Kernel Estimation for Panel Data with Heterogeneous Dynamics

Ryo Okui* and Takahide Yanagi†

February, 2018

Abstract

This paper proposes nonparametric kernel-smoothing estimation for panel data to examine the degree of heterogeneity across cross-sectional units. Our procedure is model-free and easy to implement, and provides useful visual information, which enables us to understand intuitively the properties of heterogeneity. We first estimate the sample mean, autocovariances, and autocorrelations for each unit and then apply kernel smoothing to compute estimates of their density and cumulative distribution functions. The kernel estimators are consistent and asymptotically normal under double asymptotics, i.e., when both cross-sectional and time series sample sizes tend to infinity. However, as these exhibit biases given the incidental parameter problem and the nonlinearity of the kernel function, we propose jackknife methods to alleviate any bias. We also develop bandwidth selection methods and bootstrap inferences based on the asymptotic properties. Lastly, we illustrate the success of our procedure using an empirical application of the dynamics of US prices and Monte Carlo simulation.

Keywords: panel data, heterogeneity, autocorrelation structure, nonparametric kernel smoothing, jackknife, bootstrap.

JEL Classification: C13, C14, C23.

*NYU Shanghai, 1555 Century Avenue, Pudong, Shanghai, China, 200122; and Department of Economics, University of Gothenburg, P.O. Box 640, SE-405 30 Gothenburg, Sweden. Tel: +86-21-2059-6157. Email: okui@nyu.edu

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

1 Introduction

The characteristics of heterogeneity across economic units have important implications for many empirical applications. For example, there is an interest in empirical analyses of the dynamics of price deviations or changes (e.g., Klenow and Malin, 2010). The results of structural estimation based on price panel data may then heavily depend on whether we have taken into account heterogeneity across items. Specifically, estimation assuming the absence of heterogeneity may lead to inconsistent estimates if in fact heterogeneity exists (see, e.g., Nakamura and Steinsson, 2008 for the structural estimation of price change). As another example, let us consider the identification and estimation of production functions, which are fundamental in many economic applications (e.g., Ackerberg, Benkard, Berry, and Pakes, 2007). Once again, allowing for the presence of heterogeneity may make a crucial difference in the identification and estimation results for the production function. In response, several studies examine the degree of heterogeneity across firms based on the descriptive statistics of panel data (e.g., Kasahara, Schrimpf, and Suzuki, 2017). However, there are few formal econometric tools available to examine the degree of heterogeneity in a model-free manner, despite the importance of such tools in empirical applications.

This paper contributes to the econometric literature by proposing an easy-to-implement nonparametric kernel-smoothing procedure for panel data to analyze the degree of heterogeneity across cross-sectional units.[1] We first estimate the sample mean, autocovariances, and autocorrelations of the time series of each unit. We then implement the kernel density and undertake cumulative distribution function (CDF) estimation based on the estimated quantities. Our procedure allows us to examine the degree of heterogeneity in a model-free manner. For example, the densities of the heterogeneous mean, variance, and first-order autocorrelation of price deviation visually indicate the characteristics of heterogeneity in the long run, volatility, and the persistence of price deviation across items. Further, even if our ultimate goal is structural estimation, our procedure yields useful information on how to incorporate heterogeneity in structural models. For example, if our procedure shows that the persistency of price deviation exhibits a left-skewed distribution across items, then the structural model should be designed in such a way that it can yield this distribution of persistency.

[1] An `R` package to implement the proposed procedure is available. See the authors' websites for details.

In our theoretical investigation, we prove the asymptotic properties of the proposed estimators based on the double asymptotics under which both the number of cross-sectional units, N, and the length of time series, T, tend to infinity (denoted $N, T \to \infty$).[2] We show that the proposed kernel density and CDF estimators are consistent and asymptotically normal, but exhibit asymptotic biases whose rates of convergence depend on T and the bandwidth h satisfying $h \to 0$.

For the kernel density estimation, there are three asymptotic biases. The first is the standard kernel-smoothing bias of order $O(h^2)$, which is well known in the literature on nonparametric kernel-smoothing methods (e.g., Li and Racine, 2007). The second is caused by the incidental parameter problem for panel data analyses (e.g., Neyman and Scott, 1948 and Nickell, 1981), and is of order $O(1/T)$. This means that the estimation of N individual autocovariances or autocorrelations requires the computation of N sample means, which creates the incidental parameter problem. The third results from the nonlinearity of the kernel function and the difference between the estimated quantity and the true quantity (e.g., the estimated and true means). We show that this is of order $O(1/(Th^2)) + \sum_{j=3}^{\infty} O(1/\sqrt{T^j h^{2j}})$. Similarly, the kernel CDF estimator has asymptotic biases associated with the incidental parameter problem and the nonlinearity of the kernel function, whose orders are $O(1/T)$ and $O(1/T) + \sum_{j=3}^{\infty} O(1/\sqrt{T^j h^{2j-2}})$, respectively. Remarkably, the convergence rates of the second-order nonlinearity biases for the density and CDF estimation differ, i.e., $O(1/(Th^2))$ and $O(1/T)$, which stems from the difference between the convergence rates of the estimators and the symmetry of the kernel function.

Because the asymptotic biases may be crucial, especially when T is relatively small compared with N, as in many microeconometric applications, we propose bias-correction methods based on a split-panel jackknife (Dhaene and Jochmans, 2015) to reduce these biases. In particular, we show that half-panel jackknife (HPJ) bias correction eliminates both the incidental parameter bias and the second-order nonlinearity bias for each estimator without inflating the asymptotic variance.

These asymptotic results depend upon conditions resulting from the relative magnitudes of N, T, and h and their ability to eliminate any higher-order nonlinearity biases of the estimators. Specifically, the consistency and asymptotic normality for the density estimation with bias correction require that $1/(Th^2) \to 0$ and $N/(T^3h^5) \to 0$, respectively. When using the standard

[2] Precisely, the double asymptotics $N, T \to \infty$ are any monotonic sequence $T = T(N) \to \infty$ as $N \to \infty$ and the bandwidth $h \to 0$ is any monotonic sequence $h = h(N, T(N)) = h(N) \to 0$ as $N, T \to \infty$ in our setting.

bandwidth $h \asymp N^{-1/5}$ in the kernel density estimation, the conditions require $N^2/T^5 \to 0$ and $N^2/T^3 \to 0$, respectively, which are integrated to $N^2/T^3 \to 0$. Similarly, for the consistency and asymptotic normality of the kernel CDF estimation with bias correction, we need $1/(Th^2) \to 0$ and $N/(T^3h^4) \to 0$, respectively. When using the standard bandwidth $h \asymp N^{-1/3}$ in the kernel CDF estimation, the conditions require $N^2/T^3 \to 0$ and $N^7/T^9 \to 0$, respectively, which are integrated to $N^7/T^9 \to 0$.

We can select the optimal bandwidth for the bias-corrected density or CDF estimator based on standard procedures such as plug-in methods and least-squares cross-validation. This is because the asymptotic mean squared error (AMSE) of the HPJ bias-corrected density or CDF estimator is the same as that of the infeasible density or CDF estimator based on the true mean, autocovariances, and autocorrelations under conditions on the relative magnitudes of N and T. In particular, this paper proposes the use of plug-in methods because of their fast and easy implementation. The selected plug-in bandwidth for the density or CDF estimation is $h \asymp N^{-1/5}$ or $h \asymp N^{-1/3}$, respectively.

We propose to apply the cross-sectional bootstrap (e.g., Gonçalves and Kaffo, 2015) to construct confidence intervals or confidence bands for the density and CDF estimation. We show that the bootstrap distribution fails to capture all three kinds of bias. Our recommendation is thus to use bootstrap inferences based on the HPJ bias-corrected estimators, as they do not suffer from the incidental parameter and second-order nonlinearity biases.

We develop an empirical application and Monte Carlo simulations to illustrate the use of our procedure for finite samples. The empirical application examines the degree of heterogeneity of price deviation from the law of one price (LOP) using panel data for US prices. We find that the dynamics of the LOP deviation are significantly heterogeneous across items and US cities, and that the properties of heterogeneity differ between goods and services. Monte Carlo simulations demonstrate the successful performance of our procedure involving bias correction.

Paper organization. Section 2 reviews related studies, while Sections 3 and 4 introduce our setting and the proposed procedure, respectively. Section 5 illustrates the use of our procedure in the application. Sections 6 and 7 provide the asymptotic theory for the density and CDF estimation, and Section 8 details the Monte Carlo simulations. Section 9 concludes. Appendices A and B contain the proofs of the theorems and the technical lemmas. Appendices C and D contain other technical discussions.

2 Relation to the literature

This paper contributes to the literature on panel data and nonparametric estimation. We propose a method to analyze heterogeneity in dynamics in a model-free and nonparametric manner. This contrasts with existing theoretical studies based on specific models that do not focus on the entire distribution of heterogeneity. Nonetheless, in practice, several applied studies have used nonparametric kernel estimation to assess the degree of heterogeneity (e.g., Kasahara et al., 2017, Figure 2 and Roca and Puga, 2017, Figure 8). However, to our knowledge, no existing theoretical study justifies such procedures. The present paper thus provides a theoretical foundation that reveals the importance of asymptotic bias and its correction.

Many studies develop model-based procedures to examine the degree of heterogeneity in panel data. Given that panel data analyses with heterogeneous parameters often create asymptotic biases such as the incidental parameter bias, it is important to derive the convergence rates of such biases and to develop statistical procedures that are valid even in the presence of such biases. For example, Pesaran and Smith (1995), Hsiao, Pesaran, and Tahmiscioglu (1999), and Pesaran, Shin, and Smith (1999) study regression models with random coefficients and develop asymptotically valid inferences under the condition that $N/T^2 \to 0$. More recently, Fernández-Val and Lee (2013) consider moment restriction models. Hsiao and Pesaran (2008) and Pesaran (2015, Chapter 28) review the literature on heterogeneous panel data analyses with large N and/or large T. In contrast to those studies, our approach is model-free, and this enables us to examine the entire distribution.

Some studies of panel data employ deconvolution techniques to identify and estimate the density and distribution of the heterogeneous parameters. Horowitz and Markatou (1996) is the seminal study in econometrics. Recently, Arellano and Bonhomme (2012) and Mavroeidis, Sasaki, and Welch (2015) respectively consider static and dynamic panel data models with random coefficients. However, these studies examine heterogeneity in the parameters in specified models, while our approach does not rely on a specific model. Moreover, our procedure is relatively straightforward to implement.

The present study closely relates to the analysis of model-free heterogeneous panel data in Okui and Yanagi (2017), who develop empirical distribution estimation based on the quantities estimated for each cross-sectional unit, and then estimate the CDF and moments (e.g., the mean, variance, and correlation) of the heterogeneous means, autocovariances, and autocorrelations. However, unlike the present paper, their approach does not employ smoothing, such that their approach cannot estimate the density. Moreover, it is difficult to derive the exact form of bias using their nonsmoothed CDF estimation. In contrast, the present study complements the results in Okui and Yanagi (2017) with several novel contributions by proposing smoothed density and CDF estimation and by deriving the exact forms of the asymptotic biases.

Several other studies propose model-free analyses for panel data, but do not focus on the degree of heterogeneity in the dynamics of panel data. For example, Okui (2008, 2011, 2014) consider the estimation of the autocovariance and autocorrelation structures of panel data. However, these studies do not consider the presence of heterogeneity in the dynamics. Elsewhere, Galvao and Kato (2014) investigate the properties of the fixed effects (FE) estimator in a model-free manner, but their approach does not examine the degree of heterogeneity. Lee, Okui, and Shintani (2018) develop a panel AR(∞) inference that is essentially model-free, but again, their approach does not account for heterogeneity in the dynamics.

The present study also somewhat relates to the recent literature on finite unobserved grouped heterogeneity for panel data (Bonhomme and Manresa, 2015 and Su, Shi, and Phillips, 2016). Roughly speaking, this literature assumes that heterogeneity can be modeled by discrete distributions, while the present analysis assumes that it is represented by continuous distributions.[3] Another difference is that our aim is to examine the degree of heterogeneity based on a model-free procedure, while these studies investigate heterogeneity in the model coefficients.

3 Setup

This section describes the setting. We observe panel data $\{\{y_{it}\}_{t=1}^{T}\}_{i=1}^{N}$ where y_{it} is a scalar random variable, N is the number of cross-sectional units, and T is the length of the time series. We assume that y_{it} is strictly stationary across time.

[3] Bonhomme, Lamadon, and Manresa (2017) propose grouped FE estimation in maximum likelihood settings in which individual heterogeneity is approximated to finite grouped heterogeneity based on some discretization.

To introduce unobserved heterogeneity in a model-free manner, we consider the following data-generating process, as in Galvao and Kato (2014) and Okui and Yanagi (2017). First, the unit-specific unobserved variable α_i is independent and identically generated from some unknown probability distribution. Then, each individual times series $\{y_{it}\}_{t=1}^T$ is generated from some probability distribution $L(\{y_{it}\}_{t=1}^T; \alpha_i)$ that may depend on α_i. This setting assumes that $(\{y_{it}\}_{t=1}^T, \alpha_i)$ is i.i.d. across units. However, the realization of the unit-specific unobserved variable α_i may be different across units, so that the dynamics of the individual time series $\{y_{it}\}_{t=1}^T$ may be heterogeneous depending on the value of α_i. For simplicity, we denote the conditional expectation given α_i by $E(\cdot|i)$. Intuitively, it indicates the expected value for the individual time series $\{y_{it}\}_{t=1}^T$.

Our goal is to examine the degree of heterogeneity for the dynamics of y_{it} across units in a model-free manner. To this end, we focus on the mean, k-th autocovariance, and k-th autocorrelation of the individual time series $\{y_{it}\}_{t=1}^T$:

$$\mu_i := E(y_{it}|i), \qquad \gamma_{k,i} := E\big((y_{it} - \mu_i)(y_{i,t-k} - \mu_i)\big|i\big), \qquad \rho_{k,i} := \frac{\gamma_{k,i}}{\gamma_{0,i}}. \tag{1}$$

These quantities are random variables in general and do not depend on t as we assume stationarity. We note that $\gamma_{0,i}$ is the variance for the individual time series $\{y_{it}\}_{t=1}^T$. Throughout the paper, we mostly use the notation ξ_i to represent one of μ_i, $\gamma_{k,i}$, or $\rho_{k,i}$. We denote the density and CDF of $\xi_i = \mu_i$, $\gamma_{k,i}$, or $\rho_{k,i}$ by $f_\xi(\cdot)$ and $F_\xi(\cdot) = \Pr(\xi_i \leq \cdot)$, respectively.

As an example, consider an empirical analysis of heterogeneity in price dynamics. Let y_{it} be the price change for item i at time t. Then, α_i is item-specific unobserved heterogeneity, which may depend on, for example, item-specific unobserved permanent trade costs. In this case, μ_i is the long-run level of price change, $\gamma_{0,i}$ is the time series volatility of price change, and $\rho_{1,i}$ is the time series persistency of price change for item i. The density and CDF of μ_i, $\gamma_{k,i}$, and $\rho_{k,i}$ visually indicate the degree of heterogeneity for the quantities with information on the mode, symmetry, and tail behaviors of their distributions.

Introducing the random variable α_i is a way to consider heterogeneity in the dynamics of y_{it} in a model-free manner, and it can nest existing models with heterogeneity. For example, we regard the following popular models as special cases of our setting.

Example 1. Suppose that y_{it} is generated by the following AR(1) model with random coefficients:

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

where c_i and ϕ_i are heterogeneous random coefficients and $u_{it} \sim (0, \sigma_i^2)$ is the error term with the heterogeneous variance σ_i^2. Here, $\alpha_i = (c_i, \phi_i, \sigma_i)^\top$ and the distribution of the individual time series $\{y_{it}\}_{t=1}^T$ depends on α_i. It is easy to see that $\mu_i = c_i/(1-\phi_i)$, $\gamma_{k,i} = \sigma_i^2 \phi_i^k/(1-\phi_i^2)$, and $\rho_{k,i} = \phi_i^k$.

Example 2. Our setting also includes the following general nonparametric model:

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

where m is a nonparametric structural function with the covariate x_{it}, the unobserved heterogeneity α_i, and the idiosyncratic error term ε_{it}. In this case, $\mu_i = E(m(x_{it}, \alpha_i, \varepsilon_{it})|\alpha_i)$ and $\gamma_{k,i}$ and $\rho_{k,i}$ are the k-th order autocovariance and autocorrelation of $m(x_{it}, \alpha_i, \varepsilon_{it})$ given α_i. As in Example 1, these quantities are generally heterogeneous if α_i is heterogeneous across units, so that the density and CDF of these quantities provide useful information on heterogeneity in a model-free manner.

We recognize that the primary goal in some empirical studies may be to infer the structural function or the properties of the unobserved heterogeneity α_i rather than the features of μ_i, $\gamma_{k,i}$, and $\rho_{k,i}$, but we do not focus on such inference in this paper. This is because the density and CDF of μ_i, $\gamma_{k,i}$, and $\rho_{k,i}$ can be relatively easily estimated without requiring parametric specifications and identification conditions. In addition, even if our ultimate goal is to infer the structural function or the properties of the unobserved heterogeneity α_i, our procedure can provide useful information for such inference in a model-free manner. Indeed, as an example of production function estimation, Kasahara et al. (2017) examine the histograms of the means of the ratio of intermediate input cost to output value across firms to investigate the degree of heterogeneity in the production function, before they proceed to their structural estimation with unobserved heterogeneity. The present paper aims to present a new model-free procedure for such situations to examine the degree of heterogeneity with formal statistical properties.

4 Procedure

This section introduces our procedure. We first estimate the quantities of interest for each time series. We then apply nonparametric kernel smoothing to estimate their densities and/or distributions. We also discuss the bias-correction method, the choice of bandwidths, and a bootstrap procedure for statistical inference.

The first step is to estimate the heterogeneous mean μ_i, autocovariance $\gamma_{k,i}$, and autocorrelation $\rho_{k,i}$ in (1) by the sample analogues:

$$\hat{\mu}_i \coloneqq \bar{y}_i \coloneqq \frac{1}{T}\sum_{t=1}^{T} y_{it}, \qquad \hat{\gamma}_{k,i} \coloneqq \frac{1}{T-k}\sum_{t=k+1}^{T}(y_{it}-\bar{y}_i)(y_{i,t-k}-\bar{y}_i), \qquad \hat{\rho}_{k,i} \coloneqq \frac{\hat{\gamma}_{k,i}}{\hat{\gamma}_{0,i}}. \tag{2}$$

We then estimate the density and CDF of $\xi_i = \mu_i$, $\gamma_{k,i}$, or $\rho_{k,i}$ using nonparametric kernel-smoothing estimation based on $\hat{\xi}_i = \hat{\mu}_i$, $\hat{\gamma}_{k,i}$, or $\hat{\rho}_{k,i}$. The kernel estimators for the density $f_\xi(x)$ and the CDF $F_\xi(x) = \Pr(\xi_i \leq x)$ are given by

$$\hat{f}_{\hat{\xi}}(x) \coloneqq \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x-\hat{\xi}_i}{h}\right), \qquad \hat{F}_{\hat{\xi}}(x) \coloneqq \frac{1}{N}\sum_{i=1}^{N} \mathbb{K}\left(\frac{x-\hat{\xi}_i}{h}\right), \tag{3}$$

where $x \in \mathbb{R}$ is some fixed point, $K : \mathbb{R} \to \mathbb{R}$ is a kernel function such as the Gaussian kernel, $\mathbb{K} : \mathbb{R} \to [0,1]$ is a Borel-measurable CDF such as the Gaussian CDF, and $h > 0$ is a bandwidth satisfying $h \to 0$. These are standard kernel estimators except that we replace the true ξ_i with the estimated $\hat{\xi}_i$. We note that the CDF estimator may be regarded as the integral of the density estimator, that is, $\hat{F}_{\hat{\xi}}(x) = \int_{-\infty}^{x} \hat{f}_{\hat{\xi}}(v)dv$ and $\mathbb{K}(a) = \int_{-\infty}^{a} K(v)dv$. In practice, the bandwidths should be differently specified for the density and CDF estimation, but we use the same notation h for notational convenience.

As shown in Sections 6 and 7, the density and CDF estimators exhibit asymptotic biases, and we propose the adoption of split-panel jackknife bias correction to reduce them. Let us consider an even number of T for simplicity.[4] For $\xi_i = \mu_i$, $\gamma_{k,i}$, or $\rho_{k,i}$, we first obtain the estimators

[4] Even if T is odd, we can consider a similar split-panel jackknife correction, as in Dhaene and Jochmans (2015, page 9). Specifically, the HPJ bias-corrected estimator for $f_\xi(x)$ with odd T is given by (4) with $\bar{f}_{\hat{\xi}}(x) = (\hat{f}_{\hat{\xi},(1,1)}(x) + \hat{f}_{\hat{\xi},(2,1)}(x) + \hat{f}_{\hat{\xi},(1,2)}(x) + \hat{f}_{\hat{\xi},(2,2)}(x))/4$, where $\hat{f}_{\hat{\xi},(1,1)}(x)$, $\hat{f}_{\hat{\xi},(2,1)}(x)$, $\hat{f}_{\hat{\xi},(1,2)}(x)$, and $\hat{f}_{\hat{\xi},(2,2)}(x)$ are the estimators of $f_\xi(x)$ computed using $\{\{y_{it}\}_{t=1}^{\lceil T/2 \rceil}\}_{i=1}^{N}$, $\{\{y_{it}\}_{t=\lceil T/2 \rceil+1}^{T}\}_{i=1}^{N}$, $\{\{y_{it}\}_{t=1}^{\lfloor T/2 \rfloor}\}_{i=1}^{N}$, and $\{\{y_{it}\}_{t=\lfloor T/2 \rfloor+1}^{T}\}_{i=1}^{N}$, respectively. Here, $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ are the ceiling and floor functions, respectively.

$\hat{f}_{\hat{\xi},(1)}(x)$ and $\hat{f}_{\hat{\xi},(2)}(x)$ of $f_\xi(x)$ based on half-panel data $\{y_{i,(1)}\}_{i=1}^N \coloneqq \{\{y_{it}\}_{t=1}^{T/2}\}_{i=1}^N$ and $\{y_{i,(2)}\}_{i=1}^N \coloneqq \{\{y_{it}\}_{t=T/2+1}^{T}\}_{i=1}^N$, respectively. Importantly, the bandwidths for computing $\hat{f}_{\hat{\xi},(1)}(x)$ and $\hat{f}_{\hat{\xi},(2)}(x)$ must be the same as that for the original estimator $\hat{f}_{\hat{\xi}}(x)$ to successfully reduce the biases. The HPJ bias-corrected estimator of $f_\xi(x)$ is

$$\hat{f}^H_{\hat{\xi}}(x) \coloneqq 2\hat{f}_{\hat{\xi}}(x) - \bar{f}_{\hat{\xi}}(x) = \hat{f}_{\hat{\xi}}(x) - \left(\bar{f}_{\hat{\xi}}(x) - \hat{f}_{\hat{\xi}}(x)\right), \tag{4}$$

where $\bar{f}_{\hat{\xi}}(x) \coloneqq [\hat{f}_{\hat{\xi},(1)}(x) + \hat{f}_{\hat{\xi},(2)}(x)]/2$. The term $\bar{f}_{\hat{\xi}}(x) - \hat{f}_{\hat{\xi}}(x)$ estimates the bias in the original estimator $\hat{f}_{\hat{\xi}}(x)$. Similarly, the HPJ bias-corrected CDF estimator is

$$\hat{F}^H_{\hat{\xi}}(x) \coloneqq 2\hat{F}_{\hat{\xi}}(x) - \bar{F}_{\hat{\xi}}(x), \tag{5}$$

where $\bar{F}_{\hat{\xi}}(x) \coloneqq [\hat{F}_{\hat{\xi},(1)}(x) + \hat{F}_{\hat{\xi},(2)}(x)]/2$ is analogously defined as $\bar{f}_{\hat{\xi}}(x)$. As well as the density estimation, the bandwidths for computing $\hat{F}_{\hat{\xi},(1)}(x)$ and $\hat{F}_{\hat{\xi},(2)}(x)$ should be the same as that for the original estimator $\hat{F}_{\hat{\xi}}(x)$. We show that the HPJ bias correction reduces the incidental parameter bias and the second-order nonlinearity bias of each estimator.

Standard bandwidth selection procedures are available for the HPJ estimators. In particular, we use plug-in bandwidths in the empirical application and the Monte Carlo simulations because of their easy and fast implementation.[5] Wand and Jones (1994) and Polansky and Baker (2000) develop the plug-in bandwidths for the kernel density and CDF estimators that are of order $h \asymp N^{-1/5}$ and $h \asymp N^{-1/3}$, respectively. The use of the standard procedures is justified because they minimize the integrated asymptotic mean squared errors (IAMSEs) and because the IAMSEs of the HPJ bias-corrected estimators are the same as those of the standard kernel estimators based on the true μ_i, $\gamma_{k,i}$, and $\rho_{k,i}$ under conditions on the relative magnitudes of N and T. For example, we show that the IAMSE of the HPJ estimator $\hat{f}^H_{\hat{\xi}}(x)$ is the same as that of the infeasible estimator $\hat{f}_\xi(x) \coloneqq (Nh)^{-1}\sum_{i=1}^N K((x-\xi_i)/h)$.

We can construct confidence intervals or confidence bands for the density and CDF of μ_i, $\gamma_{k,i}$, or $\rho_{k,i}$ using cross-sectional bootstrap inference. In the cross-sectional bootstrap, we regard the

[5] Nonetheless, for the empirical application and the Monte Carlo simulations, we also checked the estimation results based on other bandwidth selections, including least-squares cross-validation and rules-of-thumb. The results for these are similar to those based on the plug-in bandwidths reported in this paper.

individual time series $y_i \coloneqq \{y_{it}\}_{t=1}^T$ as the unit of observation and implement the standard non-parametric bootstrap procedure. For example, the cross-sectional bootstrap inference to construct a $1-\alpha$ confidence interval for $f_\xi(x)$ without bias correction is as follows.

1. Randomly draw $y_1^*, y_2^*, \ldots, y_N^*$ from $\{y_1, y_2, \ldots, y_N\}$ with replacement.
2. Compute the estimate $\hat{f}_{\hat{\xi}}^*(x)$ based on the bootstrap sample $y_1^*, y_2^*, \ldots, y_N^*$.
3. Repeat 1 and 2 B times to obtain B bootstrap estimates $\{\hat{f}_{\hat{\xi},b}^*(x)\}_{b=1}^B$.
4. Obtain the bootstrap $(1-\alpha)$ confidence interval for $f_\xi(x)$ by $[\hat{f}_{\hat{\xi}}(x) - \hat{q}_{1-\alpha/2}^*,\ \hat{f}_{\hat{\xi}}(x) - \hat{q}_{\alpha/2}^*]$, where $\hat{q}_\alpha^*$ is the α quantile of the empirical distribution of $\{\hat{f}_{\hat{\xi},b}^*(x) - \hat{f}_{\hat{\xi}}(x)\}_{b=1}^B$.

It is important to keep using the same bandwidth h used for the original estimate $\hat{f}_{\hat{\xi}}(x)$ during the bootstrap repetition. We show that the cross-sectional bootstrap distribution approximates the asymptotic distribution of the density or CDF estimator, but fails to capture all three types of bias. Hence, our recommendation is bootstrap inference based on the HPJ bias-corrected estimator to reduce both the incidental parameter bias and the second-order nonlinearity bias.

Remark 1. We estimate the joint density and CDF for μ_i, $\gamma_{k,i}$, and $\rho_{k,i}$ in the same manner. For example, the joint density estimator for μ_i and $\gamma_{0,i}$ is

$$\hat{f}_{\hat{\mu},\hat{\gamma}_0}(x_1, x_2) \coloneqq \frac{1}{Nh^2} \sum_{i=1}^N K\left(\frac{x_1 - \hat{\mu}_i}{h}\right) K\left(\frac{x_2 - \hat{\gamma}_{0,i}}{h}\right).$$

We prove the asymptotic properties of this joint estimator in the same manner in our theoretical investigation below. In particular, the joint estimator exhibits incidental parameter and nonlinearity biases, such that we should implement HPJ bias correction.

Remark 2. We may extend our procedure to examine the degree of heterogeneity in other quantities. For example, if we are interested in the degree of heterogeneity of the median, we first compute the sample median for each unit and then estimate the density and CDF based on the estimated medians. As another example, if we are interested in the degree of heterogeneity of random coefficients in an AR model, as in Example 1, we first estimate the model using each individual time series and then implement the kernel estimation based on the estimated random coefficients. Even in these situations, as the estimators exhibit incidental parameter and/or nonlinearity biases, we should employ HPJ bias correction.

Remark 3. While the CDF estimator $\hat{F}_{\hat{\xi}}(x)$ without bias correction is a nondecreasing function in x, the HPJ bias-corrected CDF estimator $\hat{F}^{H}_{\hat{\xi}}(x)$ is nonmonotonic in finite samples as a consequence of the bias correction. However, we can correct the problem by obtaining a rearranged version of $\hat{F}^{H}_{\hat{\xi}}(x)$ that is a nondecreasing function in x, based on the rearrangement method proposed by Chernozhukov, Fernandez-Val, and Galichon (2009).

Remark 4. Given that the bootstrap distribution for the density estimation cannot capture the kernel-smoothing bias of order $O(h^2)$, the $1-\alpha$ bootstrap confidence interval above covers the density $f_{\xi}(x)$ with asymptotic probability smaller than $1-\alpha$. This problem is well known in the literature, and several solutions to correct it are available (see, e.g., Hall and Horowitz, 2013, Section 1). In particular, we can adopt the modified bootstrap confidence bands estimation proposed by Hall and Horowitz (2013) coupled with the cross-sectional bootstrap procedure based on the estimated $\hat{\xi}_i$. Based on this procedure, we estimate confidence bands that cover the density $f_{\xi}(x)$ with asymptotic probability of at least $1-\alpha$ for at least a proportion $1-\delta$ of values of x for some δ satisfying $0<\delta<1$. See Hall and Horowitz (2013, Section 2.7) for details. We note that unlike the density estimation, the standard cross-sectional bootstrap confidence interval is valid for the CDF estimation, such that we do not need to adopt such modified confidence bands for the CDF estimation. The reason is that the kernel-smoothing bias of order $O(h^2)$ does not appear in the asymptotic normality of the CDF estimation (see Theorems 5 and 6). As a result, the standard confidence interval based on the cross-sectional bootstrap covers the CDF $F_{\xi}(x)$ with asymptotic probability $1-\alpha$.

Remark 5. When we observe other variables in addition to y_{it}, we can implement several extensions of the proposed procedure. For example, if we observe panel data for consumption y_{it} and income x_{it}, we can estimate the joint density of the heterogeneous autocorrelations for y_{it} and x_{it}, which indicates the degree of heterogeneity of the persistence of the dynamics of consumption and income. As another example, if y_{it} is an outcome variable and x_{it} is a vector of covariates, we can implement our procedure based on residuals obtained by regressing y_{it} on x_{it} to control the effects of the covariates. Given that these procedures may also suffer from biases arising from the incidental parameter problem and the nonlinearity of the kernel function, a practical recommendation is to implement HPJ bias correction.

5 Empirical application

We apply our procedure to panel data on prices of items in US cities. Our procedure allows us to examine the heterogeneous properties of the deviation of prices from the LOP across items and cities, and the difference in the degree of heterogeneity between goods and services.

Many empirical studies examine the heterogeneous properties of the level and volatility of price deviation and the speed of price adjustment toward the long-run LOP deviation (see Anderson and Van Wincoop, 2004 for a review). For example, Engel and Rogers (2001), Parsley and Wei (2001), and Crucini, Shintani, and Tsuruga (2015) examine such heterogeneous properties and find that the LOP deviation dynamics are significantly heterogeneous across items and cities based on regression models. Our investigation below complements such empirical analyses by using our model-free procedure, as it provides visual information concerning the degree of heterogeneity.

We consider the kernel density and CDF estimation with and without bias correction for the heterogeneous mean μ_i, variance $\gamma_{0,i}$, and first-order autocorrelation $\rho_{1,i}$. The density and CDF of μ_i, $\gamma_{0,i}$, and $\rho_{1,i}$ indicate the degree of heterogeneity for the long-run level, time series volatility, and time series persistence of the LOP deviation. The density estimation uses the Gaussian kernel $K(u) = \exp(-u^2/2)/\sqrt{2\pi}$ with the plug-in bandwidth of Wand and Jones (1994). The CDF estimation is conducted with the Gaussian CDF $\mathbb{K}(u) = \int_{-\infty}^{u} K(v)dv$ with the plug-in bandwidth of Polansky and Baker (2000).

Data. We use data from the American Chamber of Commerce Researchers Association (ACCRA) Cost of Living Index produced by the Council of Community and Economic Research.[6] The same data set is used by Parsley and Wei (1996), Yazgan and Yilmazkuday (2011), Crucini et al. (2015), Lee et al. (2018), and Okui and Yanagi (2017). The data set contains quarterly price series of 48 Consumer Price Index (CPI) items (goods and services) for 52 US cities from 1990Q1 to 2007Q4.[7] The categorization of goods and services is in Table 1.

[6] Mototsugu Shintani kindly provided us with the data set ready for analysis.

[7] While the original data source contains the price information for more items in additional cities, we restrict the observations to obtain a balanced panel data set, as in Crucini et al. (2015).

We define the LOP deviation for item k in city i at time t as $y_{ikt} = \ln P_{ikt} - \ln P_{0kt}$, where P_{ikt} is the price of item k in city i at time t and P_{0kt} is that for the benchmark city of Albuquerque, NM. We regard each item–city pair as a cross-sectional unit, such that we focus on the degree of heterogeneity of the LOP deviation across item–city pairs. The number of cross-sectional units is $N = 48 \times (52 - 1) = 2448$ and the length of the time series is $T = 18 \times 4 = 72$. The number of cross-sectional units for goods and services is 1,428 and 1,020, respectively.

Results. Figure 1 depicts the density and CDF estimates for μ_i, $\gamma_{0,i}$, and $\rho_{1,i}$. In the figure, the dashed blue lines are kernel estimates without bias correction, and the solid red lines are HPJ bias-corrected kernel estimates. Table 2 reports the bandwidths selected by the plug-in methods for the density and CDF estimators.

Estimation results both with and without bias correction show that the LOP deviation dynamics are significantly heterogeneous across items. The density and CDF estimates without bias correction for μ_i are almost the same as those with bias correction. This is likely because the estimators for μ_i do not suffer from incidental parameter bias, as shown in Theorems 1 and 5. The results for μ_i also show that the mode of the heterogeneous long-run LOP deviation is close to zero, with a nearly symmetric, unimodal distribution. In contrast, the estimates without bias correction for $\gamma_{0,i}$ and $\rho_{1,i}$ are very different from the HPJ bias-corrected estimates. In evidence, the bias-corrected estimates for $\gamma_{0,i}$ demonstrate larger volatilities for the LOP deviation dynamics, while the bias-corrected estimates for $\rho_{1,i}$ show more persistent dynamics with a more left-skewed distribution. These results suggest the severe impact of the incidental parameter biases, which highlights the importance of bias correction methods.

Figure 2 illustrates the HPJ bias-corrected density and CDF estimates of μ_i, $\gamma_{0,i}$, and $\rho_{1,i}$ for goods and services separately. In the figure, the solid red lines are the HPJ estimates for goods, and the dashed blue lines are those for services. Table 3 also reports the bandwidths selected by the plug-in methods for the density and CDF estimators for goods and services separately. We note that the estimated densities for services exhibit a few points around which estimates fluctuate somewhat wildly, because there are fewer cross-sectional units for services than for goods, such that observations near such points are sparse.

The estimated densities and CDFs show that the heterogeneous properties are significantly different between goods and services. The densities for μ_i show that the long-run LOP deviation for goods generally tends to be larger than that for services (in an absolute sense). The estimation results for $\gamma_{0,i}$ and $\rho_{1,i}$ show that the LOP deviation for goods tend to be more volatile but less persistent than that for services. These results suggest that goods tend to have more volatile processes with faster adjustment speeds toward the nonnegligible long-run LOP deviation than services.

If we seek to examine the degree of heterogeneity of the LOP deviation across items and cities as in Crucini et al. (2015), our model-free results are informative in their own right. There are several possible sources of the degree of heterogeneity, including the differences in trade costs across items (e.g., Anderson and Van Wincoop, 2004) and sale and nonsale prices across goods and services (e.g., Nakamura and Steinsson, 2008). Further, our model-free results also suggest how we should model heterogeneity when implementing structural estimation for price deviation or change. For example, as our procedure demonstrates that the heterogeneous properties of goods and services differ, we should model unobserved heterogeneity differently for goods and services.

6 Asymptotic theory for the density estimation

We now develop the asymptotic properties of the kernel density estimators with and without HPJ bias correction and the asymptotic validity of the bootstrap inference. To this end, we define the notation $w_{it} \coloneqq y_{it} - \mu_i = y_{it} - E(y_{it}|i)$ and $\bar{w}_i \coloneqq T^{-1}\sum_{t=1}^{T} w_{it}$. Note that $\hat{\mu}_i = \bar{y}_i = \mu_i + \bar{w}_i$, $E(w_{it}|i) = 0$, and $\gamma_{k,i} = E(w_{it}w_{i,t-k}|i)$ using this notation.

6.1 Basic conditions

Throughout subsequent sections, we assume the following basic conditions for the data-generating process for the panel data. These formalize the conditions discussed in Section 3 and are essentially the same as the assumptions in Okui and Yanagi (2017).

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

Assumption 2. *For each i, $\{y_{it}\}_{t=1}^{\infty}$ is strictly stationary and α-mixing given α_i with mixing coefficients $\{\alpha(m|i)\}_{m=0}^{\infty}$. For any natural number $r_m \in \mathbb{N}$, there exists 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. *For any natural number $r_d \in \mathbb{N}$, it holds that $E|w_{it}|^{r_d+\delta} < \infty$ for some $\delta > 0$.*

Assumption 4. *There exists a constant $\epsilon > 0$ such that $\gamma_{0,i} > \epsilon$ almost surely.*

Assumptions 1 and 2 require that the individual time series given α_i is strictly stationary across time but i.i.d. across units. We note that the i.i.d. assumption does not exclude the presence of heterogeneity in panel data. In our setting, heterogeneity is caused by differences in the realized values of $\{\alpha_i\}_{i=1}^N$ across units, as discussed in Section 3. Assumption 2 also restricts the degree of persistence of the individual time series. The conditions for the stationarity and the degree of persistence require that the times series for each unit is not a unit root process and that the initial value of each time series is generated from a stationary distribution. Assumption 3 requires the existence of the moments of w_{it}, and it allows us to derive the asymptotic biases of the estimators. While we can develop the theoretical properties of the estimators in situations where Assumptions 2 and 3 do not hold for some numbers r_m and r_d, we cannot derive the exact forms of the higher-order biases of the estimators based on infinite-order Taylor expansions in such situations. As a result, in such situations, we require stronger conditions on the relative magnitudes of N and T than the conditions in the theorems below to establish the consistency and asymptotic normality of the estimators. Assumption 4 allows us to provide the asymptotic properties of the kernel estimators for $\rho_{k,i}$. All of the assumptions can be satisfied in popular panel data models. For example, they all hold when y_{it} follows a heterogeneous stationary panel autoregressive moving average (ARMA) model with a Gaussian error term.

6.2 Asymptotic biases

We show the presence of asymptotic biases of the kernel density estimator in (3). To this end, we assume the following additional conditions.

Assumption 5. *The kernel function $K : \mathbb{R} \to \mathbb{R}$ is bounded, symmetric, and infinitely differentiable. It satisfies $\int K(s)ds = 1$, $\int |K^{(j)}(s)|ds < \infty$, $\int |sK^{(j)}(s)|ds < \infty$, $\int |s^2K^{(j)}(s)|ds < \infty$, and $\int |s^3K^{(j)}(s)|ds < \infty$ for any nonnegative integer j, where $K^{(j)}$ denotes the j-th order derivative of K.*

Assumption 5 includes the standard conditions for the kernel function in nonparametric kernel-smoothing estimation, except for infinite differentiability. We need the differentiability to expand the kernel estimator for the estimated $\hat{\xi}_i$ at the true ξ_i based on the Taylor expansion. We note that the symmetry of K implies that $\int K^{(j)}(s)ds = 0$ for any odd j. The Gaussian kernel function satisfies Assumption 5.

Assumption 6. *The random variables $\mu_i \in \mathbb{R}$, $\gamma_{k,i} \in \mathbb{R}$, and $\rho_{k,i} \in (-1, 1)$ are continuously distributed. The densities f_ξ with $\xi = \mu$, γ_k, and ρ_k are bounded away from zero near x and three-times boundedly continuously differentiable near x.*

Assumption 6 requires that ξ_i is continuously distributed without a mass of probability. The continuity of the random variable is essential for implementing nonparametric kernel-smoothing estimation as it rules out situations where there is no heterogeneity for ξ_i (that is, the situation where $\xi_i = \xi$ for any i with some constant ξ) and where there is finite grouped heterogeneity (that is, $\xi_{i_1} = \xi_{i_2}$ for any $i_1, i_2 \in \mathbb{I}_g$ with some sets $\mathbb{I}_1, \mathbb{I}_2, \dots, \mathbb{I}_G$ satisfying $\bigoplus_{g=1}^G \mathbb{I}_g = \{1, 2, \dots, N\}$).

Assumption 7. *The following functions are twice boundedly continuously differentiable near x for any $T \in \mathbb{N}$ with the finite limits at x as $T \to \infty$:*

$$\sqrt{T^j} E\left((\bar{w}_i)^j \middle| \mu_i = \cdot\right), \qquad \sqrt{T^j} E\left((\bar{w}_i)^j \middle| \gamma_{k,i} = \cdot\right), \qquad \sqrt{T^j} E\left((\bar{w}_i)^j \middle| \rho_{k,i} = \cdot\right),$$
$$\frac{1}{\sqrt{T^j}} E\left(\left(\sum_{t=k+1}^{T} (w_{it} w_{i,t-k} - \gamma_{k,i})\right)^j \middle| \gamma_{k,i} = \cdot\right),$$
$$\frac{1}{\sqrt{T^{j_1+j_2}}} E\left(\left(\sum_{t=k+1}^{T} (w_{it} w_{i,t-k} - \gamma_{k,i})\right)^{j_1} \left(\sum_{t=1}^{T} (w_{it}^2 - \gamma_{0,i})\right)^{j_2} \frac{\gamma_{k,i}^{j_3}}{\gamma_{0,i}^{j_4}} \middle| \rho_{k,i} = \cdot\right),$$

for any nonnegative integers j, j_1, j_2, j_3, j_4.

Assumption 7 depends on the existence and smoothness of the conditional expectations. We need these conditions to derive the exact forms of the asymptotic biases of the kernel density and CDF estimators. The convergence rates of the terms are standard and guaranteed by Lemmas 1 and 3 in Appendix B. For example, the assumption requires that $T \cdot E((\bar{w}_i)^2 | \mu_i = \cdot) = O(1)$ and the convergence rate is consistent with the result in Lemma 1.

The following theorem shows that the kernel density estimators are consistent and asymptotically normal, but exhibit asymptotic biases. We define $\kappa_1 := \int s^2 K(s)ds$ and $\kappa_2 := \int K^2(s)ds$. We denote the normal distribution with mean μ and variance σ^2 by $\mathcal{N}(\mu, \sigma^2)$.

Theorem 1. *Let $x \in \mathbb{R}$ be an interior point in the support of $\xi_i = \mu_i$, $\gamma_{k,i}$, or $\rho_{k,i}$. Suppose that Assumptions 1, 2, 3, 5, 6, and 7 hold. In addition, if $\xi_i = \rho_{k,i}$, suppose that Assumption 4 also holds. Suppose that the infinite-order Taylor expansion of $\hat{f}_{\hat{\xi}}(x) = (Nh)^{-1}\sum_{i=1}^N K((x-\hat{\xi}_i)/h)$ at ξ_i holds and that the infinite series of the asymptotic biases below is well-defined. When $N, T \to \infty$ and $h \to 0$ with $Nh \to \infty$, $Nh^5 \to C \in [0, \infty)$, and $Th^2 \to \infty$, it holds that*

$$\hat{f}_{\hat{\xi}}(x) - f_\xi(x) = \frac{1}{Nh}\sum_{i=1}^N K\left(\frac{x-\xi_i}{h}\right) - f_\xi(x) + \frac{A_{\xi,1}(x)}{T} + \frac{A_{\xi,2}(x)}{Th^2} + \sum_{j=3}^{\infty} \frac{A_{\xi,j}(x)}{\sqrt{T^j h^{2j}}} + o_p(1),$$

where $A_{\xi,j}(x)$ is a nonrandom bias term that depends on x and satisfies $A_{\mu,1}(x) = 0$ for any x. As a result, when $N/(T^3h^5) \to 0$ also holds, it holds that

$$\sqrt{Nh}\left(\hat{f}_{\hat{\xi}}(x) - f_\xi(x) - h^2\frac{\kappa_1 f''_\xi(x)}{2} - \frac{A_{\xi,1}(x)}{T} - \frac{A_{\xi,2}(x)}{Th^2}\right) \xrightarrow{d} \mathcal{N}\big(0, \kappa_2 f_\xi(x)\big).$$

The theorem shows that the density estimator can be written as the sum of the infeasible estimator based on the true ξ_i and the asymptotic biases. The convergence rate of the estimator is the standard order of $O_p(1/\sqrt{Nh})$, and the asymptotic distribution in the theorem is the same as that of the infeasible kernel density estimator based on the true ξ_i, $\hat{f}_\xi(x) = (Nh)^{-1}\sum_{i=1}^N K((x - \xi_i)/h)$. However, the estimator exhibits asymptotic biases. The results also require conditions on the relative magnitudes of N, T, and h; that is, $1/(Th^2) \to 0$ and $N/(T^3h^5) \to 0$.

The theorem also assumes the infinite-order Taylor expansion and the summability of the infinite series of the asymptotic biases. While we assume them directly in the theorem, in fact, we can show their validity under unrestrictive regularity conditions. Since these are highly technical issues, we explore their discussions in Appendices C and D.

The estimator for μ_i has two main asymptotic biases given $A_{\mu,1}(x) = 0$, but the estimators for $\gamma_{k,i}$ and $\rho_{k,i}$ have three main asymptotic biases, in addition to the higher-order biases. The first bias of the form $h^2\kappa_1 f''_\xi(x)/2$ is the standard kernel-smoothing bias for the density estimation. The second bias of the form $A_{\xi,1}(x)/T$ is the incidental parameter bias caused from estimating $\gamma_{k,i}$ and $\rho_{k,i}$ by $\hat{\gamma}_{k,i}$ and $\hat{\rho}_{k,i}$, respectively. The estimation of $\hat{\gamma}_{k,i}$ and $\hat{\rho}_{k,i}$ involves estimating μ_i by $\hat{\mu}_i = \bar{y}_i$ for each i, which becomes a source of the incidental parameter bias. Note that the estimator for

μ_i does not exhibit the incidental parameter bias. The third bias of the form $A_{\xi,2}(x)/(Th^2)$ is the second-order nonlinearity bias caused by expanding $K((x-\hat{\xi}_i)/h)$ for $K((x-\xi_i)/h)$ by Taylor expansion. Moreover, the theorem shows that the j-th order nonlinearity bias exhibits the form $A_{\xi,j}(x)/\sqrt{T^j h^{2j}}$ for $j \geq 3$.

We need the two conditions, $1/(Th^2) \to 0$ and $N/(T^3 h^5) \to 0$, to ensure the asymptotic negligibility of the higher-order nonlinearity biases. If we use the standard bandwidth $h \asymp N^{-1/5}$ in the kernel density estimation, the conditions $1/(Th^2) \to 0$ and $N/(T^3h^5) \to 0$ imply that $N^2/T^5 \to 0$ and that $N^2/T^3 \to 0$, respectively, which are integrated to $N^2/T^3 \to 0$. We note that while the incidental parameter bias and the second-order nonlinearity bias are also asymptotically negligible under these conditions, the practical magnitudes of these biases would be larger than those of the higher-order nonlinearity biases.

The results may be understood better using a simple sketch of the proof for the density estimation of $\gamma_{k,i}$. By Taylor expansion and the expansion for $\hat{\gamma}_{k,i} - \gamma_{k,i}$, the proof shows that

$$
\begin{aligned}
\hat{f}_{\hat{\gamma}_k}(x) - f_{\gamma_k}(x) = & \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x-\gamma_{k,i}}{h}\right) - f_{\gamma_k}(x) \\
& - \frac{1}{Nh^2}\sum_{i=1}^{N}(\bar{w}_i)^2 K'\left(\frac{x-\gamma_{k,i}}{h}\right) \\
& + \frac{1}{2Nh^3}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T} w_{it}w_{i,t-k} - \gamma_{k,i}\right)^2 K''\left(\frac{x-\gamma_{k,i}}{h}\right) \\
& + \sum_{j=3}^{\infty}\frac{(-1)^j}{j!Nh^{j+1}}(\hat{\gamma}_{k,i}-\gamma_{k,i})^j K^{(j)}\left(\frac{x-\gamma_{k,i}}{h}\right) + o_p(1).
\end{aligned}
\tag{6}
$$

For the first line, consistency and asymptotic normality hold given the standard arguments for the kernel density estimation. The second line is a source of the incidental parameter bias and its mean is

$$
\begin{aligned}
E\left(-\frac{1}{Nh^2}\sum_{i=1}^{N}(\bar{w}_i)^2 K'\left(\frac{x-\gamma_{k,i}}{h}\right)\right) &= -\frac{1}{Th}\int T\cdot E((\bar{w}_i)^2|\gamma_{k,i}=x-sh) f_{\gamma_k}(x-sh)K'(s)ds \\
&= -\frac{1}{Th}a_T(x)\int K'(s)ds + \frac{1}{T}a_T'(x)\int sK'(s)ds + o\left(\frac{1}{T}\right) \\
&= \frac{1}{T}a_T'(x)\int sK'(s)ds + o\left(\frac{1}{T}\right)
\end{aligned}
$$

$$= \frac{A_{\gamma_k,1}(x)}{T} + o\left(\frac{1}{T}\right),$$

where we define $a_T(x) \coloneqq T \cdot E((\bar{w}_i)^2 | \gamma_{k,i} = x) f_{\gamma_k}(x)$ and $A_{\gamma_k,1}(x) \coloneqq \lim_{T\to\infty} a'_T(x) \int sK'(s)ds$. Note that this result follows from $\int K'(s)ds = 0$ based on the symmetry of K. As a result, the incidental parameter bias is of order $O(1/T)$. The third line in (6) is a source of the second-order nonlinearity bias and can be written as $A_{\gamma_k,2}(x)/(Th^2) + o_p(1)$. The last line in (6) is a source of the higher-order nonlinearity biases and can be written as $\sum_{j=3}^{\infty} A_{\gamma_k,j}(x)/\sqrt{T^j h^{2j}} + o_p(1)$.

Remark 6. There are important differences between this result and that of Okui and Yanagi (2017) in terms of the conditions on the relative magnitudes of N and T. Okui and Yanagi (2017) show that the asymptotic normality of estimators for moments without bias correction requires the condition that $N/T^2 \to 0$. As a result, some might surmise that our kernel smoothing requires a "weaker" condition such as $Nh/T^2 \to 0$ because the kernel estimation is essentially taking the average number of observations in a local neighborhood that contains Nh observations on average. However, such conjecture is not true, as the asymptotic normality of our smoothed density and CDF estimation requires somewhat stronger conditions than Okui and Yanagi (2017). The failure of the conjecture stems from the fact that the rate at which the ratio of the estimated quantity to h converges to the ratio of the true quantity to h depends on h in the kernel-smoothing estimation, unlike the empirical distribution estimation in Okui and Yanagi (2017).

6.3 HPJ bias correction

We now examine the asymptotic properties of the HPJ bias-corrected estimator $\hat{f}^H_{\hat{\xi}}(x)$ in (4) and show that this does not suffer from the incidental parameter bias or the second-order nonlinearity bias.

Before presenting the result formally, we briefly explain the mechanism behind the HPJ bias correction. We observe the following expansion for the density estimator for $\gamma_{k,i}$, as shown in Theorem 1:

$$\hat{f}_{\hat{\gamma}_k}(x) - f_{\gamma_k}(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x-\gamma_{k,i}}{h}\right) - f_{\gamma_k}(x) + \frac{A_{\gamma_k,1}(x)}{T} + \frac{A_{\gamma_k,2}(x)}{Th^2} + O_p\left(\frac{1}{\sqrt{T^3h^6}}\right).$$

This expansion implies that the HPJ bias-corrected estimator $\hat{f}^H_{\hat{\gamma}_k}(x) = 2\hat{f}_{\hat{\gamma}_k}(x) - \bar{f}_{\hat{\gamma}_k}(x)$ satisfies

$$\hat{f}^H_{\hat{\gamma}_k}(x) - f_{\gamma_k}(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x-\gamma_{k,i}}{h}\right) - f_{\gamma_k}(x) + O_p\left(\frac{1}{\sqrt{T^3h^6}}\right),$$

as $\bar{f}_{\hat{\gamma}_k}(x)$ is computed based on half-panel data with a time series length of $T/2$. This shows that the HPJ bias correction eliminates the incidental parameter bias and the second-order nonlinearity bias, but cannot eliminate the higher-order nonlinearity biases. This discussion also illustrates the reason why all $\hat{f}_{\hat{\xi}}(x)$, $\hat{f}_{\hat{\xi},(1)}(x)$, and $\hat{f}_{\hat{\xi},(2)}(x)$ must depend on the same bandwidth.

The following theorem presents the formal result for the HPJ bias-corrected density estimator.

Theorem 2. *Suppose that the assumptions in Theorem 1 hold. When $N, T \to \infty$ and $h \to 0$ with $Nh \to \infty$, $Nh^5 \to C \in [0, \infty)$, $Th^2 \to \infty$, and $N/(T^3h^5) \to 0$, it holds that*

$$\sqrt{Nh}\left(\hat{f}^H_{\hat{\xi}}(x) - f_\xi(x) - h^2\frac{\kappa_1 f''_\xi(x)}{2}\right) \xrightarrow{d} \mathcal{N}\big(0, \kappa_2 f_\xi(x)\big).$$

The theorem shows that the HPJ bias correction eliminates both the incidental parameter bias and the second-order nonlinearity bias without altering the asymptotic variance of the estimator. Similar results are observed by Galvao and Kato (2014), Dhaene and Jochmans (2015), and Okui and Yanagi (2017). However, the HPJ bias correction does not weaken the relative magnitudes condition on N, T, and h for the asymptotic normality in Theorem 1; that is, $N/(T^3h^5) \to 0$. This is because the HPJ bias correction cannot eliminate the higher-order nonlinearity biases.

The theorem also implies that the AMSE of the HPJ bias-corrected estimator $\hat{f}^H_{\hat{\xi}}(x)$ is

$$\left(h^2\frac{\kappa_1 f''_\xi(x)}{2}\right)^2 + \frac{\kappa_2 f_\xi(x)}{Nh},$$

which is identical to the AMSE of the infeasible estimator $\hat{f}_\xi(x) = (Nh)^{-1}\sum_{i=1}^N K((x-\xi_i)/h)$. As a result, we can select the optimal bandwidth for the HPJ bias-corrected estimation using standard procedures such as the plug-in method based on the estimated $\hat{\xi}_i$.

Remark 7. We consider higher-order jackknifes, such as the third-order jackknife, to eliminate higher-order biases, as discussed in Dhaene and Jochmans (2015) and Okui and Yanagi (2017). Theoretically, the third-order jackknife bias correction can eliminate the third-order nonlinearity bias, as well as the incidental parameter bias and the second-order nonlinearity bias, without inflating the asymptotic variance of the estimator. However, in practice, the HPJ bias correction can be more successful than the higher-order jackknife bias correction in the sense of the MSE, as noted in Dhaene and Jochmans (2015) and as observed in Monte Carlo simulations by Okui and Yanagi (2017). This is because the third-order jackknife may inflate higher-order biases it does not correct. For this reason, we recommend the HPJ instead of a higher-order jackknife.

6.4 Cross-sectional bootstrap

We now examine the asymptotic validity of the cross-sectional bootstrap inference for the density estimation. Let P^* be a bootstrap distribution that is identical to the empirical distribution of the original sample $\{y_i\}_{i=1}^N$ regarding the individual time series as the unit of observations.

Theorem 3. *Suppose that the assumptions in Theorem 1 hold. When $N, T \to \infty$ and $h \to 0$ with $Nh \to \infty$, $Nh^5 \to C \in [0, \infty)$, and $Th^2 \to \infty$, it holds that*

$$\sup_{a \in \mathbb{R}} \left| P^* \left(\sqrt{Nh} \left(\hat{f}^*_{\hat{\xi}}(x) - \hat{f}_{\hat{\xi}}(x) \right) \leq a \right) - \Pr \left(\mathcal{N}\left(0, \kappa_2 f_\xi(x)\right) \leq a \right) \right| \xrightarrow{p} 0.$$

This theorem shows that the cross-sectional bootstrap distribution can approximate the asymptotic distribution of the density estimator in Theorem 1 but cannot capture the asymptotic biases. We note that the theorem does not require the relative magnitudes condition $N/(T^3h^5) \to 0$ for the asymptotic normality of the density estimator in Theorem 1.

The bootstrap distribution does not capture the incidental parameter bias and the nonlinearity biases as well as the standard kernel-smoothing bias of order $O(h^2)$ for the kernel density estimator. Similar observations that the cross-sectional bootstrap does not capture the incidental parameter bias for panel data are observed in Galvao and Kato (2014), Kaffo (2014), Gonçalves and Kaffo (2015), and Okui and Yanagi (2017). Also, many existing studies observe that the bootstrap inference for nonparametric kernel-smoothing estimation does not capture the kernel-smoothing bias of order $O(h^2)$ (e.g., Hall, 1992 and Hall and Horowitz, 2013).

Because the HPJ bias correction can reduce the incidental parameter bias and the second-order nonlinearity bias, the cross-sectional bootstrap inference based on the HPJ bias-corrected estimator is preferred. The following theorem shows the asymptotic validity of the cross-sectional bootstrap inference with the HPJ bias correction. The proof of the theorem is the same as that of Theorem 3, so we omit it here.

Theorem 4. *Suppose that the assumptions in Theorem 1 hold. When $N, T \to \infty$ and $h \to 0$ with $Nh \to \infty$, $Nh^5 \to C \in [0, \infty)$, and $Th^2 \to \infty$, it holds that*

$$\sup_{a \in \mathbb{R}} \left| P^* \left(\sqrt{Nh} \left(\hat{f}_{\hat{\xi}}^{H*}(x) - \hat{f}_{\hat{\xi}}^{H}(x) \right) \le a \right) - \Pr \left(\mathcal{N}\big(0, \kappa_2 f_\xi(x)\big) \le a \right) \right| \xrightarrow{p} 0.$$

The theorem shows that the cross-sectional bootstrap distribution can approximate the asymptotic distribution of the HPJ bias-corrected density estimator in Theorem 2 but cannot capture the standard kernel-smoothing bias of order $O(h^2)$ and the higher-order nonlinearity biases.

7 Asymptotic theory for the CDF estimation

This section examines the asymptotic properties of the kernel CDF estimators with and without HPJ bias correction and demonstrates the asymptotic validity of the cross-sectional bootstrap inference. As the results for the CDF estimation are similar to those for the density estimation, the explanation of the results is brief.

7.1 Asymptotic biases

We derive the asymptotic biases of the kernel CDF estimator in (3). For the CDF estimation, we need the following instead of Assumption 6 for the density estimation. The continuity of the random variable is essential, even for the kernel-smoothing CDF estimation.

Assumption 8. *The random variables $\mu_i \in \mathbb{R}$, $\gamma_{k,i} \in \mathbb{R}$, and $\rho_{k,i} \in (-1, 1)$ are continuously distributed. The CDFs F_ξ with $\xi = \mu$, γ_k, and ρ_k are three-times boundedly continuously differentiable near x.*

The following theorem shows the presence of asymptotic biases for the kernel CDF estimator.

Theorem 5. *Let $x \in \mathbb{R}$ be an interior point in the support of $\xi_i = \mu_i$, $\gamma_{k,i}$, or $\rho_{k,i}$. Suppose that Assumptions 1, 2, 3, 5, 7, and 8 hold. In addition, if $\xi_i = \rho_{k,i}$, suppose that Assumption 4 also holds. Suppose that the infinite-order Taylor expansion of $\hat{F}_{\hat{\xi}}(x) = N^{-1} \sum_{i=1}^{N} \mathbb{K}((x - \hat{\xi}_i)/h)$ at ξ_i*

holds and that the infinite series of the asymptotic biases below is well-defined. When $N, T \to \infty$ and $h \to 0$ with $Nh^3 \to C \in [0, \infty)$ and $Th^2 \to \infty$, it holds that

$$\hat{F}_{\hat{\xi}}(x) - F_\xi(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{K}\left(\frac{x-\xi_i}{h}\right) - F_\xi(x) + \frac{B_{\xi,1}(x)}{T} + \frac{B_{\xi,2}(x)}{T} + \sum_{j=3}^{\infty} \frac{B_{\xi,j}(x)}{\sqrt{T^j h^{2j-2}}} + o_p(1),$$

where $B_{\xi,j}(x)$ is a nonrandom bias term that depends on x and that satisfies $B_{\mu,1}(x) = 0$ for any x. As a result, when $N/(T^3h^4) \to 0$ also holds, it holds that

$$\sqrt{N}\left(\hat{F}_{\hat{\xi}}(x) - F_\xi(x) - \frac{B_{\xi,1}(x)}{T} - \frac{B_{\xi,2}(x)}{T}\right) \xrightarrow{d} \mathcal{N}\big(0, F_\xi(x)[1 - F_\xi(x)]\big).$$

The theorem shows that the CDF estimator can be rearranged as the sum of the infeasible estimator based on the true ξ_i and the asymptotic biases. The result requires the infinite-order Taylor expansion and the summability of the infinite series, but they hold under technical regularity conditions as well as the density estimation in Theorem 1.

The biases of the forms $B_{\xi,1}(x)/T$ and $B_{\xi,2}(x)/T$ are the incidental parameter bias and the second-order nonlinearity bias, respectively. Note that $\hat{F}_{\hat{\mu}}(x)$ does not exhibit the incidental parameter bias as in the case of the density estimation. We also note that the standard kernel-smoothing bias of order $O(h^2)$ is not shown in the asymptotic normality because it is asymptotically negligible under $Nh^3 \to C$ (see Lemma 8 in Appendix B).

The consistency and the asymptotic normality of the estimator require the conditions $1/(Th^2) \to 0$ and $N/(T^3h^4) \to 0$, respectively, which asymptotically eliminate the higher-order biases. When using the standard bandwidth $h \asymp N^{-1/3}$, as selected by the plug-in method, the conditions $1/(Th^2) \to 0$ and $N/(T^3h^4) \to 0$ are the same as $N^2/T^3 \to 0$ and $N^7/T^9 \to 0$, respectively, which are integrated to $N^7/T^9 \to 0$.

Remark 8. We estimate the quantile function of $\xi_i = \mu_i$, $\gamma_{k,i}$, or $\rho_{k,i}$ by inverting the kernel CDF estimator. Specifically, let $q_\xi(\tau) := \inf\{x \in \mathbb{R} : F_\xi(x) \geq \tau\}$ be the τ-th quantile of ξ_i, where $\tau \in (0, 1)$. The kernel quantile estimator for $q_\xi(\tau)$ is given by

$$\hat{q}_{\hat{\xi}}(\tau) := \inf\{x \in \mathbb{R} : \hat{F}_{\hat{\xi}}(x) \geq \tau\} = \hat{F}_{\hat{\xi}}^{-1}(\tau),$$

based on a strictly increasing CDF $\mathbb{K}$, such as the Gaussian CDF. The asymptotic properties of the estimator can be shown by the results in Theorem 5 and the Delta method. For example, when $N, T \to \infty$ and $h \to 0$ with $Nh^3 \to C \in [0, \infty)$, $Th^2 \to \infty$, and $N/(T^3 h^4) \to 0$, it may hold that

$$\sqrt{N}\left(\hat{q}_{\hat{\xi}}(\tau) - q_{\xi}(\tau)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\tau(1-\tau)}{f_{\xi}^2(q_{\xi}(\tau))}\right).$$

7.2 HPJ bias correction

The following theorem shows that the HPJ bias-corrected estimator in (5) reduces the incidental parameter bias and the second-order nonlinearity bias. The idea behind the HPJ bias correction for the CDF estimation is the same as that for the density estimation.

Theorem 6. *Suppose that the assumptions in Theorem 5 hold. When $N, T \to \infty$ and $h \to 0$ with $Nh^3 \to C \in [0, \infty)$, $Th^2 \to \infty$, and $N/(T^3 h^4) \to 0$, it holds that*

$$\sqrt{N}\left(\hat{F}_{\hat{\xi}}^{H}(x) - F_{\xi}(x)\right) \xrightarrow{d} \mathcal{N}\big(0, F_{\xi}(x)[1 - F_{\xi}(x)]\big).$$

The HPJ bias-corrected kernel CDF estimator does not suffer from the incidental parameter bias or the second-order nonlinearity bias. However, because of the presence of the higher-order nonlinearity biases, the HPJ bias correction cannot weaken the relative magnitudes condition on N, T, and h for the asymptotic normality of the kernel CDF estimator in Theorem 5; that is, $N/(T^3 h^4) \to 0$. The AMSE of the bias-corrected estimator is identical to that of the infeasible estimator $\hat{F}_{\xi}(x) = N^{-1} \sum_{i=1}^{N} \mathbb{K}((x - \xi_i)/h)$, such that we can select the optimal bandwidth based on standard procedures such as the plug-in method.

7.3 Cross-sectional bootstrap

The following theorem shows that the cross-sectional bootstrap distribution approximates the asymptotic distribution of the kernel CDF estimator even though it cannot capture the incidental parameter bias or nonlinearity biases. We note that the following theorem does not require the relative magnitudes condition $N/(T^3 h^4) \to 0$ for the asymptotic normality of the CDF estimator.

Theorem 7. *Suppose that the assumptions in Theorem 5 hold. When $N, T \to \infty$ and $h \to 0$ with*

$Nh^3 \to C \in [0, \infty)$ *and* $Th^2 \to \infty$*, it holds that*

$$\sup_{a \in \mathbb{R}} \left| P^* \left(\sqrt{N} \left(\hat{F}^*_{\hat{\xi}}(x) - \hat{F}_{\hat{\xi}}(x) \right) \leq a \right) - \Pr \left(\mathcal{N}\big(0, F_\xi(x)[1 - F_\xi(x)]\big) \leq a \right) \right| \xrightarrow{p} 0.$$

Our recommendation is bootstrap inference based on the HPJ bias-corrected estimator because it does not suffer from the incidental parameter bias or second-order nonlinearity bias. The following theorem shows the validity of the cross-sectional bootstrap.

Theorem 8. *Suppose that the assumptions in Theorem 5 hold. When* $N, T \to \infty$ *and* $h \to 0$ *with* $Nh^3 \to C \in [0, \infty)$ *and* $Th^2 \to \infty$*, it holds that*

$$\sup_{a \in \mathbb{R}} \left| P^* \left(\sqrt{N} \left(\hat{F}^{H*}_{\hat{\xi}}(x) - \hat{F}^{H}_{\hat{\xi}}(x) \right) \leq a \right) - \Pr \left(\mathcal{N}\big(0, F_\xi(x)[1 - F_\xi(x)]\big) \leq a \right) \right| \xrightarrow{p} 0.$$

As the proof of the theorem is the same as that of Theorem 7, we omit it here.

8 Monte Carlo simulations

This section presents the results of Monte Carlo simulations to examine the finite-sample performance of the proposed procedure. The number of simulation replications is 5,000. The sample sizes are $N = 200$ and $T = 10, 11, \dots, 100$.

Design. We consider a data-generating process identical to that in Okui and Yanagi (2017) and motivated by their application to LOP deviation dynamics. Specifically, we generate the data using the following AR(1) process:

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

where $\epsilon_{it} \sim$ *i.i.d.* $\mathcal{N}(0, 1)$. Note that this design satisfies $\mu_i = \varsigma_i$, $\gamma_{0,i} = \sigma_i^2$, $\rho_{1,i} = \phi_i$. The unit-specific random variables ς_i, ϕ_i, and σ_i^2 are generated by the truncated normal distribution:

$$\begin{pmatrix} \varsigma_i \\ \sigma_i^2 \\ \phi_i \end{pmatrix} \sim i.i.d.\ \mathcal{N} \left(\begin{pmatrix} m_\varsigma \\ m_\sigma \\ m_\phi \end{pmatrix}, \begin{pmatrix} s_\varsigma^2 & \rho_{\varsigma,\sigma} s_\varsigma s_\sigma & \rho_{\varsigma,\phi} s_\varsigma s_\phi \\ \rho_{\varsigma,\sigma} s_\varsigma s_\sigma & s_\sigma^2 & \rho_{\sigma,\phi} s_\sigma s_\phi \\ \rho_{\varsigma,\phi} s_\varsigma s_\phi & \rho_{\sigma,\phi} s_\sigma s_\phi & s_\phi^2 \end{pmatrix} \right),$$

conditional on $|\varsigma_i - m_\varsigma| < 2.58 s_\varsigma$, $|\sigma_i^2 - m_\sigma| < 2.58 s_\sigma$, $|\phi_i - m_\phi| < 2.58 s_\phi$. We set

$$\begin{aligned}
&m_\varsigma = -0.0373, && m_\sigma = 0.0717, && m_\phi = 0.627, \\
&s_\varsigma^2 = 0.0141, && s_\sigma^2 = 0.000353, && s_\phi^2 = 0.021, \\
&\rho_{\varsigma,\sigma} = 0.082, && \rho_{\sigma,\phi} = 0.106, && \rho_{\varsigma,\phi} = -0.160.
\end{aligned}$$

We generate the initial observations (y_{i0}, ϵ_{i0}) for $i = 1, 2, \dots, N$ from the stationary distribution:

$$\begin{pmatrix} y_{i0} \\ \epsilon_{i0} \end{pmatrix} \sim i.i.d.\ \mathcal{N}\left(\begin{pmatrix} \varsigma_i \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_i^2 & 1 \\ 1 & 1 \end{pmatrix} \right).$$

Estimators. We consider three estimators for the density and CDF estimation of μ_i, $\gamma_{0,i}$, and $\rho_{1,i}$. The first is the naive estimator (NE) based on $\hat{\mu}_i$, $\hat{\gamma}_{0,i}$, and $\hat{\rho}_{1,i}$ without bias correction. The second is the HPJ bias-corrected estimator (HPJ) based on $\hat{\mu}_i$, $\hat{\gamma}_{0,i}$, and $\hat{\rho}_{1,i}$. The third is the infeasible estimator (IE) based on the true μ_i, $\gamma_{0,i}$, and $\rho_{1,i}$. For all estimators, we use the Gaussian kernel $K(u) = \exp(-u^2/2)/\sqrt{2\pi}$, the Gaussian CDF $\mathbb{K}(u) = \int_{-\infty}^{u} K(v)dv$, and the plug-in bandwidths.

Results. Figures 3 and 4 provide numerically computed values of the integrated squared bias $\int [E(\hat{f}(x)) - f(x)]^2 dx$ and the integrated MSE $\int E[(\hat{f}(x) - f(x))^2] dx$ for the estimator $\hat{f}(x)$ of the parameter $f(x)$ based on the Trapezoidal rule. In the figures, the blue dashed lines are the results for the NE, the red solid lines are for the HPJ, and the green dotted lines are for the IE. The horizontal axis of each figure is the length of the time series, T. Note that the results for the IE are almost insensitive to variations in T because it does not depend on T. Also, Figure 5 presents the mean of the plug-in bandwidths for each estimator, with the horizontal axis indicating T. We note that the selected values of the plug-in bandwidths are not sensitive to variation in moderately sized T values, as the selected bandwidths do not depend on T asymptotically.

As shown, the NE exhibits large biases, especially when T is small. In particular, the biases of the density and CDF estimators for $\gamma_{0,i}$ and $\rho_{1,i}$ are crucial. This is because of the presence of the incidental parameter biases and the nonlinearity biases. As a result, the integrated squared bias and the integrated MSE of the NE for each estimator are significantly larger than those of the HPJ and IE. However, the integrated squared bias and the integrated MSE of the NE become moderate as T increases, which is expected given our asymptotic investigations.

The performances of the HPJ are significantly better than the NE in that the integrated squared bias and the integrated MSE of the HPJ are close to those of the IE with moderate T. These results demonstrate that the HPJ bias correction successfully eliminates the incidental parameter biases and the second-order nonlinearity biases of the estimators.

In summary, the simulation results stress the importance of bias correction and corroborate the results of our theoretical investigations.

9 Conclusion

This paper presented novel panel data analyses to examine the degree of heterogeneity based on nonparametric kernel-smoothing estimation. The proposed procedure first computes the mean, autocovariances, and autocorrelations for each cross-sectional unit and then implements the kernel density and CDF estimation based on the estimated quantities. We showed that each kernel estimator exhibits incidental parameter bias and second-order nonlinearity bias in addition to the standard kernel-smoothing bias and higher-order nonlinearity biases. We thus proposed applying HPJ bias correction to reduce the incidental parameter bias and the second-order nonlinearity bias, and demonstrated the asymptotic properties of the bias-corrected estimator. We also proved the asymptotic validity of the cross-sectional bootstrap inferences. We illustrated the usefulness of the proposed procedure in finite samples using Monte Carlo simulation and an empirical application to the degree of heterogeneity in the price deviation of items in US cities.

Future work. The results in this paper suggest several future research topics. First, it would be desirable to develop methods to examine the degree of heterogeneity when the time series is nonstationary, such as in a unit root process. Second, we could extend our procedure for panel data with time-specific effects such as time trends. Third, we could also consider investigating the degree of finite unknown grouped heterogeneity in a model-free manner. Finally, as this paper focuses only on balanced panel data, it would be useful to develop an analysis for unbalanced panel data. Such an extension is conceptually easy because our procedure depends only on the quantities estimated for each unit, but the theoretical investigation would be rather more complicated.

A Appendix: Proofs of the theorems

This appendix collects the proofs of the theorems in the main body. The lemmas used to prove the theorems are in Appendix B. In the following, we denote generic constants by M.

A.1 Proof of Theorem 1

The density of μ_i. We evaluate each term in the following Taylor expansion:

$$\begin{aligned}
\hat{f}_{\hat{\mu}}(x) - f_\mu(x) = & \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x-\mu_i}{h}\right) - f_\mu(x) && \text{(A.1)}\\
& - \frac{1}{Nh^2}\sum_{i=1}^{N}(\hat{\mu}_i - \mu_i)K'\left(\frac{x-\mu_i}{h}\right) && \text{(A.2)}\\
& + \sum_{j=2}^{\infty}\frac{(-1)^j}{j!Nh^{j+1}}\sum_{i=1}^{N}(\hat{\mu}_i-\mu_i)^j K^{(j)}\left(\frac{x-\mu_i}{h}\right). && \text{(A.3)}
\end{aligned}$$

For (A.1), we can use the standard results for the kernel density estimation. Lemma 7 under Assumptions 1, 5, and 6 shows that

$$\frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x-\mu_i}{h}\right) - f_\mu(x) \xrightarrow{p} 0,$$

as $N \to \infty$ and $h \to 0$ with $Nh \to \infty$. Further, Lemma 7 also shows that

$$\sqrt{Nh}\left(\frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x-\mu_i}{h}\right) - f_\mu(x) - h^2\frac{\kappa_1 f''_\mu(x)}{2}\right) \xrightarrow{d} \mathcal{N}\big(0, \kappa_2 f_\mu(x)\big),$$

as $N \to \infty$ and $h \to 0$ with $Nh \to \infty$ and $Nh^5 \to C \in [0, \infty)$.

For (A.2), the mean is zero by the law of iterated expectations, since $\hat{\mu}_i - \mu_i = \bar{w}_i$ and $E(\bar{w}_i|i) = 0$. The variance is

$$var\left(\frac{1}{Nh^2}\sum_{i=1}^{N}\bar{w}_i K'\left(\frac{x-\mu_i}{h}\right)\right) = \frac{1}{Nh^4}E\left((\bar{w}_i)^2\left(K'\left(\frac{x-\mu_i}{h}\right)\right)^2\right) = O\left(\frac{1}{NTh^3}\right),$$

by Lemmas 1 and 6. Therefore, (A.2) is of order $O_p(1/\sqrt{NTh^3})$ by Markov inequality.

For the term in (A.3), the mean is

$$\begin{aligned}
&E\left(\frac{1}{j!Nh^{j+1}}\sum_{i=1}^{N}(\bar{w}_i)^j K^{(j)}\left(\frac{x-\mu_i}{h}\right)\right)\\
&=\frac{1}{j!h^{j+1}}E\left(E\left((\bar{w}_i)^j|\mu_i\right)K^{(j)}\left(\frac{x-\mu_i}{h}\right)\right)\\
&=\frac{1}{j!T^{j/2}h^j}E\left(T^{j/2}(\bar{w}_i)^j\Big|\mu_i=x\right)f_\mu(x)\int K^{(j)}(s)ds+o\left(\frac{1}{T^{j/2}h^j}\right)\\
&=(-1)^j\frac{A_{\mu,j}(x)}{\sqrt{T^jh^{2j}}}+o\left(\frac{1}{\sqrt{T^jh^{2j}}}\right),
\end{aligned}$$

by the law of iterated expectations and Lemma 6 with the definition of

$$A_{\mu,j}(x)\coloneqq\lim_{T\to\infty}\frac{(-1)^j}{j!}E\left(\sqrt{T^j}(\bar{w}_i)^j\Big|\mu_i=x\right)f_\mu(x)\int K^{(j)}(s)ds.$$

The variance is

$$var\left(\frac{1}{j!Nh^{j+1}}\sum_{i=1}^{N}(\bar{w}_i)^jK^{(j)}\left(\frac{x-\mu_i}{h}\right)\right)=O\left(\frac{1}{NT^jh^{2j+1}}\right),$$

by Lemmas 1 and 6. Thus, it holds that

$$\sum_{j=2}^{\infty}\frac{(-1)^j}{j!Nh^{j+1}}\sum_{i=1}^{N}(\hat{\mu}_i-\mu_i)^jK^{(j)}\left(\frac{x-\mu_i}{h}\right)=\sum_{j=2}^{\infty}\left(\frac{A_{\mu,j}(x)}{\sqrt{T^jh^{2j}}}+o_p\left(\frac{1}{\sqrt{T^jh^{2j}}}\right)\right).$$

Consequently, we obtain the desired result for $\hat{f}_{\hat{\mu}}(x)$ by Slutsky's theorem.

The density of $\gamma_{k,i}$. We evaluate each term in the following Taylor expansion:

$$\begin{aligned}
\hat{f}_{\hat{\gamma}_k}(x)-f_{\gamma_k}(x)=&\frac{1}{Nh}\sum_{i=1}^{N}K\left(\frac{x-\gamma_{k,i}}{h}\right)-f_{\gamma_k}(x) &\text{(A.4)}\\
&-\frac{1}{Nh^2}\sum_{i=1}^{N}(\hat{\gamma}_{k,i}-\gamma_{k,i})K'\left(\frac{x-\gamma_{k,i}}{h}\right) &\text{(A.5)}\\
&+\sum_{j=2}^{\infty}\frac{(-1)^j}{j!Nh^{j+1}}\sum_{i=1}^{N}(\hat{\gamma}_{k,i}-\gamma_{k,i})^jK^{(j)}\left(\frac{x-\gamma_{k,i}}{h}\right). &\text{(A.6)}
\end{aligned}$$

For (A.4), the consistency and asymptotic normality of the term are established by the same argument as for the density of μ_i.

For (A.5), we have the following equation based on the expansion for $\hat{\gamma}_{k,i}$:

$$\begin{aligned}
&\frac{1}{Nh^2}\sum_{i=1}^{N}(\hat{\gamma}_{k,i}-\gamma_{k,i})K'\left(\frac{x-\gamma_{k,i}}{h}\right)\\
=&\frac{1}{Nh^2}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T}w_{it}w_{i,t-k}-\gamma_{k,i}\right)K'\left(\frac{x-\gamma_{k,i}}{h}\right) \qquad (A.7)\\
&-\frac{1}{Nh^2}\sum_{i=1}^{N}\frac{T+k}{T-k}(\bar{w}_i)^2K'\left(\frac{x-\gamma_{k,i}}{h}\right) \qquad (A.8)\\
&+\frac{1}{Nh^2}\sum_{i=1}^{N}\frac{1}{T-k}\sum_{t=1}^{k}w_{it}\bar{w}_iK'\left(\frac{x-\gamma_{k,i}}{h}\right) \qquad (A.9)\\
&+\frac{1}{Nh^2}\sum_{i=1}^{N}\frac{1}{T-k}\sum_{t=T-k+1}^{T}w_{it}\bar{w}_iK'\left(\frac{x-\gamma_{k,i}}{h}\right). \qquad (A.10)
\end{aligned}$$

For (A.7), the mean is zero by the law of iterated expectations given $E(w_{it}w_{i,t-k}|i)=\gamma_{k,i}$. The variance is

$$var\left(\frac{1}{Nh^2}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T}w_{it}w_{i,t-k}-\gamma_{k,i}\right)K'\left(\frac{x-\gamma_{k,i}}{h}\right)\right)=O\left(\frac{1}{NTh^3}\right),$$

by Lemmas 3 and 6. Thus, (A.7) is of order $O_p(1/\sqrt{NTh^3})$. For (A.8), denoting $a_T(x)=E(T(\bar{w}_i)^2|\gamma_{k,i}=x)f_{\gamma_k}(x)$, the mean is expanded as

$$\begin{aligned}
E\left(\frac{1}{Nh^2}\sum_{i=1}^{N}\frac{T+k}{T-k}(\bar{w}_i)^2K'\left(\frac{x-\gamma_{k,i}}{h}\right)\right)=&\frac{T+k}{T(T-k)h^2}E\left(E\left(T(\bar{w}_i)^2|\gamma_{k,i}\right)K'\left(\frac{x-\gamma_{k,i}}{h}\right)\right)\\
=&\frac{a_T(x)}{Th}\int K'(s)ds-\frac{a_T'(x)}{T}\int sK'(s)ds+o\left(\frac{1}{T}\right)\\
=&-\frac{a_T'(x)}{T}\int sK'(s)ds+o\left(\frac{1}{T}\right)\\
=&-\frac{A_{\gamma_k,1}(x)}{T}+o\left(\frac{1}{T}\right),
\end{aligned}$$

by the law of iterated expectations, Lemma 6, and $\int K'(s)ds=0$ with the definition of

$$A_{\gamma_k,1}(x)=\lim_{T\to\infty}a_T'(x)\int sK'(s)ds.$$

The variance is

$$
\begin{aligned}
var\left(\frac{1}{Nh^2}\sum_{i=1}^{N}\frac{T+k}{T-k}(\bar{w}_i)^2 K'\left(\frac{x-\gamma_{k,i}}{h}\right)\right) &\le \frac{(T+k)^2}{Nh^4(T-k)^2}\sqrt{E\left((\bar{w}_i)^8\right)}\sqrt{E\left(\left(K'\left(\frac{x-\gamma_{k,i}}{h}\right)\right)^4\right)} \\
&= O\left(\frac{1}{Nh^4}\right)\cdot O\left(\frac{1}{T^2}\right)\cdot O\left(\sqrt{h}\right) = O\left(\frac{1}{\sqrt{N^2T^4h^7}}\right),
\end{aligned}
$$

by the Cauchy–Schwarz inequality and Lemmas 1 and 6. Thus, (A.8) is $A_{\gamma_k,1}(x)/T + o_p(1/T)$. For (A.9), the triangle inequality and the Cauchy–Schwarz inequality lead to

$$
\begin{aligned}
&E\left|\frac{1}{Nh^2}\sum_{i=1}^{N}\frac{1}{T-k}\sum_{t=1}^{k} w_{it}\bar{w}_i K'\left(\frac{x-\gamma_{k,i}}{h}\right)\right| \\
&\le \frac{1}{(T-k)h^2}E\left|\sum_{t=1}^{k} w_{it}\bar{w}_i K'\left(\frac{x-\gamma_{k,i}}{h}\right)\right| \\
&\le \frac{1}{(T-k)h^2}\sqrt{E\left(\left(\sum_{t=1}^{k} w_{it}\bar{w}_i\right)^2\right)}\sqrt{E\left(\left(K'\left(\frac{x-\gamma_{k,i}}{h}\right)\right)^2\right)} \\
&= O\left(\frac{1}{Th^2}\right)\cdot O\left(\frac{1}{\sqrt{T}}\right)\cdot O\left(\sqrt{h}\right) = O\left(\frac{1}{\sqrt{T^3h^3}}\right),
\end{aligned}
$$

by Lemmas 1 and 6. Thus, (A.9) is of order $O_p(1/\sqrt{T^3h^3})$. In the same manner, we can show that (A.10) is also of order $O_p(1/\sqrt{T^3h^3})$. These results mean that (A.5) is $A_{\gamma_k,1}(x)/T + o_p(1/T)$.

For (A.6), it is easy to see that

$$
\begin{aligned}
&\frac{1}{j!Nh^{j+1}}\sum_{i=1}^{N}(\hat{\gamma}_{k,i}-\gamma_{k,i})^j K^{(j)}\left(\frac{x-\gamma_{k,i}}{h}\right) \\
=&\frac{1}{j!Nh^{j+1}}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T}(w_{it}w_{i,t-k}-\gamma_{k,i})\right)^j K^{(j)}\left(\frac{x-\gamma_{k,i}}{h}\right) + o_p\left(\frac{1}{\sqrt{T^jh^{2j}}}\right),
\end{aligned}
$$

by the same procedures to show the order of the terms in (A.5). The mean of the term is

$$
\begin{aligned}
&E\left(\frac{1}{j!Nh^{j+1}}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T}(w_{it}w_{i,t-k}-\gamma_{k,i})\right)^j K^{(j)}\left(\frac{x-\gamma_{k,i}}{h}\right)\right) \\
&= \frac{1}{j!(T-k)^{j/2}h^{j+1}}E\left(E\left(\frac{1}{(T-k)^{j/2}}\left(\sum_{t=k+1}^{T}(w_{it}w_{i,t-k}-\gamma_{k,i})\right)^j \middle|\gamma_{k,i}\right) K^{(j)}\left(\frac{x-\gamma_{k,i}}{h}\right)\right)
\end{aligned}
$$

$$= \frac{1}{j!T^{j/2}h^j} f_{\gamma_k}(x) E\left(\frac{1}{T^{j/2}} \left(\sum_{t=k+1}^{T} (w_{it}w_{i,t-k} - \gamma_{k,i}) \right)^j \middle| \gamma_{k,i} = x \right) \int K^{(j)}(s)ds + o\left(\frac{1}{T^{j/2}h^j} \right)$$

$$= (-1)^j \frac{A_{\gamma_k,j}(x)}{\sqrt{T^j h^{2j}}} + o\left(\frac{1}{\sqrt{T^j h^{2j}}} \right),$$

by the law of iterated expectations and Lemma 6 with the definition of

$$A_{\gamma_k,j}(x) \coloneqq \lim_{T\to\infty} (-1)^j \frac{f_{\gamma_k}(x)}{j!} E\left(\frac{1}{\sqrt{T^j}} \left(\sum_{t=k+1}^{T} (w_{it}w_{i,t-k} - \gamma_{k,i}) \right)^j \middle| \gamma_{k,i} = x \right) \int K^{(j)}(s)ds.$$

The variance of the term is

$$var\left(\frac{1}{Nh^{j+1}} \sum_{i=1}^{N} \left(\frac{1}{T-k} \sum_{t=k+1}^{T} (w_{it}w_{i,t-k} - \gamma_{k,i}) \right)^j K^{(j)}\left(\frac{x-\gamma_{k,i}}{h} \right) \right) = O\left(\frac{1}{NT^j h^{2j+1}} \right),$$

by Lemmas 3 and 6. Thus, it holds that

$$\sum_{j=2}^{\infty} \frac{(-1)^j}{j!Nh^{j+1}} \sum_{i=1}^{N} (\hat{\gamma}_{k,i} - \gamma_{k,i})^j K^{(j)}\left(\frac{x-\gamma_{k,i}}{h} \right) = \sum_{j=2}^{\infty} \left(\frac{A_{\gamma_k,j}(x)}{\sqrt{T^j h^{2j}}} + o_p\left(\frac{1}{\sqrt{T^j h^{2j}}} \right) \right).$$

Consequently, we obtain the desired result for $\hat{f}_{\hat{\gamma}_k}(x)$ by Slutsky's theorem.

The density of $\rho_{k,i}$. We regard $K((x - \hat{\rho}_{k,i})/h) = K((x - \hat{\gamma}_{k,i}/\hat{\gamma}_{0,i})/h)$ as a function of two variables $(\hat{\gamma}_{k,i}, \hat{\gamma}_{0,i})$. Taylor's theorem for multivariate functions leads to

$$\hat{f}_{\hat{\rho}_k}(x) - f_{\rho_k}(x)$$

$$= \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x-\rho_{k,i}}{h} \right) - f_{\rho_k}(x) \tag{A.11}$$

$$- \frac{1}{Nh} \sum_{i=1}^{N} \sum_{j_1+j_2=1} (\hat{\gamma}_{k,i} - \gamma_{k,i})^{j_1} (\hat{\gamma}_{0,i} - \gamma_{0,i})^{j_2} \frac{\partial^{j_1+j_2}}{\partial a^{j_1} \partial b^{j_2}} K\left(\frac{x-a/b}{h} \right) \Big|_{a=\gamma_{k,i}, b=\gamma_{0,i}} \tag{A.12}$$

$$+ \sum_{j=2}^{\infty} \frac{(-1)^j}{Nh} \sum_{i=1}^{N} \sum_{j_1+j_2=j} \frac{1}{j_1!j_2!} (\hat{\gamma}_{k,i} - \gamma_{k,i})^{j_1} (\hat{\gamma}_{0,i} - \gamma_{0,i})^{j_2} \frac{\partial^{j_1+j_2}}{\partial a^{j_1} \partial b^{j_2}} K\left(\frac{x-a/b}{h} \right) \Big|_{a=\gamma_{k,i}, b=\gamma_{0,i}}. \tag{A.13}$$

We evaluate each term below.

For (A.11), the consistency and asymptotic normality of the term are established by the same argument as for the density of μ_i.

(A.12) contains two terms. Of these, we consider only

$$\frac{1}{Nh^2}\sum_{i=1}^{N}\frac{1}{\gamma_{0,i}}(\hat{\gamma}_{k,i}-\gamma_{k,i})K'\left(\frac{x-\rho_{k,i}}{h}\right),$$

because the other term can be evaluated by the same argument. However, this term is analogous to that in (A.5), so it can be evaluated by the same argument. This means that (A.12) can be written as $A_{\rho_k,1}(x)/T+o_p(1/T)$ for a nonrandom $A_{\rho_k,1}(x)$.

For (A.13), we evaluate the mean of the term

$$\sum_{j_1+j_2=j}\frac{1}{j_1!j_2!}(\hat{\gamma}_{k,i}-\gamma_{k,i})^{j_1}(\hat{\gamma}_{0,i}-\gamma_{0,i})^{j_2}\frac{\partial^{j_1+j_2}}{\partial a^{j_1}\partial b^{j_2}}K\left(\frac{x-a/b}{h}\right)\Big|_{a=\gamma_{k,i},b=\gamma_{0,i}},$$

which contains $j+1$ terms. Among these, we consider only

$$\frac{1}{j!Nh^{j+1}}\sum_{i=1}^{N}\frac{1}{\gamma_{0,i}^{j}}(\hat{\gamma}_{k,i}-\gamma_{k,i})^{j}K^{(j)}\left(\frac{x-\rho_{k,i}}{h}\right),$$

as the other terms can be evaluated in the same manner. However, this term is analogous to that in (A.6), so it can be evaluated by the same argument. This means that (A.13) can be written as

$$\begin{aligned}&\sum_{j=2}^{\infty}\frac{(-1)^j}{Nh}\sum_{i=1}^{N}\sum_{j_1+j_2=j}\frac{1}{j_1!j_2!}(\hat{\gamma}_{k,i}-\gamma_{k,i})^{j_1}(\hat{\gamma}_{0,i}-\gamma_{0,i})^{j_2}\frac{\partial^{j_1+j_2}}{\partial a^{j_1}\partial b^{j_2}}K\left(\frac{x-a/b}{h}\right)\Big|_{a=\gamma_{k,i},b=\gamma_{0,i}}\\&=\sum_{j=2}^{\infty}\left(\frac{A_{\rho_k,j}(x)}{\sqrt{T^jh^{2j}}}+o_p\left(\frac{1}{\sqrt{T^jh^{2j}}}\right)\right),\end{aligned}$$

for a nonrandom $A_{\rho_k,j}(x)$.

Consequently, we have the desired result for $\hat{f}_{\hat{\rho}_k}(x)$ by Slutsky's theorem.

□

A.2 Proof of Theorem 2

We show the proof for the density estimator of $\gamma_{k,i}$ only. Those of μ_i and $\rho_{k,i}$ are the same. The proof of Theorem 1 has shown that

$$\hat{f}_{\hat{\gamma}_k}(x) - f_{\gamma_k}(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x-\gamma_{k,i}}{h}\right) - f_{\gamma_k}(x) + \frac{A_{\gamma_k,1}(x)}{T} + \frac{A_{\gamma_k,2}(x)}{Th^2} + O_p\left(\frac{1}{\sqrt{T^3h^6}}\right).$$

This result implies that the estimators based on the half-panel data are

$$\hat{f}_{\hat{\gamma}_k,(l)}(x) - f_{\gamma_k}(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x-\gamma_{k,i}}{h}\right) - f_{\gamma_k}(x) + \frac{2A_{\gamma_k,1}(x)}{T} + \frac{2A_{\gamma_k,2}(x)}{Th^2} + O_p\left(\frac{1}{\sqrt{T^3h^6}}\right),$$

for $l = 1, 2$. As a result, the HPJ bias-corrected estimator satisfies

$$\hat{f}^{H}_{\hat{\gamma}_k}(x) - f_{\gamma_k}(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x-\gamma_{k,i}}{h}\right) - f_{\gamma_k}(x) + O_p\left(\frac{1}{\sqrt{T^3h^6}}\right).$$

Therefore, the same argument as for the term in (A.4) leads to the desired result.

□

A.3 Proof of Theorem 3

We only prove the result of the density estimation for μ_i. The proofs for $\gamma_{k,i}$ and $\rho_{k,i}$ are the same. We denote the mean and variance of the bootstrap sample by E^* and var^*, which are identical to the expectation and variance operators under the bootstrap distribution P^* when the original sample $\{y_i\}_{i=1}^N$ is fixed. We note that we can write $\hat{f}^*_{\hat{\mu}}(x) = (Nh)^{-1}\sum_{i=1}^N K((x-\hat{\mu}_i^*)/h)$ where $\hat{\mu}_i^*$ is the estimator of μ_i based on the bootstrap sample y_i^*.

If we show the Lyapunov condition for $\hat{f}^*_{\hat{\mu}}(x)$ in probability, then we have the statement. In the following, we show that

$$E^*\left(\hat{f}^*_{\hat{\mu}}(x)\right) = \hat{f}_{\hat{\mu}}(x), \qquad var^*\left(\hat{f}^*_{\hat{\mu}}(x)\right) = \frac{1}{Nh}\kappa_2 f_\mu(x) + o_p\left(\frac{1}{Nh}\right),$$

and the Lyapunov condition in probability.

For the mean, it is easy to see that

$$E^* \left(\hat{f}^*_{\hat{\mu}}(x) \right) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x - \hat{\mu}_i}{h} \right) = \hat{f}_{\hat{\mu}}(x).$$

For the variance, it holds that

$$var^* \left(\hat{f}^*_{\hat{\mu}}(x) \right) = \frac{1}{N^2h^2} \sum_{i=1}^{N} K^2 \left(\frac{x - \hat{\mu}_i}{h} \right) - \frac{1}{N} \left(\hat{f}_{\hat{\mu}}(x) \right)^2.$$

Given that $\hat{f}_{\hat{\mu}}(x) = O_p(1)$ by Theorem 1, the second term is $o_p(1/(Nh))$. For the first term, it is easy to see that

$$\frac{1}{N^2h^2} \sum_{i=1}^{N} K^2 \left(\frac{x - \hat{\mu}_i}{h} \right) = \frac{1}{N^2h^2} \sum_{i=1}^{N} K^2 \left(\frac{x - \mu_i}{h} \right) \tag{A.14}$$

$$- \frac{2}{N^2h^3} \sum_{i=1}^{N} K\left(\frac{x - \mu_i}{h} \right) K'\left(\frac{x - \mu_i}{h} \right) (\hat{\mu}_i - \mu_i) \tag{A.15}$$

$$+ o_p\left(\frac{1}{Nh} \right),$$

by Taylor's theorem. For the term in (A.14), the mean is

$$\begin{aligned} E\left(\frac{1}{N^2h^2} \sum_{i=1}^{N} K^2 \left(\frac{x - \mu_i}{h} \right) \right) &= \frac{1}{Nh^2} E\left(K^2 \left(\frac{x - \mu_i}{h} \right) \right) \\ &= \frac{1}{Nh} f_\mu(x) \int K^2(s) ds + o\left(\frac{1}{Nh} \right) \\ &= \frac{1}{Nh} \kappa_2 f_\mu(x) + o\left(\frac{1}{Nh} \right), \end{aligned}$$

by Lemma 6. The variance is

$$var\left(\frac{1}{N^2h^2} \sum_{i=1}^{N} K^2 \left(\frac{x - \mu_i}{h} \right) \right) \leq \frac{1}{N^3h^4} E\left(K^4 \left(\frac{x - \mu_i}{h} \right) \right) = O\left(\frac{1}{N^3h^3} \right) = o\left(\frac{1}{N^2h^2} \right),$$

by Lemma 6. Therefore, the term in (A.14) can be written as

$$\frac{1}{N^2h^2} \sum_{i=1}^{N} K^2 \left(\frac{x - \mu_i}{h} \right) = \frac{1}{Nh} \kappa_2 f_\mu(x) + o_p\left(\frac{1}{Nh} \right),$$

by Markov inequality. For the term in (A.15), the mean is zero and the variance is

$$var\left(\frac{2}{N^2h^3}\sum_{i=1}^{N}K\left(\frac{x-\mu_i}{h}\right)K'\left(\frac{x-\mu_i}{h}\right)(\hat{\mu}_i-\mu_i)\right)\leq\frac{M}{N^3Th^6}E\left(K'\left(\frac{x-\mu_i}{h}\right)T(\bar{w}_i)^2\right)$$
$$=O\left(\frac{1}{N^3Th^5}\right)=o\left(\frac{1}{N^3h^3}\right),$$

where we have used the boundedness of K and Lemma 6. Therefore, the term in (A.15) is $o_p(1/(Nh))$. Summing up, we have shown that

$$var^*\left(\hat{f}^*_{\hat{\mu}}(x)\right)=\frac{1}{Nh}\kappa_2 f_\mu(x)+o_p\left(\frac{1}{Nh}\right).$$

Finally, we check the Lyapunov condition:

$$\sqrt{N^3h^3}\sum_{i=1}^{N}E^*\left(\left|\frac{1}{Nh}K\left(\frac{x-\hat{\mu}_i^*}{h}\right)-\frac{1}{N}\hat{f}_{\hat{\mu}}(x)\right|^3\right)=o_p(1).$$

The term on the left-hand side is expanded as

$$E^*\left(\left|\frac{1}{Nh}K\left(\frac{x-\hat{\mu}_i^*}{h}\right)-\frac{1}{N}\hat{f}_{\hat{\mu}}(x)\right|^3\right)=\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{Nh}K\left(\frac{x-\hat{\mu}_i}{h}\right)-\frac{1}{N}\hat{f}_{\hat{\mu}}(x)\right|^3$$
$$\leq\frac{4}{N^4h^3}\sum_{i=1}^{N}\left|K\left(\frac{x-\hat{\mu}_i}{h}\right)\right|^3+\frac{4}{N^4}\left|\hat{f}_{\hat{\mu}}(x)\right|^3$$
$$=O_p\left(\frac{1}{N^3h^3}\right),$$

by Loéve's c_r inequality, the boundedness of K, and $\hat{f}_{\hat{\mu}}(x)=O_p(1)$ by Theorem 1. Accordingly, as $Nh\to\infty$, we have shown the Lyapunov condition in probability, which completes the proof.

□

A.4 Proof of Theorem 5

The CDF of μ_i. We evaluate each term in the following Taylor expansion:

$$\hat{F}_{\hat{\mu}}(x)-F_\mu(x)=\frac{1}{N}\sum_{i=1}^{N}\mathbb{K}\left(\frac{x-\mu_i}{h}\right)-F_\mu(x) \quad (A.16)$$

$$-\frac{1}{Nh}\sum_{i=1}^{N}(\hat{\mu}_i-\mu_i)K\left(\frac{x-\mu_i}{h}\right) \tag{A.17}$$

$$+\frac{1}{2Nh^2}\sum_{i=1}^{N}(\hat{\mu}_i-\mu_i)^2K'\left(\frac{x-\mu_i}{h}\right) \tag{A.18}$$

$$+\sum_{j=3}^{\infty}\frac{(-1)^j}{j!Nh^j}\sum_{i=1}^{N}(\hat{\mu}_i-\mu_i)^jK^{(j-1)}\left(\frac{x-\mu_i}{h}\right). \tag{A.19}$$

For the term in (A.16), Lemma 8 under Assumptions 1, 5, and 8 shows that

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{K}\left(\frac{x-\mu_i}{h}\right)-F_\mu(x)\xrightarrow{p}0,$$

as $N\to\infty$ and $h\to0$. Moreover, Lemma 8 also shows that

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{K}\left(\frac{x-\mu_i}{h}\right)-F_\mu(x)\right)\xrightarrow{d}\mathcal{N}\big(0,F_\mu(x)[1-F_\mu(x)]\big),$$

as $N\to\infty$ and $h\to0$ with $Nh^4\to0$.

For (A.17), the mean is zero given $\hat{\mu}_i-\mu_i=\bar{w}_i$ and $E(\bar{w}_i|i)=0$. The variance is

$$var\left(\frac{1}{Nh}\sum_{i=1}^{N}(\hat{\mu}_i-\mu_i)K\left(\frac{x-\mu_i}{h}\right)\right)=O\left(\frac{1}{NTh}\right),$$

by Lemmas 2 and 6. Thus, (A.17) is of order $O_p(1/\sqrt{NTh})$ by Markov inequality.

For (A.18), we define $c_T(x)\coloneqq E(T(\bar{w}_i)^2|\mu_i=x)f_\mu(x)$. The mean is

$$\begin{aligned}
E\left(\frac{1}{2Nh^2}\sum_{i=1}^{N}(\hat{\mu}_i-\mu_i)^2K'\left(\frac{x-\mu_i}{h}\right)\right)&=\frac{1}{2Th^2}E\left(E[T(\bar{w}_i)^2|\mu_i]K'\left(\frac{x-\mu_i}{h}\right)\right)\\
&=\frac{1}{2Th}c_T(x)\int K'(s)ds-\frac{1}{2T}c_T'(x)\int sK'(s)ds+o\left(\frac{1}{T}\right)\\
&=-\frac{1}{2T}c_T'(x)\int sK'(s)ds+o\left(\frac{1}{T}\right)\\
&=\frac{B_{\mu,2}(x)}{T}+o\left(\frac{1}{T}\right),
\end{aligned}$$

by the law of iterated expectations, Lemma 6, and $\int K'(s) = 0$ with the definition of

$$B_{\mu,2}(x) := -\lim_{T\to\infty} \frac{c_T'(x)}{2} \int sK'(s)ds.$$

The variance is

$$var\left(\frac{1}{2Nh^2}\sum_{i=1}^{N}(\hat{\mu}_i-\mu_i)^2 K'\left(\frac{x-\mu_i}{h}\right)\right) = O\left(\frac{1}{NT^2h^3}\right),$$

by Lemmas 1 and 6. Thus, (A.18) can be written as $B_{\mu,2}(x)/T + o_p(1/T)$.

For the term in (A.19), the mean is

$$\begin{aligned}
&E\left(\frac{1}{j!Nh^j}\sum_{i=1}^{N}(\hat{\mu}_i-\mu_i)^j K^{(j-1)}\left(\frac{x-\mu_i}{h}\right)\right) \\
&= \frac{1}{j!T^{j/2}h^j} E\left(E\left(T^{j/2}(\bar{w}_i)^j \middle| \mu_i\right) K^{(j-1)}\left(\frac{x-\mu_i}{h}\right)\right) \\
&= \frac{1}{j!\sqrt{T^j h^{2j-2}}} E\left(\sqrt{T^j}(\bar{w}_i)^j \middle| \mu_i = x\right) f_\mu(x) \int K^{(j-1)}(s)ds + o\left(\frac{1}{\sqrt{T^j h^{2j-2}}}\right) \\
&= (-1)^j \frac{B_{\mu,j}(x)}{T^j h^{2j-2}} + o\left(\frac{1}{\sqrt{T^j h^{2j-2}}}\right),
\end{aligned}$$

by the law of iterated expectations and Lemmas 1 and 6 with the definition of

$$B_{\mu,j}(x) := \lim_{T\to\infty} \frac{(-1)^j}{j!} E\left(\sqrt{T^j}(\bar{w}_i)^j \middle| \mu_i = x\right) f_\mu(x) \int K^{(j-1)}(s)ds.$$

The variance is

$$var\left(\frac{1}{j!Nh^j}\sum_{i=1}^{N}(\hat{\mu}_i-\mu_i)^j K^{(j-1)}\left(\frac{x-\mu_i}{h}\right)\right) = O\left(\frac{1}{NT^j h^{2j-1}}\right),$$

by Lemmas 2 and 6. Thus, (A.19) can be written as

$$\sum_{j=3}^{\infty} \frac{(-1)^j}{j!Nh^j}\sum_{i=1}^{N}(\hat{\mu}_i-\mu_i)^j K^{(j-1)}\left(\frac{x-\mu_i}{h}\right) = \sum_{j=3}^{\infty}\left(\frac{B_{\mu,j}(x)}{T^j h^{2j-2}} + o_p\left(\frac{1}{\sqrt{T^j h^{2j-2}}}\right)\right).$$

Consequently, we get the desired result for $\hat{F}_{\hat{\mu}}(x)$ by Slutsky's theorem.

The CDF of $\gamma_{k,i}$. We evaluate each term in the following Taylor expansion.

$$\hat{F}_{\hat{\gamma}_k}(x) - F_{\gamma_k}(x) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{K}\left(\frac{x-\gamma_{k,i}}{h}\right) - F_{\gamma_k}(x) \tag{A.20}$$

$$- \frac{1}{Nh}\sum_{i=1}^{N}(\hat{\gamma}_{k,i}-\gamma_{k,i})K\left(\frac{x-\gamma_{k,i}}{h}\right) \tag{A.21}$$

$$+ \frac{1}{2Nh^2}\sum_{i=1}^{N}(\hat{\gamma}_{k,i}-\gamma_{k,i})^2K'\left(\frac{x-\gamma_{k,i}}{h}\right) \tag{A.22}$$

$$+ \sum_{j=3}^{\infty}\frac{(-1)^j}{j!Nh^j}\sum_{i=1}^{N}(\hat{\gamma}_{k,i}-\gamma_{k,i})^jK^{(j-1)}\left(\frac{x-\gamma_{k,i}}{h}\right). \tag{A.23}$$

For (A.20), the consistency and asymptotic normality are established by the same arguments as for the CDF of μ_i.

For (A.21), we have the following equation by the expansion for $\hat{\gamma}_{k,i}$:

$$\frac{1}{Nh}\sum_{i=1}^{N}(\hat{\gamma}_{k,i}-\gamma_{k,i})K\left(\frac{x-\gamma_{k,i}}{h}\right) = \frac{1}{Nh}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T}w_{it}w_{i,t-k}-\gamma_{k,i}\right)K\left(\frac{x-\gamma_{k,i}}{h}\right) \tag{A.24}$$

$$- \frac{1}{Nh}\sum_{i=1}^{N}\frac{T+k}{T-k}(\bar{w}_i)^2K\left(\frac{x-\gamma_{k,i}}{h}\right) \tag{A.25}$$

$$+ \frac{1}{Nh}\sum_{i=1}^{N}\frac{1}{T-k}\sum_{t=1}^{k}w_{it}\bar{w}_iK\left(\frac{x-\gamma_{k,i}}{h}\right) \tag{A.26}$$

$$+ \frac{1}{Nh}\sum_{i=1}^{N}\frac{1}{T-k}\sum_{t=T-k+1}^{T}w_{it}\bar{w}_iK\left(\frac{x-\gamma_{k,i}}{h}\right). \tag{A.27}$$

For the term in (A.24), the mean is zero and the variance is

$$var\left(\frac{1}{Nh}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T}w_{it}w_{i,t-k}-\gamma_{k,i}\right)K\left(\frac{x-\gamma_{k,i}}{h}\right)\right) = O\left(\frac{1}{NTh}\right),$$

by Lemmas 3 and 6. Thus, (A.24) is of order $O_p(1/\sqrt{NTh})$. For (A.25), the mean is

$$E\left(\frac{1}{Nh}\frac{T+k}{T-k}\sum_{i=1}^{N}(\bar{w}_i)^2K\left(\frac{x-\gamma_{k,i}}{h}\right)\right) = \frac{T+k}{hT(T-k)}E\left(E[T(\bar{w}_i)^2|\gamma_{k,i}]K\left(\frac{x-\gamma_{k,i}}{h}\right)\right)$$
$$= \frac{1}{T}E(T(\bar{w}_i)^2|\gamma_{k,i}=x)f_{\gamma_k}(x)\int K(s)ds + o\left(\frac{1}{T}\right)$$

$$= -\frac{B_{\gamma_k,1}(x)}{T} + o\left(\frac{1}{T}\right),$$

by the law of iterated expectations and Lemma 6 with the definition of

$$B_{\gamma_k,1}(x) := -\lim_{T\to\infty} T \cdot E((\bar{w}_i)^2|\gamma_{k,i} = x) f_{\gamma_k}(x) \int K(s)ds.$$

The variance is

$$var\left(\frac{1}{Nh}\frac{T+k}{T-k}\sum_{i=1}^{N}(\bar{w}_i)^2 K\left(\frac{x-\gamma_{k,i}}{h}\right)\right) = O\left(\frac{1}{NT^2h}\right),$$

by Lemmas 1 and 6. Thus, (A.25) is $B_{\gamma_k,1}(x)/T + o_p(1/T)$. For (A.26), the absolute mean is

$$\begin{aligned}
E\left|\frac{1}{Nh}\sum_{i=1}^{N}\frac{1}{T-k}\sum_{t=1}^{k} w_{it}\bar{w}_i K\left(\frac{x-\gamma_{k,i}}{h}\right)\right| &\leq \frac{1}{(T-k)h}\sqrt{E\left(\left(\sum_{t=1}^{k} w_{it}\bar{w}_i\right)^2\right)}\sqrt{E\left(K^2\left(\frac{x-\gamma_{k,i}}{h}\right)\right)} \\
&= O\left(\frac{1}{Th}\right)\cdot O\left(\frac{1}{\sqrt{T}}\right)\cdot O\left(\sqrt{h}\right) \\
&= O\left(\frac{1}{\sqrt{T^3h}}\right),
\end{aligned}$$

by Lemmas 1 and 6. Thus, (A.26) is of order $O_p(1/\sqrt{T^3h})$. For (A.27), we can show that it is of order $O_p(1/\sqrt{T^3h})$ by the same argument. Thus, (A.21) is $B_{\gamma_k,1}(x)/T + o_p(1/T)$.

For (A.22), it is easy to see that

$$\begin{aligned}
&\frac{1}{2Nh^2}\sum_{i=1}^{N}(\hat{\gamma}_{k,i}-\gamma_{k,i})^2 K'\left(\frac{x-\gamma_{k,i}}{h}\right) \\
&= \frac{1}{2Nh^2}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T} w_{it}w_{i,t-k}-\gamma_{k,i}\right)^2 K'\left(\frac{x-\gamma_{k,i}}{h}\right) + o_p\left(\frac{1}{T}\right),
\end{aligned}$$

by similar procedures, to show the orders of (A.24), (A.25), (A.26), and (A.27). Introducing the shorthand notation $d_T(x) := E[(\sum_{t=k+1}^{T}[w_{it}w_{i,t-k}-\gamma_{k,i}])^2|\gamma_{k,i} = x] f_{\gamma_k}(x)/(T-k)$, the mean of the term is

$$E\left(\frac{1}{2Nh^2}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T} w_{it}w_{i,t-k}-\gamma_{k,i}\right)^2 K'\left(\frac{x-\gamma_{k,i}}{h}\right)\right)$$

$$
\begin{aligned}
=&\frac{1}{2(T-k)h^2}E\left(E\left(\frac{1}{T-k}\left(\sum_{t=k+1}^{T}w_{it}w_{i,t-k}-\gamma_{k,i}\right)^2\middle|\gamma_{k,i}\right)K'\left(\frac{x-\gamma_{k,i}}{h}\right)\right)\\
=&\frac{1}{2(T-k)h}d_T(x)\int K'(s)ds-\frac{1}{2(T-k)}d'_T(x)\int sK'(s)ds+o\left(\frac{1}{T}\right)\\
=&-\frac{1}{2T}d'_T(x)\int sK'(s)ds+o\left(\frac{1}{T}\right)\\
=&\frac{B_{\gamma_k,2}(x)}{T}+o\left(\frac{1}{T}\right),
\end{aligned}
$$

by Lemma 6 and $\int K'(s)ds=0$ with the definition of

$$
B_{\gamma_k,2}(x):=-\lim_{T\to\infty}\frac{d'_T(x)}{2}\int sK'(s)ds.
$$

The variance of the term is

$$
var\left(\frac{1}{2Nh^2}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T}w_{it}w_{i,t-k}-\gamma_{k,i}\right)^2K'\left(\frac{x-\gamma_{k,i}}{h}\right)\right)=O\left(\frac{1}{NT^2h^3}\right),
$$

by Lemmas 3 and 6. Thus, (A.22) is $B_{\gamma_k,2}(x)/T+o_p(1/T)$.

For (A.23), it is easy to see that

$$
\begin{aligned}
&\frac{1}{j!Nh^j}\sum_{i=1}^{N}(\hat{\gamma}_{k,i}-\gamma_{k,i})^jK^{(j-1)}\left(\frac{x-\gamma_{k,i}}{h}\right)\\
&=\frac{1}{j!Nh^j}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T}w_{it}w_{i,t-k}-\gamma_{k,i}\right)^jK^{(j-1)}\left(\frac{x-\gamma_{k,i}}{h}\right)+o_p\left(\frac{1}{\sqrt{T^jh^{2j-2}}}\right),
\end{aligned}
$$

by the same argument as for (A.22). The mean of the term is

$$
\begin{aligned}
&E\left(\frac{1}{j!Nh^j}\sum_{i=1}^{N}\left(\frac{1}{T-k}\sum_{t=k+1}^{T}w_{it}w_{i,t-k}-\gamma_{k,i}\right)^jK^{(j-1)}\left(\frac{x-\gamma_{k,i}}{h}\right)\right)\\
&=\frac{1}{j!(T-k)^{j/2}h^j}E\left(\frac{1}{(T-k)^{j/2}}E\left(\left(\sum_{t=k+1}^{T}(w_{it}w_{i,t-k}-\gamma_{k,i})\right)^j\middle|\gamma_{k,i}\right)K^{(j-1)}\left(\frac{x-\gamma_{k,i}}{h}\right)\right)\\
&=\frac{1}{j!\sqrt{T^jh^{2j-2}}}\left(\frac{1}{\sqrt{T^j}}E\left(\left(\sum_{t=k+1}^{T}(w_{it}w_{i,t-k}-\gamma_{k,i})\right)^j\middle|\gamma_{k,i}=x\right)f_{\gamma_k}(x)\right)\int K^{(j-1)}(s)ds+o_p\left(\frac{1}{\sqrt{T^jh^{2j-2}}}\right)\\
&=(-1)^j\frac{B_{\gamma_k,j}(x)}{\sqrt{T^jh^{2j-2}}}+o_p\left(\frac{1}{\sqrt{T^jh^{2j-2}}}\right),
\end{aligned}
$$

by Lemma 6 with the definition of

$$B_{\gamma_k,j}(x) := \lim_{T\to\infty} \frac{(-1)^j}{j!} E\left(\frac{1}{\sqrt{T^j}} \left(\sum_{t=k+1}^{T} (w_{it}w_{i,t-k} - \gamma_{k,i}) \right)^j \middle| \gamma_{k,i} = x \right) f_{\gamma_k}(x) \int K^{(j-1)}(s)ds.$$

The variance is

$$var\left(\frac{1}{j!Nh^j} \sum_{i=1}^{N} \left(\frac{1}{T-k} \sum_{t=k+1}^{T} w_{it}w_{i,t-k} - \gamma_{k,i} \right)^j K^{(j-1)}\left(\frac{x-\gamma_{k,i}}{h} \right) \right) = O\left(\frac{1}{NT^j h^{2j-1}} \right),$$

by Lemmas 4 and 6. Thus, (A.23) can be written as

$$\sum_{j=3}^{\infty} \frac{(-1)^j}{j!Nh^j} \sum_{i=1}^{N} (\hat{\gamma}_{k,i} - \gamma_{k,i})^j K^{(j-1)}\left(\frac{x-\gamma_{k,i}}{h} \right) = \sum_{j=3}^{\infty} \left(\frac{B_{\gamma_k,j}(x)}{\sqrt{T^j h^{2j-2}}} + o_p\left(\frac{1}{\sqrt{T^j h^{2j-2}}} \right) \right).$$

Consequently, we obtain the desired result for $\hat{F}_{\hat{\gamma}_k}(x)$ by Slutsky's theorem.

The CDF of $\rho_{k,i}$. We regard $\mathbb{K}((x-\hat{\rho}_{k,i})/h) = \mathbb{K}((x-\hat{\gamma}_{k,i}/\hat{\gamma}_{0,i})/h)$ as a function of two variables $(\hat{\gamma}_{k,i}, \hat{\gamma}_{0,i})$. Taylor's theorem for multivariate functions leads to

$$\begin{aligned}
&\hat{F}_{\hat{\rho}_k}(x) - F_{\rho_k}(x) \\
=&\frac{1}{N}\sum_{i=1}^{N} \mathbb{K}\left(\frac{x-\rho_{k,i}}{h} \right) - F_{\rho_k}(x) && \text{(A.28)} \\
&+ \frac{1}{N}\sum_{i=1}^{N} \sum_{j_1+j_2=1} (\hat{\gamma}_{k,i} - \gamma_{k,i})^{j_1} (\hat{\gamma}_{0,i} - \gamma_{0,i})^{j_2} \frac{\partial^{j_1+j_2}}{\partial a^{j_1} \partial b^{j_2}} \mathbb{K}\left(\frac{x-a/b}{h} \right) \Big|_{a=\gamma_{k,i}, b=\gamma_{0,i}} && \text{(A.29)} \\
&+ \frac{1}{N}\sum_{i=1}^{N} \sum_{j_1+j_2=2} \frac{1}{j_1!j_2!} (\hat{\gamma}_{k,i} - \gamma_{k,i})^{j_1} (\hat{\gamma}_{0,i} - \gamma_{0,i})^{j_2} \frac{\partial^{j_1+j_2}}{\partial a^{j_1} \partial b^{j_2}} \mathbb{K}\left(\frac{x-a/b}{h} \right) \Big|_{a=\gamma_{k,i}, b=\gamma_{0,i}} && \text{(A.30)} \\
&+ \sum_{j=3}^{\infty} \frac{1}{N}\sum_{i=1}^{N} \sum_{j_1+j_2=j} \frac{1}{j_1!j_2!} (\hat{\gamma}_{k,i} - \gamma_{k,i})^{j_1} (\hat{\gamma}_{0,i} - \gamma_{0,i})^{j_2} \frac{\partial^{j_1+j_2}}{\partial a^{j_1} \partial b^{j_2}} \mathbb{K}\left(\frac{x-a/b}{h} \right) \Big|_{a=\gamma_{k,i}, b=\gamma_{0,i}}. && \text{(A.31)}
\end{aligned}$$

For (A.28), the consistency and asymptotic normality of the term are established by the same argument as for the CDF of μ_i.

(A.29) contains two terms. Of these, we focus only on

$$\frac{1}{Nh}\sum_{i=1}^{N}\frac{1}{\gamma_{0,i}}(\hat{\gamma}_{k,i}-\gamma_{k,i})K\left(\frac{x-\rho_{k,i}}{h}\right),$$

as the other term can be evaluated in the same manner. However, this term is analogous to that in (A.21), so it can be evaluated by the same argument. This means that (A.29) can be written as $B_{\rho_k,1}(x)+o_p(1/T)$ for a nonrandom $B_{\rho_k,1}(x)$.

(A.30) contains three terms. Of these, we focus only on

$$\frac{1}{2Nh^2}\sum_{i=1}^{N}\frac{1}{\gamma_{0,i}^2}(\hat{\gamma}_{k,i}-\gamma_{k,i})^2K'\left(\frac{x-\rho_{k,i}}{h}\right),$$

as the other terms can be evaluated in the same manner. However, this term is analogous to that in (A.22), so it can be evaluated by the same argument. This means that (A.30) is also $B_{\rho_k,2}(x)/T+o_p(1/T)$ for a nonrandom $B_{\rho_k,2}(x)$.

For (A.31), we evaluate the mean of the term

$$\sum_{j_1+j_2=j}\frac{1}{j_1!j_2!}(\hat{\gamma}_{k,i}-\gamma_{k,i})^{j_1}(\hat{\gamma}_{0,i}-\gamma_{0,i})^{j_2}\frac{\partial^{j_1+j_2}}{\partial a^{j_1}\partial b^{j_2}}\mathbb{K}\left(\frac{x-a/b}{h}\right)\Big|_{a=\gamma_{k,i},b=\gamma_{0,i}},$$

which contains $j+1$ terms. Of these terms, we consider only

$$\frac{1}{j!Nh^j}\sum_{i=1}^{N}\frac{1}{\gamma_{0,i}^j}(\hat{\gamma}_{k,i}-\gamma_{k,i})^jK^{(j-1)}\left(\frac{x-\rho_{k,i}}{h}\right),$$

as the other terms can be evaluated in the same manner. However, this term is analogous to that in (A.23), so we can evaluate it using the same argument. This means that (A.31) can be written as

$$\begin{aligned}&\sum_{j=3}^{\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{j_1+j_2=j}\frac{1}{j_1!j_2!}(\hat{\gamma}_{k,i}-\gamma_{k,i})^{j_1}(\hat{\gamma}_{0,i}-\gamma_{0,i})^{j_2}\frac{\partial^{j_1+j_2}}{\partial a^{j_1}\partial b^{j_2}}\mathbb{K}\left(\frac{x-a/b}{h}\right)\Big|_{a=\gamma_{k,i},b=\gamma_{0,i}}\\&=\sum_{j=3}^{\infty}\left(\frac{B_{\rho_k,j}(x)}{\sqrt{T^jh^{2j-2}}}+o\left(\frac{1}{\sqrt{T^jh^{2j-2}}}\right)\right),\end{aligned}$$

for a nonrandom $B_{\rho_k,j}(x)$.

Consequently, we have the desired result for $\hat{F}_{\hat{\rho}_k}(x)$ by Slutsky's theorem.

□

A.5 Proof of Theorem 6

We only show the proof for the CDF estimator of $\hat{\gamma}_{k,i}$, as those for $\hat{\mu}_i$ and $\hat{\rho}_{k,i}$ are identical. The proof of Theorem 5 has shown that

$$\hat{F}_{\hat{\gamma}_k}(x) - F_{\gamma_k}(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{K}\left(\frac{x-\gamma_{k,i}}{h}\right) - F_{\gamma_k}(x) + \frac{B_{\gamma_k,1}(x)}{T} + \frac{B_{\gamma_k,2}(x)}{T} + O_p\left(\frac{1}{\sqrt{T^3h^4}}\right).$$

This result implies that for $l = 1, 2$,

$$\hat{F}_{\hat{\gamma}_k,(l)}(x) - F_{\gamma_k}(x) = \frac{1}{Nh}\sum_{i=1}^{N} \mathbb{K}\left(\frac{x-\gamma_{k,i}}{h}\right) - F_{\gamma_k}(x) + \frac{2B_{\gamma_k,1}(x)}{T} + \frac{2B_{\gamma_k,2}(x)}{T} + O_p\left(\frac{1}{\sqrt{T^3h^4}}\right).$$

As a result, the HPJ bias-corrected estimator satisfies

$$\hat{F}^H_{\hat{\gamma}_k}(x) - F_{\gamma_k}(x) = \frac{1}{Nh}\sum_{i=1}^{N} \mathbb{K}\left(\frac{x-\gamma_{k,i}}{h}\right) - F_{\gamma_k}(x) + O_p\left(\frac{1}{\sqrt{T^3h^4}}\right).$$

Therefore, the same argument for the term in (A.20) leads to the desired result.

□

A.6 Proof of Theorem 7

We only show the result for $\hat{\mu}_i$, as that for $\hat{\gamma}_{k,i}$ or $\hat{\rho}_{k,i}$ can be shown in a similar manner. We note that we can write $\hat{F}^*_{\hat{\mu}}(x) = N^{-1}\sum_{i=1}^{N} \mathbb{K}((x-\hat{\mu}^*_i)/h)$ where $\hat{\mu}^*_i$ is the estimator of μ_i based on the bootstrap sample y^*_i.

As in the proof of Theorem 3, we obtain the statement if we show the Lyapunov condition in probability. As below, we show that

$$E^*\left(\hat{F}^*_{\hat{\mu}}(x)\right) = \hat{F}_{\hat{\mu}}(x), \qquad var^*\left(\hat{F}^*_{\hat{\mu}}(x)\right) = \frac{1}{N}F_{\mu}(x)[1-F_{\mu}(x)] + o_p\left(\frac{1}{N}\right),$$

and the Lyapunov condition in probability.

For the mean, it holds that

$$E^* \left(\hat{F}^*_{\hat{\mu}}(x) \right) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{K} \left(\frac{x - \hat{\mu}_i}{h} \right) = \hat{F}_{\hat{\mu}}(x).$$

For the variance, we observe that

$$var^* \left(\hat{F}^*_{\hat{\mu}}(x) \right) = \frac{1}{N^2} \sum_{i=1}^{N} \mathbb{K}^2 \left(\frac{x - \hat{\mu}_i}{h} \right) - \frac{1}{N} \left(\hat{F}_{\hat{\mu}}(x) \right)^2.$$

For the second term, we have $(\hat{F}_{\hat{\mu}}(x))^2/N = (F_\mu(x))^2/N + o_p(1/N)$ by Theorem 5. For the first term, Taylor's theorem leads to

$$\frac{1}{N^2} \sum_{i=1}^{N} \mathbb{K}^2 \left(\frac{x - \hat{\mu}_i}{h} \right) = \frac{1}{N^2} \sum_{i=1}^{N} \mathbb{K}^2 \left(\frac{x - \mu_i}{h} \right) \tag{A.32}$$

$$- \frac{2}{N^2 h} \sum_{i=1}^{N} \mathbb{K} \left(\frac{x - \mu_i}{h} \right) K \left(\frac{x - \mu_i}{h} \right) (\hat{\mu}_i - \mu_i) \tag{A.33}$$

$$+ o_p \left(\frac{1}{N} \right).$$

For the term in (A.32), the mean is

$$\begin{aligned} E \left(\frac{1}{N^2} \sum_{i=1}^{N} \mathbb{K}^2 \left(\frac{x - \mu_i}{h} \right) \right) &= \frac{1}{N} \int \mathbb{K}^2 \left(\frac{x - z}{h} \right) f_\mu(z) dz \\ &= \frac{1}{N} \int \mathbb{K}^2(s) f_\mu(x - sh) ds \\ &= \frac{1}{N} \left(\left[-\mathbb{K}^2(s) F_\mu(x - sh) \right]_{-\infty}^{\infty} + 2 \int \mathbb{K}(s) K(s) F_\mu(x - sh) ds \right) \\ &= \frac{2}{N} \int \mathbb{K}(s) K(s) F_\mu(x - sh) ds \\ &= \frac{2}{N} \int \mathbb{K}(s) K(s) \left(F_\mu(x) - f_\mu(x) sh + \frac{1}{2} f'_\mu(\tilde{x}) s^2 h^2 \right) ds \\ &= \frac{1}{N} F_\mu(x) + o \left(\frac{1}{N} \right), \end{aligned}$$

where we have used integration by parts, the property of the CDF, Taylor's theorem with $\tilde{x}$ between x and $x - sh$, and $2 \int \mathbb{K}(s) K(s) ds = \int d\mathbb{K}^2(s) = \mathbb{K}^2(\infty) - \mathbb{K}^2(-\infty) = 1$. The variance of the term is

$$var\left(\frac{1}{N^2}\sum_{i=1}^{N}\mathbb{K}^2\left(\frac{x-\mu_i}{h}\right)\right)=\frac{1}{N^3}var\left(\mathbb{K}^2\left(\frac{x-\mu_i}{h}\right)\right)=O\left(\frac{1}{N^3}\right)=o\left(\frac{1}{N^2}\right),$$

by the fact that $0\leq\mathbb{K}\leq 1$. Thus, (A.32) is $F_\mu(x)/N+o_p(1/N)$ by Markov inequality. For (A.33), the absolute mean is

$$\begin{aligned}E\left|\frac{2}{N^2h}\sum_{i=1}^{N}\mathbb{K}\left(\frac{x-\mu_i}{h}\right)K\left(\frac{x-\mu_i}{h}\right)(\hat{\mu}_i-\mu_i)\right|\leq&\frac{2}{Nh}E\left|K\left(\frac{x-\mu_i}{h}\right)(\hat{\mu}_i-\mu_i)\right|\\ \leq&\frac{2}{Nh}\sqrt{E\left(K^2\left(\frac{x-\mu_i}{h}\right)\right)}\sqrt{E\left((\hat{\mu}_i-\mu_i)^2\right)}\\ =&O\left(\frac{1}{N\sqrt{Th}}\right)=o\left(\frac{1}{N}\right),\end{aligned}$$

where we have used the triangle inequality, $0\leq\mathbb{K}\leq 1$, the Cauchy–Schwarz inequality, and Lemma 6. Thus, (A.33) is of order $o_p(1/N)$ by Markov inequality. To sum up, we have shown that

$$var^*\left(\hat{F}^*_{\hat{\mu}}(x)\right)=\frac{1}{N}F_\mu(x)[1-F_\mu(x)]+o_p\left(\frac{1}{N}\right).$$

Finally, we show the following Lyapunov condition in probability:

$$\sqrt{N^3}\sum_{i=1}^{N}E^*\left|\frac{1}{N}\mathbb{K}\left(\frac{x-\hat{\mu}^*_i}{h}\right)-\frac{1}{N}\hat{F}_{\hat{\mu}}(x)\right|^3\xrightarrow{p}0.$$

The left-hand side can be written as

$$\frac{1}{\sqrt{N}}E^*\left|\mathbb{K}\left(\frac{x-\hat{\mu}^*_i}{h}\right)-\hat{F}_{\hat{\mu}}(x)\right|^3\leq\frac{4}{\sqrt{N}}E^*\left(\mathbb{K}^3\left(\frac{x-\hat{\mu}^*_i}{h}\right)\right)+\frac{4}{\sqrt{N}}\left|\hat{F}_{\hat{\mu}}(x)\right|^3,$$

by Loéve's c_r inequality and the fact that $0\leq\mathbb{K}\leq 1$. For the second term, Theorem 5 and the continuous mapping theorem imply that $4/\sqrt{N}|\hat{F}_{\hat{\mu}}(x)|^3\xrightarrow{p}0$. For the first term, it holds that

$$\frac{4}{\sqrt{N}}E^*\left(\mathbb{K}^3\left(\frac{x-\hat{\mu}^*_i}{h}\right)\right)=\frac{4}{\sqrt{N}}\frac{1}{N}\sum_{i=1}^{N}\mathbb{K}^3\left(\frac{x-\hat{\mu}_i}{h}\right)=O_p\left(\frac{1}{\sqrt{N}}\right)=o_p(1),$$

where we have used the fact that $0\leq\mathbb{K}\leq 1$. We thus have shown the Lyapunov condition in probability, which completes the proof.

□

B Appendix: Lemmas

This appendix contains the technical lemmas used to demonstrate the theorems in the main body.

We first present the lemmas for which the proofs are given in Okui and Yanagi (2017).

Lemma 1. *Suppose that Assumptions 1, 2, and 3 hold for $r_m = r$ and $r_d = r$ with a natural number r. Then, it holds that $E((\bar{w}_i)^r) = O(T^{-r/2})$.*

Lemma 2. *Suppose that Assumptions 1, 2, and 3 hold for $r_m = r$ and $r_d = r$ with a natural number r. Then, it holds that $E\left((\hat{\mu}_i - \mu_i)^r\right) = O(T^{-r/2})$.*

Lemma 3. *Suppose that Assumptions 1, 2, and 3 hold for $r_m = r$ and $r_d = 2r$ with a natural number r. Then, it holds that $E((\sum_{t=k+1}^{T}(w_{it}w_{i,t-k} - \gamma_{k,i}))^r) = O(T^{r/2})$.*

Lemma 4. *Suppose that Assumptions 1, 2, and 3 hold for $r_m = 2r$ and $r_d = 2r$ with a natural number r. Then, it holds that $E((\hat{\gamma}_{k,i} - \gamma_{k,i})^r) = O(T^{-r/2})$.*

Lemma 5. *Suppose that Assumptions 1, 2, 3, and 4 hold for $r_m = 2r$ and $r_d = 2r$ with a natural number r. Then, it holds that $E((\hat{\rho}_{k,i} - \rho_{k,i})^r) = O(T^{-r/2})$.*

We repeatedly use the following lemmas to prove our theorems. The proofs are similar to those in Pagan and Ullah (1999) and Li and Racine (2007), and are omitted.

Lemma 6. *Consider a continuous random variable $X \in \mathbb{R}$, a random vector $Y = (Y_1, Y_2, \dots, Y_d)^\top \in \mathbb{R}^d$, and an interior point $x \in \mathbb{R}$. Suppose that a function $g_1 : \mathbb{R} \to \mathbb{R}$ satisfies $\int |g_1(s)|ds < \infty$, $\int |sg_1(s)|ds < \infty$, and $\int |s^2 g_1(s)|ds < \infty$, and that $E[g_2(X,Y)|X = \cdot] : \mathbb{R} \to \mathbb{R}$ and the density $f_X : \mathbb{R} \to \mathbb{R}$ are twice boundedly continuously differentiable at x. It holds that*

$$E\left(g_1\left(\frac{x-X}{h}\right)g_2(X,Y)\right) = hA(x)\int g_1(s)ds - h^2 A'(x)\int sg_1(s)ds + o(h^2)$$
$$= O(h) + O(h^2) + o(h^2),$$

where $A(x) \coloneqq E[g_2(X,Y)|X = x]f_X(x)$.

We note that the above result implies that

$$E\left(g_1\left(\frac{x-X}{h}\right)\right) = hf_X(x)\int g_1(s)ds - h^2 f'_X(x)\int sg_1(s)ds + o(h^2)$$
$$= O(h) + O(h^2) + o(h^2),$$

if we set $g_2(x,y) = 1$ (constant).

Suppose that $\{X_i\}_{i=1}^N$ is a random sample of a continuous random variable $X \in \mathbb{R}$. We denote the density and CDF of X by $f_X(\cdot)$ and $F_X(\cdot) = \Pr(X \le \cdot)$, respectively.

Lemma 7. *Let $\hat{f}_X(x) \coloneqq (Nh)^{-1}\sum_{i=1}^N K((x-X_i)/h)$ be the kernel density estimator. Let $x \in \mathbb{R}$ be a fixed interior point in the support of X. Suppose that the kernel function $K : \mathbb{R} \to \mathbb{R}$ is symmetric and satisfies $\int K(s)ds = 1$, $\kappa_1 = \int s^2 K(s)ds < \infty$, $\kappa_2 = \int K^2(s)ds < \infty$, and $\int |s^3 K(s)|ds < \infty$, and that f_X is bounded away from zero and three-times boundedly continuously differentiable near x. When $N \to \infty$ and $h \to 0$ with $Nh \to \infty$, it holds that*

$$E\big(\hat{f}_X(x)\big) = f_X(x) + \frac{h^2}{2}\kappa_1 f''_X(x) + o(h^2), \qquad var\big(\hat{f}_X(x)\big) = \frac{1}{Nh}\kappa_2 f_X(x) + o\left(\frac{1}{Nh}\right).$$

Moreover, when $\int |K(s)|^3 ds < \infty$ and $Nh^5 \to C \in [0,\infty)$, it holds that

$$\sqrt{Nh}\left(\hat{f}_X(x) - f_X(x) - h^2\frac{\kappa_1 f''_X(x)}{2}\right) \xrightarrow{d} \mathcal{N}\big(0, \kappa_2 f_X(x)\big).$$

Lemma 8. *Let $\hat{F}_X(x) \coloneqq N^{-1}\sum_{i=1}^N \mathbb{K}((x-X_i)/h)$ be the kernel CDF estimator. Let $x \in \mathbb{R}$ be a fixed interior point in the support of X. Let $K(s) = d\mathbb{K}(s)/ds$ be the derivative. Suppose that $K : \mathbb{R} \to \mathbb{R}$ is symmetric and satisfies $\int K(s)ds = 1$, $\kappa_1 = \int s^2 K(s)ds < \infty$, $\int |s^3 K(s)|ds < \infty$, and $\int |sK(s)|\mathbb{K}(s)ds < \infty$, and that F_X is three-times boundedly continuously differentiable near x. When $N \to \infty$ and $h \to 0$, then*

$$E\left(\hat{F}_X(x)\right) = F_X(x) + \frac{h^2}{2} f'_X(x)\kappa_1 + o(h^2), \qquad var\left(\hat{F}_X(x)\right) = \frac{1}{N}F_X(x)[1-F_X(x)] + o\left(\frac{1}{N}\right).$$

Moreover, when $Nh^4 \to 0$ also holds, it holds that

$$\sqrt{N}\left(\hat{F}_X(x) - F_X(x)\right) \xrightarrow{d} \mathcal{N}\big(0, F_X(x)[1-F_X(x)]\big).$$

C Appendix: The validity of the infinite-order Taylor expansion

This appendix discusses the validity of the infinite-order Taylor expansion for the density estimation in Theorem 1. The discussion for the expansion of the CDF estimation in Theorem 5 is similar, and we omit it.

The infinite-order Taylor expansion of $\hat{f}_{\hat{\xi}}(x) = (Nh)^{-1} \sum_{i=1}^{N} K((x-\hat{\xi}_i)/h)$ is

$$\hat{f}_{\hat{\xi}}(x) = \frac{1}{Nh} \sum_{i=1}^{N} \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \frac{(\hat{\xi}_i - \xi_i)^j}{h^j} K^{(j)} \left(\frac{x - \xi_i}{h} \right).$$

It holds if the remainder term of the finite-order Taylor expansion converges to zero with probability approaching one. The remainder term is given by

$$\frac{1}{Nh^{j+1}} \frac{(-1)^j}{j!} \sum_{i=1}^{N} (\hat{\xi}_i - \xi_i)^j K^{(j)} \left(\frac{x - \tilde{\xi}_i}{h} \right),$$

where $\tilde{\xi}_i$ is between $\hat{\xi}_i$ and ξ_i. It is sufficient to argue that it converges to zero, as $j \to \infty$, with probability approaching one. We observe that

$$\begin{aligned} &\left| \frac{1}{Nh^{j+1}} \frac{(-1)^j}{j!} \sum_{i=1}^{N} (\hat{\xi}_i - \xi_i)^j K^{(j)} \left(\frac{x - \tilde{\xi}_i}{h} \right) \right| \\ &\leq \left(\frac{1}{h^j} \max_{1 \leq i \leq N} |\hat{\xi}_i - \xi_i|^j \right) \left(\frac{1}{j!h} \max_{1 \leq i \leq N} \left| K^{(j)} \left(\frac{x - \tilde{\xi}_i}{h} \right) \right| \right). \end{aligned} \tag{A.34}$$

We argue that the term in the first parenthesis of (A.34) converges to zero, as $j \to \infty$, with probability approaching one. Note that the convergence hold when $\max_{1 \leq i \leq N} |\hat{\xi}_i - \xi_i|/h < 1$. For this, we observe that for any fixed $\varepsilon > 0$ and positive integer $r \geq 1$

$$\begin{aligned} \Pr\left(\frac{1}{h} \max_{1 \leq i \leq N} |\hat{\xi}_i - \xi_i| \leq \varepsilon \right) = \Pr\left(\frac{1}{h^r} \max_{1 \leq i \leq N} |\hat{\xi}_i - \xi_i|^r \leq \varepsilon^r \right) &= \left(\Pr\left(|\hat{\xi}_i - \xi_i|^r \leq \varepsilon^r h^r \right) \right)^N \\ &\geq \left(1 - \frac{E|\hat{\xi}_i - \xi_i|^r}{\varepsilon^r h^r} \right)^N \\ &\geq \left(1 - \frac{M}{\sqrt{T^r h^{2r}}} \right)^N, \end{aligned} \tag{A.35}$$

by Assumption 1, Markov's inequality, and Lemma 2, 4, or 5 with fixed $M > 0$. The probability in the left hand side of (A.35) thus converges to one if $(1-1/\sqrt{T^r h^{2r}})^N \to 1$. Based on the binomial theorem, we observe that

$$\begin{aligned}\left(1-\frac{1}{\sqrt{T^r h^{2r}}}\right)^N &= \sum_{l=0}^{N}\binom{N}{l}\left(-\frac{1}{\sqrt{T^r h^{2r}}}\right)^l \\ &= 1-\frac{N}{\sqrt{T^r h^{2r}}}+\frac{N(N-1)}{2!(\sqrt{T^r h^{2r}})^2}-\frac{N(N-1)(N-2)}{3!(\sqrt{T^r h^{2r}})^3}+\cdots+\left(-\frac{1}{\sqrt{T^r h^{2r}}}\right)^N.\end{aligned}$$

As a result, the probability in the left hand side of (A.35) converges to one if $N/\sqrt{T^r h^{2r}} \to 0$ for sufficiently large r as $N \to \infty$ and $Th^2 \to \infty$. By taking $\varepsilon < 1$, we obtain the desired result. We note that the condition is significantly weaker than the relative magnitudes condition in Theorem 1. Hence, the term in the first parenthesis of (A.34) converges to zero, as $j \to \infty$, with probability approaching one.

In a similar manner, we can observe that the term in the second parenthesis of (A.34) converges to zero with probability approaching one under regularity conditions.

Therefore, the infinite-order Taylor expansion in Theorem 1 holds under regularity conditions.

D Appendix: The infinite series of the asymptotic biases

This appendix discusses conditions under which the infinite series of the asymptotic biases is well-defined (i.e., summable and convergent). We focus on the density estimator for μ_i only, since the discussions for the other estimators are similar.

To examine the series of the asymptotic biases, we focus on the nonlinearity biases of the density estimator $\hat{f}_{\hat{\mu}}(x)$. Let $e_{T,j}(x) \coloneqq E(T^{j/2}(\bar{w}_i)^j|\mu_i = x)f_\mu(x)$. For the nonlinearity bias in (A.3) of the proof of the density estimation, we observe that

$$\begin{aligned}&E\left(\frac{1}{j!Nh^{j+1}}\sum_{i=1}^{N}(\bar{w}_i)^j K^{(j)}\left(\frac{x-\mu_i}{h}\right)\right) \\ &= \frac{1}{j!h^{j+1}}E\left(E\left((\bar{w}_i)^j|\mu_i\right)K^{(j)}\left(\frac{x-\mu_i}{h}\right)\right) \\ &= \frac{1}{j!T^{j/2}h^j}e_{T,j}(x)\int K^{(j)}(s)ds-\frac{1}{j!T^{j/2}h^{j-1}}e'_{T,j}(x)\int sK^{(j)}(s)ds+\frac{1}{j!T^{j/2}h^{j-2}}\int e''_{T,j}(\tilde{x})s^2K^{(j)}(s)ds,\end{aligned}$$

by the law of iterated expectations, the change of variables, and Taylor's theorem with $\tilde{x}$ located between $x - sh$ and x. The equation for any odd j is equal to

$$-\frac{1}{j!T^{j/2}h^{j-1}}e'_{T,j}(x)\int sK^{(j)}(s)ds + \frac{1}{j!T^{j/2}h^{j-2}}\int e''_{T,j}(\tilde{x})s^2K^{(j)}(s)ds,$$

because of the symmetry of K. On the contrary, the equation for any even j is equal to

$$\frac{1}{j!T^{j/2}h^{j}}e_{T,j}(x)\int K^{(j)}(s)ds + \frac{1}{j!T^{j/2}h^{j-2}}\int e''_{T,j}(\tilde{x})s^2K^{(j)}(s)ds.$$

We focus on the summability of the series of the biases for odd j only, since the discussion for even j is the same. The partial sum of the series of the biases for odd j can be written as

$$\begin{aligned}
&\sum_{j=3}^{2n+1}\left(-\frac{1}{j!T^{j/2}h^{j-1}}e'_{T,j}(x)\int sK^{(j)}(s)ds + \frac{1}{j!T^{j/2}h^{j-2}}\int e''_{T,j}(\tilde{x})s^2K^{(j)}(s)ds\right)\\
&=\sum_{l=1}^{n}\left(-\frac{1}{(2l+1)!T^{(2l+1)/2}h^{2l}}e'_{T,2l+1}(x)\int sK^{(2l+1)}(s)ds\right.\\
&\quad\left.+\frac{1}{(2l+1)!T^{(2l+1)/2}h^{2l-1}}\int e''_{T,2l+1}(\tilde{x})s^2K^{(2l+1)}(s)ds\right)\\
&=-\left(\sum_{l=1}^{n}S_{1l}\right)+\left(\sum_{l=1}^{n}S_{2l}\right),
\end{aligned}$$

where we define

$$\begin{aligned}
S_{1l} &:= \frac{1}{(2l+1)!T^{(2l+1)/2}h^{2l}}e'_{T,2l+1}(x)\int sK^{(2l+1)}(s)ds,\\
S_{2l} &:= \frac{1}{(2l+1)!T^{(2l+1)/2}h^{2l-1}}\int e''_{T,2l+1}(\tilde{x})s^2K^{(2l+1)}(s)ds.
\end{aligned}$$

We examine the series of S_{1l} only, since the discussion for S_{2l} is the same. The ratio test means that the series $\sum_{l=1}^{\infty} S_{1l}$ is summable and convergent if $\lim_{l\to\infty}|S_{1,l+1}/S_{1l}| < 1$. We observe that

$$\left|\frac{S_{1,l+1}}{S_{1,l}}\right| = \left|\frac{1}{Th^2}\frac{1}{(2l+3)(2l+2)}\frac{e'_{T,2l+3}(x)}{e'_{T,2l+1}(x)}\frac{\int sK^{(2l+3)}(s)ds}{\int sK^{(2l+1)}(s)ds}\right|,$$

for any $l \geq 1$. It converges to zero as $l \to \infty$ if $e'_{T,2l+3}(x)/e'_{T,2l+1}(x) = O(1)$ over l and if $\int sK^{(2l+3)}(s)ds/\int sK^{(2l+1)}(s)ds = O(l)$. The former is a regularity condition, and we can simply assume it. The latter is also a regularity condition, and we can easily show its validity by assuming that K is the Gaussian kernel function. Hence, the condition for the ratio test is satisfied, implying that the series $\sum_{l=1}^{\infty} S_{1l}$ is summable and convergent.

The above discussions imply that the infinite series of the asymptotic biases is well-defined under regularity conditions.

Acknowledgments

The authors greatly appreciate the assistance of Mototsugu Shintani in providing the price panel data. The authors would also like to thank Kazuhiko Hayakawa, Shin Kanaya, Hiroyuki Kasahara, Yukitoshi Matsushita, and Yohei Yamamoto for helpful discussions and comments. All remaining errors are our own. Part of this research was conducted while Okui was at Kyoto University. Okui acknowledges financial support from the Japan Society for the Promotion of Science (JSPS) under KAKENHI Grant Nos. 25780151, 25285067, 15H03329, and 16K03598. Yanagi recognizes the financial support of KAKENHI Grant Nos. 15H06214 and 17K13715.

References

D. Ackerberg, C. L. Benkard, S. Berry, and A. Pakes. Econometric tools for analyzing market outcomes. In J. J. Heckman and E. E. Leamer, editors, *Handbook of Econometrics*, volume 6, chapter 63, pages 4171–4276. Elsevier, 2007.

J. E. Anderson and E. Van Wincoop. Trade costs. *Journal of Economic Literature*, 42(3):691–751, 2004.

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

S. Bonhomme and E. Manresa. Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184, 2015.

S. Bonhomme, T. Lamadon, and E. Manresa. Discretizing unobserved heterogeneity. mimeo, 2017.

V. Chernozhukov, I. Fernandez-Val, and A. Galichon. Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, 96(3):559–575, 2009.

M. J. Crucini, M. Shintani, and T. Tsuruga. Noisy information, distance and law of one price dynamics across us cities. *Journal of Monetary Economics*, 74:52–66, 2015.

G. Dhaene and K. Jochmans. Split-panel jackknife estimation of fixed effects models. *Review of Economic Studies*, 82:991–1030, 2015.

C. Engel and J. H. Rogers. Deviations from purchasing power parity: causes and welfare costs. *Journal of International Economics*, 55(1):29–57, 2001.

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

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

S. Gonçalves and M. Kaffo. Bootstrap inference for linear dynamic panel data models with individual fixed effects. *Journal of Econometrics*, 186:407–426, 2015.

P. Hall. On bootstrap confidence intervals in nonparametric regression. *The Annals of Statistics*, 20(2):695–711, 1992.

P. Hall and J. Horowitz. A simple bootstrap method for constructing nonparametric confidence bands for functions. *The Annals of Statistics*, 41(4):1892–1921, 2013.

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

C. Hsiao and M. H. Pesaran. Random coefficient models. In L. Matyas and P. Sevestre, editors, *The Econometrics of Panel Data*, pages 185–213. Springer, 2008.

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

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

H. Kasahara, P. Schrimpf, and M. Suzuki. Identification and estimation of production function with unobserved heterogeneity. mimeo, 2017.

P. J. Klenow and B. A. Malin. Microeconomic evidence on price-setting. *Handbook of Monetary Economics*, 3:231–284, 2010.

Y.-J. Lee, R. Okui, and M. Shintani. Asymptotic inference for dynamic panel estimators of infinite order autoregressive processes. *forthcoming in Journal of Econometrics*, 2018.

Q. Li and J. S. Racine. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, 2007.

S. Mavroeidis, Y. Sasaki, and I. Welch. Estimation of heterogenous autoregressive parameters using short panel data. *Journal of Econometrics*, 188:219–235, 2015.

E. Nakamura and J. Steinsson. Five facts about prices: A reevaluation of menu cost models. *The Quarterly Journal of Economics*, 123(4):1415–1464, 2008.

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

S. Nickell. Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–1426, 1981.

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

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

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

R. Okui and T. Yanagi. Panel data analysis with heterogeneous dynamics. mimeo, 2017.

A. Pagan and A. Ullah. *Nonparametric Econometrics*. Cambridge University Press, 1999.

D. C. Parsley and S.-J. Wei. Convergence to the law of one price without trade barriers or currency fluctuations. *Quarterly Journal of Economics*, 111(4):1211–1236, 1996.

D. C. Parsley and S.-J. Wei. Explaining the border effect: the role of exchange rate variability, shipping costs, and geography. *Journal of International Economics*, 55(1):87–105, 2001.

M. H. Pesaran. *Time Series and Panel Data Econometrics*. Oxford University Press, 2015.

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

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

A. M. Polansky and E. R. Baker. Multistage plug-in bandwidth selection for kernel distribution function estimates. *Journal of Statistical Computation and Simulation*, 65(1-4):63–80, 2000.

J. D. L. Roca and D. Puga. Learning by working in big cities. *The Review of Economic Studies*, 84(1):106–142, 2017.

L. Su, Z. Shi, and P. C. B. Phillips. Identifying latent structures in panel data. *Econometrica*, 84 (6):2215–2264, 2016.

M. P. Wand and M. C. Jones. Multivariate plug-in bandwidth selection. *Computational Statistics*, 9(2):97–116, 1994.

M. E. Yazgan and H. Yilmazkuday. Price-level convergence: New evidence from U.S. cities. *Economics Letters*, 110(2):76–78, 2011.

Table 1: The categorization of goods and services

Goods	T-bone steak, Ground beef, Frying chicken, Chunk light tuna, Whole milk, Eggs, Margarine, Parmesan cheese, Potatoes, Bananas, Lettuce, Bread, Coffee, Sugar, Corn flakes, Sweat peas, Peaches, Shortening, Frozen corn, Soft drink, Beer, Wine, Facial tissues, Dishwashing powder, Men's dress shirt, Shampoo, Toothpaste, Tennis balls.
Services	Hamburger sandwich, Pizza, Fried chicken, Total home energy cost, Telephone, Apartment, Home purchase price, Mortgage rate, Monthly payment, Dry cleaning, Major appliance repair, Auto maintenance, Gasoline, Doctor office visit, Dentist office visit, Haircut, Beauty salon, Newspaper subscription, Movie, Bowling.

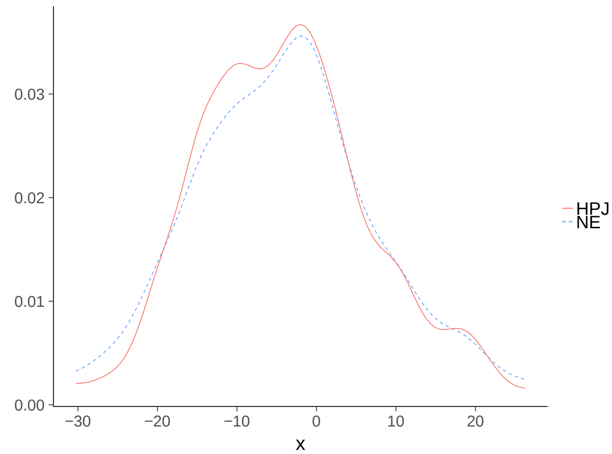
Note: The service category includes those considered as goods but whose prices are likely to include the cost of a service.

Table 2: Plug-in bandwidths for all items

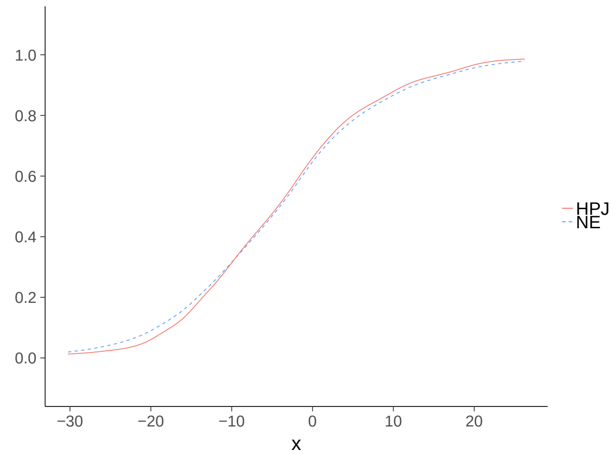
	μ	γ_0	ρ_1
Density	2.159	18.935	0.045
CDF	1.300	10.768	0.025

Table 3: Plug-in bandwidths for goods and services separately

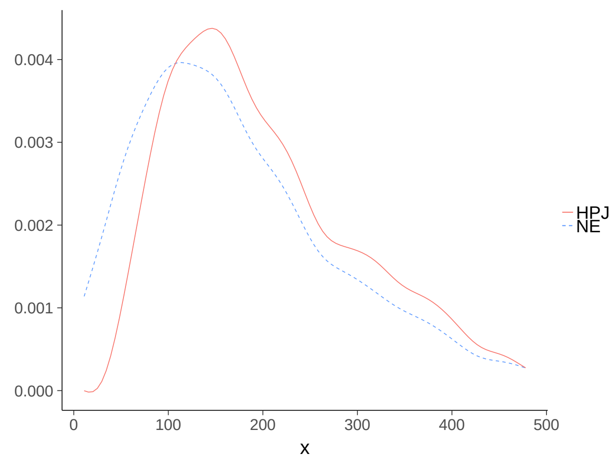
	Goods			**Services**		
	μ	γ_0	ρ_1	μ	γ_0	ρ_1
Density	2.592	19.820	0.049	2.347	15.009	0.038
CDF	1.531	12.367	0.027	1.649	9.820	0.025



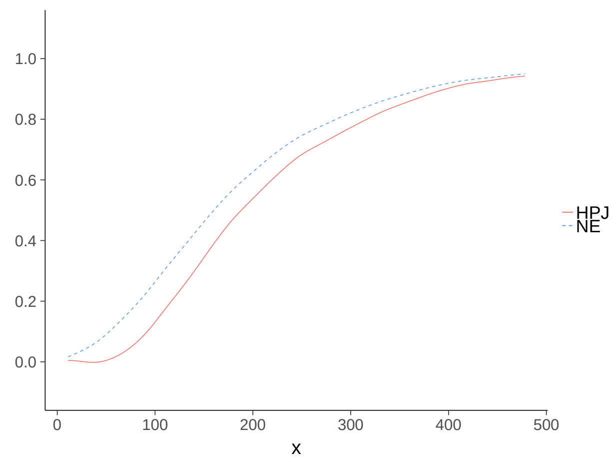
(a) Density of μ_i



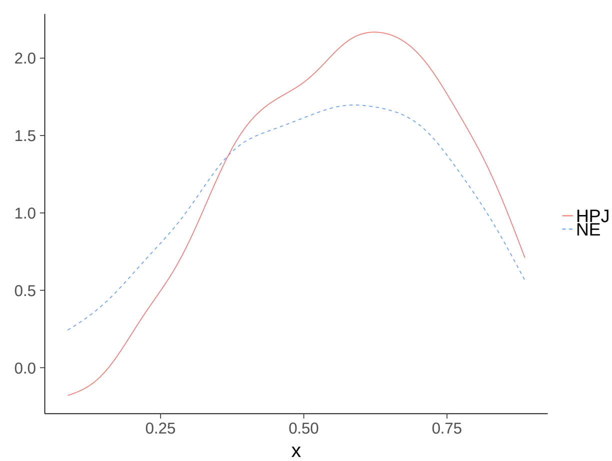
(b) CDF of μ_i



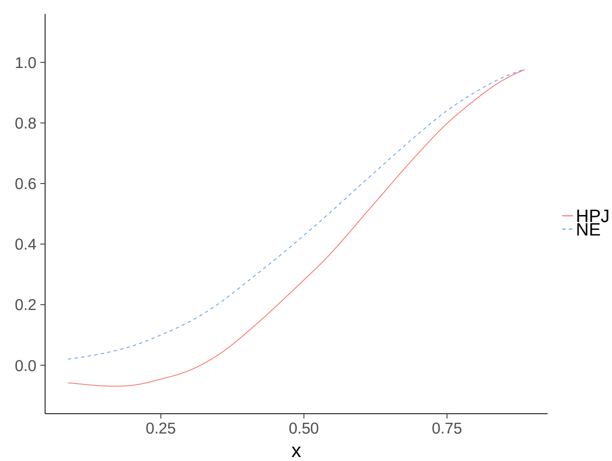
(c) Density of $\gamma_{0,i}$



(d) CDF of $\gamma_{0,i}$

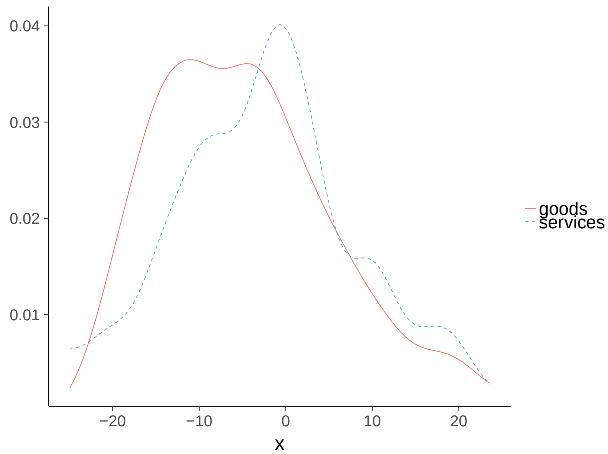


(e) Density of $\rho_{1,i}$

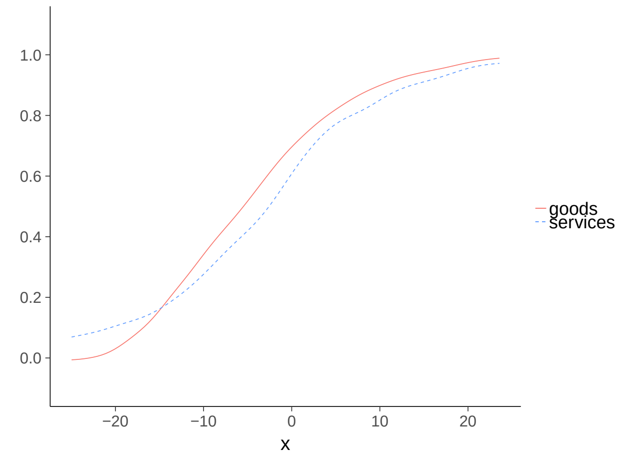


(f) CDF of $\rho_{1,i}$

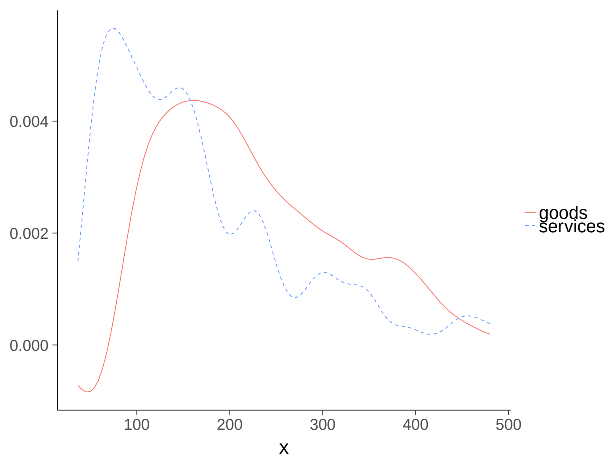
Figure 1: The estimated density and CDF for all items



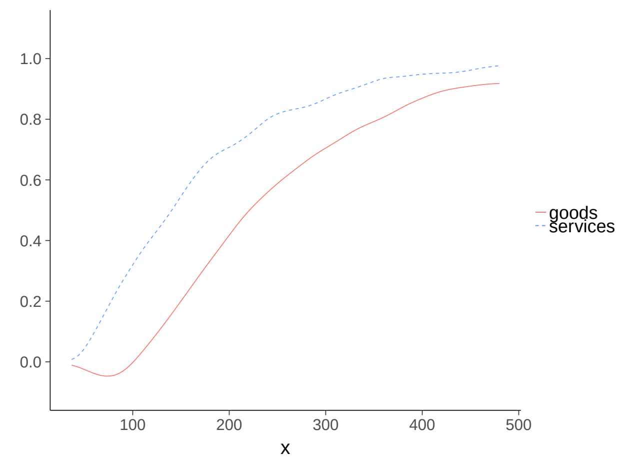
(a) Density of μ_i



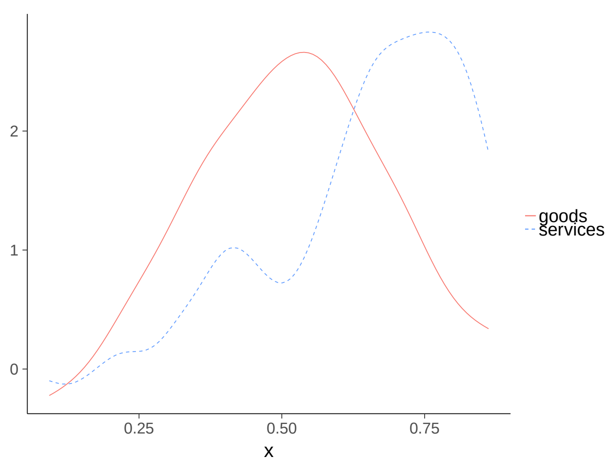
(b) CDF of μ_i



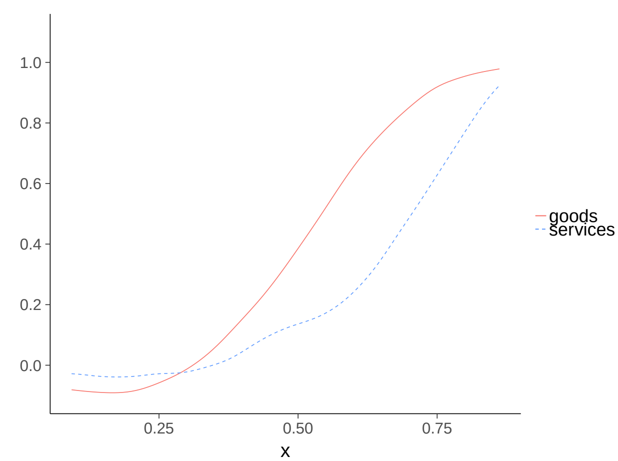
(c) Density of $\gamma_{0,i}$



(d) CDF of $\gamma_{0,i}$

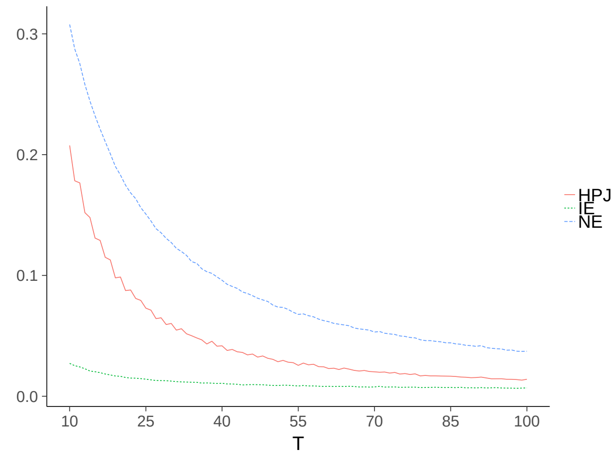


(e) Density of $\rho_{1,i}$

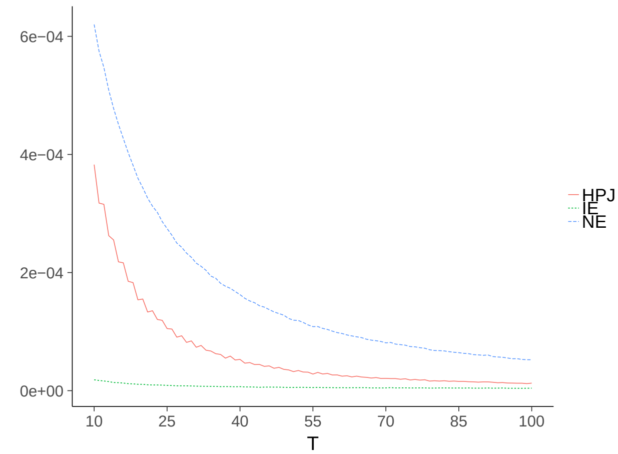


(f) CDF of $\rho_{1,i}$

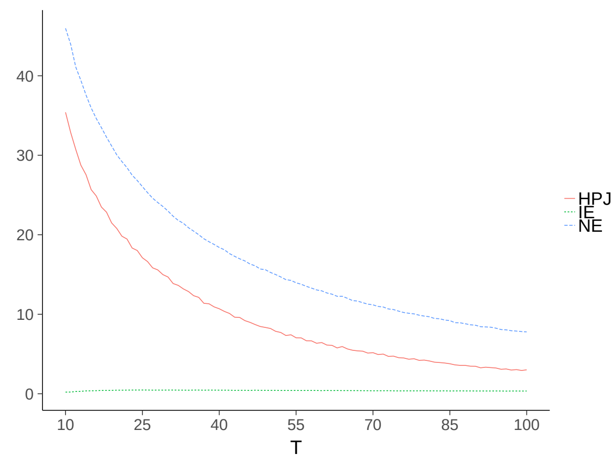
Figure 2: The estimated HPJ density and CDF for goods and services



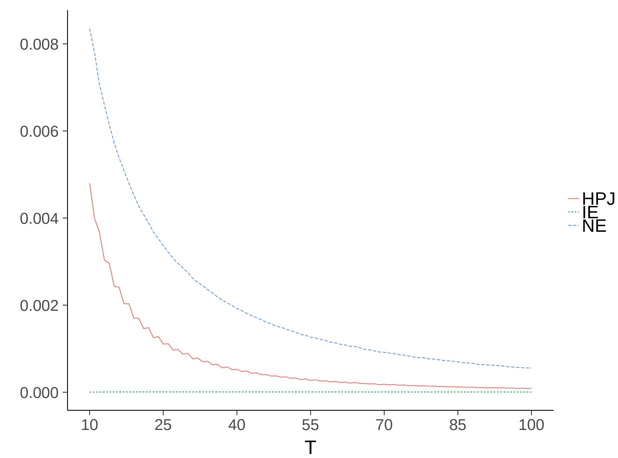
(a) Density of μ_i



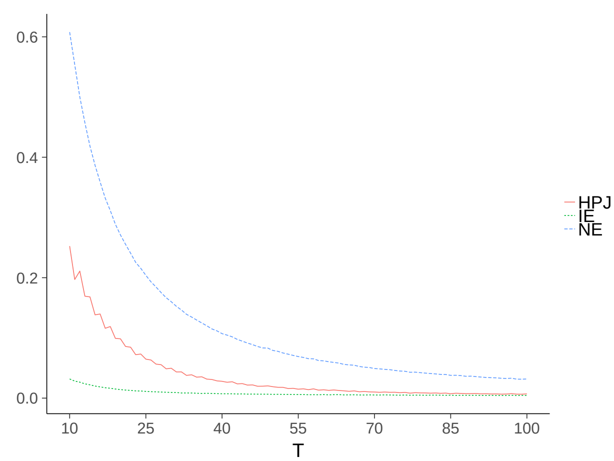
(b) CDF of μ_i



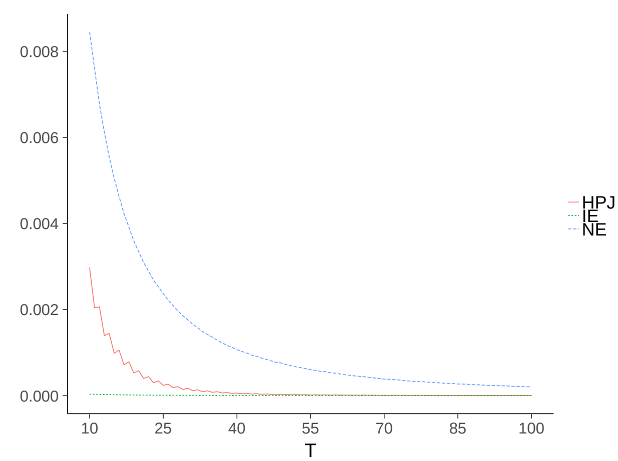
(c) Density of $\gamma_{0,i}$



(d) CDF of $\gamma_{0,i}$

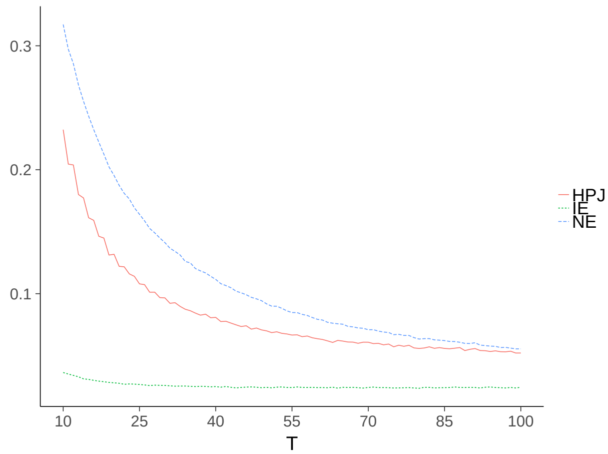


(e) Density of $\rho_{1,i}$

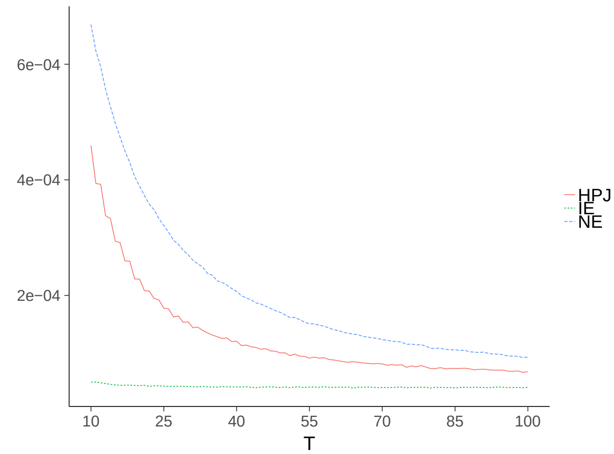


(f) CDF of $\rho_{1,i}$

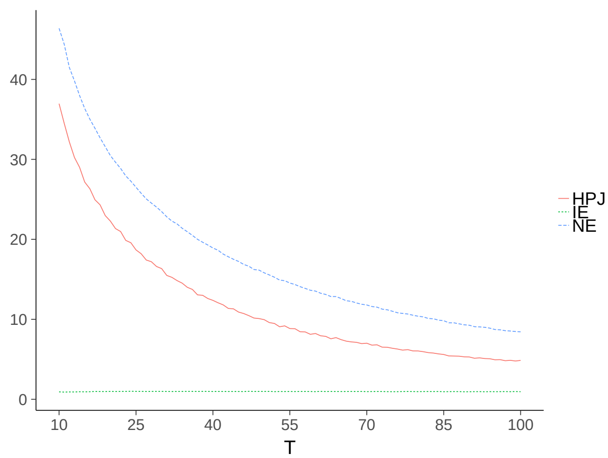
Figure 3: Integrated squared bias of each estimator



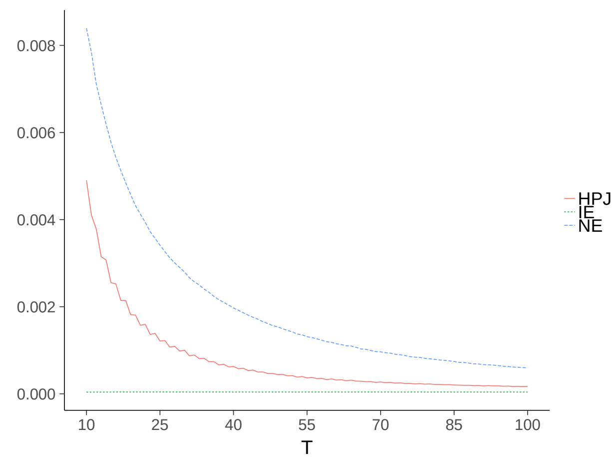
(a) Density of μ_i



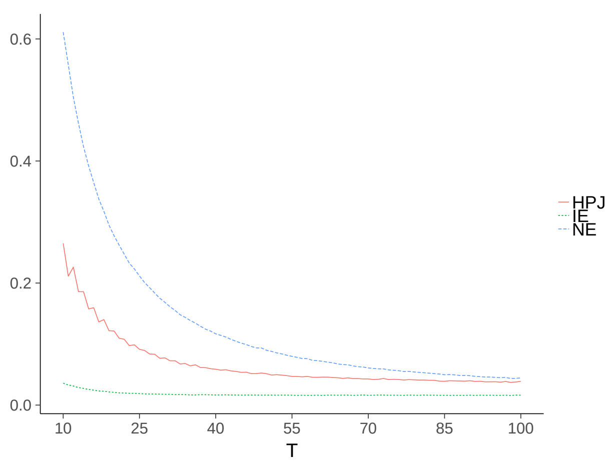
(b) CDF of μ_i



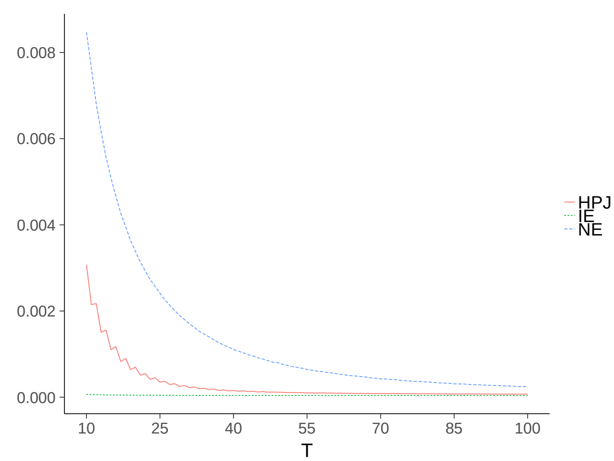
(c) Density of $\gamma_{0,i}$



(d) CDF of $\gamma_{0,i}$

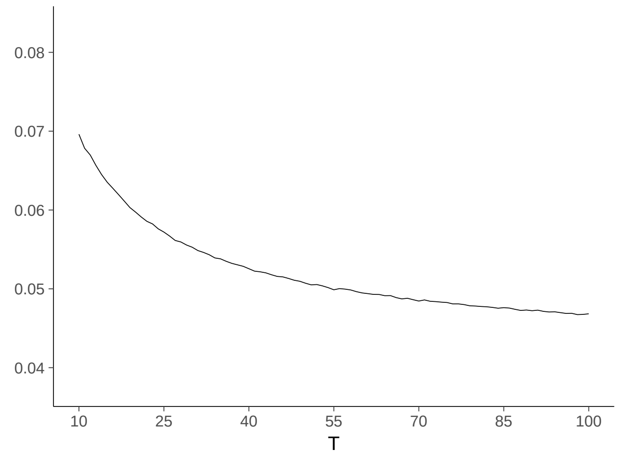


(e) Density of $\rho_{1,i}$

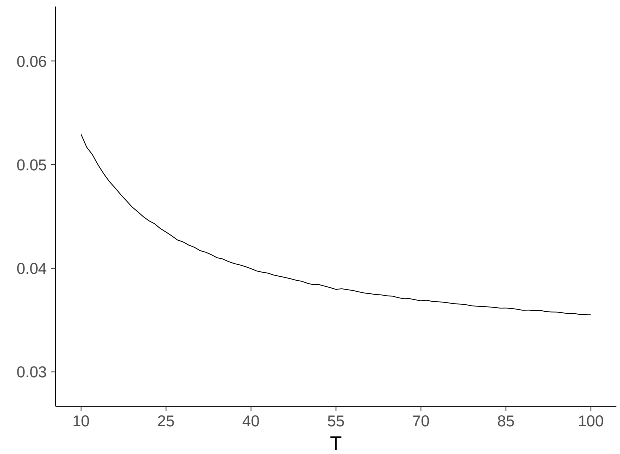


(f) CDF of $\rho_{1,i}$

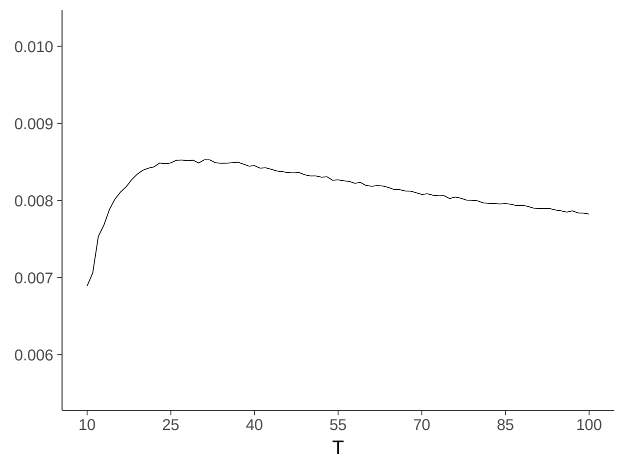
Figure 4: Integrated MSE of each estimator



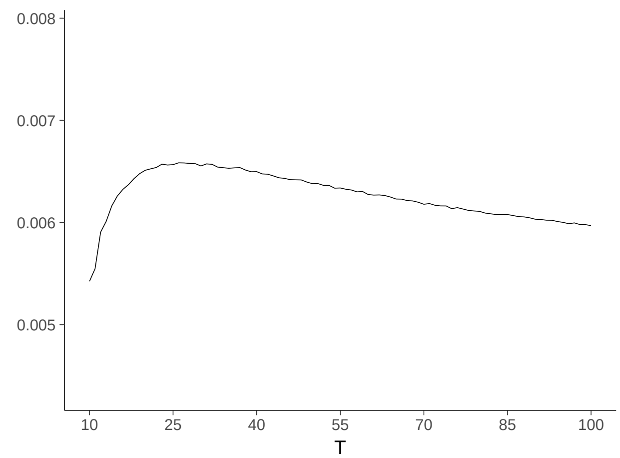
(a) Density of μ_i



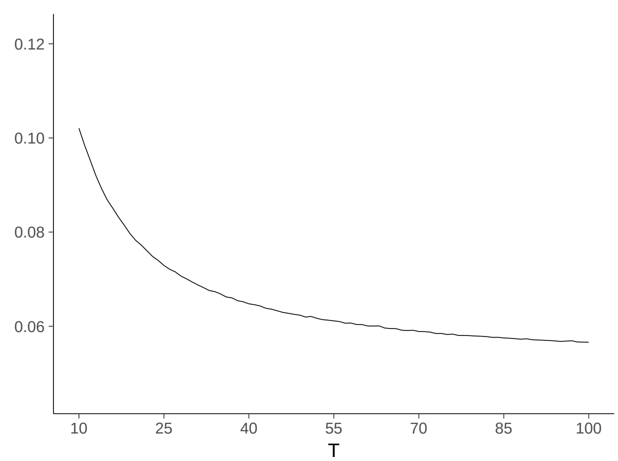
(b) CDF of μ_i



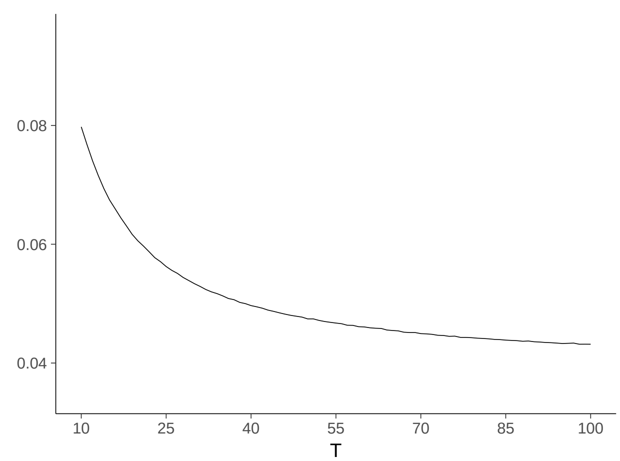
(c) Density of $\gamma_{0,i}$



(d) CDF of $\gamma_{0,i}$



(e) Density of $\rho_{1,i}$



(f) CDF of $\rho_{1,i}$

Figure 5: The means of plug-in bandwidths for each estimator