mathcal{F}} |\rho_{P_T}(f,g) - \rho_{P_0}(f,g)| \\
\leq \sup_{a,a' \in \mathbb{R}} \left| (P_T1_{(-\infty,a]} - P_01_{(-\infty,a]} - (P_T1_{(-\infty,a]} - P_01_{(-\infty,a]})(P_T1_{(-\infty,a']} - P_01_{(-\infty,a']}) \right| \\
+ (P_T1_{(-\infty,a']} - P_01_{(-\infty,a']}) - ((P_T1_{(-\infty,a]})^2 - (P_01_{(-\infty,a]}))^2 \\
- ((P_T1_{(-\infty,a]})^2 - (P_01_{(-\infty,a]}))^2 + 2(P_T1_{(-\infty,a]}P_T1_{(-\infty,a']) - P_T1_{(-\infty,a]}P_01_{(-\infty,a']}) \\
+ 2(P_T1_{(-\infty,a]}P_01_{(-\infty,a']) - P_01_{(-\infty,a]}P_01_{(-\infty,a']})^{1/2}
\]

44
\[ \leq 11 \sup_{a \in \mathbb{R}} \left| P_T \mathbf{1}_{(-\infty,a]} - P_0 \mathbf{1}_{(-\infty,a]} \right|^{1/2} \]
\[ \rightarrow 0, \]
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 6 (for the case of \( \hat{\mu}_i \), Corollary 2 and for the case of \( \hat{\rho}_{k,i} \), Lemma 7). Therefore, condition (2.8.5) in VW is satisfied.

Therefore, by Lemma 2.8.7 in VW, we have shown that
\[ G_{N,P_T} \xrightarrow{\ell^\infty(F)} G_{P_0} \]
(12)
Note that (12) holds for all monotonic diagonal paths \( T(N) \rightarrow \infty \) as \( N \rightarrow \infty \). As in REMARKS (a) of Phillips and Moon (1999), (12) thus holds under double asymptotics \( N,T \rightarrow \infty \).

Next, we study the asymptotic behavior of (5): \( \sqrt{N}(P_T f - P_0 f) \). Because the non-stochastic function sequence \( P_T f - P_0 f \) is uniformly bounded in \( f \in F \), we should consider the convergence rate of \( \sup_{f \in F} |P_T f - P_0 f| \). Lemmas 6 and 8 (for the case of \( \mathbb{P}^\hat{\mu}_N \), Corollary 2 and Lemma 8 and for the case of \( \mathbb{P}^\hat{\rho}_N \), Lemmas 7 and 8) imply that
\[ \sup_{f \in F} \left| P_T f - P_0 f \right| = O \left( \frac{1}{T^{r/(2+r)}} \right) . \]

Therefore, given \( N^{2+r}/T^{2r} \rightarrow 0 \), the desired result holds by Slutsky’s theorem.

\[ \square \]

A.1.3 Proof of Theorem 3

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 continuous mapping theorem. Let \( \hat{G} := N^{-1} \sum_{i=1}^N g(\hat{\theta}_i) \), \( G := E(g(\theta_i)) \), \( \hat{\theta}_i := (\hat{\theta}_{1,i}, \hat{\theta}_{2,i}, \ldots, \hat{\theta}_{l,i})^\top \), and \( \theta_i := (\theta_{1,i}, \theta_{2,i}, \ldots, \theta_{l,i})^\top \). If we show \( \sqrt{N}(\hat{G} - G) \sim N(0, \Gamma) \), the asymptotic normality of \( \sqrt{N}(\hat{H} - H) \) follows by the delta method. We thus focus on the proof for \( \sqrt{N}(\hat{G} - G) \sim N(0, \Gamma) \).

By Taylor’s theorem, we have
\[ \sqrt{N}(\hat{G} - G) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \left( g(\hat{\theta}_i) - G \right) = \frac{1}{\sqrt{N}} \sum_{i=1}^N (g(\theta_i) - G) \]
\[ + \frac{1}{\sqrt{N}} \sum_{i=1}^N \nabla g(\theta_i)(\hat{\theta}_i - \theta_i) \]
(14)
\[ + \frac{1}{2\sqrt{N}} \sum_{i=1}^N Q_i, \]
(15)
where \( \nabla g(\theta_i) := (\nabla g_1(\theta_i)^\top, \ldots, \nabla g_m(\theta_i)^\top)^\top \) is the first derivative (a \( m \times l \) matrix) of \( g(\cdot) \) evaluated at \( \theta_i \), \( Q_i \) is the \( m \times 1 \) matrix whose \( p \)th element is \( (\hat{\theta}_i - \theta_i)^\top \mathcal{H}(g_p)(\hat{\theta}_i - \theta_i) \) for \( p = 1, 2, \ldots, m \) where \( \mathcal{H}(g_p)(\cdot) \) is the Hessian matrix of \( g_p(\cdot) \), and \( \hat{\theta}_i^{(p)} \) is between \( \theta_i \) and \( \hat{\theta}_i \) (see Feng, Wang, Chen, and Tu, 2014). We examine each term below.

For (13), under Assumptions 1 and 9, it holds that

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (g(\theta_i) - G) \sim N(0, \Gamma),
\]

by the central limit theorem for i.i.d. random vectors.

For (14), we show that it is of order \( O_p(\sqrt{N/T}) \). By simple algebra, we have

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \nabla g(\theta_i)(\hat{\theta}_i - \theta_i) = \sum_{j=1}^{l} \left( \begin{array}{c} N^{-1/2} \sum_{i=1}^{N} a_{1,j,i}(\hat{\theta}_j,i - \theta_j,i) \\ N^{-1/2} \sum_{i=1}^{N} a_{2,j,i}(\hat{\theta}_j,i - \theta_j,i) \\ \vdots \\ N^{-1/2} \sum_{i=1}^{N} a_{m,j,i}(\hat{\theta}_j,i - \theta_j,i) \end{array} \right),
\]

where \( a_{p,j,i} := (\partial/\partial z_j)g_p(z)|_{z=\theta_i} \) for \( p = 1, 2, \ldots, m \). Since \( l \) and \( m \) are fixed, it is sufficient to show that for \( p = 1, 2, \ldots, m \)

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} a_{p,j,i}(\hat{\theta}_j,i - \theta_j,i) = O_p\left( \frac{\sqrt{N}}{T} \right),
\]

thanks to Chebyshev’s inequality for random vectors and Slutsky’s theorem. To this end, we consider the following two cases: (i) \( \theta_j,i = \mu_i \) and (ii) \( \theta_j,i = \gamma_{k,i} \) for some \( k \).

If \( \theta_j,i = \mu_i \) in the left-hand side of (16), we have

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} a_{p,j,i}(\hat{\theta}_j,i - \theta_j,i) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} a_{p,j,i} \bar{w}_i.
\]

This term has zero mean and the variance is

\[
var\left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} a_{p,j,i} \bar{w}_i \right) = E\left( a_{p,j,i}^2 \right) \leq \sqrt{E(a_{p,j,i}^4)} \sqrt{E((\bar{w}_i)^4)} = O(T^{-1}),
\]

where the first equality follows from Assumption 1, the inequality is the Cauchy–Schwarz inequality, and the last equality follows from Assumption 9 and Lemma 3 for \( r_m = 4 \) and \( r_d = 4 \). Thus, we have (16) in this case.
If $\theta_{j,i} = \gamma_{k,i}$ for some $k$ in the left-hand side of (16), we have the following expansion.

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} a_{p,j,i} (\hat{\gamma}_{k,i} - \gamma_{k,i})
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} a_{p,j,i} \left( \frac{1}{T-k} \sum_{t=k+1}^{T} w_{it}w_{i,t-k} - \gamma_{k,i} \right) - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} a_{p,j,i} (\bar{w}_{i})^2
+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{T-k} a_{p,j,i} \sum_{t=1}^{k} w_{it} \bar{w}_{i} + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{T-k} a_{p,j,i} \sum_{t=T-k+1}^{T} w_{it} \bar{w}_{i}.
\]

(17)

The term in (17) has zero mean and the variance is

\[
\text{var} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} a_{p,j,i} \left( \frac{1}{T-k} \sum_{t=k+1}^{T} w_{it}w_{i,t-k} - \gamma_{k,i} \right) \right)
= E \left( a_{p,j,i}^2 \left( \frac{1}{T-k} \sum_{t=k+1}^{T} w_{it}w_{i,t-k} - \gamma_{k,i} \right)^2 \right)
\leq \sqrt{E(a_{p,j,i}^4)} \sqrt{E \left( \left( \frac{1}{T-k} \sum_{t=k+1}^{T} w_{it}w_{i,t-k} - \gamma_{k,i} \right)^4 \right)} = O \left( \frac{1}{T} \right),
\]

where the first equality follows from Assumption 1, the inequality is the Cauchy–Schwarz inequality, and the last equality follows from Assumption 9 and Lemma 5 for $r_m = 4$ and $r_d = 8$. Thus, the term in (17) is of order $O_p(1/\sqrt{T})$ by Markov’s inequality. For the term in (18), the absolute mean is

\[
E \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} a_{p,j,i} (\bar{w}_{i})^2 \right| \leq \frac{T+k}{T-k} \sqrt{N} E|a_{p,j,i}(\bar{w}_{i})^2|
\leq \frac{T+k}{T-k} \sqrt{N} \sqrt{E(a_{p,j,i}^2)} \sqrt{E((\bar{w}_{i})^4)} = O \left( \frac{\sqrt{N}}{T} \right),
\]

where the first and second inequalities are the triangle inequality and the Cauchy–Schwarz inequality, respectively, and the last result follows from Assumption 9 and Lemma 3 for $r_m = 4$ and $r_d = 4$. Thus, the term in (18) is of order $O_p(\sqrt{N}/T)$ by Markov’s inequality. In the same manner, we can show that the terms in (19) are of order $o_p(\sqrt{N}/T)$. Therefore, in this case, we have (16) by Slutsky’s theorem.

Accordingly, we have shown (16) for any $\theta_{j,i}$, which means that the term in (14) is of order $O_p(\sqrt{N}/T)$.
For (15), we show that it is of order $O_p(\sqrt{N/T})$. We observe that the (absolute) mean of the Euclidean norm is

$$E\left\|\frac{1}{2\sqrt{N}} \sum_{i=1}^{N} Q_i\right\| \leq \frac{\sqrt{N}}{2} E\|Q_i\|$$

$$\leq \frac{\sqrt{N}}{2} \sum_{p=1}^{m} E|\hat{\theta}_i - \theta_i| \mathcal{H}(g_p)(\hat{\theta}_i)(\hat{\theta}_i - \theta_i)|$$

$$\leq \sqrt{N} M \sum_{p=1}^{m} \sum_{j_1=1}^{l} \sum_{j_2=1}^{l} E|(\hat{\theta}_{j_1,i} - \theta_{j_1,i})(\hat{\theta}_{j_2,i} - \theta_{j_2,i})|$$

$$= O(\sqrt{N}/T),$$

where the first inequality follows from the triangle inequality, the second follows from the triangle inequality, the third follows from the triangle inequality and Assumption 9, and the last follows from the Cauchy–Schwarz inequality, Corollary 3, Lemma 6, and a fixed $m$ and $l$. Therefore, (15) is $O_p(\sqrt{N}/T)$ by Markov’s inequality.

Consequently, we obtain the desired result by Slutsky’s theorem.

□

A.1.4 Proof of Theorem 4

We first examine the estimation of $G := E(g(\theta_i))$. We focus on cases in which $m = 1$ so that $g(\cdot)$ is a univariate function. Note that it is easy to extend our result to cases with $m > 1$.

The Taylor expansion gives

$$\sqrt{N}(\hat{G} - G)$$

$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (g(\theta_i) - G)$$

$$+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{j=1}^{l} \frac{\partial}{\partial z_j} g(z) \bigg|_{z=\theta_i} (\hat{\theta}_{j,i} - \theta_{j,i})$$

$$+ \frac{1}{2\sqrt{N}} \sum_{i=1}^{N} \sum_{j_1=1}^{l} \sum_{j_2=1}^{l} \frac{\partial^2}{\partial z_{j_1} \partial z_{j_2}} g(z) \bigg|_{z=\theta_i} (\hat{\theta}_{j_1,i} - \theta_{j_1,i})(\hat{\theta}_{j_2,i} - \theta_{j_2,i})$$

$$+ \frac{1}{6\sqrt{N}} \sum_{i=1}^{N} \sum_{j_1=1}^{l} \sum_{j_2=1}^{l} \sum_{j_3=1}^{l} \frac{\partial^3}{\partial z_{j_1} \partial z_{j_2} \partial z_{j_3}} g(z) \bigg|_{z=\theta_i} (\hat{\theta}_{j_1,i} - \theta_{j_1,i})(\hat{\theta}_{j_2,i} - \theta_{j_2,i})(\hat{\theta}_{j_3,i} - \theta_{j_3,i}),$$

where $\hat{\theta}_i$ is between $\theta_i$ and $\hat{\theta}_i$ (see Feng et al., 2014).
The proof follows the arguments made by Dhaene and Jochmans (2015). We shall show that
\[
\sqrt{N}(\hat{G} - G) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (g(\theta_i) - G) + \frac{\sqrt{N}}{T} B + o_p \left( \frac{\sqrt{N}}{T} \right),
\]
for some $B$. This result also implies that
\[
\sqrt{N}(\bar{G} - G) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (g(\theta_i) - G) + 2 \frac{\sqrt{N}}{T} B + o_p \left( \frac{\sqrt{N}}{T} \right),
\]
where $\bar{G} := (\hat{G}(1) + \hat{G}(2))/2$. Consequently, when $N, T \to \infty$ and $N/T^2 \to \nu$, we have
\[
\sqrt{N}(\hat{G}^H - G) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (g(\theta_i) - E(g(\theta_i))) + o_p \left( \frac{1}{\sqrt{T}} \right) \xrightarrow{N \to \infty} N(0, \Gamma),
\]
where $\hat{G}^H := 2\hat{G} - \bar{G}$.

We first consider the term in (20). Let $a_{j,i} := \frac{\partial}{\partial z} g(z) \mid_{z=\theta_i}$. Suppose that $\theta_{j,i} = \mu_i$. In this case, we have
\[
1 \sqrt{N} \sum_{i=1}^{N} a_{j,i}(\hat{\mu}_i - \mu_i) = 1 \sqrt{N} \sum_{i=1}^{N} a_{j,i} \bar{w}_i = O_p \left( \frac{1}{\sqrt{T}} \right),
\]
by the same argument in the proof of Theorem 3.

Suppose next that $\theta_{j,i} = \gamma_{k,i}$ for some $k$. In this case, we have
\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} a_{j,i}(\hat{\gamma}_{k,i} - \gamma_{k,i})
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} a_{j,i} \left( \frac{1}{T-k} \sum_{t=k+1}^{T} w_{it}w_{i,t-k} - \gamma_{k,i} \right) \tag{23}
\]
\[
- \frac{1}{\sqrt{N}} \frac{T+k}{T-k} \sum_{i=1}^{N} a_{j,i} \bar{w}_i^2 \tag{24}
\]
\[
+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{T-k} a_{j,i} \sum_{t=1}^{k} w_{it} \bar{w}_i + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{T-k} a_{j,i} \sum_{t=T-k+1}^{T} w_{it} \bar{w}_i. \tag{25}
\]

The term (23) is of order $O_p(T^{1/2})$ by the same argument in the proof of Theorem 3. For the term in (24), we observe that
\[
E \left( \frac{1}{\sqrt{N}} \frac{T+k}{T-k} \sum_{i=1}^{N} a_{j,i} \bar{w}_i^2 \right) = \frac{T+k}{T-k} E(a_{j,i} V_{T,i}) = \frac{\sqrt{N}}{T} E(a_{j,i} V_i) + o \left( \frac{\sqrt{N}}{T} \right),
\]
for
where \( V_{T,i} := TE((\bar{w}_i)^2|i) = \sum_{j=-T}^T \gamma_{j,i}(T-|j|)/T \) and \( V_i := \sum_{j=-\infty}^{\infty} \gamma_{j,i} \). The variance is

\[
\text{var} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^N a_{j,i}(\bar{w}_i)^2 \right) = \text{var}(a_{j,i}(\bar{w}_i)^2) \leq E(a_{j,i}^4(\bar{w}_i)^4) \leq \sqrt{E(a_{j,i}^4)} \sqrt{E((\bar{w}_i)^8)} = O \left( \frac{1}{T^2} \right),
\]

where the second inequality is the Cauchy–Schwarz inequality and the third inequality follows from Assumptions 2, 3, and 10 and Lemma 3. Therefore, we have

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^N a_{j,i} \frac{T+k}{T-k}(\bar{w}_i)^2 = \frac{\sqrt{N}}{T} E(a_{j,i} V_i) + o_p \left( \frac{\sqrt{N}}{T} \right).
\]

Given this result, it is easy to see that the terms in (25) are of order \( o_p(\sqrt{N}/T) \).

We now have that the term in (20) can be written as

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^N \sum_{j=1}^l \frac{\partial}{\partial z_j} g(z) \big|_{z=\theta} (\hat{\theta}_{j,i} - \theta_{j,i}) = \frac{\sqrt{N}}{T} \sum_{j=1}^l 1_{\{\theta_{j,i} \neq \mu_i\}} E(a_{j,i} V_i) + o_p \left( \frac{\sqrt{N}}{T} \right).
\]

Next, we consider the term in (21). Let \( b_{j_1,j_2,i} := \frac{\partial^2}{\partial z_j \partial z_2} g(z) \big|_{z=\theta} \). We consider four patterns for \( \theta_{j_1,i} \) and \( \theta_{j_2,i} \): \( \theta_{j_1,i} = \theta_{j_2,i} = \mu_i \), \( \theta_{j_1,i} = \theta_{j_2,i} = \gamma_{k,i} \) for some \( k \), \( \theta_{j_1,i} = \mu_i \) and \( \theta_{j_2,i} = \gamma_{k,i} \) for some \( k \), and \( \theta_{j_1,i} = \gamma_{k_1,i} \) and \( \theta_{j_2,i} = \gamma_{k_2,i} \) for some \( k_1, k_2 \). We first consider the case in which \( j_1 = j_2 = j \) and \( \theta_{j,i} = \mu_i \). In this case, we have

\[
\frac{1}{2\sqrt{N}} \sum_{i=1}^N b_{j,j,i}(\hat{\mu}_i - \mu_i)^2 = \frac{1}{2\sqrt{N}} \sum_{i=1}^N b_{j,j,i}(\bar{w}_i)^2.
\]

The argument used to analyze the term in (24) gives

\[
\frac{1}{2\sqrt{N}} \sum_{i=1}^N b_{j,j,i}(\hat{\mu}_i - \mu_i)^2 = \sqrt{\frac{N}{2T}} E(b_{j,j,i} V_i) + o_p \left( \frac{\sqrt{N}}{T} \right).
\]

Second, we consider the case in which \( j_1 = j_2 = j \) and \( \theta_{j,i} = \gamma_{k,i} \) for some \( k \). It is easy to see that the mean is

\[
E \left( \frac{1}{2\sqrt{N}} \sum_{i=1}^N b_{j,j,i}(\hat{\gamma}_{k,i} - \gamma_{k,i})^2 \right) = \frac{\sqrt{N}}{2} E \left( b_{j,j,i} \left( \frac{1}{T-k} \sum_{t=k+1}^T w_{it}w_{i,t-k} - \gamma_{k,i} \right)^2 \right) + o \left( \frac{\sqrt{N}}{T} \right).
\]

The i.i.d. assumption and the Cauchy–Schwarz inequality imply that

\[
\frac{\sqrt{N}}{2} E \left( b_{j,j,i} \left( \frac{1}{T-k} \sum_{t=k+1}^T w_{it}w_{i,t-k} - \gamma_{k,i} \right)^2 \right) \leq \frac{\sqrt{N}}{2(T-k)^2} \sqrt{E \left( \left( \sum_{t=k+1}^T (w_{it}w_{i,t-k} - \gamma_{k,i}) \right)^4 \right)} \sqrt{E(b_{j,j,i}^2)} = O \left( \frac{\sqrt{N}}{T} \right),
\]

\[50\]
by Lemma 5. Moreover, \(TE\left(b_{j,j,i} \left( \sum_{t=k+1}^{T} w_{i,t} w_{i,t-k} - \gamma_{k,i} \right)^2 \right)/2 \) converges by the dominated convergence theorem. We denote this limit by \(B_k\). Next, we consider the variance term. We observe that

\[
\text{var} \left( \frac{1}{2\sqrt{N}} \sum_{i=1}^{N} b_{j,j,i}(\hat{\gamma}_{k,i} - \gamma_{k,i})^2 \right) = \frac{1}{4} \text{var} \left( b_{j,j,i}(\hat{\gamma}_{k,i} - \gamma_{k,i})^2 \right) \leq \frac{1}{4} E \left( b_{j,j,i}^2 \right) \leq \frac{1}{4} \sqrt{E \left( \hat{\gamma}_{k,i} - \gamma_{k,i} \right)^8} = O \left( \frac{1}{T^2} \right),
\]

by Lemma 6. Thus, we have

\[
\frac{1}{2\sqrt{N}} \sum_{i=1}^{N} b_{j,j,i}(\hat{\gamma}_{k,i} - \gamma_{k,i})^2 = B_k \frac{\sqrt{N}}{T} + o_p \left( \frac{\sqrt{N}}{T} \right).
\]

Third, we consider the case in which \(\theta_{j_1,i} = \mu_i\) and \(\theta_{j_2,i} = \gamma_{k,i}\) for some \(k\). The mean of this term is

\[
E \left( \frac{1}{2\sqrt{N}} \sum_{i=1}^{N} b_{j_1,j_2,i}(\hat{\mu}_{i} - \mu_i)(\hat{\gamma}_{k,i} - \gamma_{k,i}) \right) = \frac{\sqrt{N}}{2} E \left( b_{j_1,j_2,i} \bar{w}_i \left( \frac{1}{T-k} \sum_{t=k+1}^{T} w_{i,t} w_{i,t-k} \right) \right) + o \left( \frac{\sqrt{N}}{T} \right).
\]

As in the second case, \(TE\left( b_{j_1,j_2,i} \bar{w}_i \left( \sum_{t=k+1}^{T} w_{i,t} w_{i,t-k} \right) \right)\) converges by the dominated convergence theorem and we denote the limit by \(B_{\mu,k}\). The variance is

\[
\text{var} \left( \frac{1}{2\sqrt{N}} \sum_{i=1}^{N} b_{j_1,j_2,i}(\hat{\mu}_{i} - \mu_i)(\hat{\gamma}_{k,i} - \gamma_{k,i}) \right) = \frac{1}{4} \text{var} \left( b_{j_1,j_2,i} \bar{w}_i (\hat{\gamma}_{k,i} - \gamma_{k,i}) \right) \leq \frac{1}{4} E \left( b_{j_1,j_2,i}^2 \bar{w}_i^2 (\hat{\gamma}_{k,i} - \gamma_{k,i})^2 \right) \leq \frac{1}{4} \left( E(b_{j_1,j_2,i}^4) \right)^{1/2} \left( E(\bar{w}_i^8) \right)^{1/4} \left( E \left( (\hat{\gamma}_{k,i} - \gamma_{k,i})^8 \right) \right)^{1/4} = O \left( \frac{1}{T^2} \right),
\]

by Lemmas 3 and 6 and Assumption 10. Thus we have

\[
\frac{1}{2\sqrt{N}} \sum_{i=1}^{N} b_{j_1,j_2,i}(\hat{\mu}_{i} - \mu_i)(\hat{\gamma}_{k,i} - \gamma_{k,i}) = B_{\mu,k} \frac{\sqrt{N}}{T} + o_p \left( \frac{\sqrt{N}}{T} \right).
\]
A similar argument shows that when \( \theta_{j_1,i} = \gamma_{k_1,i} \) and \( \theta_{j_2,i} = \gamma_{k_2,i} \) for some \( k_1 \) and \( k_2 \), we have

\[
\frac{1}{2\sqrt{N}} \sum_{i=1}^{N} b_{j_1,j_2,i}(\hat{\gamma}_{k_1,i} - \gamma_{k_1,i})(\hat{\gamma}_{k_2,i} - \gamma_{k_2,i}) = B_{k_1,k_2} \frac{\sqrt{N}}{T} + o_p \left( \frac{\sqrt{N}}{T} \right).
\]

for some \( B_{k_1,k_2} \).

It therefore holds that

\[
\frac{1}{2\sqrt{N}} \sum_{i=1}^{N} \sum_{j_1=1}^{l} \sum_{j_2=1}^{l} \frac{\partial^2 g(z)}{\partial z_{j_1} \partial z_{j_2}} \bigg|_{z=\theta_i} \left( \hat{\theta}_{j_1,i} - \theta_{j_1,i} \right) \left( \hat{\theta}_{j_2,i} - \theta_{j_2,i} \right) = B \frac{\sqrt{N}}{T} + o_p \left( \frac{\sqrt{N}}{T} \right),
\]

for some \( B \).

Lastly, we consider the term in (22). Let \( c_{j_1,j_2,j_3,i} := \frac{\partial^3}{\partial z_{j_1} \partial z_{j_2} \partial z_{j_3}} g(z) \big|_{z=\theta_i} \). For any \( j_1, j_2, j_3 = 1, 2, \ldots, l \), we have

\[
E \left| \frac{1}{6\sqrt{N}} \sum_{i=1}^{N} c_{j_1,j_2,j_3,i}(\hat{\theta}_{j_1,i} - \theta_{j_1,i})(\hat{\theta}_{j_2,i} - \theta_{j_2,i})(\hat{\theta}_{j_3,i} - \theta_{j_3,i}) \right| \leq \frac{\sqrt{N}}{6} E \left| c_{j_1,j_2,j_3,i}(\hat{\theta}_{j_1,i} - \theta_{j_1,i})(\hat{\theta}_{j_2,i} - \theta_{j_2,i})(\hat{\theta}_{j_3,i} - \theta_{j_3,i}) \right| \leq M \frac{\sqrt{N}}{6} \left( E((\hat{\theta}_{j_1,i} - \theta_{j_1,i})^4) \right)^{1/4} \left( E((\hat{\theta}_{j_2,i} - \theta_{j_2,i})^4) \right)^{1/4} \left( E((\hat{\theta}_{j_3,i} - \theta_{j_3,i})^2) \right)^{1/2} = O \left( \frac{\sqrt{N}}{T^{3/2}} \right),
\]

where the second inequality follows from Assumption 10 and the repeated application of the Cauchy–Schwarz inequality and the last equality follows from Assumptions 2 and 3 and Corollary 2 and/or Lemma 6. Thus, the term in (22) is of order \( O_p(\sqrt{N}/T^{3/2}) \).

Summing up, we have shown the asymptotic normality and unbiasedness for the HPJ estimation of \( G \). When \( m > 1 \), we analyze each element of \( G \) and obtain the asymptotic normality and unbiasedness.

Next, we consider the estimation of \( H := h(G) \). Taylor’s theorem gives

\[
\sqrt{N} \left( \hat{H} - H \right) = \sqrt{N} \nabla h(G)(\hat{G} - G) + \sqrt{N}(\hat{G} - G)^\top H(G)(\hat{G} - G),
\]

where \( H \) is the Hessian matrix of \( h \) and \( \hat{G} \) is between \( \hat{G} \) and \( G \). We note that

\[
\sqrt{N} \nabla h(G)(\hat{G} - G) = \frac{1}{N} \sum_{i=1}^{N} \nabla h(G)(g(\theta_i) - G) + \frac{\sqrt{N}}{T} \nabla h(G) B + O_p \left( \frac{\sqrt{N}}{T^{3/2}} \right) + o_p \left( \frac{\sqrt{N}}{T} \right).
\]

Given

\[
\hat{G} - G = O_p \left( \frac{1}{\sqrt{N}} + \frac{1}{T} \right),
\]

52
we have
\[
\sqrt{N}(\hat{G} - G)\top H(\tilde{G})(\hat{G} - G) = O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{T} + \frac{\sqrt{N}}{T^2}\right) = o_p\left(\frac{\sqrt{N}}{T}\right).
\]
Thus, the condition for the HPJ estimation of $H$ also holds and the proof is complete.

\[\square\]

**A.1.5 Proof of Theorem 5**

We first consider $\hat{G}^* := N^{-1} \sum_{i=1}^{N} z_{Ni}g(\hat{\theta}_i)$. We initially fix a diagonal path $N \to \infty$, $T(N) \to \infty$ and then prove the statement. The statement turns out to hold for any diagonal path, which implies that the result holds under double asymptotics. Let $E_z$ and $\text{var}_z$ denote the expectation and variance operators under the probability measure $P_z$ while holding $\{\{y_{jt}\}_{j=1}^{N}\}_{t=1}^{T}$ fixed, respectively.

For each diagonal path, we first show that the moments of $\hat{G}^*$ under the bootstrap distribution satisfy Lyapunov’s conditions. We then prove that $\sqrt{N}(\hat{G}^* - \hat{G})$ converges in distribution to $N(0, \Gamma)$ almost surely under a subsequence of any subsequences of the original sequence. This implies that the bootstrap distribution of $\sqrt{N}(\hat{G}^* - \hat{G})$ 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 $\hat{G}^*$. It is thus easy to see, conditional on the data, that the mean and variance of $z_{Ni}g(\hat{\theta}_i)$ are
\[
E_z\left(z_{Ni}g(\hat{\theta}_i)\right) = \frac{1}{N} \sum_{i=1}^{N} g(\hat{\theta}_i) = \hat{G},
\]
\[
\text{var}_z\left(z_{Ni}g(\hat{\theta}_i)\right) = \frac{1}{N} \sum_{i=1}^{N} g(\hat{\theta}_i)g(\hat{\theta}_i)\top - \hat{G}\hat{G}\top,
\]
by the fact that $z_{Ni}$ follows the binomial distribution with parameters $(N, N^{-1})$. The conditional variance converges to
\[
E(g(\theta_i)g(\theta_i)\top) - GG\top = \Gamma
\]
in probability by Theorem 3 under Assumptions 1, 2, 3, and 11 and the continuous mapping theorem. We then consider the third-order moment. We note that
\[
E_z\left(\frac{1}{\sqrt{N}} \|z_{Ni}g(\hat{\theta}_i) - \hat{G}\|^3\right) = \frac{1}{N^{3/2}} \sum_{i=1}^{N} \|g(\hat{\theta}_i) - \hat{G}\|^3.
\]
We have
\[ \| g(\hat{\theta}_i) - \hat{G} \|^3 \leq 4\| g(\hat{\theta}_i) \|^3 + 4\| \hat{G} \|^3. \]

As \( \hat{G} \) converges by Theorem 3, the continuous mapping theorem leads to
\[ \frac{1}{N^{1/2}} \| \hat{G} \|^3 \xrightarrow{p} 0. \]

By Hölder’s inequality, it follows that
\[ \frac{1}{N^{3/2}} \sum_{i=1}^{N} \| g(\hat{\theta}_i) \|^3 \leq \frac{1}{\sqrt{N}} \left( \frac{1}{N} \sum_{i=1}^{N} \| g(\hat{\theta}_i) \|^4 \right)^{\frac{3}{4}} = \frac{1}{\sqrt{N}} \left( \frac{1}{N} \sum_{i=1}^{N} (g(\hat{\theta}_i)\top g(\hat{\theta}_i))^2 \right)^{\frac{3}{4}} \xrightarrow{p} 0, \]
as \( N^{-1} \sum_{i=1}^{N} (g(\hat{\theta}_i)\top g(\hat{\theta}_i))^2 = O_p(1) \) by Theorem 3 under Assumptions 1, 2, 3, and 11. Thus,
\[ \frac{1}{N^{3/2}} \sum_{i=1}^{N} \| g(\hat{\theta}_i) - \hat{G} \|^3 \xrightarrow{p} 0. \]

We argue that for any subsequence of the original sequence, there exists a further subsequence under which \( \sqrt{N}(\hat{G}^* - \hat{G}) \) converges in distribution conditionally on \( \{y_{it}\}_{t=1}^{N} \) almost surely. We have shown that the first, second, and third moments of \( z_N g(\hat{\theta}_i) \) 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, \( \sqrt{N}(\hat{G}^* - \hat{G}) \) converges in distribution to \( X \sim N(0, \Gamma) \) conditionally almost surely. This implies that for any subsequence, there exists a further subsequence under which \( \sup_{x \in \mathbb{R}} \left| P_z \left( \sqrt{N}(\hat{G}^* - \hat{G}) \leq x \right) - \Pr (X \leq x) \right| \) converges to 0 almost surely. It thus holds that, for the original sequence,
\[ \sup_{x \in \mathbb{R}} \left| P_z \left( \sqrt{N}(\hat{G}^* - \hat{G}) \leq x \right) - \Pr (X \leq x) \right| \xrightarrow{p} 0. \]

We note that this argument holds for all monotonic diagonal paths \( N \to \infty, T(N) \to \infty \). Hence, as stated in REMARKS (a) in Phillips and Moon (1999), it also holds under double asymptotics \( N, T \to \infty \).

Next, we consider \( \hat{H}^* \). By Taylor’s theorem, we have
\[ \sqrt{N}(\hat{H}^* - \hat{H}) = (\nabla h_1(\hat{G}^{(1)*}\top, \nabla h_2(\hat{G}^{(2)*}\top), \ldots, \nabla h_n(\hat{G}^{(n)*}\top)\top)\top \sqrt{N}(\hat{G}^* - \hat{G}), \]
where \( \hat{G}^{(p)*} \) is between \( \hat{G}^* \) and \( \hat{G} \) for \( p = 1, 2, \ldots, n \). It is easy to see that \( \hat{G}^{(p)*} = G + o_p(1) \) conditionally and this implies that \( \hat{G}^{(p)*} = G + o_p(1) \) unconditionally as well (Cheng and Huang, 2010, Lemma 3). It thus holds that \( \sqrt{N}(\hat{H}^* - \hat{H}) = \nabla h(G)\sqrt{N}(\hat{G}^* - \hat{G}) + o_p(1) \).

The rest of the proof follows the same argument for \( \hat{G}^* \) and we obtain the desired result.
A.1.6 Proof of Theorem 7

The proof is almost identical to the proof of Corollary 19.21 in van der Vaart (1998). We first note that, under $H_0$, $\sqrt{N} (P_N - Q) \Rightarrow G_Q$ in $\ell^\infty(F)$ given $N, T \to \infty$ with $N^{2+r}/T^{2r} \to 0$ by Theorem 2. Therefore, because the norm $\| \cdot \|_\infty$ on $D[\mathbb{R}]$ is continuous with respect to the uniform norm, we have $KS_1 \Rightarrow \|G_Q\|_\infty$ under $H_0$ by the continuous mapping theorem.

□

A.1.7 Proof of Theorem 8

We first observe that

$$KS_2 = \left\| \sqrt{\frac{N_1 N_2}{N_1 + N_2}} (P_{N_1}(1) - P_0(1)) - \sqrt{\frac{N_1 N_2}{N_1 + N_2}} (P_{N_2}(2) - P_0(2)) + \sqrt{\frac{N_1 N_2}{N_1 + N_2}} (P_0(1) - P_0(2)) \right\|_\infty.$$

We note that, under Assumption 12, $\sqrt{N_1} (P_{N_1}(1) - P_0(1))$ and $\sqrt{N_2} (P_{N_2}(2) - P_0(2))$ jointly converge in distribution to independent Brownian processes $G_{P_0(1)}$ and $G_{P_0(2)}$ given $N_1, T_1 \to \infty$ with $N_1^{2+r}/T_1^{2r} \to 0$ and $N_2, T_2 \to \infty$ with $N_2^{2+r}/T_2^{2r} \to 0$ by Theorem 2. Therefore, under $H_{0k}^k : P_0(1) = P_0(2)$, $KS_2$ converges in distribution to

$$\left\| \sqrt{1 - \lambda} G_{P_0(1)} - \sqrt{\lambda} G_{P_0(2)} \right\|_\infty,$$

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 $\sqrt{1 - \lambda} G_{P_0(1)} - \sqrt{\lambda} G_{P_0(2)}$ is identical to that of $G_{P_0(1)}$ under $H_0$. 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 $\alpha$-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

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

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 \( \alpha \)-mixing process taking values in \( \mathbb{R} \), and let \( \alpha (m) \) denote its \( \alpha \)-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. Let \( r \) be an even natural number. Suppose that Assumptions 1, 2, and 3 hold for \( r = r_m = r \) and \( r_d = r \). Then, it holds that \( E((\bar{w}_i)^r) \leq CT^{-r/2} \).

Proof. We first consider the case with \( r = 2 \). Given \( E(\bar{w}_i|i) = 0 \), Lemma 1 states that
\[
E \left( \left( \bar{w}_i \right)^2 \right) \leq C_i / T,
\]
where \( C_i = 12(E(|w_{it}|^{(4+\delta)/2}|i))^{4/(4+\delta)} \sum_{m=0}^{\infty} \alpha (m)^{\delta/(4+\delta)} \). Assumption 2 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
\[
E \left( \left( \bar{w}_i \right)^2 \right) = E \left( E \left( \left( \bar{w}_i \right)^2 \right)|i \right) \leq 12E \left( E \left( |w_{it}|^{(4+\delta)/2}|i \right) \right)^{4/(4+\delta)} \sum_{m=0}^{\infty} \alpha (m)^{\delta/(4+\delta)} / T \leq 12 \left( E \left( E \left( |w_{it}|^{(4+\delta)/2}|i \right) \right)^{4/(4+\delta)} \sum_{m=0}^{\infty} \alpha (m)^{\delta/(4+\delta)} / T \right) = 12 \left( E \left( |w_{it}|^{(4+\delta)/2} \right) \right)^{4/(4+\delta)} \sum_{m=0}^{\infty} \alpha (m)^{\delta/(4+\delta)} / T = O \left( \frac{1}{T} \right)
\]
where the second inequality is Jensen’s inequality and the last equality follows from Assumptions 2 and 3. Hence, the desired result holds for \( r = 2 \).

Next, we consider the case with \( r > 2 \). We use Lemma 2. From the proof of Lemma 2 available in Yokoyama (1980), we have
\[
E \left( \left| \sum_{t=1}^{T} w_{it} \right|^r \right) \leq K_{r,i} \left( E \left( |w_{it}|^{r+\delta}|i \right) \right)^{r/(r+\delta)} T^{r/2}.
\]
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 2, there exists a constant $K_r < \infty$ such that $K_{r,i} < K_r$ for all $i$. Thus, we have

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

where the second inequality is Jensen’s inequality and the last equality follows from Assumption 3. The proof for $r > 2$ is complete.

Because $\hat{\mu}_i - \mu_i = \bar{y}_i - \mu_i = \tilde{w}_i$, we obtain the following result as a corollary.

**Corollary 2.** Let $r$ be an even natural number. Suppose that Assumptions 1, 2, and 3 hold for $r_m = r$ and $r_d = r$. Then we have

$$E((\hat{\mu}_i - \mu_i)^r) = O(T^{-r/2}).$$

**Lemma 4.** Let $r$ be an even natural number. Suppose that Assumptions 1 and 2 hold for $r_m = r$. Then, $\{w_{it}w_{i,t-k}\}_{i=0}^{\infty}$ for a fixed $k$ given $\alpha_i$ is strictly stationary and $\alpha$-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 $\sum_{m=0}^{\infty}(m+1)^{r/2-1}\alpha_k(m)\delta/(r+\delta) < \infty$ for some $\delta > 0$.

**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 $\alpha$-mixing coefficients and Assumption 2. Thus, we have $\sum_{m=0}^{\infty}(m+1)^{r/2-1}\alpha_k(m)\delta/(r+\delta) \leq \sum_{m=0}^{k-1}(m+1)^{r/2-1} + \sum_{m=k}^{\infty}(m+1)^{r/2-1}\alpha(m-k)^{\delta/(r+\delta)} < \infty$ under Assumption 2.

**Lemma 5.** Let $r$ be an even natural number. Suppose that Assumptions 1, 2, and 3 hold for $r_m = r$ and $r_d = 2r$. 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 $C$.

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

**Lemma 6.** Let $r$ be an even natural number. Suppose that Assumptions 1, 2, and 3 hold for $r_m = 2r$ and $r_d = 2r$. Then we have

$$E((\hat{\gamma}_{k,i} - \gamma_{k,i})^r) = O(T^{-r/2}).$$
Proof. We have

\[
E((\hat{\gamma}_{k,i} - \gamma_{k,i})^r) = E\left(\left(\frac{1}{T-k} \sum_{t=k+1}^{T} (w_{it}w_{i,t-k} - \gamma_{k,i}) - \frac{T+k}{T-k} (\bar{w}_i)^2 \right)
+ \frac{1}{T-k} \sum_{t=1}^{k} w_{it} \bar{w}_i + \frac{1}{T-k} \sum_{t=T-k+1}^{T} w_{it} \bar{w}_i\right)^r .
\]

(26)

Thanks to Loève’s cr inequality, we only need examine the \(r\)-order moment of each term in parentheses on the right-hand side of (26). For the first term, Lemma 5 leads to

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

For the second term in (26), we first note that \((T+k)/(T-k) = O(1)\). We observe that

\[
E((\bar{w}_i)^2) = E((\bar{w}_i)^{2r}) = O \left(T^{-r}\right),
\]

by Lemma 3. We thus have that

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

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

\[
E \left(\left(\frac{1}{T-k} \sum_{t=1}^{k} w_{it} \bar{w}_i\right)^r\right) = E \left(\left(\bar{w}_i\right)^r \left(\frac{1}{T-k} \sum_{t=1}^{k} w_{it}\right)^r\right)
\leq \left(E \left(\left(\bar{w}_i\right)^{2r}\right)\right)^{1/2} \left(E \left(\left(\frac{1}{T-k} \sum_{t=1}^{k} w_{it}\right)^{2r}\right)\right)^{1/2}.
\]

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

\[
E \left(\left(\frac{1}{T-k} \sum_{t=1}^{k} w_{it}\right)^{2r}\right) = O \left(T^{-2r}\right).
\]

Therefore, it holds that

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

We can use the same argument to show that

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

Thus, we have shown the desired result.

\[\square\]
Lemma 7. Let \( r \) be an even natural number. Suppose that Assumptions 1, 2, and 3 hold for \( r_m = 2r \) and \( r_d = 2r \) and that \( \gamma_{0,i} > \epsilon \) almost surely for some constant \( \epsilon > 0 \). We have

\[
E((\hat{\rho}_{k,i} - \rho_{k,i})^r) = O(T^{-r/2}).
\]

Proof. We observe that

\[
E((\hat{\rho}_{k,i} - \rho_{k,i})^r) = E\left(\left(\frac{1}{\gamma_{0,i}} (\hat{\gamma}_{k,i} - \gamma_{k,i}) - \frac{1}{\gamma_{0,i}} \hat{\rho}_{k,i} (\hat{\gamma}_{0,i} - \gamma_{0,i})\right)^r\right).
\]

By Loève’s \( c_r \) inequality, we only need to examine the \( r \)-order moment of each term in parentheses on the right-hand side. We have

\[
E\left(\left(\frac{1}{\gamma_{0,i}} (\hat{\gamma}_{k,i} - \gamma_{k,i})\right)^r\right) \leq M E((\hat{\gamma}_{k,i} - \gamma_{k,i})^r),
\]

for some \( M < \infty \) by the assumption that \( \gamma_{0,i} > \epsilon \). Lemma 6 implies that \( E((\hat{\gamma}_{k,i} - \gamma_{k,i})^r) = T^{-r/2} \). For the second term, it holds that

\[
E\left(\left(\frac{1}{\gamma_{0,i}} \hat{\rho}_{k,i} (\hat{\gamma}_{0,i} - \gamma_{0,i})\right)^r\right) \leq M E((\hat{\gamma}_{0,i} - \gamma_{0,i})^r),
\]

for some \( M < \infty \) by the assumption that \( \gamma_{0,i} > \epsilon \) and the fact that \( |\hat{\rho}_{k,i}| \leq 1 \). Lemma 6 implies that \( E((\hat{\gamma}_{0,i} - \gamma_{0,i})^r) = T^{-r/2} \). We thus have the desired result.

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

\[
E(|a_T - b_T|^p) = O(T^c),
\]

for some integer \( p \) and real number \( c < 0 \). It then holds that

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

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

\[
\Pr(a_T < x, b_T \geq x) = \Pr(a_T < x, b_T \geq x, |a_T - b_T| > \epsilon) + \Pr(a_T < x, b_T \geq x, |a_T - b_T| \leq \epsilon).
\]
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
\[
\sup_x \Pr (a_T < x, b_T \geq x, |a_T - b_T| \leq \epsilon) \leq \sup_x \Pr (x - \epsilon \leq a_T < x, x \leq b_T \leq x + \epsilon) \leq \sup_x \sup_{x - \epsilon \leq a < x, \epsilon \leq b \leq x + \epsilon} f_{a_T,b_T}(a,b) \leq \epsilon^2 C,
\]
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
\[
\sup_x |\Pr (a_T < x) - \Pr (b_T < x)| \leq \sup_x |\Pr (a_T < x) - \Pr (a_T < x, b_T < x)| + \sup_x |\Pr (b_T < x) - \Pr (a_T < x, b_T < x)| = O \left( T^{2c/(2+p)} \right).
\]

\begin{flushright}
\square
\end{flushright}

**Acknowledgments**

The authors greatly appreciate the assistance of Yuya Sasaki in providing the Chilean panel data and in helpful discussions. The authors would also like to thank Debopam Bhattacharya, Richard Blundell, Maurice Bun, Yoosoon Chang, Andrew Chesher, Yasunori Fujikoshi, Kazuhiko Hayakawa, Toshio Honda, Hidehiko Ichimura, Koen Jochmans, Artūras
Juodis, Hiroaki Kaido, Hiroyuki Kasahara, Kengo Kato, Toru Kitagawa, Yuichi Kitamura, Frank Kleibergen, Roger Koenker, Eiji Kurozumi, Simon Lee, Oliver Linton, Alexei Onatski, Taisuke Otsu, Hashem Pesaran, Katsumi Shimotsu, Richard Smith, Mans Soderbom, Liangjun Su, Martin Weidner, Yoon-Jae Whang, Takashi Yamagata, Yohei Yamamoto and seminar participants at the University of Tokyo, Seoul National University, Hiroshima University, Hitotsubashi University, the University of Gothenburg, Tinbergen Institute, Cemmap, the University of York and the University of Cambridge 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, the Summer Workshop on Economic Theory 2014, Netherlands Econometric Society Group 2015, the Princeton-QUT-SJTU-SMU conference and the Econometric Society World Congress 2015 held in Montreal, for their helpful comments and useful discussions. All remaining errors are of course ours. Okui acknowledges financial support from the Japan Society for the Promotion of Science (JSPS) under KAKENHI Grant Nos. 22330067, 25780151, 25285067, and 15H03329. Yanagi recognizes the financial support of Grant-in-Aid for JSPS Fellows No. 252035 and KAKENHI Grant No. 15H06214.

References


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


