GMM ESTIMATION WITH BROWNIAN KERNELS APPLIED TO INCOME INEQUALITY MEASUREMENT

By

Jin Seo Cho and Peter C. B. Phillips

October 2024

COWLES FOUNDATION DISCUSSION PAPER NO. 2411

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

GMM Estimation with Brownian Kernels Applied to Income Inequality Measurement

JIN SEO CHO

School of Economics, Yonsei University, Seoul 03722, Korea
Email: jinseocho@yonsei.ac.kr

PETER C. B. PHILLIPS

Yale University, University of Auckland
& Singapore Management University
Email: peter.phillips@yale.edu

October 2024

Abstract

In GMM estimation, it is well known that if the moment dimension grows with the sample size, the asymptotics of GMM differ from the standard finite dimensional case. The present work examines the asymptotic properties of infinite dimensional GMM estimation when the weight matrix is formed by inverting Brownian motion or Brownian bridge covariance kernels. These kernels arise in econometric work such as minimum Cramér-von Mises distance estimation when testing distributional specification. The properties of GMM estimation are studied under different environments where the moment conditions converge to a smooth Gaussian or non-differentiable Gaussian process. Conditions are also developed for testing the validity of the moment conditions by means of a suitably constructed J-statistic. In case these conditions are invalid we propose another test called the U-test. As an empirical application of these infinite dimensional GMM procedures the evolution of cohort labor income inequality indices is studied using the Continuous Work History Sample database. The findings show that labor income inequality indices are maximized at early career years, implying that economic policies to reduce income inequality should be more effective when designed for workers at an early stage in their career cycles.

Key Words: Infinite-dimensional GMM estimation; Brownian motion kernel; Brownian bridge kernel; Gaussian process; Infinite-dimensional MCMD estimation; Labor income inequality.

Subject Classification: C13, C18, C32, C55, D31, O15, P36.

Acknowledgments: A comment of a referee on one of our earlier papers prompted us to pursue our thinking on the research presented here. We have further benefited from discussions with Chirok Han, Dukpa Kim, Seo Jeong Lee, and seminar participants at the Allied Economic Associations Annual Meeting in Korea (2024). Phillips acknowledges research support from the Kelly Fund at the University of Auckland and a KLC Fellowship at Singapore Management University.

1 Introduction

"*Science is the Differential Calculus of the mind. Art the Integral Calculus; they may be beautiful when apart, but are greatest only when combined.*" — Ronald Ross.

The generalized method of moments (GMM) approach to estimation and inference has been adapted to several nonstandard environments. A particular extension that has proved to be important theoretically and empirically useful involves the application of GMM when the number of moment conditions is allowed to pass to infinity with the sample size. An example of such high dimensional GMM occurs with the use of minimum Cramér-von Mises distance (MCMD) estimation, which is useful in estimating unknown parameters in a model's distribution and can be extended to test distributional specification. Early work by Pollard (1980) and later Cho, Park, and Phillips (2018) showed how to develop limit theory for MCMD estimation in GMM form with a particular weight matrix. The latter paper applied MCMD to test the Pareto distributional form for income data in Korea, an approach that is commonly used in estimating top income shares (see Piketty, 2003; Piketty and Saez, 2003; Atkinson, Piketty, and Saez, 2011, among others). In another context, Angrist and Keueger (1991) estimated the monetary return to education in the labor market using two-stage least squares on U.S. census data with a large number of instruments, leading to another high dimensional GMM problem.

The large sample properties of high dimensional GMM with many moment conditions rely on the asymptotic behavior of the GMM components and these are case-dependent. So, unless the GMM environment is fully characterized, the asymptotic properties of GMM can be difficult to derive and remain elusive. For example, when persistently correlated moment conditions are employed in GMM estimation, large sample analysis differs from that when the moment conditions are weakly correlated. As discussed below, the moment conditions for MCMD converge to a Brownian bridge (BB) process, and GMM using these moment conditions is very different from cases of GMM estimation with weakly correlated moment conditions that are employed in instrumental variable or two-stage least squares estimation, as investigated by Carrasco (2012)

Different weight matrices in GMM may also produce different limit properties. But when the matrix dimension grows, inversion inevitably becomes imprecise as the smallest eigenvalue of the matrix approaches zero. Indeed, the limiting inverse may not exist and computation becomes case-dependent and difficult even when the limit exists. For example, the BB process for MCMD estimation allows GMM to be computed using the inverse of the BB covariance matrix for asymptotic optimality, but it is unknown how the inverse matrix affects GMM asymptotically, irrespective of the existence of the limit inverse. Likewise, if the moment conditions form a unit-root process (another persistently correlated process) that converges to Brownian motion (BM) upon standarization, the size of the moment conditions grows and an asymptotically optimal GMM is obtained by inverting the BM covariance matrix. Nonetheless, its influence on GMM limit behavior is presently unknown in the literature. The reason is that the limit of the inverse of the BM covariance matrix does not exist in any standard form of the type that assumes a finite number of moment conditions. These restrictions effectively diminish the scope of GMM applications in the literature to reliance on stationary moment conditions. Challenges of the type just described associated with high dimensional matrix inversions relate to the so-called ill-posedness problem associated with the inversion of operators in functional analysis, a problem that is typically resolved by various forms of regularization.

A primary goal of the present paper is to tackle directly and without regularization the challenges of high dimensional GMM that are associated with BM and BB covariance matrices, which are in turn induced by persistently correlated moment conditions. Direct inversion of BM and BB covariance matrices and the manner of doing so is quite new and opens up a wide range of potential applications. To mention a few of them here: (i) For empirical processes that converge to a BB process, as for the Kolmogorov-Smirnov (KS) test, it is necessary to invert the BB covariance matrix when testing a distributional hypothesis within the GMM framework; (ii) BM and BB processes are widely assumed in finance for series such as stock prices, interest rates, and option prices among others (e.g., Andersen and Piterbarg, 2010; Hirsa and Neftci, 2014). If the sample paths of these processes are employed as a series of moment conditions for GMM estimation it becomes necessary to invert BM or BB covariance matrices; (iii) More generally, the use of covariance matrices based on BM and BB processes and kernels is common because of convenience of form and ease of analysis, yet they are rarely applied in empirical work because of the difficulties induced by ill-posedness problems in inversion.

To meet these needs the main aim of the present study is to provide a unified framework for delivering the asymptotic properties of GMM when BM or BB covariance matrix inversions are involved that is applicable in a wide range of different models and circumstances where moment conditions are persistently correlated in a manner that yields a continuous sample path and the model is specified to cope with this complexity. In particular, our approach allows the moment dimension to grow infinitely large and makes explicit the variate space and inner product framework that provides the mechanism for embodying the GMM limit theory. We call the GMM procedures associated with these techniques BM-GMM and BB-GMM.

The asymptotic properties of BM-GMM and BB-GMM depend intimately on two key component elements – the weight matrix and the moment conditions – both of which become infinite dimensional in the limit. We briefly explain here how our theory is developed using these elements. First, the properties of the weight matrix in BM-GMM and BB-GMM inevitably affect the limit properties of the estimation procedure, just as they do in the finite dimensional case. The BM and BB covariance matrices become infinite dimensional as the moment size grows, and they are often called the BM-kernel and BB-kernel, in accord with their use as kernels for integral operators. More specifically, if the moment index is adjusted to fit to the unit interval, the BM and BB covariance matrices can be positioned on the unit square. Further, as the moment size increases, the corresponding BM and BB covariance matrices converge to continuous functions as shown in Figure 1. Hence, when they are both suitably standardized, the product of a BM (or BB) covariance matrix and a vector (whose standardized form has a continuous function limit on the unit interval) converges to a double integral formed with the BM (or BB) kernel and the limiting continuous function of the standardized vector. As is detailed fully below, we can also represent this limit using an inner product between two continuous functions such that the first is the integral transform of the limiting continuous function using the BM (or BB) kernel, and the second is the limiting continuous function. For our GMM analysis, the GMM distance is constructed using the inverted BM (or BB) covariance matrix as the weight matrix instead of the BM (or BB) covariance matrix itself and our development shows how this inversion affects GMM estimation. Specifically, the analysis reveals that under some regularity conditions this inversion ensures the GMM distance converges to an inner product between the derivatives or differentials defined by the limit process of the moment conditions. For this development we explicitly derive the inverse kernel functions from the BM-kernel and BB-kernel for the case of a set of finite moments, and then let the moment size tend to infinity to obtain the limiting inverse kernel operators. Heuristically, given that the BM covariance matrix is the covariance matrix of an integrated process, the quadratic form product using the 'inverted' BM covariance matrix yields an inner product between the 'dis-integrated' processes. Put simply, this means that inverting integration operators leads to differentiation. This approach delivers directly what is implied in the prior literature that the inverse BM-kernel operator is a second-order differential kernel operator (e.g., Carrasco, Florens, and Renault, 2007).

Second, infinite dimensional moment conditions also affect the limit properties of GMM. As mentioned above, there may be two types of infinite dimensional moment conditions: persistently correlated moment conditions and weakly correlated moment conditions. We distinguish these by their sample path characteristics. If the sample path is continuous, we say that the moment conditions are persistently correlated. Otherwise, we call them weakly correlated moment conditions. For the current study, we assume persistently correlated moment conditions and leave the case of weakly correlated moment conditions for future research. This approach is convenient because the large sample analysis of GMM based on the two types of moment conditions are distinct and better treated in separate work. Further, continuous BM and BB kernels are incompatible with weakly correlated moment conditions, although we do not necessarily require that the moment conditions themselves converge to BM or BB processes. Instead, our framework requires that the moment conditions converge to a twice continuously differentiable Gaussian process or an Itô process for which BM and BB processes are just special cases. Using this wide class of infinite dimensional moment conditions, we have a unified framework for investigating the large sample properties of GMM formed by inverted BM or BB covariance matrices leading to asymptotically normal GMM estimation limit theory.

In two studies that relate to the present work, Carrasco and Florens (2000) and Amengual, Carrasco, and Sentana (2020) considered infinite-dimensional GMM estimation by using Tikhonov regularization methods. Specifically, these authors obtained the limit of an infinite-dimensional weight matrix by combining its spectrum with asymptotically negligible bias in a manner analogous to ridge regression so that the methodology is applicable even when the weight matrix is not bounded in the limit. Picard (1910) provided necessary and sufficient conditions for the existence of a bounded inverted kernel function. But many popular kernel functions for empirical applications do not satisfy those conditions. In consequence, it is generally believed that it is necessary to apply regularization techniques to enable analysis of inverse kernels in such cases. Indeed, Carrasco and Florens (2000) used regularized kernel inversion to obtain the limit distribution of the estimators defined in terms of the inverse kernel, and Amengual et al. (2020) applied that approach in testing distributional assumptions by GMM (see also Kirsch, 1996; Carrasco et al., 2007). Neither BM-GMM nor BB-GMM kernels satisfy Picard's conditions, so the option of using regularized kernel inversion is available in the present study. But our approach is instead to develop explicit derivations of the inverse BM-kernel and inverse BB-kernel. These explicit inverse kernels enable us to develop and analyze GMM asymptotics without having to resort to the use of regularized kernel inversions.

In addition to providing a unifying framework for BM-GMM and BB-GMM asymptotics, the paper addresses overidentification testing. We provide regularity conditions under which the Sargan J-test statistic (Sargan, 1958; Hansen, 1982) can be validly used for testing overidentification in the high-dimensional moment case. In case the regularity conditions for the J-test do not hold, we revisit the T-test approach taken in Donald, Imbens, and Newey (2003) and provide a new test for the present setting called the U-test, developing its asymptotic theory and showing how the two different testing methods supplement each other according to the context.

High dimensional BM-GMM and BB-GMM methods can be applied in many areas where large data sets are available to test relevant economic hypotheses. We demonstrate their use in labor economics, focusing on BB-GMM estimation to measure top labor income shares over time. Among others, Piketty (2003), Piketty and Saez (2003), and Atkinson et al. (2011) have estimated top income shares in many countries over time using income data. A key assumption in their approach is that income observations in the right tail of the distribution closely follow the form of a Pareto tail. If the hypothesis is invalid, the estimated top income shares are biased. In our approach BB-GMM estimation is conducted under the Pareto tail hypothesis. For this purpose our empirical application employs the Continuous Work History Sample (CWHS) database, which collects labor income data from individuals born in the U.S. between 1960 and 1962. In addition, the Pareto tail distributional condition is tested using the U-test developed in the present study. When the Pareto tail hypothesis is not rejected, the top income shares are estimated under this condition using the BB-GMM approach.

A further goal of our empirical study is to examine the evolution of income inequality within the same cohort. Previous research has examined income inequality over time using country-level data, which may not adequately capture structural factors involved in the evolution. Instead, we compute income inequality indices using observations from the same cohort in the CWHS database over time, which enables identification of a standard pattern in income inequality evolution. By constructing several cohorts from the database, we estimate the labor income inequality indices within the same cohort over time and derive policy implications to reduce income inequality.

The plan of the present study is as follows. Section 2 develops limit theory in our high dimensional setting for GMM estimation and the tests for overidentification. A particular focus in this discussion is the large sample behavior of BB-GMM estimation when it is applied to MCMD estimation, which is subsequently treated as a running example in the rest of the paper. Section 3 reports the results of a simulation study that employs this running example and corroborates the large sample behavior established in Section 2. Section 4 examines the CWHS database and measures various labor income inequality indices, focusing on data classified by gender, education, race, and birth year. Conclusions are drawn in Section 5. All the main results of the paper are proved in the Online Supplement, together with some additional technical results and empirical evidence.

For ease of reference we introduce some notation. For an arbitrary function $f(\cdot)$ and $j = 1, 2, \ldots$, we use $(d^j/d^j x)f(\bar{x})$ for $(d^j/dx^j)f(x)|_{x=\bar{x}}$. Integral operators are shown in boldface, and $(a(\cdot), b(\cdot))$ is the L^2 inner product of $a(\cdot)$ and $b(\cdot)$, so that $(a(\cdot), b(\cdot)) := \int a(u)b(u)du$. If $A(\cdot) \in \mathbb{R}^a$ and $B(\cdot) \in \mathbb{R}^b$, then $[A(\cdot), B(\cdot)]$ denotes the Gramian matrix of $A(\cdot)$ and $B(\cdot)$, viz., the matrix of inner products between the elements of $A(\cdot)$ and $B(\cdot)$, so that $[A(\cdot), B(\cdot)]$ is an $a \times b$ matrix with (i, j)-th element $(A_i(\cdot), B_j(\cdot))$. Finally, for i and $j = 1, 2, \ldots, n$, we let $i_n := \frac{i}{n}$ and $j_n := \frac{j}{n}$. Other notation in the paper is standard.

2 Estimation and Inference for BM-GMM and BB-GMM

This section develops a large sample asymptotic theory for BM-GMM and BB-GMM estimation and inference. To illustrate the theory we use the MCMD framework as a running example and BB-GMM is applied in this MCMD setting.

2.1 Environments of BM-GMM and BB-GMM

To fix ideas the standard framework for finite dimensional GMM involves extremum estimation with an objective function to be minimized that has the form

$$\bar{q}_n(\cdot) := \bar{G}_n(\cdot)'\widehat{\Sigma}_n^{-1}\bar{G}_n(\cdot) \quad \text{with} \quad q_n(\cdot) := n\bar{q}_n(\cdot)$$

for which there is assumed to be a unique vector $\theta_* \in \Theta$ satisfying the moment condition $\mathbb{E}[\bar{G}_n(\theta_*)] = 0$. Here Θ is a convex and compact parameter space that is a subset of $\mathbb{R}^d$ ($d \in \mathbb{N}$) and the sample moment vector is given by

$$\bar{G}_n(\cdot) := \frac{1}{n}\sum_{t=1}^{n} U_n(W_t, \cdot) \;\; \text{and} \;\; G_n(\cdot) := n\bar{G}_n,$$

with $U_n(W_t, \cdot) : \Theta \mapsto \mathbb{R}^s$ continuously differentiable on Θ with probability (prob.) 1, $\widehat{\Sigma}_n \in \mathbb{R}^{s\times s}$ is a symmetric, positive definite matrix for large enough n, and $\{W_t \in \mathbb{R}^r : t = 1, 2, \ldots, n\}$ is a sequence of strictly stationary and ergodic random variables defined on a complete probability space. Let $\widehat{\theta}_n$ denote the GMM estimator obtained as $\widehat{\theta}_n := \arg\min_{\theta\in\Theta} \bar{q}_n(\theta)$. The dimension of the moment conditions is the dimension, s, of $G_n(\cdot)$.

Typically in GMM limit theory the number of moment conditions s is invariant to the sample size, although this is not always the case in practical work as in many problems the underlying theory provides a large number of possible moment conditions. Hansen (1982) and Bates and White (1985) among many others explored the asymptotic behavior of the GMM estimator in fixed dimensional settings. Specifically, if the GMM estimator is approximated as

$$\sqrt{n}(\widehat{\theta}_n - \theta_*) = -\left[\nabla_\theta \bar{G}_n(\theta_*)\widehat{\Sigma}_n^{-1}\nabla_\theta' \bar{G}_n(\theta_*)\right]^{-1}\left[\nabla_\theta \bar{G}_n(\theta_*)\widehat{\Sigma}_n^{-1}\widetilde{G}_n(\theta_*)\right] + o_{\mathbb{P}}(1) \tag{1}$$

by way of Taylor expansion, where $\widetilde{G}_n(\theta_*) := \sqrt{n}\bar{G}_n(\theta_*)$, the limit distribution of the GMM estimator is obtained by deriving the limit behavior of the components on the right side of (1). If $\nabla_\theta \bar{G}_n(\theta_*)$ and $\widehat{\Sigma}_n$ converge to $H_* := \mathbb{E}[\nabla_\theta U_{n,t}(\theta_*)]$ and Σ with probability converging to 1, and $\widetilde{G}_n(\theta_*) \overset{\text{A}}{\sim} \mathcal{N}(0, \Sigma_*)$, then

$$\sqrt{n}(\widehat{\theta}_n - \theta_*) \overset{\text{A}}{\sim} \mathcal{N}(0, (H_*\Sigma^{-1}H_*')^{-1}(H_*\Sigma^{-1}\Sigma_*\Sigma^{-1}H_*')(H_*\Sigma^{-1}H_*')^{-1}),$$

under the conditions that Σ_* is the covariance matrix of $U_{n,t}(\theta_*) := U_n(W_t, \cdot)$ and $H_*\Sigma^{-1}H'_*$ and Σ_* are positive definite.

The current study differs from the standard GMM frameworkn as the number of moment conditions s is allowed to increase with $n \to \infty$. For this purpose we let $s = s_n$ be the dimension of the moment conditions with consequent implications for the weight matrices in GMM estimation. Our particular focus in estimating θ when $\widehat{\Sigma}_n$ has the following possible forms

$$\ddot{\Sigma}_n := \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \frac{1}{n} & \frac{2}{n} & \cdots & \frac{2}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{2}{n} & \cdots & 1 \end{bmatrix} \quad \text{or} \quad \widetilde{\Sigma}_n := \begin{bmatrix} \frac{1}{n}(1-\frac{1}{n}) & \frac{1}{n}(1-\frac{2}{n}) & \cdots & \frac{1}{n}(1-\frac{n-1}{n}) \\ \frac{1}{n}(1-\frac{2}{n}) & \frac{2}{n}(1-\frac{2}{n}) & \cdots & \frac{2}{n}(1-\frac{2}{n}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n}(1-\frac{n-1}{n}) & \frac{2}{n}(1-\frac{n-1}{n}) & \cdots & \frac{n-1}{n}(1-\frac{n-1}{n}) \end{bmatrix}.$$

Here the dimension of $\ddot{\Sigma}_n$ is n and the (i,j)-th element is $\min(i_n, j_n)$, which corresponds to the finite sample analog of a Brownian motion kernel. We refer to the GMM driven by $\ddot{\Sigma}_n$ as Brownian motion GMM (BM-GMM) and we denote the corresponding estimator as $\ddot{\theta}_n$. Similarly, $\widetilde{\Sigma}_n$ is a sample analog of the Brownian bridge kernel, and the (i,j)-th element is $\min(i_n, j_n)(1 - \max(i_n, j_n))$. We call the estimator based on $\widetilde{\Sigma}_n$ the Brownian bridge GMM (BB-GMM), and it is denoted $\widetilde{\theta}_n$. A primary goal of the paper is to derive the limit properties of BM-GMM and BB-GMM.

Estimation in this context is referred to as an infinite-dimensional GMM. A number of earlier studies examine econometric models under similar divergence conditions. Bontemps and Meddahi (2012) test distributional assumptions by GMM methods using the moment conditions implied by the assumption. Use of BM-GMM and BB-GMM is more related to MCMD estimation. In that connection, Pollard (1980) and Cho et al. (2018) examined estimating an unknown parameter θ_* in the distribution function, $F(\cdot, \theta_*)$, of a variable x_t by minimizing the Cramér-von Mises distance. Specifically, it was assumed that grouped data $\{[c_{j-1}, c_j), \#\{x_t \in [c_{j-1}, c_j)\} : j = 1, 2, \ldots, s; t = 1, \ldots, n\}$ are available on x_t and θ_* is estimated by minimizing the objective function $\bar{q}_n^{(s)}(\theta) := \sum_{j=1}^{s}\{\widehat{p}_{n,j} - F(c_j, \theta)\}^2$ with respect to θ, where for each $j = 1, 2, \ldots, s$, $\widehat{p}_{n,j} := n^{-1}\sum_{t=1}^{n}\mathbb{I}(x_t \in [c_0, c_j])$, and $\{x_t : t = 1, 2, \ldots, n\}$ is a sequence of independent identically distributed (IID) random variables. In this formulation the quantity $\widehat{p}_{n,j}$, giving the proportion of the data in interval j, is treated as the dependent variable, and $F(c_j, \cdot)$ serves as a nonlinear model for $\widehat{p}_{n,j}$. When $F(\cdot, \theta_*)$ correctly matches the distribution of x_t, the MCMD estimator of θ_* is consistent, and its distribution is asymptotically normal. In particular, if the MCMD estimator is used to construct the Kolmogorov-Smirnov (KS) statistic to test a distributional hypothesis with an unknown parameter, the null limit distribution is a functional of a linearly transformed Brownian bridge. This property is particularly appealing for empirical applications because of the difficulty in obtaining the null limit distribution of the KS test when other consistent estimators, such as the maximum likelihood estimator, are used in the construction, as pointed out by Durbin (1973).

MCMD estimation of this type can be formulated in GMM format. For each $j = 1, 2, \ldots, s$, the parameter θ_* satisfies the moment condition $\mathbb{E}[\widehat{p}_{n,j} - F(c_j, \theta_*)] = 0$. Let $\widehat{\theta}_n^{(s)}$ be the GMM extremum estimator satisfying

$$\widehat{\theta}_n^{(s)} := \arg\min_{\theta \in \Theta} (\widehat{P}_n^{(s)} - F^{(s)}(\theta))' W^{(s)} (\widehat{P}_n^{(s)} - F^{(s)}(\theta)),$$

where $\widehat{P}_n^{(s)} := [\widehat{p}_{n,1}, \ldots, \widehat{p}_{n,s}]'$, $F^{(s)}(\theta) := [F(c_1, \theta), \ldots, F(c_s, \theta)]'$, and $W^{(s)}$ is an $s \times s$ positive definite matrix that converges to a positive definite matrix as n tends to infinity. The MCMD estimator is therefore a GMM estimator with a structure of generalized least squares. The GMM estimator $\widehat{\theta}_n^{(s)}$ is consistent for θ_* and asymptotically normal under the standard framework that applies when the number of groups s is fixed. The MCMD estimator is then the special case where the weight matrix is $W^{(s)} = I_s$. As another case, if the data is grouped and the group dimension is $s_n = n - 1$ with $W^{(s_n)} = \widetilde{\Sigma}_n^{-1}$, then MCMD falls within the framework of BB-GMM. Note that $\sqrt{n}(\widehat{P}_n^{(s)} - F^{(s)}(\theta_*))$ weakly converges to a Brownian bridge process $\mathcal{B}^0(\cdot)$, motivating $\widetilde{\Sigma}_n^{-1}$ as the weight matrix for GMM. Accordingly from now on we refer to MCMD estimation driven by $W^{(s_n)} = \widetilde{\Sigma}_n^{-1}$ as infinite-dimensional MCMD estimation and use it as a running example of BB-GMM. In this setup the weight matrix is not parameterized by θ. For such a case, BB-GMM methodology should be understood as an effective way of deriving the limit behavior of the GMM estimator with a weight matrix that estimates Brownian kernels consistently.

In addition to infinite dimension MCMD estimation there are other examples where the moment dimension is determined by the sample size n. Grenander (1981) developed a theory of abstract inference in function space that deals with parameters in models involving various stochastic processes and studied best linear unbiased estimation in that context. The abstract space setting relates to the approach taken in the current study where our focus involves inverse BM and BB kernels. Carrasco and Florens (2000) also worked with a model that falls into the infinite-dimensional GMM framework. They noted the important limitation that the limiting form of the usual weight matrix (in this setting $\ddot{\Sigma}_n^{-1}$ and $\widetilde{\Sigma}_n^{-1}$) does not necessarily exist as a bounded linear operator, because the associated covariance operator does not satisfy Picard's (1910) conditions for the existence of a linear inverse operator in the limit. Their approach instead uses Tikhonov's regularization, as in ridge regression, so that the inverse operator can be represented in terms of an approximate spectral decomposition of the covariance operator, which involves bias that vanishes asymptotically (see also Kirsch, 1996). Amengual et al. (2020) applied this method to test distributional assumptions by

GMM. Our approach, in contrast to these methods, focuses on BM and BB kernels and explicitly obtains their limiting inverse kernels, which enables the limit behavior of BM-GMM and BB-GMM to be obtained without applying Tikhonov regularization.

Several other high-dimensional studies relate to the present work. Shi (2016) examines estimating a nonlinear structural model similar to that of the present paper but with an approach that maximizes a constrained empirical likelihood and selects only informative moment conditions. Donald and Newey (2001), Bai and Ng (2010) and Belloni, Chen, Chernozhukov, and Hansen (2012) also examine estimating linear structural parameters using high-dimensional moment conditions in environments that differ from the current study, supposing different dimension sizes for the moments compared to the current study. Donald and Newey (2001) assume $s_n^2 = o(n)$ and select the set of instruments to produce an asymptotically efficient GMM estimator based on mean squared error. Bai and Ng (2010) let $s_n = o(n)$ and estimate the parameter by exploiting a factor structure in the data. Belloni et al. (2012) select informative instrumental variables by means of Lasso and Post-Lasso regressions to estimate the unknown parameter by two-stage least squares estimation for $\log(s_n) = o(n^{1/2})$.

2.2 Limit Properties of BM-GMM and BB-GMM

The limit distribution of the BM-GMM and BB-GMM estimators are obtained by deriving the limit behaviors of the components on the right side of (1), viz., $(\bar{G}_n(\theta_*), \nabla_\theta \bar{G}_n(\theta_*), \widehat{\Sigma}_n)$. For this purpose, it is convenient to translate them using functional representation. We first transform $\bar{G}_n(\theta_*)$ into a càdlàg function defined on $[0, 1]$. For each θ, we let $\bar{G}_{n,j}(\theta)$ be the j-th row element of $\bar{G}_n(\theta) \in \mathbb{R}^{s_n}$ and define

$$g_n(u, \theta) := \begin{cases} \bar{G}_{n,j}(\theta), & \text{if } u \in [(j-1)/n, j/n),\ j = 1, 2, \ldots, s_n; \text{ and} \\ 0, & \text{if } u \in [s_n/n, 1]. \end{cases}$$

Note that $\bar{G}_n(\theta)$ has s_n rows, and the above function $g_n(\cdot, \theta)$ is defined by translating $\bar{G}_n(\theta)$ into a function defined on the unit interval, which represents the space of the standardized index.

This standardization makes the weak limit analysis of $\bar{G}_n(\theta_*)$ straightforward. As s_n grows to infinity, it is appropriate to apply the functional central limit theorem (FCLT). Many applications of FCLT methods have appeared in the econometric literature since, in one context, the work of Phillips (1987) on limit theory in unit root time series regression using partial sums, and in another, Cho and White (2011) on limit theory for generalized runs tests. Here we suppose that $\widetilde{g}_n(\cdot, \theta_*) := \sqrt{n} g_n(\cdot, \theta_*)$ weakly converges to a Gaussian stochastic process $\mathcal{G}(\cdot)$ say, such that for a continuous function $\omega(\cdot, \cdot)$ defined on $[0, 1]^2$, $\mathbb{E}[\mathcal{G}(u_1)\mathcal{G}(u_2)] = \omega(u_1, u_2)$. Note that $\mathcal{G}(\cdot)$ is also continuous with probability one if $\omega(\cdot, \cdot)$ is a continuous function. From this weak convergence, the uniform law of large numbers (ULLN) follows for $g_n(\cdot, \theta_*)$ so that $\sup_{u\in[0,1]} |g_n(u, \theta_*)| \to 0$ with probability converging to 1.

We focus on two types of Gaussian processes in this work. We suppose that $\mathcal{G}(\cdot)$ is either a Gaussian process in $\mathcal{C}^{(2)}([0, 1])$ with probability 1, or an Itô process satisfying a stochastic differential equation so that for some $\mu : [0, 1] \times \Omega \mapsto \mathbb{R}$ and $\sigma : [0, 1] \times \Omega \mapsto \mathbb{R}$, $d\mathcal{G}(u) = \mu(u, \mathcal{G}(u))du + \sigma(u, \mathcal{G}(u))d\mathcal{W}(u)$, where $\mathcal{W}(\cdot)$ is a standard Wiener process. As discussed below, BM-GMM and BB-GMM estimators have different asymptotic behavior depending on the path properties of $\mathcal{G}(\cdot)$. If $\mathcal{G}(\cdot)$ is differentiable with probability 1, $\omega(\cdot, \cdot)$ is also differentiable on $[0, 1]^2$, and $\mathcal{G}'(\cdot)$ becomes another Gaussian process. This further implies that $\mathcal{G}''(\cdot)$ is also a Gaussian process. On the other hand, if $\mathcal{G}(\cdot)$ is an Itô process, $\omega(\cdot, \cdot)$ is not differentiable. There are many such Gaussian examples, including Brownian motion and Brownian bridge processes. To simplify notation from now on, we suppress θ_* in $\widetilde{g}_n(\cdot, \theta_*)$, writing $\widetilde{g}_n(\cdot) \equiv \widetilde{g}_n(\cdot, \theta_*)$; and we let $\mu(u)$ and $\sigma(u)$ denote $\mu(u, \mathcal{G}(u))$ and $\sigma(u, \mathcal{G}(u))$, respectively.

We next rewrite $\nabla_\theta \bar{G}_n(\theta_*)$ as a set of functions defined on the unit interval that uniformly converges to a continuous function on the same interval. For $j = 1, 2, \ldots, s_n$ and $j = 1, 2, \ldots, d$, first let $H_n^{(j,i)}$ be the j-th row and i-th column element of $\nabla'_\theta \bar{G}_n(\theta_*) \in \mathbb{R}^{s_n \times d}$, and then for each $i = 1, 2, \ldots, d$ further let

$$H_{n,i}(u) := \begin{cases} H_n^{(j,i)}, & \text{if } u \in [(j-1)/n, j/n),\ j = 1, 2, \ldots, s_n; \text{ and} \\ 0, & \text{if } u \in [s_n/n, 1], \end{cases}$$

and $H_n(\cdot) := [H_{n,1}(\cdot), H_{n,2}(\cdot), \ldots, H_{n,d}(\cdot)]'$. As for $g_n(\cdot)$, $H_{n,i}(\cdot)$ has a jump at each increment of j/n, where $j = 1, 2, \ldots, s_n$. Here, we suppose that the ULLN holds for this stochastic function, so that as n tends to infinity for each j and for a continuous function $H_j(\cdot)$, $\sup_{u\in[0,1]} |H_{n,j}(u) - H_j(u)| \to 0$ with probability converging to 1. We also let $H(\cdot) := [H_1(\cdot), H_2(\cdot), \ldots, H_d(\cdot)]'$. This condition typically holds for many regular cases.

As a final functional reformulation, we translate $\widehat{\Sigma}_n$ as a two-dimensional càdlàg function defined on $[0, 1]^2$ by defining

$$\widehat{\sigma}_n(u_1, u_2) := \begin{cases} \widehat{\Sigma}_n^{(j,i)}, & \text{if } u_1 \in [(j-1)/n, j/n),\ u_2 \in [(i-1)/n, i/n), \text{ and } j, i = 1, 2, \ldots, s_n; \text{ and} \\ 0, & \text{if } u_1 \in [s_n/n, 1] \text{ and } u_2 \in [s_n/n, 1], \end{cases}$$

where $\widehat{\Sigma}_n^{(j,i)}$ is the j-th row and i-th column element of $\widehat{\Sigma}_n$. Note that $\widehat{\sigma}_n(\cdot, \circ)$ is a càdlàg function on $[0, 1]^2$ such that $\widehat{\sigma}_n(\cdot, \circ)$ has a jump at each increment of $(\cdot, j/s_n)$ or $(i/s_n, \circ)$, where $i, j = 1, 2, \ldots, s_n$. For the case of $\ddot{\Sigma}_n$, $\widehat{\Sigma}_n^{(j,i)} = \min[j_n, i_n]$, and we let $\ddot{\sigma}_n(\cdot, \circ)$ denote $\widehat{\sigma}_n(\cdot, \circ)$, which converges to $\ddot{\sigma}(\cdot, \circ) := \min[\cdot, \circ]$ uniformly on $[0, 1] \times [0, 1]$. Figure 1 (a) shows the functional shape of $\ddot{\Sigma}_n$ for $n = 25$, and Figure 1 (b) shows it for $n = 100$. Although it is not continuous for finite n, the limit kernel $\ddot{\sigma}(\cdot, \circ)$ is a continuous function. For the case of $\widetilde{\Sigma}_n$, $\widehat{\Sigma}_n^{(j,i)} = \min[j_n, i_n](1 - \max[j_n, i_n])$, and we let $\widetilde{\sigma}_n(\cdot, \circ)$ denote $\widehat{\sigma}_n(\cdot, \circ)$, which converges to $\widetilde{\sigma}(\cdot, \circ) := \min[\cdot, \circ](1 - \max[\cdot, \circ])$ uniformly on $[0, 1] \times [0, 1]$. Figures 1 (c) and (d) show the functional shapes of $\widetilde{\Sigma}_n$ for $n = 25$ and $n = 100$, respectively. Although it is not continuous for finite n, the limit kernel $\ddot{\sigma}(\cdot, \circ)$ is a continuous function.

The functional representations discussed above make it convenient to represent the associated statistics by an integral transform. For example, if we let $B_n := [b_n(\frac{1}{n}), \ldots, b_n(\frac{s_n}{n})]'$ and $C_n := [c_n(\frac{1}{n}), \ldots, c_n(\frac{s_n}{n})]'$, where $b_n(\cdot) := b(\lfloor n(\cdot) \rfloor / n)$ and $c_n(\cdot) := c(\lfloor n(\cdot) \rfloor / n)$ with $b(\cdot)$ and $c(\cdot)$ being continuous on $[0, 1]$,

$$B_n' \widehat{\Sigma}_n C_n = \sum_{j=1}^{s_n} \sum_{i=1}^{s_n} b_n(j_n) c_n(i_n) \widehat{\Sigma}_n^{(j,i)} = n^2 \int_0^1 \int_0^1 b_n(u_1) \widehat{\sigma}_n(u_1, u_2) c_n(u_2) du_1 du_2.$$

Here, $\lfloor \cdot \rfloor$ denotes the smallest integer greater than the given argument. If we further let $\widehat{\mathbf{\Sigma}}_n$ be an integral operator with kernel $n^2 \widehat{\sigma}_n(\cdot, \circ)$, viz., $\widehat{\mathbf{\Sigma}}_n b_n(\cdot) = n^2 \int_0^1 b_n(u_1) \widehat{\sigma}_n(u_1, \circ) du_1$, we also have $B_n' \widehat{\Sigma}_n C_b = (\widehat{\mathbf{\Sigma}}_n b_n(\cdot), c_n(\cdot))$. Likewise, the inner product representation for the integral transform of the quadratic product can apply to a quadratic product with the weight matrix $\widehat{\Sigma}_n^{-1}$. For an integral operator $\widehat{\mathbf{\Xi}}_n$ with a kernel $\widehat{\xi}_n(\cdot, \circ)$, we suppose that

$$\bar{q}_n(\theta_*) = \bar{G}_n(\theta_*)' \widehat{\Sigma}_n^{-1} \bar{G}_n(\theta_*) = \int_0^1 \int_0^1 g_n(u_1) \widehat{\xi}_n(u_1, u_2) g_n(u_2) du_1 du_2 = (\widehat{\mathbf{\Xi}}_n g_n(\cdot), g_n(\cdot)).$$

Note that $\widehat{\xi}_n(\cdot, \circ)$ corresponds to the kernel of $\widehat{\mathbf{\Sigma}}_n$, viz., $n^2 \widehat{\sigma}_n(\cdot, \circ)$. Likewise, it follows that

$$\nabla_\theta \bar{G}_n(\theta_*) \widehat{\Sigma}_n^{-1} \nabla_\theta' \bar{G}_n(\theta_*) = [\widehat{\mathbf{\Xi}}_n H_n(\cdot), H_n(\cdot)] \quad \text{and} \quad \nabla_\theta \bar{G}_n(\theta_*) \widehat{\Sigma}_n^{-1} \widetilde{G}_n(\theta_*) = [\widehat{\mathbf{\Xi}}_n H_n(\cdot), \widetilde{g}_n(\cdot)]$$

by analogy. Here, $[\widehat{\mathbf{\Xi}}_n H_n(\cdot), H_n(\cdot)]$ denotes the Gramian matrix constructed by $\widehat{\mathbf{\Xi}}_n H_n(\cdot)$ and $H_n(\cdot)$, viz., its j-th row and i-column element is $(\widehat{\mathbf{\Xi}}_n H_{n,j}(\cdot),\ H_{n,i}(\cdot))$. The same interpretation also applies to $[\widehat{\mathbf{\Xi}}_n H_n(\cdot), \widetilde{g}_n(\cdot)]$. Note that these inner products involving the integral transforms are employed to handle the large size of the moment conditions tending to infinity. From now, we specifically let $\widetilde{\mathbf{\Xi}}_n$ and $\ddot{\mathbf{\Xi}}_n$ be the integral operators associated with $\widehat{\Sigma}_n^{-1}$ and $\ddot{\Sigma}_n^{-1}$, respectively, and their limit operators are also denoted as $\widetilde{\mathbf{\Xi}}$ and $\ddot{\mathbf{\Xi}}$, respectively.

The inverse kernel $\widehat{\xi}_n(\cdot, \circ)$ exhibits various asymptotic properties, and the asymptotic behavior of $\widehat{\mathbf{\Xi}}_n$ critically depends on them. Before examining the asymptotic properties of BM-GMM and BB-GMM, we first examine the asymptotic behaviors of $\widetilde{\xi}_n(\cdot, \cdot)$ and $\ddot{\xi}_n(\cdot, \circ)$. The inverse kernel property is substantially different from being continuous. We first focus on $\widetilde{\Sigma}_n^{-1}$ by letting $\widetilde{\xi}_n(\cdot, \circ)$ denote the kernel function of $\widetilde{\mathbf{\Xi}}_n$.

If n is finite, it is not difficult to obtain $\widetilde{\Sigma}_n^{-1}$ and we can obtain the following inverse kernel function:

$$\widetilde{\xi}_n(\cdot,\circ) = n^3\left\{2\widetilde{\mathbb{I}}_n(\cdot,\circ) - \widetilde{\mathbb{J}}_n(\cdot,\circ) - \widetilde{\mathbb{J}}_n(\circ,\cdot)\right\},$$

where

$$\widetilde{\mathbb{I}}_n(\cdot,\circ) := \mathbb{I}\left[(\cdot,\circ)\in\bigcup_{i=1}^{s_n}\left[\frac{i-1}{n}, i_n\right)\times\left[\frac{i-1}{n}, i_n\right)\right] \quad \text{and}$$

$$\widetilde{\mathbb{J}}_n(\cdot,\circ) := \mathbb{I}\left[(\cdot,\circ)\in\bigcup_{i=1}^{s_n}\left[i_n, \frac{i+1}{n}\right)\times\left[\frac{i-1}{n}, i_n\right)\right].$$

Note that $\widetilde{\xi}_n(\cdot,\circ)$ is not uniformly bounded with respect to n. The following lemma delivers the limit behavior of $(\widetilde{\Xi}_n b_n(\cdot), c_n(\cdot))$:

Lemma 1. *Given that $b(\cdot)$ and $c(\cdot)$ are such that $b(0) = c(0) = b(1) = c(1) = 0$,*

(i) (i.a) if $b(\cdot)\in\mathcal{C}^{(2)}([0,1])$, $\widetilde{\Xi}_n b_n(\cdot) = -b''(\cdot) + o(1)$;

(i.b) if it further holds that $c(\cdot)\in\mathcal{C}^{(1)}([0,1])$, $(\widetilde{\Xi}_n b_n(\cdot), c_n(\cdot)) = (b'(\cdot), c'(\cdot)) + o(1)$;

(ii) if $b(\cdot)$ and $c(\cdot)$ are continuous functions with finite second variations, $n^{-1}(\widetilde{\Xi}_n b_n(\cdot), c_n(\cdot)) = (db(\cdot), dc(\cdot)) + o(1)$, where $db(\cdot)$ and $dc(\cdot)$ denote the differentials of $b(\cdot)$ and $c(\cdot)$, respectively. □

Remarks 1. (i) Lemma 1 shows that the limit properties and convergence rate of $(\widehat{\Xi}_n b_n(\cdot), c_n(\cdot))$ depend on the functions attached to the operator.

(ii) Lemma 1 (i.a) shows that the kernel function of $\widetilde{\Xi}$ is given by $-\delta''(\cdot - \circ)$, viz., the negative second-order derivative of the Dirac delta generalized function. Under the condition of Lemma 1 (i.a), we show that $n\widetilde{\Sigma}_n^{-1}B_n \to -b''(\cdot)$ which can be written as $-\int_0^1 \delta''(u-\cdot)b(u)du$.

(iii) Lemma 1 (i.b) follows by the integration by parts. That is, $c(1)b'(1) - c(0)b'(0) = \int_0^1 d\{c(u)b'(u)\} = \int_0^1 c'(u)b'(u)du + \int_0^1 c(u)b''(u)du$. Note that $(b'(\cdot), c'(\cdot)) = -(c(\cdot), b''(\cdot))$ as $c(0) = c(1) = 0$ implies that.

(iv) By the bounded second variation condition, $\int_0^1 (db(u))^2 < \infty$ and $\int_0^1 (dc(u))^2 < \infty$. The sample path generated by the Brownian bridge process satisfies the conditions in Lemma 1 (ii). For example, if we let $\mathcal{B}^0(\cdot)$ be the Brownian bridge process and $b(\cdot) = c(\cdot) = \mathcal{B}^0(\cdot)$, Lemma 1 (ii) implies that $n^{-1}(\widehat{\Xi}_n b_n(\cdot), c_n(\cdot)) \Rightarrow \int_0^1 (d\mathcal{B}^0(u))^2$, which is identical to 1 by noting that $d\mathcal{B}^0(u) = -(1-u)^{-1}\mathcal{B}^0(u)du + d\mathcal{W}(u)$, so that $(d\mathcal{B}^0(u))^2 = du$. If $b(\cdot)$ and $c(\cdot)$ are continuously differentiable, it simply follows that $(db(\cdot), dc(\cdot)) = 0$.

(v) The boundary conditions at zero and one are needed for the desired results. For example, if $b(0)\neq 0$ and $b(1)\neq 0$ such that for some $b_0\neq 0$ and $b_1\neq 0$, $n^2 b_n(0) = b_0 + o(1)$ and $n^2 b_n(1) = b_1 + o(1)$, then it follows that $\ddot{\Xi}_n b_n(\cdot) = g(\cdot) + o(1)$ such that $g(0) = b_0 - b''(0)$, $g(1) = b_1 - b''(1)$, and for $x\in(0,1)$, $g(x) = -b''(x)$ under the conditions in Lemma 1 (i.a). Note that the non-zero boundary conditions modify the limit. □

We next examine $\ddot{\Sigma}_n^{-1}$ by letting $\ddot{\xi}_n(\cdot,\circ)$ denote the kernel function of $\ddot{\Xi}_n$. When n is finite, it follows from $\ddot{\Sigma}_n^{-1}$ that

$$\ddot{\xi}_n(\cdot,\circ) = n^3\left\{\ddot{\mathbb{I}}_n(\cdot,\circ) - \ddot{\mathbb{J}}_n(\cdot,\circ) - \ddot{\mathbb{J}}_n(\circ,\cdot)\right\},$$

where

$$\ddot{\mathbb{I}}_n(\cdot,\circ) := 2\mathbb{I}\left[(\cdot,\circ) \in \bigcup_{i=1}^{s_n-1}\left[\frac{i-1}{n}, i_n\right) \times \left[\frac{i-1}{n}, i_n\right)\right] + \mathbb{I}\left[(\cdot,\circ) \in \left[\frac{n-1}{n}, 1\right] \times \left[\frac{n-1}{n}, 1\right]\right] \quad \text{and}$$

$$\ddot{\mathbb{J}}_n(\cdot,\circ) := \mathbb{I}\left[(\cdot,\circ) \in \bigcup_{i=1}^{s_n-2}\left[i_n, \frac{i+1}{n}\right) \times \left[\frac{i-1}{n}, i_n\right)\right] + \mathbb{I}\left[(\cdot,\circ) \in \left[\frac{n-1}{n}, 1\right] \times \left[\frac{n-2}{n}, \frac{n-1}{n}\right)\right].$$

Note that the structure of $\ddot{\xi}(\cdot,\circ)$ is similarly defined to $\widehat{\xi}_n(\cdot,\circ)$ and it is not bounded, although $\ddot{\mathbb{I}}(\cdot,\circ)$ is not exactly the same as $\widetilde{\mathbb{I}}(\cdot,\circ)$. Mainly, the coefficient of the final diagonal block $\left[\frac{n-1}{n}, 1\right] \times \left[\frac{n-1}{n}, 1\right]$ differs from 1, and the functional values of $\ddot{\xi}_n(1,\circ)$ and $\ddot{\xi}_n(\cdot,1)$ are defined by noting that $s_n = n$, so that it follows that

$$\begin{aligned}\ddot{\xi}_n(\cdot,\circ) &= \widetilde{\xi}_n(\cdot,\circ) + n^3\mathbb{I}\left[(\cdot,\circ) \in \left[\frac{n-1}{n}, 1\right] \times \left[\frac{n-1}{n}, 1\right]\right] \\ &\quad - n^3\mathbb{I}\left[(\cdot,\circ) \in \left[\frac{n-1}{n}, 1\right] \times \left[\frac{n-2}{n}, \frac{n-1}{n}\right)\right] - n^3\mathbb{I}\left[(\circ,\cdot) \in \left[\frac{n-1}{n}, 1\right] \times \left[\frac{n-2}{n}, \frac{n-1}{n}\right)\right].\end{aligned}$$

The following lemma reveals the asymptotic properties of $(\ddot{\Xi}_n b_n(\cdot), c_n(\cdot))$:

Lemma 2. *Given that $b(\cdot)$ and $c(\cdot)$ are such that $b(0) = c(0) = 0$,*

(i) (i.a) if $b(\cdot) \in \mathcal{C}^{(2)}([0,1])$ and $b'(1) = 0$, $\ddot{\Xi}_n b_n(\cdot) = -b''(\cdot) + o(1)$;

(i.b) if it further holds that $c(\cdot) \in \mathcal{C}^{(1)}([0,1])$, $(\ddot{\Xi}_n b_n(\cdot), c_n(\cdot)) = (b'(\cdot), c'(\cdot)) + o(1)$;

(ii) if $b(\cdot)$ and $c(\cdot)$ are continuous functions with finite second variations, $n^{-1}(\ddot{\Xi}_n b_n(\cdot), c_n(\cdot)) = (db(\cdot), dc(\cdot)) + o(1)$. □

Remarks 2. (i) Lemma 2 derives the results of Lemma 1 under different conditions for $c_n(\cdot)$ and $b_n(\cdot)$.

(ii) The result in Lemma 2 (i.a) is consistent with the standard result on the inverse of the Brownian motion kernel. That is, the inverse of the BM-kernel is a kernel for second-order differentiation (e.g., Carrasco et al., 2007). Lemma 2 (i.b) elaborates Lemma 2 (i.a) and obtains the inner product between the derivatives using the inverse BM-kernel.

(iii) The sample path generated by the Brownian motion satisfies the conditions in Lemma 2 (ii). If we let $\mathcal{B}(\cdot)$ be the Brownian motion and $b(\cdot) = c(\cdot) = \mathcal{B}(\cdot)$, we obtain that $n^{-1}(\ddot{\Xi}_n b_n(\cdot), c_n(\cdot)) \Rightarrow \int_0^1 (d\mathcal{B}(u))^2 = \int_0^1 du = 1$.

(iv) Lemma 2 (ii) provides an intuitive interpretation of the BM-kernel and BB-kernel. For example, the BM-kernel is the covariance kernel of an integrated process, so that if we use the integral transform operator using its inverse kernel, it should deliver anti-integrated processes, viz., differentials. Therefore, the quadratic transform using $\ddot{\Sigma}_n^{-1}$ should converge to the inner product between the differentials of $b(\cdot)$ and $c(\cdot)$, viz., $(db(\cdot), dc(\cdot))$. This interpretation also applies to the BB-kernel, and Lemmas 1 (ii) and 2 (ii) provide conditions for the desired result. Furthermore, if $b_n(\cdot)$ and $c_n(\cdot)$ converge to differentiable functions, the quadratic transform converges to the inner product between the derivatives of $b(\cdot)$ and $c(\cdot)$, viz., $(b'(\cdot), c'(\cdot))$. Lemma 2 (i.b) provides conditions for this result, and Lemma 1 (i.b) provides other conditions for the inverse BB-kernel.

(v) As for Lemma 1, the boundary condition at zero is needed for the desired results. For example, if $b(0) \neq 0$ such that for some $b_0 \neq 0$, $n^2 b_n(0) = b_0 + o(1)$, then it follows that $\ddot{\Xi}_n b_n(\cdot) = f(\cdot) + o(1)$

such that $f(0) = b_0 - b''(0)$ and for $x \in (0,1]$, $f(x) = -b''(x)$ under the conditions in Lemma 2 (i.a). Hence, a non-zero boundary condition modifies the limit. $\square$

Lemmas 1 and 2 are key vehicles for delivering the limit properties of BM-GMM and BB-GMM. Note that the two lemmas derive the same results by imposing different conditions on $b(\cdot)$ and $c(\cdot)$. The limits are inner product of derivatives or differentials. In addition, $\widehat{\xi}_n(\cdot,\circ)$ does not converge to a continuous and uniformly bounded function, even if $\widehat{\sigma}_n(\cdot,\circ)$ converges to a uniformly bounded continuous function. For this reason, it becomes difficult to apply the dominated convergence theorem to $(\widehat{\Xi}_n a_n(\cdot), b_n(\cdot))$. Put another way, the BM-kernel and BB-kernel do not satisfy Picard's (1910) conditions requiring a bounded inverse kernel (e.g., Kirsch, 1996; Carrasco et al., 2007).

The properties of $\widetilde{\Sigma}_n^{-1}$ and $\ddot{\Sigma}_n^{-1}$ established in Lemmas 1 and 2 facilitate the derivation of the limit behavior of the BM-GMM and BB-GMM estimators for which the following assumptions are employed.

Assumption 1. *(i) $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space on which the strictly stationary and ergodic sequence $\{W_t \in \mathbb{R}^p : t = 1, 2, \ldots, n\}$ is defined;*

(ii) for each $n \in \mathbb{N}$, $U_n : \mathbb{R}^p \times \Theta \mapsto \mathbb{R}^{s_n}$ ($s_n = n$ or $n-1$) defines the component elements of the moment conditions such that for each n and $\theta \in \Theta$, $U_n(\cdot,\theta)$ is a measurable function, and for each $\omega \in \Omega_0 \in \mathcal{F}$, $U_n(W_t(\omega),\cdot) \in \mathcal{C}^{(2)}(\Theta)$ and $\mathbb{P}(\omega \in \Omega_0) = 1$;

(iii) for each n, there is a unique $\theta_* \in \Theta$ such that θ_* is invariant to n, $\mathbb{E}[U_n(\theta_*)] = 0$, where $\Theta \subset \mathbb{R}^d$ ($d \in \mathbb{N}$) is compact and convex, and $U_{n,t}(\cdot) := U_n(W_t,\cdot)$;

(iv) for each n, $H_{n,*}\widehat{\Sigma}_n^{-1}H_{n,*}'$ is positive definite, where $H_{n,*} := \mathbb{E}[\nabla_\theta U_n(W_t,\theta_*)]$ and $\widehat{\Sigma}_n = \ddot{\Sigma}_n$ or $\widetilde{\Sigma}_n$. $\square$

Assumption 2. *$\widetilde{g}_n(\cdot) \Rightarrow \mathcal{G}(\cdot)$, where $\mathcal{G}(\cdot)$ is a Gaussian stochastic process defined on $[0,1]$ with a continuous covariance kernel $\omega(\cdot,\circ) : [0,1]^2 \mapsto \mathbb{R}$ such that $\mathcal{G}(0) = 0$ with probability 1 and*

(i) $\mathcal{G}(\cdot) \in \mathcal{C}^{(2)}([0,1])$ with prob. 1; or

(ii) $\mathcal{G}(\cdot)$ is an Itô process satisfying the following stochastic differential equation: for some $\mu : [0,1] \times \Omega \mapsto \mathbb{R}$ and $\sigma : [0,1] \times \Omega \mapsto \mathbb{R}$, $d\mathcal{G}(u) = \mu(u)du + \sigma(u)d\mathcal{W}(u)$ such that

(ii.a) $\mu(\cdot,\omega)$ and $\sigma(\cdot,\omega)$ are continuous for each $\omega \in \Omega$;

(ii.b) for each $i = 1, 2, \ldots, n$, $\sigma(i_n)$ is stationary, ergodic, adapted mixingale of size -1 such that for a continuous function $\gamma_1 : [0,1] \mapsto \mathbb{R}^+$ and a continuous and symmetric function $\gamma_2 : [0,1]^2 \mapsto \mathbb{R}$,

$$\mathrm{cov}\left[(\Delta\widetilde{g}_n(i_n))^2, (\Delta\widetilde{g}_n(j_n))^2\right] = \begin{cases} n^{-2}\gamma_1(i_n) + o(n^{-2}), & \text{if } i = j; \\ n^{-3}\gamma_2(i_n, j_n) + o(n^{-3}), & \text{if } i \neq j \end{cases}$$

uniformly in n, where for each $i = 1, 2, \ldots, n$, $\Delta\widetilde{g}_n(i_n) := \widetilde{g}_n(i_n) - \widetilde{g}_n((i-1)/n)$;

(ii.c) $\mathbb{E}[\sigma^2(\cdot)]$ is finite uniformly on $[0,1]$. $\square$

Assumption 3. *For $j = 1, 2, \ldots, d$, there is $H_j(\cdot) \in L^2([0,1])$ such that $H_n(\cdot) \to H(\cdot)$ uniformly on $[0,1]$ with probability converging to 1 with $H(0) = 0$, where $H(\cdot) \in \mathcal{C}^{(2)}([0,1])$ and for continuous $C_j(\cdot) : [0,1] \mapsto \mathbb{R}^d$ ($j = 1, 2, 3, 4$), $\sqrt{n}\Delta H_n(\cdot) = n^{-1/2}C_1(\cdot) + C_2(\cdot)\Delta\widetilde{g}_n(\cdot) + n^{-1/2}C_3(\cdot)\widetilde{g}_n(\cdot)\Delta\widetilde{g}_n(\cdot) + n^{-1}C_4(\cdot)\widetilde{g}_n(\cdot) + O_{\mathbb{P}}(n^{-3/2})$.* $\square$

Remarks 3. (i) The conditions stated in Assumption 2 (i and ii) imply different convergence rates for the variation of $\widetilde{g}_n(\cdot)$, viz., $\Delta\widetilde{g}_n(\cdot)$. Assumption 2 (i) supposes that $\Delta\widetilde{g}_n(\cdot) = O_{\mathbb{P}}(n^{-1})$, whereas Assumption 2 (ii) supposes that $\Delta\widetilde{g}_n(\cdot) = O_{\mathbb{P}}(n^{-1/2})$. Specifically, we can derive that $n\Delta\widetilde{g}_n(\cdot) = \mathcal{G}'(\cdot) + o_{\mathbb{P}}(1)$ and $n(\Delta\widetilde{g}_n(\cdot))^2 = \sigma^2(\cdot) + o_{\mathbb{P}}(1)$ under Assumptions 2 (i and ii), respectively.

(ii) As detailed below, the asymptotic distribution of the BM-GMM or BB-GMM estimator is determined by applying central limit theory (CLT) to the sequence $\{(\Delta\widetilde{g}_n(i_n))^2\}$. Infinite-dimensional MCMD estimation belongs to this case and Assumption 2 (ii.b) is imposed to handle this case. More specifically it provides the condition required to deliver the asymptotic variance of $\sum_{i=1}^{s_n}(\Delta\widetilde{g}_n(i_n))^2$.

(iii) Assumption 3 imposes continuity conditions on $\mu(\cdot)$, $\sigma(\cdot)$, $H(\cdot)$, $C_1(\cdot)$, $C_2(\cdot)$, $C_3(\cdot)$, and $C_4(\cdot)$ to ensure finite integrals of these functions, continuous functions being integrable on a compact set.

(iv) The approximation of $\Delta H_n(\cdot)$ in Assumption 3 is obtained by generalizing the infinite-dimensional MCMD estimation environment. For such a case, it follows that $H_n(\cdot) = H(\widehat{p}_n(\cdot))$, where $\widehat{p}_n(\cdot)$ is the empirical process estimating the cumulative distribution function (CDF) of the data, and it can be treated as a special case of Assumption 3 (ii). Appendix A.1 verifies this assumption for the infinite-dimensional MCMD estimator.

(v) The approximation $\sqrt{n}\Delta H_n(\cdot)$ given in Assumption 3 produces different asymptotic behavior under Assumptions 2 (i and ii). Under Assumption 2 (i), $C_2(\cdot)\Delta\widetilde{g}_n(\cdot) = O_{\mathbb{P}}(n^{-1})$, $n^{-1/2}C_3(\cdot)\widetilde{g}_n(\cdot)\Delta\widetilde{g}_n(\cdot) = O_{\mathbb{P}}(n^{-3/2})$, and $n^{-1}C_4(\cdot)\widetilde{g}_n(\cdot) = O_{\mathbb{P}}(n^{-1})$. Therefore, $n\Delta H_n(\cdot) = C_1(\cdot) + o_{\mathbb{P}}(1)$. We also note that $n\Delta H_n(\cdot) = H'(\cdot) + o_{\mathbb{P}}(1)$ by differentiation, which implies that $H'(\cdot) = C_1(\cdot)$. Further, Assumption 2 (ii) implies that $C_2(\cdot)\Delta\widetilde{g}_n(\cdot) = O_{\mathbb{P}}(n^{-1/2})$, $n^{-1/2}C_3(\cdot)\widetilde{g}_n(\cdot)\Delta\widetilde{g}_n(\cdot) = O_{\mathbb{P}}(n^{-1})$, and $n^{-1}C_4(\cdot)\widetilde{g}_n(\cdot) = O_{\mathbb{P}}(n^{-1})$. Therefore, $n\Delta H_n(\cdot) = C_1(\cdot)+C_2(\cdot)\sqrt{n}\Delta\widetilde{g}_n(\cdot)+o_{\mathbb{P}}(1)$. These different asymptotic behaviors affect the limit distributions of the BM-GMM and BM-GMM estimators in different ways.

(vi) Assumption 3 implies that $n\Delta H_n(\cdot) = C_1(\cdot) + o_{\mathbb{P}}(1)$. We also note that $n\Delta H_n(\cdot) = H'(\cdot) + o_{\mathbb{P}}(1)$ by differentiation, implying that $H'(\cdot) = C_1(\cdot)$. □

We now examine the limit properties of BM-GMM and BB-GMM. For notational simplicity, we first let $\bar{A}_n := [\widehat{\mathbf{\Xi}}_n H_n(\cdot), H_n(\cdot)]$ and $D_n := [\widehat{\mathbf{\Xi}}_n H_n(\cdot), \widetilde{g}_n(\cdot)]$ and give their limit behavior in the following lemma:

Lemma 3. *Given that Assumption 1 holds for $\widehat{\Sigma}_n = \ddot{\Sigma}_n$ or for $\widehat{\Sigma}_n = \widetilde{\Sigma}_n$ with $\mathcal{G}(1) = 0$ and $H(1) = 0$,*

(i) if Assumptions 2 (i) and 3 hold such that $H'(1) = 0$ or $\mathcal{G}'(1) = 0$ for the BM-GMM estimator,

(i.a) $q_n(\theta_*) \Rightarrow \mathcal{Q}_d := (\mathcal{G}'(\cdot), \mathcal{G}'(\cdot))$;

(i.b) $\bar{A}_n \to A_d := [C_1(\cdot), C_1(\cdot)]$ with probability converging to 1;

(i.c) $D_n \Rightarrow \mathcal{D}_d := [C_1(\cdot), \mathcal{G}'(\cdot)]$;

(ii) if Assumptions 2 (ii) and 3 hold,

(ii.a) $n^{-1}q_n(\theta_*) \to q_u := (\sigma(\cdot), \sigma(\cdot))$ with probability converging to 1;

(ii.b) $\bar{A}_n \to A_u := [C_1(\cdot), C_1(\cdot)] + [\sigma(\cdot)C_2(\cdot), \sigma(\cdot)C_2(\cdot)]$ with probability converging to 1;

(ii.c) $D_n - n^{1/2}\sum_{i=1}^{s_n} C_2(i_n)(\Delta\widetilde{g}_n(i_n))^2 \Rightarrow \mathcal{D}_u := [C_1(\cdot), d\mathcal{G}(\cdot)] + [\sigma(\cdot)C_3(\cdot), \sigma(\cdot)\,\mathcal{G}(\cdot)]$;

(ii.d) if it further holds that $[\sigma^2(\cdot), C_2(\cdot)] = 0$ with probability 1 and Γ is positive definite, $D_n \Rightarrow \mathcal{D}_w := \mathcal{Z} + \mathcal{D}_u$, where $\mathcal{Z} \sim \mathcal{N}(0, \Gamma)$, $\Gamma := [\gamma_1^{1/2}(\cdot)C_2(\cdot), \gamma_1^{1/2}(\cdot)C_2(\cdot)] + [\mathbf{\Gamma}_2 C_2(\cdot), C_2(\cdot)]$, and $\mathbf{\Gamma}_2$ is the integral operator with kernel function $\gamma_2(\cdot, \circ)$. □

Remarks 4. (i) Lemma 3 (i) corresponds to the results in Lemmas 1 (ii) and 2 (ii). Therefore, it also follows that $\mathcal{Q}_d = -(\mathcal{G}''(\cdot), \mathcal{G}(\cdot))$, $A_d = -[H''(\cdot), H(\cdot)]$, and $\mathcal{D}_d = -[H''(\cdot), \mathcal{G}(\cdot)]$ by Lemma 1 (i) and 2 (i).

(ii) The process $\mathcal{G}'(\cdot)$ in Lemma 3 (i) is a continuous Gaussian process because the derivative of a Gaussian process is Gaussian, here with covariance kernel $\dot{\omega}(u_1, u_2) := (\partial^2/\partial u_1 \partial u_2)\omega(u_1, u_2)$.

(iii) The variable $\mathcal{D}_d$ is normally distributed. That is, $\mathcal{D}_d \sim \mathcal{N}(0, B_d)$, where $B_d := [\dot{\mathbf{\Omega}} C_1(\cdot), C_1(\cdot)] = \int_0^1 \int_0^1 C_1(u_1)\dot{\omega}(u_1, u_2)C_1(u_2)' du_1 du_2$ by letting $\dot{\mathbf{\Omega}}$ be the integral transform operator with the kernel function $\dot{\omega}(\cdot, \circ)$.

(iv) By virtue of Lemma 3 (ii.c) the BM-GMM and BB-GMM estimators are asymptotically biased unless $\sum_{i=1}^{s_n} (\Delta\widetilde{g}_n(i_n))^2 C_2(i_n) \to 0$ with probability converging to 1. Note that $\sum_{i=1}^{s_n}(\Delta\widetilde{g}_n(i_n))^2 C_2(i_n) \to \int_0^1 \sigma^2(u)C_2(u)du$ with probability converging to 1, so that if the final entity is zero, we can apply the CLT to obtain the limit distribution of $n^{1/2}\sum_{i=1}^{s_n}(\Delta\widetilde{g}_n(i_n))^2 C_2(i_n)$, giving the normal random variable $\mathcal{Z}$ in Lemma 3 (ii.d). Assumption 2 (ii.b) provides regularity conditions to apply the CLT.

(v) The limit distribution of the infinite-dimensional MCMD estimator is obtained by applying Lemma 3 (ii.d). For this application, the functions corresponding to $C_1(\cdot)$, $C_2(\cdot)$, and $C_3(\cdot)$ are found from the model assumption. Further, in the environment of infinite-dimensional MCMD estimation, $\sigma(\cdot) \equiv 1$, $C_1'(\cdot) = C_3(\cdot)$, so that it holds that $\mathcal{D}_u = 0$ by integration by parts. Hence, if $\int_0^1 C_2(u)du = 0$, $\gamma_1(\cdot) = c_1$, and $\gamma_2(\cdot, \circ) = c_2$ for some constants c_1 and c_2, then $\Gamma = c_1[C_2(\cdot), C_2(\cdot)]$ because $\int_0^1\int_0^1 \gamma_2(u, v)C_2(u)C_2(v)' dudv = c_2 \int_0^1 C_2(u)du \int_0^1 C_2(v)'dv = 0$. We demonstrate these properties in Section 2.4. □

In the next step we derive the limit distribution of the BM-GMM and BB-GMM estimators using Lemma 3. Note that standard GMM is typically characterized asymptotically by a Gaussian probability law that is a consequence of applying a CLT to the first-order condition. The limit distributions are obtained by noting that $\sqrt{n}(\widehat{\theta}_n - \theta_*) = -\bar{A}_n^{-1}D_n + o_{\mathbb{P}}(1)$. We provide the asymptotics in the following result.

Theorem 1. *Let Assumption 1 hold for $\widehat{\Sigma}_n = \ddot{\Sigma}_n$ or for $\widehat{\Sigma}_n = \widetilde{\Sigma}_n$ with $\mathcal{G}(1) = 0$ and $H(1) = 0$,*

(i) if Assumptions 2 (i) and 3 hold and A_d is positive definite such that $H'(1) = 0$ or $\mathcal{G}'(1) = 0$ for the BM-GMM estimator, $\sqrt{n}(\widehat{\theta}_n - \theta_*) \Rightarrow -A_d^{-1}\mathcal{D}_d$;

(ii) if Assumptions 2 (ii) and 3 hold and A_u is positive definite such that $\int_0^1 \sigma^2(u)C_2(u)du = 0$ with prob. 1, $\sqrt{n}(\widehat{\theta}_n - \theta_*) \Rightarrow -A_u^{-1}\mathcal{D}_w$. □

Before moving to the next section we consider the effect of permuting moment conditions in the application of BM-GMM and BB-GMM. To fix ideas we focus on BM-GMM estimation by considering $\widetilde{g}_n(\cdot, \theta)$ as the partial sum of $\{\Delta\widetilde{g}_n(\cdot, \theta)\}$, and the BM-GMM estimator $\widehat{\theta}_n$ which minimizes the GMM distance

$$\frac{1}{n}\bar{G}_n(\cdot)'\ddot{\Sigma}_n^{-1}\bar{G}_n(\cdot) = \Delta\bar{G}_n(\cdot)'\Delta\bar{G}_n(\cdot),$$

which holds by (A.36) in the Supplement, where $\Delta\bar{G}_n(\cdot) := [\Delta g_n(\frac{1}{n}, \cdot), \ldots, \Delta g_n(1, \cdot)]'$. Consider another partial sum process denoted $\widetilde{g}_n^p(\cdot, \theta)$ constructed from the variations $\{\Delta\widetilde{g}_n^p(i_n, \theta) : i = 1, 2, \ldots, s_n\}$ by permuting the sequence of the variations, obtaining $\widehat{\theta}_n^p$ by minimizing

$$\frac{1}{n}\bar{G}_n^p(\theta)'\ddot{\Sigma}_n^{-1}\bar{G}_n^p(\theta) = \Delta\bar{G}_n^p(\cdot)'\Delta\bar{G}_n^p(\cdot).$$

By definition, for any permutation, there is a permutation matrix Q such that every row and column contains a single 1 with 0's elsewhere, so that $\Delta\bar{G}_n^p(\cdot) = Q\Delta\bar{G}_n(\cdot)$ and $Q'Q = I$. This implies that $\Delta\bar{G}_n(\cdot)'\Delta\bar{G}_n(\cdot) = \Delta\bar{G}_n^p(\cdot)'\Delta\bar{G}_n^p(\cdot)$, so that $\widehat{\theta}_n = \widehat{\theta}_n^p$, showing that BM-GMM estimation is independent of permutations of the moments. A similar argument applies for BB-GMM. In infinite-dimensional MCMD estimation, moment conditions are sorted from the smallest to the largest, so that no issue of permutation arises in that case.

2.3 Testing Overidentification Using BM-GMM and BB-GMM

For testing overidentification using BM-GMM and BB-GMM we consider the following hypotheses: for every t,

$$\mathcal{H}_0: \text{ for some } \theta_* \in \Theta, \;\; \mathbb{E}[U_{n,t}(\theta_*)] = 0 \;\text{ versus }\; \mathcal{H}_1: \text{ for each } \theta \in \Theta, \mathbb{E}[U_{n,t}(\theta)] \neq 0.$$

Note that $\mathcal{H}_0$ is one of the regularity conditions given in Assumption 1. We consider two types of tests.

First, we follow Sargan (1958) and Hansen (1982) with their motivation for an overidentification test defined as $J_n := q_n(\widehat{\theta}_n)$. In standard cases, the J-test follows a chi-squared null distribution asymptotically under $\mathcal{H}_0$ and is unbounded under $\mathcal{H}_1$. But this null limit distribution is modified if the number of moment conditions tends to infinity. Second, we examine the following standardized J-test:

$$U_n := \frac{J_n - n \cdot q_u}{\sqrt{v_n^2 n}},$$

where v_n^2 is a consistent estimator for $v^2 := \int_0^1 \gamma_1(u)du + \int_0^1\int_0^1 \gamma_2(u,v)dudv$. Here, q_u and v_n can be determined by the model assumptions. For example, in infinite-dimensional MCMD estimation, $q_u = 1$ and $v^2 = 4$ as we show in Section 2.4. The U-test is motivated from Donald et al. (2003) by noting that the J-test may not be bounded under the null hypothesis. Specifically, they examined

$$T_n := \frac{J_n - (s_n - d)}{\sqrt{2(s_n - d)}},$$

and showed that its null limit distribution is a standard normal under their model setup. Note that T_n is defined by supposing that $q_u = 1$ and $v^2 = 2$, but the U-test supposes that q_u and v^2 do not necessarily satisfy this condition. The following result gives the null limit behavior of the tests.

Theorem 2. *Given Assumptions 1,*

(i) if Assumptions 2 (i) and 3 hold and A_d is positive definite, $J_n \Rightarrow \mathcal{J}_d := (\mathbf{\Pi}_d\mathcal{G}(\cdot), \mathcal{G}(\cdot))$ under $\mathcal{H}_0$, where $\mathbf{\Pi}_d := \mathbf{\Xi} - \mathbf{\Lambda}_d$, $\mathbf{\Lambda}_d$ is an integral transform operator with kernel $\lambda(\cdot)'A_d^{-1}\lambda(\circ)$, $\lambda(\circ) := \mathbf{\Xi}H(\cdot)$, and $\mathbf{\Xi}$ is an integral operator such that for $b(\cdot) \in \mathcal{C}^{(2)}([0,1])$, $\mathbf{\Xi}b(\cdot) = -b''(\cdot)$; and

(ii) if Assumptions 2 (ii) and 3 hold for $\widehat{\Sigma}_n = \ddot{\Sigma}_n$ or $\widehat{\Sigma}_n = \widetilde{\Sigma}_n$ with $\mathcal{G}(1) = 0$ and $H(1) = 0$, and A_u is positive definite such that $\int_0^1 \sigma^2(u) C_2(u) du = 0$ with probability 1 and for some v_n, $v_n^2 \to v^2 < \infty$ with probability converging to 1, then $U_n \overset{\text{A}}{\sim} \mathcal{N}(0,1)$ under $\mathcal{H}_0$. □

Remarks 5. (i) Carrasco and Florens (2000) also provide an overidentification test having a structure similar to U_n under their GMM estimation framework.

(ii) The null limit distribution of the J-test in Theorem 2 (i) is provided under the same structure as in Theorem 1 (i). The only difference lies in the fact that the limiting inverse kernel operator is fixed to $\mathbf{\Pi}_d$ in Theorem 2 (i). Note that $\mathcal{J}_d = (\mathcal{G}'(\cdot), \mathcal{G}'(\cdot)) - [H'(\cdot)', \mathcal{G}'(\cdot)] A_d^{-1} [H'(\cdot), \mathcal{G}'(\cdot)]$ by the definition of $\mathbf{\Pi}_d$.

(iii) We cannot apply the J-test when the conditions in Theorem 2 (ii) hold as it converges to a constant q_u as stated in Lemma 3 (ii.a). For this reason we need to apply the U-test for overidentification.

(iv) Finally, the J-test acquires asymptotic power when the weak limit of $\widetilde{g}(\cdot)$ is unbounded in probability. Under $\mathcal{H}_1$, there is no θ such that $\mathbb{E}[U_{n,t}(\theta)] \neq 0$, and it is reasonable to suppose that for some $\nu(\cdot) \in L^2([0,1])$ and $\theta_o \in \Theta$, $\sqrt{n}(g_n(\cdot, \theta_o) - \nu(\cdot)) = O_{\mathbb{P}}(1)$, where θ_o is the probability limit of the BM-GMM or BB-GMM estimator. This supposition lets $J_n = O_{\mathbb{P}}(n)$ under $\mathcal{H}_1$ so that the J-test has nontrivial asymptotic power. The power of U-test is acquired in a similar way to the J-test. □

2.4 Infinite-Dimensional MCMD Estimation

This section uses infinite-dimensional MCMD estimation to illustrate BB-GMM estimation and its limit theory. The large sample properties of the MCMD estimator are modified if observations are from a continuous distribution and GMM estimation is applied. We let x_t be a continuous random variable as before, and $p_t(\theta) := F(x_{(t)}, \theta)$, where $x_{(t)}$ is the t-th smallest realization of an IID data set: $\{x_t : t = 1, \ldots, n\}$.

We first focus on $\sqrt{n}\{\widehat{p}_{n,t} - F(x_{(t)}, \theta)\}$. Note that $\widehat{p}_{n,t} = \widehat{p}_n(p_t) := n^{-1} \sum_{j=1}^n \mathbb{I}(F(x_j, \theta_*) \le p_t)$, where $p_t := p_t(\theta_*)$. Now let $c(\cdot, \theta) := F(F^{-1}(\cdot, \theta_*), \theta)$, and then $F(x_{(t)}, \theta) = c(p_t, \theta)$, so that we have $\sqrt{n}\{\widehat{p}_{n,t} - F(x_{(t)}, \theta)\} = \sqrt{n}\{\widehat{p}_n(p_t) - c(p_t, \theta)\}$. Further, suppose that $c(\cdot, \theta) = (\cdot)$, if and only if $\theta \neq \theta_*$. Therefore, if we let $\sqrt{n}\{\widehat{p}_n(\cdot) - c(\cdot, \theta)\} = \sqrt{n}\{\widehat{p}_n(\cdot) - (\cdot)\} - \sqrt{n}\{c(\cdot, \theta) - (\cdot)\}$, then we have $\widetilde{g}_n(\cdot) := \sqrt{n}(\widehat{p}_n(\cdot) - (\cdot)) \Rightarrow \mathcal{B}^o(\cdot)$, and $\sqrt{n}\{c(\cdot, \theta) - (\cdot)\}$ is not bounded for $\theta \neq \theta_*$.

Next estimate the unknown parameter θ_* by BB-GMM. For this, we let $\widehat{P}_n := [\widehat{p}_{n,1}, \ldots, \widehat{p}_{n,n-1}]'$ and $F_n(\theta) := [F(x_{(1)}, \theta), \ldots, F(x_{(n-1)}, \theta)]'$, and further let

$$\widetilde{\theta}_n := \arg\min_{\theta \in \Theta} \bar{q}_n(\theta), \quad \text{where} \quad \bar{q}_n(\cdot) := \bar{G}_n(\cdot)' \widetilde{\Sigma}_n^{-1} \bar{G}_n(\cdot) \quad \text{and} \quad \bar{G}_n(\cdot) := \widehat{P}_n - F_n(\cdot).$$

Therefore, letting $g_n(\cdot, \theta) := \widehat{p}_n(\cdot) - c_n(\cdot, \theta)$, where

$$c_n(u, \theta) := \begin{cases} c(j_n, \theta), & \text{if } u \in [(j-1)/n, j_n) \text{ for } j = 1, 2, \ldots, n-1; \\ 1, & \text{if } u \in [(n-1)/n, 1], \end{cases}$$

it follows that $\sqrt{n} g_n(\cdot, \theta_*) = \widetilde{g}_n(\cdot)$ and $\bar{q}_n(\theta) = \int_0^1 \int_0^1 g_n(u_1, \theta) \widetilde{\xi}_{n,h}(u_1, u_2) g_n(u_2, \theta) du_1 du_2$. We further note that $c_n(\cdot,\circ) \to c(\cdot,\circ)$ uniformly on $[0,1]\times\Theta$ by the definition of $c_n(\cdot,\circ)$, so that $g_n(\cdot,\circ) \to g(\cdot,\circ) := (\cdot) - c(\cdot,\circ)$ uniformly on $[0,1]\times\Theta$ such that $g_n(0,\cdot) \equiv 0$ and $g_n(1,\cdot) \equiv 0$ uniformly in n, and $g(0,\cdot) \equiv 0$ and $g(1,\cdot) \equiv 0$.

These properties are useful in characterizing the limit properties of infinite-dimensional MCMD estimation. Note that Lemma 1 (i) implies that

$$\begin{aligned} q(\cdot) &= -\int_0^1\int_0^1 \delta''(u_1-u_2)g(u_1,\cdot)g(u_2,\cdot)du_1du_2 = \int_0^1 \{(\partial/\partial u)g(u,\cdot)\}^2du \\ &= \int_0^1\{1-(\partial/\partial u)c(u,\cdot)\}^2du = \int_0^1\{(\partial/\partial u)c(u,\cdot)\}^2du - 1, \end{aligned} \tag{2}$$

where the last equality holds because $\int_0^1 (\partial/\partial u)c(u,\cdot)du = c(1,\cdot)-c(0,\cdot) \equiv 1$. Here, $\int_0^1\{(\partial/\partial u)c(u,\theta)\}^2 du - 1 = \int_0^1\{(\partial/\partial u)u\}^2du - 1 = 0$ if and only if $\theta=\theta_*$. Furthermore, $\int_0^1\{1-(\partial/\partial u)c(u,\cdot)\}^2du$ cannot be less than zero, so that $q(\cdot)$ is minimized at θ_*. From this fact, the GMM estimator must converge to θ_* with probability converging to 1, which implies that

$$\int_0^1\int_0^1 \delta''(u_1-u_2)g(u_1,\theta_*)g(u_2,\theta_*)du_1du_2 = 0. \tag{3}$$

Second, the derivation in (2) can be applied to $\bar{q}_n(\theta_*)$. Note that Lemma 1 (i) implies that

$$\bar{q}_n(\theta_*) = n^{-1}q_n(\theta_*) = \bar{G}_n(\theta_*)'\widehat{\Sigma}_n^{-1}\bar{G}_n(\theta_*) \to \int_0^1\{d\mathcal{B}^0(u)\}^2 = 1 \tag{4}$$

with probability converging to 1, because $\sqrt{n}\bar{G}_n(\theta_*)$ can be translated to $\widetilde{g}_n(\cdot)$ which converges weakly to $\mathcal{B}^o(\cdot)$ so that $\mathcal{B}^0(0) = \mathcal{B}^0(1) = 0$ with probability 1, where the last equality of (4) follows from the fact that $d\mathcal{B}^0(u) = -(1-u)^{-1}\mathcal{B}^0(u)du + d\mathcal{W}(u)$, implying that $\mu(\cdot) = -\mathcal{B}^0(\cdot)(1-(\cdot))^{-1}$ and $\sigma(\cdot)\equiv 1$. Therefore, $n^{-1}q_n(\theta_*) \to 1$ with probability converging to 1, that is the same result as Lemma 3 (ii.a) delivers by noting that $(\sigma(\cdot),\sigma(\cdot)) = 1$. Also, note that $q_n(\theta_*) = O_{\mathbb{P}}(n)$ instead of being $O_{\mathbb{P}}(1)$, due to the non-differentiability of $G_n(\theta_*)$. Furthermore, this limit differs from (3), which is obtained from the limit of $\widehat{\Xi}_n g_n(\cdot,\theta_*)$ first, whereas (4) is obtained by letting $n\to\infty$ after the inner product is first computed.

Third, we examine the asymptotic distribution of the infinite-dimensional MCMD estimator. For finite n, $\sqrt{n}(\widehat{\theta}_n-\theta_*) = -\bar{A}_n^{-1}D_n + o_{\mathbb{P}}(1)$, and $\bar{A}_n := \nabla_\theta' F_n(\theta_*)\widetilde{\Sigma}_n^{-1}\nabla_\theta F_n(\theta_*) = [\widetilde{\Xi}_n H_n(\cdot), H_n(\cdot)]$ such that the j'-th row and j-th column element of $[\widetilde{\Xi}_n H_n(\cdot), H_n(\cdot)]$ is obtained as $(\widetilde{\Xi}_n H_{n,j'}(\cdot), H_{n,j}(\cdot)) = n\sum_{i=1}^{n-1}\Delta H_{n,j'}(i_n)\Delta H_{n,j}(i_n)$ by applying Lemma 1 (ii), where $H_{n,j}(\cdot)$ denotes the j-th row function of $H_n(\cdot) := \nabla_\theta g_n(\cdot,\theta_*)$. In Appendix A.1.2, we separately show that $n\sum_{i=1}^{n-1}\Delta H_{n,j'}(i_n)\Delta H_{n,j}(i_n) \to \int_0^1 H_j'(u)H_{j'}'(u)\{\frac{dc(u,\theta_*)}{du} + \sigma^2(u)\}du$ with probability converging to 1, where $H_j(\cdot)$ and $H_{j'}(\cdot)$ are the j-th row and j'-th row functions of $H(\cdot) := -\nabla_\theta c(\cdot,\theta_*)$, respectively, and $c(\cdot,\theta_*) = (\cdot)$ and $\sigma(\cdot)\equiv 1$, so that it follows that $\frac{dc(\cdot,\theta_*)}{du} + \sigma(\cdot) \equiv 2$, implying that $n\sum_{i=1}^{n-1} \Delta H_{n,j'}(i_n)\Delta H_{n,j}(i_n) \to 2\int_0^1 H_j'(u)H_{j'}'(u)du$ with probability converging to 1. Here, the coefficient 2 exists particularly because $\widetilde{g}(\cdot)$ weakly converges to $\mathcal{B}^0(\cdot)$. If another Gaussian process is associated with the weak limit of $\widetilde{g}_n(\cdot)$, we could have a different limit. It therefore follows that

$$\bar{A}_n = [\widehat{\Xi}_n H_n(\cdot), H_n(\cdot)] \to A := 2[H'(\cdot), H'(\cdot)] \tag{5}$$

with probability converging to 1. This limit result can also be related to Lemma 3 (ii.b). In Appendix A.1.3, we derive the expansion

$$\sqrt{n}\Delta H_n(\cdot) = n^{-1/2}H'(\cdot) + H'(\cdot)\Delta\widetilde{g}_n(\cdot) + n^{-1/2}H''(\cdot)\widetilde{g}_n(\cdot)\Delta\widetilde{g}_n(\cdot) + o_{\mathbb{P}}(1), \tag{6}$$

so that Assumption 3 (ii) holds by letting $C_1(\cdot) = C_2(\cdot) = H'(\cdot)$ and $C_3(\cdot) = H''(\cdot)$. Lemma 3 (ii.b) now leads to $\bar{A}_n \to [C_1(\cdot), C_1(\cdot)] + [\sigma(\cdot)C_2(\cdot), \sigma(\cdot)C_2(\cdot)] = [H'(\cdot), H'(\cdot)] + [H'(\cdot), H'(\cdot)] = 2[H'(\cdot), H'(\cdot)]$ by noting that $\sigma(\cdot) \equiv 1$.

Fourth, the asymptotic distribution of the infinite-dimensional MCMD estimator is obtained from the limit distribution of the GMM score. Note that $D_n := \nabla_\theta' F_n(\theta_*)\widehat{\Sigma}_n^{-1}\sqrt{n}(\widehat{P}_n - F_n(\theta_*)) = [\widehat{\Xi}_n H_n(\cdot), \widetilde{g}_n(\cdot)]$, and the j-th row element of $[\widehat{\Xi}_n H_n(\cdot), \widetilde{g}_n(\cdot)]$ is obtained as

$$\begin{aligned}
(\widehat{\Xi}_n H_{n,j}(\cdot), \widetilde{g}_n(\cdot)) &= n\sum_{i=1}^{n-1} \Delta H_{n,j}(i_n)\,\Delta\widetilde{g}_n(i_n) \\
&= \sum_{i=1}^{n-1} H_j'(i_n)\,\Delta\widetilde{g}_n(i_n) + \sum_{i=1}^{n-1} H_j''(i_n)\,\widetilde{g}_n(i_n)\,(\Delta\widetilde{g}_n(i_n))^2 + \sqrt{n}\sum_{i=1}^{n-1} H_j'(i_n)\,(\Delta\widetilde{g}_n(i_n))^2 + o_{\mathbb{P}}(1)
\end{aligned} \tag{7}$$

by applying (A.6) and (A.8) in Appendix A.1.3 and the fact that $\widehat{p}_n(\cdot)$ converges to $(\cdot)$ uniformly on $[0, 1]$. Here, we note that

$$\sum_{i=1}^{n-1} H_j'(i_n)\,\Delta\widetilde{g}_n(i_n) \Rightarrow \mathcal{Z}_j^{(1)} := \int_0^1 H_j'(u)d\mathcal{B}^0(u) \sim \mathcal{N}(0, (H_j(\cdot), H_j(\cdot))), \tag{8}$$

and

$$\sum_{i=1}^{n-1} H_j''(i_n)\,\widetilde{g}_n(i_n)\,(\Delta\widetilde{g}_n(i_n))^2 \Rightarrow \mathcal{Z}_j^{(2)} := \int_0^1 H_j''(u)\mathcal{B}^0(u)du, \tag{9}$$

using the fact that $\widetilde{g}_n(\cdot) \Rightarrow \mathcal{B}^0(\cdot)$ and $\sum_{i=1}^{n-1} H_j''(i_n)\widetilde{g}_n(i_n)(\Delta\widetilde{g}_n(i_n))^2 = n^{-1}\sum_{i=1}^{n-1} H_j''(i_n)\widetilde{g}_n(i_n) + o_{\mathbb{P}}(1)$. We also note that $\int_0^1 H_j''(u)\mathcal{B}^0(u)du + \int_0^1 H_j''(u)\mathcal{B}^0(u)du = 0$ by integration by parts, so that $\mathcal{Z}_j^{(1)} = -\mathcal{Z}_j^{(2)}$, from which it follows that $\mathcal{Z}_j^{(2)} \sim \mathcal{N}(0, (H_j(\cdot), H_j(\cdot)))$ and the sum of the first two terms on the right side of (7) is negligible in probability. In addition, if we can apply the dominated convergence theorem to obtain

$$\nabla_\theta \int_0^1 \frac{f(F^{-1}(u,\theta_*),\theta)}{f(F^{-1}(u,\theta_*),\theta_*)}du = \int_0^1 \frac{\nabla_\theta f(F^{-1}(u,\theta_*),\theta)}{f(F^{-1}(u,\theta_*),\theta_*)}du,$$

uniformly in θ, then

$$H'(u) = -\nabla_\theta (\partial/\partial u) F(F^{-1}(u,\theta_*),\theta)|_{\theta=\theta_*} = -\left.\frac{\nabla_\theta f(F^{-1}(u,\theta_*),\theta)}{f(F^{-1}(u,\theta_*),\theta_*)}\right|_{\theta=\theta_*},$$

so that applying the dominated convergence theorem implies that

$$\int_0^1 H'(u)du = -\int_0^1 \left.\frac{\nabla_\theta f(F^{-1}(u,\theta_*),\theta)}{f(F^{-1}(u,\theta_*),\theta_*)}\right|_{\theta=\theta_*} du = -\nabla_\theta \int_0^1 \frac{f(F^{-1}(u,\theta_*),\theta_*)}{f(F^{-1}(u,\theta_*),\theta_*)}du = 0, \tag{10}$$

viz., $\int_0^1 H'(u)du = 0$. By this fact, $n^{-1}\sum_{i=1}^{n-1} H'(i_n) \to \int_0^1 H'(u)du = 0$ with probability converging to 1, so that $\sum_{i=1}^{n-1} H'(i_n) = o(n^{-1})$ by theorem 1 (c) of Chui (1971) under the given conditions for the infinite-dimensional MCMD estimator. Further note that $(\Delta\widetilde{g}_n(\cdot))^2 = n\,(\Delta\widehat{p}_n(\cdot))^2 - 2\Delta\widehat{p}_n(\cdot) + n^{-1}$ and that each $\Delta\widehat{p}_n(i_n)$ is an increment of the order statistics constructed by IID uniform random variables (e.g., David and Nagaraja, 2003, section 6.4) which is referred to as the elementary coverage or the spacing (e.g., Wilks, 1948; Rao and Kuo, 1984). In addition, $(\Delta\widehat{p}_n(\frac{1}{n}),\ldots,\Delta\widehat{p}_n(1))'$ follows a Dirichlet distribution with parameter ι_n. Using this condition, Appendix A.3.2 shows that for each i, $\mathbb{E}[(\Delta\widetilde{g}_n(i_n)^2] = \frac{n-1}{n(n+1)}$. Hence, we can rewrite the second term of (7) as

$$\begin{aligned}\sqrt{n}\sum_{i=1}^{n-1} H_j'(i_n)\,(\Delta\widetilde{g}_n(i_n))^2 &= \sqrt{n}\sum_{i=1}^{n-1} H_j'(i_n)\left\{(\Delta\widetilde{g}_n(i_n))^2 - \frac{n-1}{n(n+1)}\right\} + o_{\mathbb{P}}(1)\\ &= \sqrt{n}\sum_{i=1}^{n-1} H_j'(i_n)\left[\left(\sqrt{n}\Delta\widehat{p}_n(i_n)\right)^2 - \frac{2n}{n(n+1)}\right] - 2\sqrt{n}\sum_{i=1}^{n-1} H_j'(i_n)\left(\Delta\widehat{p}_n(i_n) - \frac{1}{n}\right) + o_{\mathbb{P}}(1).\end{aligned}$$

Here, (8) implies that $\sum_{i=1}^{n-1} H_j'(i_n)\sqrt{n}\left(\Delta\widehat{p}_n(i_n) - \frac{1}{n}\right) = \sum_{i=1}^{n-1} H_j'(i_n)\,\Delta\widetilde{g}_n(i_n) \Rightarrow \mathcal{Z}_j^{(1)} := \int_0^1 H_j'(u)\,d\mathcal{B}^o(u)$, and we show that $\mathrm{var}[\sqrt{n}\sum_{i=1}^{n-1} H_j'(i_n)\{(\sqrt{n}\Delta\widehat{p}_n(i_n))^2 - \frac{2n}{n(n+1)}\}] \to 20(H_j'(\cdot), H_j'(\cdot))$ in Appendix A.3.1. Therefore, if we apply the CLT, it follows that

$$\sqrt{n}\sum_{i=1}^{n-1} H_j'(i_n)\,(\Delta\widetilde{g}_n(i_n))^2 \Rightarrow \mathcal{Z}_j^{(3)} - 2\mathcal{Z}_j^{(1)}, \tag{11}$$

where $\mathcal{Z}_j^{(3)} \sim \mathcal{N}(0, v_j^2)$, and $v_j^2 := 20(H_j'(\cdot), H_j'(\cdot))$. We now plug (8), (9), (11) into (7) to deduce that $(\widehat{\boldsymbol{\Xi}}_n H_{n,j}(\cdot), \widetilde{g}_n(\cdot)) \Rightarrow \mathcal{Z}_j^{(3)} - 2\mathcal{Z}_j^{(1)} \sim \mathcal{N}(0, 8(H_j'(\cdot), H_j'(\cdot)))$ by noting that $\mathbb{E}[\mathcal{Z}_j^{(1)}, \mathcal{Z}_j^{(3)}] = 4(H_j'(\cdot), H_j'(\cdot))$ as verified in Appendix A.3.3.

We now extend this result to $[\widehat{\Xi}_n H_n(\cdot), \widetilde{g}_n(\cdot)]$:

$$[\widehat{\Xi}_n H_n(\cdot), \widetilde{g}_n(\cdot)] \Rightarrow \mathcal{D} := \mathcal{Z}^{(3)} - 2\mathcal{Z}^{(1)} \sim \mathcal{N}(0, 8[H'(\cdot), H'(\cdot)]), \tag{12}$$

where $\mathcal{Z}^{(1)}$ and $\mathcal{Z}^{(3)}$ are the weak limits of $\sum H'(i_n)\Delta\widetilde{g}_n(i_n)$ and $\sqrt{n}\sum H'(i_n)\{(\sqrt{n}\Delta\widehat{p}_n(i_n))^2 - \frac{2n}{n(n+1)}\}$, respectively. The weak limit on the right side can be associated with Lemmas 3 (ii.c and ii.d). Given that $\sigma^2(\cdot) \equiv 1$, $\widetilde{g}_n(\cdot) \Rightarrow \mathcal{B}^0(\cdot)$, and $C_1(\cdot) = C_2(\cdot) = H'(\cdot)$ and $C_3(\cdot) = H''(\cdot)$ from (6), we note that $[\sigma^2(\cdot), C_2(\cdot)] = 0$ by (10). Therefore, $[\widehat{\Xi}_n H_n(\cdot), \widetilde{g}_n(\cdot)] \Rightarrow \mathcal{Z} + [H'(\cdot), d\mathcal{B}^0(\cdot)] + [H''(\cdot), \mathcal{B}^0(\cdot)]$ by Lemmas 3 (ii.c and ii.d). Here, $[H'(\cdot), d\mathcal{B}^0(\cdot)] + [H''(\cdot), \mathcal{B}^0(\cdot)] = 0$ by the integration by parts, and this implies that $\mathcal{Z} = \mathcal{D}$ in (12), so that $\Gamma = 8[H'(\cdot), H'(\cdot)]$. This result can be affirmed by deriving that $\gamma_1(\cdot) \equiv 8$ and $\gamma_2(\cdot, \circ) \equiv 4$ through some tedious algebra. Using this, $\int_0^1\int_0^1 \gamma_2(u,v)C_2(u)C_2(v)'dudv$ of Lemma 3 (ii.d) is identical to $4\int_0^1 H'(u)du\int_0^1 H'(u)'du = 0$ by (10), so that $\Gamma = 8\int_0^1 H'(u)H'(u)'du$, which is $8[H'(\cdot), H'(\cdot)]$.

Fifth, using the third and fourth findings above the limit distribution of the infinite-dimensional MCMD estimator can be obtained as follows:

$$\sqrt{n}(\widehat{\theta}_n - \theta_*) = -\bar{A}_n^{-1}D_n + o_{\mathbb{P}}(1) \Rightarrow -A^{-1}\mathcal{D} \sim \mathcal{N}(0, 2[H'(\cdot), H'(\cdot)]^{-1}), \tag{13}$$

where we combined (5) and (12) to obtain the limit.

Finally, as the number of moment conditions tends to infinity, we cannot use the J-test statistic in this case. Specifically, we note that $J_n = q_n(\widehat{\theta}_n) = q_n(\theta_*) - \frac{1}{2}\bar{D}_n'\bar{A}_n^{-1}\bar{D}_n + o_{\mathbb{P}}(1)$ by applying a second-order Taylor expansion, where $\bar{D}_n := \nabla_\theta\bar{G}_n(\theta_*)\widehat{\Sigma}_n^{-1}\bar{G}_n(\theta_*) = O_{\mathbb{P}}(n^{-1/2})$ by (12) and $\bar{A}_n = O_{\mathbb{P}}(1)$, so that $J_n = q_n(\theta_*) + O_{\mathbb{P}}(n^{-1})$. Furthermore, note that $q_n(\theta_*) - n = n\{\sum_{i=1}^{n-1}(\Delta\widetilde{g}_n(i_n))^2 - 1\} = n\{n\sum_{i=1}^{n-1}(\Delta\widehat{p}_n(i_n))^2 - 2\} + o_{\mathbb{P}}(n^{-1})$, where the first equality follows from (A.31) in Appendix A.1, and the second equality holds by the fact that $\widetilde{g}_n(\cdot) := \sqrt{n}(\widehat{p}_n(\cdot) - (\cdot))$. We here note that $n^2\sum_{i=1}^{n-1}(\Delta\widehat{p}_n(i_n))^2$ is the goodness-of-fit test proposed by Greenwood (1946) with each $\Delta\widehat{p}_n(i_n)$ being the elementary coverage or spacing described above. Using the distributional condition of the elementary coverage, Appendix A.3.2 shows that $\mathrm{var}(\sum_{i=1}^{n-1}(\Delta\widetilde{g}_n(i_n))^2 - 1) = 4n^{-1} + o(n^{-1})$, implying that $\sqrt{n}\{\sum_{i=1}^{n-1}(\Delta\widetilde{g}_n(i_n))^2 - 1\} \overset{\text{A}}{\sim} \mathcal{N}(0,4)$, so that $(q_n(\theta_*) - n)/\sqrt{n} \overset{\text{A}}{\sim} \mathcal{N}(0,4)$. Therefore, it now follows that $(J_n - n)/\sqrt{4n} \overset{\text{A}}{\sim} \mathcal{N}(0,1)$ under $\mathcal{H}_0$. Note that this is nothing other than the U-test obtained by letting $q_u = 1$ and $v_n^2 = 4$, viz., $U_n = (J_n - n)/\sqrt{4n}$, whose null limit distribution is $\mathcal{N}(0,1)$ as Theorem 2 (ii) affirms.

Regarding overidentification testing we make two remarks. First, the T-test proposed by Donald et al. (2003) does not follow the standard normal distribution under the null hypothesis. That is, T_n follows

$\mathcal{N}(0,2)$ because $T_n = \sqrt{2}U_n + o_{\mathbb{P}}(1)$ under the null. We also note that the T-test could be useful if the martingale difference array (MDA) CLT could have been applied to $J_n - n$. Although this is not feasible under the current data generating process (DGP) condition, if we could let n of $(\Delta\widetilde{g}_n(\cdot))^2$ tend to infinity before applying the CLT, we could approximate $(\Delta\widetilde{g}_n(\cdot))^2$ by $(\Delta\mathcal{W}(\cdot))^2 + o_{\mathbb{P}}(1)$, and it would follow that $\sqrt{n}\{\sum_{i=1}^{n-1}(\Delta\mathcal{W}(i_n))^2 - 1\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n-1}\{n(\Delta\mathcal{W}(i_n))^2 - 1\}$. Note here that $n(\Delta\mathcal{W}(i_n))^2 - 1$ is an MDA, so that the MDA CLT leads to $\sqrt{n}\{\sum_{i=1}^{n-1}(\Delta\mathcal{W}(i_n))^2 - 1\} \overset{\text{A}}{\sim} \mathcal{N}(0,2)$. This property implies that if $J_n - n$ were an MDA, the T-test would follow a standard normal asymptotically under the null hypothesis. Nonetheless, the sample size n of $(\Delta\widetilde{g}_n(\cdot))^2$ has to tend to infinity along with the application of the CLT, making it impossible to apply the MDA CLT here. It is necessary to take the serial correlation structure of $(\Delta\widetilde{g}_n(\cdot))^2$ into account when applying the CLT, ensuring that the U-test is a valid test asymptotically.

Second, the U-test is distribution-free. Note that the same null hypothesis is commonly tested by the KS-test, and its null limit distribution is affected by parameter estimation error, which makes its application inconvenient (e.g., Durbin, 1973). It is more difficult to compute the critical values of the test than computing test value itself as a resampling method has to be applied. In contrast, the null limit distribution of U-test is standard normal, and so its critical value can be conveniently obtained.

3 Monte Carlo Simulation

Using the infinite-dimensional MCMD estimator simulations were conducted designed to corroborate the properties of the GMM estimation, addressing specifically Theorems 1 (ii) and 2 (ii). For this purpose, three different DGP conditions were considered, with x_t following exponential, Pareto, or normal distributions. The corresponding hypotheses were tested using the U-test and the limit distribution of the infinite-dimensional MCMD estimator was leveraged to test hypotheses on the unknown parameter.

The plan for the simulation follows. First, if x_t follows an exponential distribution, then $\mathbb{P}(x_t \leq x) = 1 - \exp(-\theta_* x)$, denoted as $x_t \sim \text{Exp}(\theta_*)$. Likewise, if x_t follows a Pareto distribution bounded from 1, we have $\mathbb{P}(x_t \leq x) = 1 - (1/x)^{\theta_*}$, denoted as $x_t \sim \text{Pa}(\theta_*, 1)$. Finally, for $x_t \sim \mathcal{N}(\theta_*, 1)$, we have $\mathbb{P}(x_t \leq x) = \Phi(x - \theta_*)$, where $\Phi(\cdot)$ is the standard normal CDF. The unknown parameter is estimated by infinite-dimensional MCMD using

$$\widehat{\theta}_n := \underset{\theta\in\Theta}{\arg\min}(\widehat{P}_n - F_n(\theta))'\widehat{\Sigma}_n^{-1}(\widehat{P}_n - F_n(\theta)).$$

For the exponential distribution the j-th row element of $F_n(\theta)$ is given be $1 - \exp(-\theta x_{(j)})$; for the Pareto distribution the j-th row element of $F_n(\theta)$ is $1 - \left(1/x_{(j)}\right)^{\theta}$; and for the normal distribution the j-th row element of $F_n(\theta)$ is $\Phi(x_{(j)}-\theta)$. Here, $x_{(j)}$ denotes the j-th smallest realization of the set of IID observations $\{x_1, x_2, \ldots, x_n\}$.

Second, using these distributional hypotheses we obtain

$$H(p) = \begin{cases} \frac{1}{\theta_*}(1-p)\log(1-p), & \text{for exponential and Pareto distributions;} \\ \phi(\Phi^{-1}(p)), & \text{for the normal distribution,} \end{cases}$$

if the models are correctly specified, where $\phi(\cdot)$ is the probability density function (PDF) of a standard normal distribution. In all cases, $\lim_{p\to 0} H(p) = 0$, $\lim_{p\to 1} H(1) = 0$. Furthermore,

$$\int_0^1 (H'(u))^2 du = \begin{cases} \frac{1}{\theta_*^2}, & \text{for exponential and Pareto distributions;} \\ 1, & \text{for the normal distribution.} \end{cases}$$

Therefore, it follows that

$$\sqrt{n}(\widehat{\theta}_n - \theta_*) \overset{\text{A}}{\sim} \begin{cases} \mathcal{N}(0, 2\theta_*^2), & \text{for exponential and Pareto distributions;} \\ \mathcal{N}(0, 2), & \text{for the normal distribution} \end{cases}$$

by (13), so that if we let

$$t_n := \begin{cases} \frac{\sqrt{n}(\widehat{\theta}_n - c)}{\sqrt{2\widehat{\theta}_n^2}}, & \text{for exponential and Pareto distributions;} \\ \frac{\sqrt{n}(\widehat{\theta}_n - c)}{\sqrt{2}}, & \text{for the normal distribution,} \end{cases}$$

it follows $\mathcal{N}(0,1)$ under the joint hypothesis that $\theta_* = c$ and that the distributional condition is correct. For all cases, $U_n \overset{\text{A}}{\sim} \mathcal{N}(0,1)$ under the same conditions. If the t-test rejects the null, it is not evident which condition is violated out of the jointed hypotheses. To address this issue, we apply both the U-test and the t-test to test the distributional hypothesis. If the U-test rejects the distributional hypothesis, inference from the t-test is not informative. But if the t-test rejects the null while the U-test does not, it is evident that $\theta_* \neq c$. This combined approach provides a more comprehensive assessment of the hypotheses.

Third, we conduct simulations by supposing null DGPs for testing the hypotheses of the U- and t-tests. For this purpose, we let $\theta_* = 1$, 1, and 0, so that $x_t \sim \text{Exp}(1)$, $x_t \sim \text{Pa}(1, 1)$, and $x_t \sim \mathcal{N}(0, 1)$ for the exponential, Pareto, and normal cases. These parameter values are selected arbitrarily for the simulation study. In each DGP environment, we compute the empirical rejection rates of the U- and t-tests for significance levels of 1%, 5%, and 10% with 10,000 independent repetitions. The empirical rejection rates under the null hypothesis for each test are reported in Tables 1 and 2.

Fourth, we compare the U- and t-tests with corresponding tests defined by Tikhonov's methodology. In particular, applying theorem 10 in Carrasco and Florens (2000), we have the following τ-test as the one that corresponds to the U-test:

$$\tau_n := \frac{\dot{J}_n - \dot{p}_n}{\sqrt{\dot{q}_n}},$$

where $\dot{J}_n := n\dot{q}_n(\dot{\theta}_n; \alpha_n)$, $\dot{\theta}_n := \arg\min_{\theta\in\Theta} \dot{q}_n(\theta; \alpha_n)$,

$$\dot{q}_n(\theta; \alpha_n) := (\widehat{P}_n - F_n(\theta))'\widehat{\Sigma}_n^{1/2}(\widehat{\Sigma}_n + \alpha_n I)^{-1}\widehat{\Sigma}_n^{1/2}(\widehat{P}_n - F_n(\theta)), \quad \dot{p}_n := \sum_{j=1}^{n-1} \widehat{a}_j, \quad \dot{q}_n := 2\sum_{j=1}^{n-1} \widehat{a}_j^2,$$

and $\widehat{a}_j := \widehat{\lambda}_j^2/(\widehat{\lambda}_j^2 + \alpha_n)$ such that $\widehat{\lambda}_j$ is the j-th largest eigenvalue of $\widehat{\Sigma}_n$. Note that this test is identical to T_n if $\alpha_n = 0$ for every n. Following theorem 10 in Carrasco and Florens (2000), we let $\alpha_n = n^{-1/4}$ for asymptotic optimality of the test and it is then asymptotically standard normal under the null. In addition, we define the following test as the one corresponding to the t-test:

$$t_n' := \begin{cases} \frac{\sqrt{n}(\dot{\theta}_n - c)}{\sqrt{\widehat{\theta}_n^2}}, & \text{for exponential and Pareto distributions;} \\ \sqrt{n}(\dot{\theta}_n - c), & \text{for the normal distribution,} \end{cases}$$

This test is defined by noting that theorem 8 in Carrasco and Florens (2000) implies that $\sqrt{n}(\dot{\theta}_n - \theta_*) \overset{\text{A}}{\sim} \mathcal{N}(0, \int_0^1 (H'(u))^2 du)$. Our earlier derivations show that $\int_0^1 (H'(u))^2 du = \theta_*^2$ for exponential and Pareto distributions, and $\int_0^1 (H'(u))^2 du = 1$ for the normal distribution. The asymptotic variances are straightforwardly obtained by applying Lemma 1 (i.b) to theorem 8 of Carrasco and Florens (2000), and letting $\alpha_n = n^{-1/4}$ satisfies the condition in theorem 8. Under the null, t_n' is then asymptotically standard normal. The null simulation results are summarized as follows.

(a) For each case, as the sample size n increases, the distribution of the U-test converges to the standard normal. Table 1 demonstrates that the empirical rejection rates are close to 1%, 5%, and 10% for the exponential, Pareto, and normal distribution cases when $n = 1,000$. This observation confirms that the U-test follows the null limit distribution predicted in Theorem 2 (ii). The empirical distributions of the U-test provide further support: the left column of Figure 2 displays these distributions and in each case the empirical distribution approaches the CDF of the standard normal as n increases.

(b) As the sample size n increases the distribution of the t-test converges to the standard normal. Table 2 shows that the empirical rejection rates are close to 1%, 5%, and 10% for the exponential, Pareto, and normal distribution cases when $n = 200$, 300, 400, 500, and $1,000$. These results corroborate the limit theory of the t-test under the null given in Theorem 1 (ii). Furthermore, the standard normal provides a better approximation of the distribution of the t-test compared to the U-test.

(c) The empirical distribution of the infinite-dimensional MCMD estimator provides further support. The right column of Figure 2 displays the empirical distributions of the infinite-dimensional MCMD estimators and these evidently closely approach the $\mathcal{N}(0,2)$ CDF as n increases. It is worth noting that in the normal distribution case, even when n is as small as 100 the empirical distribution of the infinite-dimensional MCMD estimator is well approximated by the distribution of $\mathcal{N}(0,2)$.

(d) When comparing the U-test with the τ-test, it is apparent that the empirical rejection rates of the U-test converge to the nominal levels faster than the τ-test. When n is small, the level distortions of the τ-test are large. Even for $n = 1,000$, the empirical rejection rates of τ-test are still far from nominal levels, although they appear to be converging to nominal levels. These findings indicate that the U-test controls type-I errors better than the τ-test.

(e) Comparison of the t-test and t'-test results shows that the empirical rejection rates of the t-test also converge to the nominal levels faster than the t'-test. Although the finite sample distortions are not as severe as the U-test, the t-test controls type-I errors better than the t'-test. □

Finally, simulations were conducted to examine the local power properties of the U- and t-tests. For this purpose, the DGP conditions were modified as follows: (i) for the exponential case, $x_{t,n} := y_t + \frac{1}{2}\sqrt{z_t/n}$, with $y_t \sim \text{Exp}(1)$ and $z_t \sim \mathcal{U}[0.5, 1.5]$; (ii) for the Pareto case, $x_{t,n} := y_t + \frac{1}{2}\sqrt{z_t/n}$, with $y_t \sim \text{Pa}(1,1)$ and $z_t \sim \mathcal{U}[0.5, 1.5]$; and (iii) for the normal case, $x_{t,n} := y_t + \frac{1}{4}y_t^4/\sqrt{n}$, with $y_t \sim \mathcal{N}(0,1)$. Importantly, as n increases the empirical distribution of $x_{t,n}$ gets closer to that of y_t but the finite sample distribution of $x_{t,n}$ is not the same as that of y_t in each case. Local powers of the U- and t-tests are examined together with those of the τ- and t'-tests. Similar to the null simulations, 10,000 independent experiments were conducted under the local alternatives and empirical rejection rates of the tests are reported in Tables 3 and 4. These results provide insights into the performance of the tests under local departures from the null and are summarized as follows.

(a) The U-test demonstrates non-negligible local power in each case. As the sample size n increases, the empirical rejection rates of the U-test exceed the nominal significance levels, indicating that the U-test exhibits local power. Notably, the empirical local power of the U-test remains relatively stable as n increases for the normal case compared to the exponential and Pareto cases. This suggests that the U-test performs well in detecting local departures from the null hypothesis, especially for the normal distribution case.

(b) The t-test also exhibits non-negligible local power in each case. As the sample size n increases, the empirical rejection rates of the t-test tend to exceed the nominal significance levels, again indicating that the t-test exhibits local power. Similar to the U-test, the empirical local power of the t-test remains relatively stable for the normal distribution case across different sample sizes, suggesting that the t-test performs well in detecting local departures from the null hypothesis, particularly for the normal distribution case.

(c) When comparing powers of the U- and τ-tests, the empirical rejection rates of the U-test are higher than those of the τ-test. The same outcome is observed in the comparison between the t- and t'-tests. Notably, the empirical local powers of the t-test are higher than t'-test for all three distributions. □

This simulation exercise comparing our approach with tests based on Tikhonov regularization shows that use of the exact inverse operator for the BB-kernel reduces finite sample size distortion and increases local power.

4 Empirical Application

This section reports the findings of an empirical implementation of infinite-dimensional MCMD estimation to examine distributional hypotheses concerning labor income data in the U.S. The Pareto distribution has been popular in research on income distributions throughout a large body literature. Since Kuznets (1953, 1955) first examined top income shares in U.S. income data, this statistic has commonly been used for an income inequality index supplementing the Gini coefficient, as the latter does not focus on income inequality associated with the tail of the distribution. In particular, Piketty (2003), Piketty and Saez (2003), Atkinson (2005, 2007), Atkinson and Leigh (2007, 2008), and Moriguchi and Saez (2008), among others, use the Pareto distribution in measuring the top x-percent income shares of many countries such as Australia, France, Japan, New Zealand, U.K., and the U.S. over long periods to reveal how the estimated top income shares have evolved over time. The findings indicate that the top income shares have increased since the 1970s, signaling a general deterioration in income equality. The results are based on the Pareto distribution assumption and top income share estimates obtained from the methodologies in these studies could be biased unless the Pareto distribution condition holds for the data.

Our empirical application is motivated by much ongoing research on income distributions and inequality where there is a need to test underlying distributional assumptions on which empirical findings are often based. In particular, we utilize this paper's infinite-dimensional MCMD methodology to test the Pareto distribution hypothesis and investigate the evolution of income inequality in the U.S over time. Previous studies, such as Piketty and Saez (2003), have pointed out that the recent increase in top income shares is primarily driven by the rise of the capital income share. This means that a small segment of the population has a significant proportion of total income, mainly by way of capital income. Apart from capital income share, Piketty and Saez (2003) highlight a persistent rise of labor income inequality in the U.S. since the 1970s. This aspect of labor income inequality has also been studied by Katz and Murphy (1995); Katz and Autor (1999); Ciccone and Peri (2005); Eisenbarth and Chen (2022) and others, who examine labor market inequality by analyzing the distribution of wage structures. Their findings consistently show that wage inequality has continuously increased in the U.S. labor market. For instance, Katz and Autor (1999) report that earnings inequality has risen for both males and females, and wage disparities based on education, occupation, and age have also widened. Moreover, wage dispersion has expanded within demographic and skill groups. These studies provide valuable insights into the dynamics of labor income inequality in the U.S.

Our main focus is to investigate how labor income inequality has evolved within the same cohort over time. The overall distribution of the wage structure is influenced not only by the inherent demand characteristics of the labor market but also by various heterogeneous factors, such as race, gender, education, and other determinants. This observation leads us to examine labor market inequality by isolating and removing the effects of heterogeneity from the data. By doing so, we aim to gain a better understanding of the contributions of these heterogeneous effects to overall income inequality. This involves comparing inequality indices obtained from different cohorts to discern the changes in labor income inequality over time while accounting for the impact of various demographic and socioeconomic factors. By controlling for these factors and focusing on the evolution of labor income inequality within specific cohorts, we can uncover important insights into the dynamics of income inequality in the U.S. labor market.

Our empirics utilize the Continuous Work History Sample (CWHS) database, which contains annual labor income data before tax for 15,000 individuals in the U.S. from 1980 to 2018. The individuals in the CWHS database were born between 1960 and 1962, and their gender, education, and race information is also provided. Leveraging this information, we classify the observations into cohorts based on gender, education, and race to ensure that individuals within the same cohort share certain degrees of homogeneity without losing a significant number of observations.

Table 5 presents the distribution of the cohorts categorized by gender, education, and race. A similar data analysis was conducted by d'Albis and Badji (2022) using French data, albeit with a different research objective. Their study used aggregate data pertaining to the same generation to estimate the functional shape of Gini coefficients over time and evaluate income inequality within that generation. Our empirical research goal differs in that we utilize the cohort data sets to focus on labor income inequality dynamics within specific cohorts in the U.S. labor market.

To do so the infinite-dimensional MCMD methodology of Section 2.4 is used to examine the labor income data sets. Our focus is top income shares as in Piketty and Saez (2003) and Atkinson et al. (2011). But prior to computing the top income shares we test the hypothesis that the labor income data sets follow a Pareto law. The test is applied for each annual labor income data set that belongs to the same cohort and the Pareto parameter is fitted using the infinite-dimensional MCMD estimator. If the U-test does not reject the Pareto distribution condition, then we proceed to compute the top 5% income share of each data set and apply the methodology developed by Piketty and Saez (2003) and Atkinson et al. (2011). This approach ensures that we apply the top income shares methodology only when the data satisfy the Pareto distribution condition as revealed by the U-test.

We conduct specific procedures for each cohort in the CWHS data, classified by the following characteristics: gender (female and male); education (high school or below, which also includes some college but no degree cases, Bachelor (BA) or equivalent degrees which includes associate degrees, Master (MA) or equivalent degrees, and Doctorate or equivalent degrees, which includes professional degrees); and race (white or Caucasian, black or African American, Asian, and others including American Indian, Native Hawaiian or other Pacific Islander, and two or more race individuals).

For each cohort data, we test the Pareto distribution hypothesis using the U-test. The null hypothesis $\mathcal{H}_0$ is formulated as follows:

$$\mathcal{H}_0 : \mathbb{P}(y_t \leq y) = 1 - \left(\frac{b_x}{y}\right)^{\theta},$$

where y_t denotes the t-th individual's labor income, and b_x represents the minimum value of the income variable, ensuring that y_t is distributed on $[b_x, \infty)$. The Pareto distribution hypothesis is intended to capture the right tail distribution of the income data. Naturally, if all observations were used in hypothesis testing, it is likely that most income data would reject the Pareto distribution due to the differences in the left tail distribution that are not directly material to the top income shares. Hence, our focus is on the top 10% of income observations and the top 5% income shares are estimated from each cohort data.

Application of the U-test for each cohort using the data sets from 1980 to 2018 leads to the findings reported in Table 6. The table provides information for each cohort, including the sample size of the top 10% labor income data sets and the number of data sets that do not reject the Pareto distribution hypothesis. For instance, for the female cohort born in 1960, there are 260 top 10% labor income observations, and out of the 39 data sets between 1980 and 2018, 27 of them do not reject the Pareto distribution hypothesis at the 1% significance level. We summarize the inferential findings as follows.

(a) The analysis shows that gender, education, and race all have an impact on inference. When individuals are categorized based on their gender, education, and race, the number of data sets not rejecting the Pareto distribution hypothesis tends to increase. This trend is evident in the bottom panel of Table 6, where observations are not classified. The numbers in the upper panels are overall consistently higher than the corresponding numbers in the bottom panel for each cohort born in 1960, 1961, and 1962. This finding indicates that the Pareto distribution hypothesis becomes more appropriate when labor income data are collected from more homogeneous sectors of individuals.

(b) Although detailed results are not reported here, the Pareto distributional hypothesis was found to becomes even more appropriate at higher income levels. If we increase b_x to represent the top 5% labor income for each cohort, the number of data sets not rejecting the Pareto distributional hypothesis increases. This suggests that the Pareto distribution is more suitable for capturing the right tail of the income distribution as we move to higher income levels. However, it is essential to note that increasing b_x reduces sample size and therefore precision. For our analysis, we have selected b_x to be the top 10% labor income level of each cohort data, allowing us to obtain moderate sample sizes from all cohorts. □

The next sep in the analysis involved estimating the Pareto parameter θ_* using infinite-dimensional MCMD. When the Pareto hypothesis is supported for characterizing the upper labor income distribution, the estimated Pareto parameter provides a measure of heavy-tailedness and thereby income inequality. The upper panel of Figure 3 displays the functional shapes of the Pareto PDFs and CDFs for $\theta_* = 1$, 2, and 3 for $b_x = 1$. Evidently the density levels of higher incomes decrease as θ_* increases, and heavy-tailed distributions (with fewer finite moments) occur for lower values of θ_*. Additionally, CDFs with higher θ_* uniformly dominate CDFs with lower θ_*, so that income distributions with a lower θ_* are more unequally distributed than those with higher θ_*. This relationship can be linked to traditional income inequality indices. In fact, under the Pareto distribution, the top x-percent income share $S(\theta_*, x)$ and the Gini coefficient $G(\theta_*)$ can be represented directly by functions of θ_* as follows:

$$S(\theta_*, x) := \left(\frac{x}{100}\right)^{\frac{\theta_* - 1}{\theta_*}} \qquad \text{and} \qquad G(\theta_*) := \frac{1}{2\theta_* - 1},$$

The lower panel of Figure 3 shows the functional shapes of $S(\cdot, x)$ and $G(\cdot)$ for $x = 5$ and 10. The functions have a negative slope, indicating that income equality indices improve as θ_* increases. We leverage this characteristic by estimating the Pareto parameter from the top 10% labor income observations. Specifically, using the results in Table 6, we compute the top 5% income shares and Gini coefficients by $S(\widehat{\theta}_n, 5)$ and $G(\widehat{\theta}_n)$, where $\widehat{\theta}_n$ denotes the infinite-dimensional MCMD estimates for the data of each cohort that do not reject the Pareto distribution condition.

Some caution is needed in using this approach as the Gini coefficient obtained in this manner should be considered an approximate value based on the upper end of the distribution. It may not be accurate to assume that the entire labor income distribution follows a Pareto distribution, as only the top 10% incomes are tested for the right tail distribution. On the contrary, the top 5% income shares can be reliably estimated through this method since they are computed from the right tail income distributions. Therefore, we focus attention on the top 5% labor income shares.

Estimation results are presented in Figures 4, 5, 6, and 7. Figure 4 displays the estimated top 5% income shares as functions of time between 1980 and 2018 for female and male cohorts. Each series is classified based on individual birth years. Missing points in the figures indicate that the U-test rejects the Pareto hypothesis for the data of that year. The level of significance is set to 1% for the U-test. Most missing values occur in the early 1980s, which is the time before the majority of individuals entered the job market. Similarly, Figures 5 and 6 show the estimated top 5% labor income shares for data sets classified by education and race. Additionally, Figure 7 illustrates the estimated top 5% income share when the data sets are not classified. We summarize the findings and implications as follows.

(a) In general, the top 5% income share functions depicted in Figures 4 to 7 exhibit a hump-shaped pattern. For each cohort, the top income share index reaches its lowest values around 1980 and then sharply rises until around 1992. However, since then, it gradually decreases. This trend is consistently observed across all cohorts, and the peak level of top 5% labor income inequality is typically reached when most workers are in the early stages of their careers and are actively seeking jobs. During this transitional period, it is expected that labor income inequality would increase, which can be termed as "frictional inequality." Notably, labor income inequality tends to reach its highest level before workers reach the age of 60, reflecting the eventual passage to retirement during the latter period of a working life.

(b) The hump-shaped top 5% income share trends observed in Figures 4 to 7 have important implications for reducing labor income inequality. In addition to the policy implications derived from the labor market inequality literature (e.g., Katz and Murphy, 1995; Katz and Autor, 1999; Ciccone and Peri, 2005; Eisenbarth and Chen, 2022), the hump-shaped indices suggest that labor income inequality can be significantly reduced by targeting the frictional inequality observed during early career years. To achieve this, economic policies aimed at reducing unemployment, setting a minimum wage, or increasing welfare benefits could be more effective for workers in the early stages of their careers, as frictional income inequality is commonly observed across all cohorts. By implementing such targeted policies during this transitional period overall labor income inequality could be mitigated.

(c) When analyzing the gender effect, we observe from Figure 4 that males generally exhibit more volatile top 5% income shares than females. After reaching the maximum top income share, it decreases gradually for females, while it declines more rapidly for males. For instance, focusing on the 1960 cohort, the maximum top 5% income share is around 0.26 for females and 0.24 for males, and it decreases at a slower rate for females compared to males. This finding suggests that labor income inequality is more pronounced for females than males within the 1960 cohort. Similar results are found for the 1962 cohort, with males showing a higher maximum index value than females for the 1961 cohort.

(d) When analyzing the education effect, we observe from Figure 5 that individuals with an MA or equivalent degree show the most volatile top 5% income shares compared to other degree holders. The maximum top income share values of individuals with an MA or equivalent degree are higher than those of other degree holders, and they maintain relatively higher top 5% income share values for some time. In contrast, the top 5% income share values of individuals with a doctorate or equivalent degree are generally less volatile and smaller than those of other degree holders. The top income share values obtained using all observations, as shown in Figure 7, are roughly between those of individuals with an MA and doctorate or equivalent degree. This finding implies that labor income for individuals with an MA or equivalent degree is more unequally distributed than that of individuals with a doctorate or equivalent degree.

(e) Another notable feature of the education effect is observed from Figure 5, where we see that it takes more years for individuals with a high school diploma or lower education levels to reach the maximum top income share compared to individuals with higher education levels. Additionally, the maximum values are not reached rapidly for the former group. For instance, the 1960 cohort reaches its maximum in 2002, and there are other years before 2002 with slightly smaller index values. This aspect implies that unequal labor income is persistently distributed over a long period for individuals with a high school diploma or lower education levels. A similar feature is observed for individuals with a BA or equivalent degree, although it is not as strong as for those with a high school diploma or lower education levels.

(f) When examining the race effect, we observe from Figure 7 that different races exhibit different patterns in their top 5% labor income shares. White and Caucasian cohorts generally show lower top income shares compared to the other races, and their coefficients remain more or less stable across different birth years, indicating a relatively stable pattern. On the other hand, black and African American cohorts show varying patterns depending on the birth year. Specifically, the cohort born in 1960 maintains lower and stable top income shares, whereas the cohort born in 1962 shows more volatile top income shares. For the Asian case, the cohort born in 1960 exhibits a distinct pattern from the others, with its index sharply increasing between 1999 and 2000, and again in 2008, during the Asian and the subprime financial crises. The other Asian cohorts are less sensitive to the financial crises and show more stable top income shares.

(g) When examining the birth year effect, we observe that the top income share values of the 1960 cohort are generally lower than those of the other cohorts. For each cohort, the index of the 1960 cohort remains persistently lower than the other cohorts, while the 1961 cohort shows consistently but slightly higher values than the 1960 cohort. This finding suggests that the income inequality is influenced by the birth year, and there are differences in income distribution patterns across different cohorts. □

From this empirical analysis we infer that labor income inequality is influenced by various heterogeneous factors such as gender, education, race, and year of birth. The influence of these factors suggest different policy implications for reducing labor income inequality. In Appendix A.5, we also provide estimated Gini coefficients in parallel with Figures 4 to 7. These results demonstrate the usefulness of the infinite-dimensional MCMD estimation in identifying how labor income inequality has evolved over time. Specifically, we can effectively test the Pareto distribution hypothesis using the U-test, and the estimated Pareto parameter allows us to measure the income inequality index.

5 Concluding Summary

If the moment dimension in GMM estimation expands to infinity proportional to the sample size, the limit properties of GMM differ from standard case where the number of moment conditions is fixed. Specifically, the limit properties are influenced by the stochastic properties of the moment conditions and the weight matrix that is employed in GMM estimation. This study has derived the asymptotic properties of GMM when inverse Brownian motion or Brownian bridge kernels are used for the weight matrix. These kernels arise in a natural way in econometric work such as minimum Cramér-von Mises distance estimation, which arises in testing distributional specification. We consider different scenarios where the moment conditions converge to either a smooth Gaussian or a non-differentiable Gaussian process. By leveraging the individual properties of the Brownian motion and Brownian bridge kernels, we show how asymptotic behavior can be fully characterized using the inner products of functionals derived from these Gaussian processes.

The paper also explores conditions under which the standard J-test can serve as an appropriate statistic for testing overidentification. In cases where the standard conditions do not hold, we propose an alternative test called the U-test, inspired by the T-test introduced by Donald et al. (2003). Throughout the discussions on GMM estimation, we use the infinite-dimensional MCMD estimation as the running example, extending the application of the MCMD estimation method introduced in Pollard (1980) and Cho et al. (2018). We illustrate the usefulness of this approach through Monte Carlo simulations and apply it in an empirical study.

Our empirical application utilizes the infinite-dimensional MCMD methodology to analyze labor income based on the CWHS database. We estimate the top 5% income shares of labor income as a function of time, covering the period from 1980 to 2018. These cohort data sets are classified based on gender, education, race, and birth year. These data are analyzed using the new U-test to test the Pareto distribution hypothesis and estimate the Pareto parameter using infinite-dimensional MCMD estimation. The results show that labor income inequality within the same cohort tends to be maximized during early career years for most of the cohort data. This observation suggests that economic policies aimed at reducing income inequality will likely be more effective if they specifically target workers in their early career years. Such policies can play a crucial role in reducing frictional inequality and contribute to a more equitable distribution of labor income.

References

AMENGUAL, D., M. CARRASCO, AND E. SENTANA (2020): "Testing Distributional Assumptions Using a Continuum of Moments," *Journal of Econometrics*, 218, 655–689.

ANDERSEN, L. B. G. AND V. V. PITERBARG (2010): *Interest Rate Modeling*, London: Atlantic Financial Press.

ANGRIST, J. D. AND A. B. KEUEGER (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 106, 979–1014.

ATKINSON, A. (2005): "Top Incomes in the UK over the 20th Centry," *Journal of the Royal Statistical Society. Series A*, 168, 325–343.

——— (2007): "Measuring Top Incomes: Methodological Issues," in *Top Incomes over the Twentieth Century : a Contrast between Continental European and English-Speaking Countries*, ed. by A. B. Atkinson and T. Piketty, Oxford, UK: Oxford University Press, 18–42.

ATKINSON, A. AND A. LEIGH (2007): "The Distribution of Top Incomes in Australia," *Economic Record*, 83, 247–261.

——— (2008): "Top Incomes in New zealand 1921–2005: Understanding the Effects of Marginal Tax Rates, Migration Threat, and the Macroeconomy," *Review of Income and Wealth*, 54, 149–165.

ATKINSON, A., T. PIKETTY, AND E. SAEZ (2011): "Top Incomes in the Long Run of History," *Journal of Economic Literature*, 49, 3–71.

BAI, J. AND S. NG (2010): "Instrumental Variable Estimation in a Data Rich Environment," *Econometric Theory*, 26, 1577–1606.

BATES, C. AND H. WHITE (1985): "A Unified Theory of Consistent Esetimation for Parametric Model," *Econometric Theory*, 1, 151–178.

BELLONI, A., D. CHEN, V. CHERNOZHUKOV, AND C. HANSEN (2012): "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain," *Econometrica*, 80, 2369–2429.

BONTEMPS, C. AND N. MEDDAHI (2012): "Testing Distributional Assumptions: A GMM aproach," *Journal of Applied Econometrics*, 27, 978–1012.

CARRASCO, M. (2012): "A Regularization Approach to the Many Instruments Problem," *Journal of Econometrics*, 170, 383–398, thirtieth Anniversary of Generalized Method of Moments.

CARRASCO, M. AND J.-P. FLORENS (2000): "Generalization of GMM to a Continuum of Moment Conditions," *Econometric Theory*, 16, 797–834.

CARRASCO, M., J.-P. FLORENS, AND E. RENAULT (2007): "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, vol. 6, chap. 77, 5633–5751.

CHO, J. S., M.-H. PARK, AND P. C. B. PHILLIPS (2018): "Practical Kolmogorov-Smirnov Testing by Minimum Distance Applied to Measure Top Income Shares in Korea," *Journal of Business & Economic Statistics*, 36, 523–537.

CHO, J. S. AND H. WHITE (2011): "Generalized Runs Tests for the IID Hypothesis," *Journal of Econometrics*, 162, 326–344.

CHUI, C. K. (1971): "Concerning Rates of Convergence of Riemann Sums," *Journal of Approximation Theory*, 4, 279–287.

CICCONE, A. AND G. PERI (2005): "Long-Run Substitutability netween More and Less Educated Workers: Evidence from U.S. States," *Review of Economcis & Statistics*, 87, 652–663.

D'ALBIS, H. AND I. BADJI (2022): "Inequality within generation: Evidence from France," *Research in Economics*, 76, 69–83.

DAVID, H. A. AND H. N. NAGARAJA (2003): *Order Statistics*, New Jersey: John Wiley & Sons.

DONALD, S., G. IMBENS, AND N. NEWEY (2003): "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions," *Econometrica*, 69, 1161–1191.

DONALD, S. AND W. NEWEY (2001): "Choosing the Number of Instruments," *Econometrica*, 69, 1161–1191.

DURBIN, J. (1973): "Weak Convergence of the Sample Distribution Function when Parameters are Estimated," *Annals of Statistics*, 1, 279–290.

EISENBARTH, A. AND Z. CHEN (2022): "The Evolution of Wage Inequality within Local U.S. Labor Markets," *Journal for Labor Market Research*, 56, 2.

GREENWOOD, M. (1946): "The Statistical Study of Infectious Diseases," *Journal of the Royal Statistical Society. Series A*, 109, 185–186.

GRENANDER, U. (1981): *Abstract Inference*, New York: John Wiley & Sons.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HIRSA, A. AND S. N. NEFTCI, eds. (2014): *Copyright*, San Diego: Academic Press, third edition ed.

KATZ, L. AND D. AUTOR (1999): "Changes in the Wage Structure and Earnings Inequality," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Amsterdam: Elsevier, 1453–1555.

KATZ, L. AND K. MURPHY (1995): "Changes in Relative Wages: Supply and Demand Factors," *Quarterly Journal of Economics*, 107, 35–78.

KIRSCH, A. (1996): *An Introduction to the Mathematical Theory of Inverse Problems*, Applied Mathematical Sciences, Springer New York.

KUZNETS, S. (1953): *Shares of Upper Income Groups in Income and Savings*, New York: National Bureau of Economic Research.

——— (1955): "Economic Growth and Economic Inequality," *American Economic Review*, 45, 1–28.

LIGHTILL, M. (1959): "Introduction to Fourier analysis and generalized functions," *Cambridge University*.

MORIGUCHI, C. AND E. SAEZ (2008): "The Evolution of Income Concentration in Japan, 1886-2005: Evidence from Income Tax Statistics," *Review of Economics and Statistics*, 90, 713–734.

PHILLIPS, P. C. B. (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

PICARD, E. (1910): "Sur un Théorème Général Relatif aux Équations Intégrales de Premiére Espéce et sur Quelques Problèmes de Physique Mathématique," *Rendiconti del Cicolo Matematico di Palermo*, 29, 79–97.

PIKETTY, T. (2003): "Income Inequality in France, 1901–1998," *Journal of Political Economy*, 111, 1004–1042.

PIKETTY, T. AND E. SAEZ (2003): "Income Inequality in the United States, 1913–1998," *Quarterly Journal of Economics*, 118, 1–39.

POLLARD, D. (1980): "The Minimum Distance Method of Testing," *Metrika*, 27, 43–70.

RAO, J. AND M. KUO (1984): "Asymptotic Results on the Greenwood Statistic and Some of its Generalizations," *Journal of the Royal Statistical Society. Series B*, 46, 228–237.

SARGAN, J. D. (1958): "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393–415.

SHI, Z. (2016): "Econometric Estimation with High-Dimensional Moment Equalities," *Journal of Econometrics*, 195, 104–119.

WHITE, H. (2001): *Asymptotic Theory for Econometricians*, Orlando, FL: Academic Press.

WILKS, S. S. (1948): "Order Statistics," *Bulletin of the American Mathematical Statistics*, 5, 6–50.

Test	Distribution	Level \ n	50	100	200	300	400	500	$1,000$
U-test	Exponential	1%	0.50	0.58	0.87	1.06	1.02	0.76	1.04
		5%	1.70	2.43	3.70	3.73	4.00	4.09	4.62
		10%	4.19	6.28	7.75	8.26	8.65	8.50	9.44
	Pareto	1%	0.47	0.74	0.85	0.89	0.84	0.92	0.97
		5%	1.38	2.71	3.48	4.33	3.87	4.10	4.42
		10%	4.14	6.74	7.69	8.73	8.20	9.16	9.48
	Normal	1%	0.34	0.79	0.76	0.88	0.95	0.93	1.02
		5%	1.27	2.85	3.24	3.72	3.91	4.00	4.72
		10%	3.92	6.26	7.51	8.23	8.30	8.97	9.54
τ-test	Exponential	1%	0.31	0.60	0.79	0.82	0.92	1.04	0.93
		5%	0.84	1.46	2.06	2.31	2.56	2.74	3.60
		10%	1.42	2.34	4.42	5.54	6.24	6.94	8.00
	Pareto	1%	0.35	0.51	0.83	0.94	0.89	0.94	0.91
		5%	0.84	1.58	2.25	2.20	2.68	2.81	3.42
		10%	1.34	2.46	4.52	5.29	6.84	7.29	7.85
	Normal	1%	0.29	0.64	0.88	0.84	0.87	0.84	0.94
		5%	0.67	1.45	2.16	2.49	2.64	2.74	3.46
		10%	1.18	2.35	4.58	5.83	6.37	6.39	8.17

Table 1: EMPIRICAL REJECTION RATES OF THE U- AND τ-TESTS UNDER THE NULL (IN PERCENT). This table shows the empirical rejection rates of the U- and τ-tests under the distributional hypothesis.

Test	Distribution	Level \ n	50	100	200	300	400	500	$1,000$
t-test	Exponential	1%	2.57	1.72	1.62	1.22	1.32	1.22	1.03
		5%	7.87	6.39	5.86	5.50	5.40	5.66	5.14
		10%	13.16	11.52	11.01	10.40	9.91	10.40	10.17
	Pareto	1%	2.68	1.87	1.34	1.19	1.23	1.13	1.27
		5%	7.64	6.75	5.60	5.44	5.42	5.58	5.21
		10%	13.24	11.64	10.38	10.58	10.44	10.92	9.66
	Normal	1%	0.89	1.02	0.73	1.11	0.91	0.97	1.04
		5%	4.35	4.89	4.46	4.51	4.91	4.88	4.89
		10%	8.67	9.26	9.07	9.18	9.87	9.48	9.74
t'-test	Exponential	1%	2.93	2.05	1.50	1.47	1.45	1.38	1.31
		5%	8.82	7.32	6.48	6.04	6.30	6.12	5.61
		10%	14.54	12.66	12.09	11.44	11.61	11.66	11.02
	Pareto	1%	2.65	2.09	1.86	1.46	1.51	1.35	1.32
		5%	8.76	7.15	6.60	6.11	6.16	6.21	5.85
		10%	14.55	12.99	12.23	11.55	11.76	11.53	10.74
	Normal	1%	1.43	1.32	1.17	1.25	1.08	1.03	1.13
		5%	6.54	6.23	5.97	5.69	5.48	5.08	5.25
		10%	11.80	11.48	10.88	10.92	10.83	10.25	10.34

Table 2: EMPIRICAL REJECTION RATES OF THE t- AND t'-TESTS UNDER THE NULL (IN PERCENT). This table shows the empirical rejection rates of the t- and t'-tests under the joint hypothesis that $\theta_* = c$ and that the distributional condition is correct. For the exponential, Pareto, and normal cases, we let $\theta_* = 1$, 1, and 0, respectively.

Test	Distribution	Level \ n	50	100	200	300	400	500	$1,000$
U-test	Exponential	1%	0.59	1.86	4.44	7.53	12.30	16.42	39.23
		5%	1.61	4.65	10.93	17.62	24.90	31.88	59.70
		10%	3.33	7.71	16.82	25.33	33.88	41.61	70.51
	Pareto	1%	0.61	1.53	4.38	8.04	11.45	15.58	40.30
		5%	2.01	4.33	10.47	17.34	24.28	31.34	60.64
		10%	3.76	7.47	15.99	25.39	34.41	41.38	70.77
	Normal	1%	1.63	2.96	4.90	5.81	5.50	6.05	4.11
		5%	4.15	7.83	12.06	13.41	13.84	14.53	11.91
		10%	6.96	12.23	18.25	20.17	21.02	22.02	18.69
τ-test	Exponential	1%	1.10	2.85	4.64	5.75	7.61	9.18	14.20
		5%	2.53	5.85	9.10	11.24	14.25	16.63	24.95
		10%	3.87	7.76	12.92	15.74	19.48	22.68	32.07
	Pareto	1%	0.80	2.35	4.30	5.52	6.95	8.30	13.90
		5%	2.02	4.93	8.48	10.22	13.58	15.43	24.10
		10%	3.19	7.29	12.56	14.69	18.65	21.24	32.01
	Normal	1%	0.56	1.28	1.81	2.33	2.46	2.99	3.83
		5%	1.26	2.71	4.00	5.02	5.91	6.79	9.13
		10%	1.96	3.92	6.37	7.97	9.43	11.01	14.31

Table 3: EMPIRICAL REJECTION RATES OF THE U-TEST AND τ-TESTS UNDER THE LOCAL ALTERNATIVE (IN PERCENT). This table shows the empirical rejection rates of the U- and τ-tests under local alternatives. For the exponential case, $x_{t,n} := y_t + \frac{1}{2}\sqrt{z_t/n}$, where $y_t \sim \text{Exp}(1)$ and $z_t \sim \mathcal{U}[0.5, 1.5]$; for the Pareto case, $x_{t,n} := y_t + \frac{1}{2}\sqrt{z_t/n})$, where $y_t \sim \text{Pa}(1,1)$ and $z_t \sim \mathcal{U}[0.5, 1.5]$; and for the normal case, $x_{t,n} := y_t + \frac{1}{4}y_t^4/\sqrt{n}$, where $y_t \sim \mathcal{N}(0,1)$.

Test	Distribution	Level \ n	50	100	200	300	400	500	$1,000$
t-test	Exponential	1%	4.53	5.66	7.56	10.61	13.37	15.77	29.30
		5%	12.94	15.59	20.37	26.26	30.33	35.19	52.49
		10%	20.47	24.35	31.05	37.23	43.03	47.09	64.67
	Pareto	1%	3.20	3.71	5.20	6.98	9.28	11.72	22.57
		5%	9.83	11.82	15.67	19.52	24.78	27.84	44.11
		10%	16.28	19.00	25.06	30.01	35.82	39.11	56.60
	Normal	1%	0.64	0.86	2.14	3.62	5.09	5.84	6.49
		5%	2.68	5.06	11.68	17.30	20.88	23.26	23.67
		10%	5.55	11.31	23.09	31.65	36.58	39.04	39.04
t'-test	Exponential	1%	5.65	5.19	4.67	4.67	4.54	4.61	4.34
		5%	14.03	14.05	12.80	13.32	13.62	13.38	13.66
		10%	20.99	21.28	20.20	20.73	21.54	20.84	21.81
	Pareto	1%	3.60	3.26	2.91	2.61	2.60	2.58	2.48
		5%	10.02	9.07	9.28	8.75	8.81	8.77	9.00
		10%	16.35	14.75	15.98	14.78	15.24	15.74	15.15
	Normal	1%	1.04	1.33	1.27	1.40	1.17	1.58	1.91
		5%	5.32	5.41	5.57	6.22	6.23	6.36	7.01
		10%	10.19	10.61	10.82	11.10	11.69	12.00	12.81

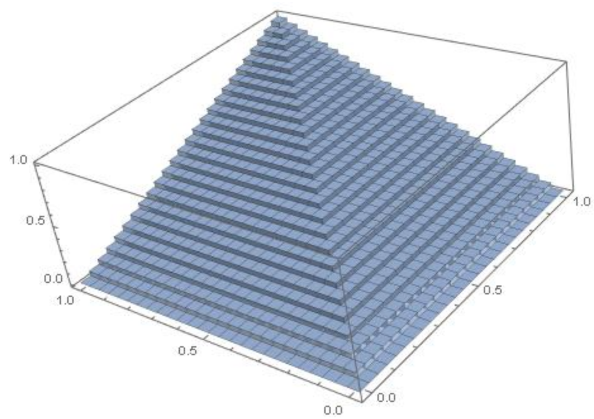
Table 4: EMPIRICAL REJECTION RATES OF THE t- AND t'-TESTS UNDER THE LOCAL ALTERNATIVE (IN PERCENT). This table shows the empirical rejection rates of the t- and τ'-tests under local alternatives. For the exponential case, $x_{t,n} := y_t + \frac{1}{2}\sqrt{z_t/n}$, where $y_t \sim \text{Exp}(1)$ and $z_t \sim \mathcal{U}[0.5, 1.5]$; for the Pareto case, $x_{t,n} := y_t + \frac{1}{2}\sqrt{z_t/n})$, where $y_t \sim \text{Pa}(1,1)$ and $z_t \sim \mathcal{U}[0.5, 1.5]$; and for the normal case, $x_{t,n} := y_t + \frac{1}{4}y_t^4/\sqrt{n}$, where $y_t \sim \mathcal{N}(0,1)$.

Classification	1960	1961	1962	Sum
Female	2,591	2,576	2,540	7,707
Male	2,479	2,456	2,358	7,293
Sum	5,070	5,032	4,898	15,000
High School or below	1,108	1,187	1,094	3,389
BA or equivalent	2,659	2,646	2,689	7,994
MA or equivalent	535	515	487	1,537
Doctorate or equivalent	768	684	628	2,080
Sum	5,070	5,032	4,898	15,000
White or Caucasian	4,106	4,085	3,954	12,145
Black or African American	644	627	619	1,890
Asian	258	250	251	759
Etc.	62	70	74	206
Sum	5,070	5,032	4,898	15,000

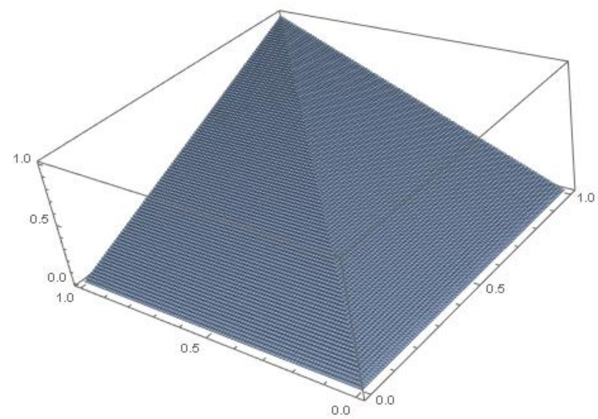
Table 5: SAMPLE SIZES OF CLASSIFIED DATA SETS. This table shows the sample sizes of the classified data sets.

Classification	Size \ Obs.	1960	1961	1962
Female	n	260	258	255
	1%	27	28	26
	5%	25	24	24
	10%	22	23	20
Male	n	248	246	236
	1%	33	31	32
	5%	31	26	28
	10%	27	23	27
High School or below	n	111	119	110
	1%	28	38	39
	5%	32	36	37
	10%	31	36	35
BA or equivalent	n	266	265	269
	1%	34	34	33
	5%	32	33	32
	10%	31	33	28
MA or equivalent	n	54	52	49
	1%	38	39	39
	5%	38	39	38
	10%	38	36	36
Doctorate or equivalent	n	77	69	63
	1%	34	28	34
	5%	32	27	32
	10%	29	24	32
White or Caucasian	n	411	409	396
	1%	29	33	29
	5%	28	31	24
	10%	25	28	19
Black or African American	n	65	63	62
	1%	38	37	39
	5%	37	36	38
	10%	35	36	37
Asian	n	26	26	26
	1%	36	39	37
	5%	36	38	37
	10%	36	35	36
Etc.	n	7	8	8
	1%	39	39	38
	5%	39	39	38
	10%	39	38	38
All	n	508	504	490
	1%	30	29	29
	5%	23	28	26
	10%	18	23	21

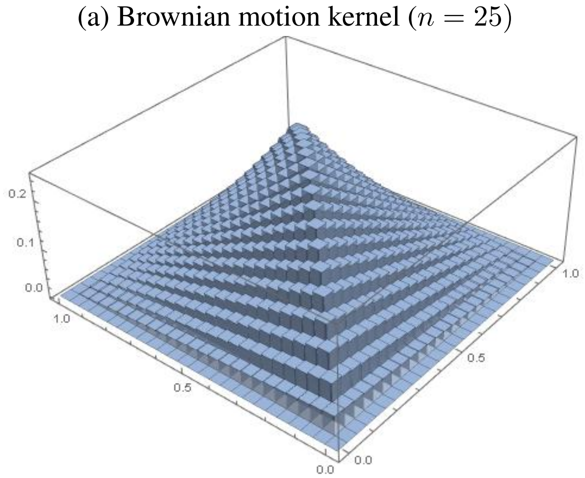
Table 6: NUMBER OF DATA SETS NOT REJECTING THE PARETO DISTRIBUTION HYPOTHESIS. This table shows the number of the top 10% CWHS data sets between 1980 and 2018 that do not reject the Pareto distribution hypothesis by the U-test. As an example, when the females are restricted to top 10% individuals who are born in 1960 and the level of significance is 1%, 27 data sets between 1980 and 2017 do not reject the Pareto distribution hypothesis. Here, n denotes the average sample size of the top 10% individuals in the data sets.



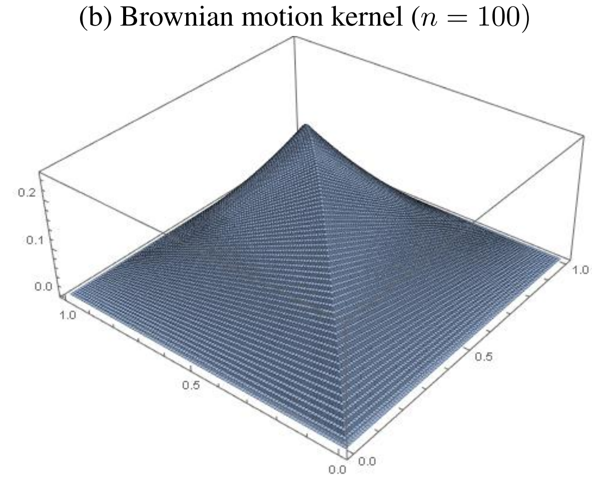
(a) Brownian motion kernel ($n = 25$)



(b) Brownian motion kernel ($n = 100$)



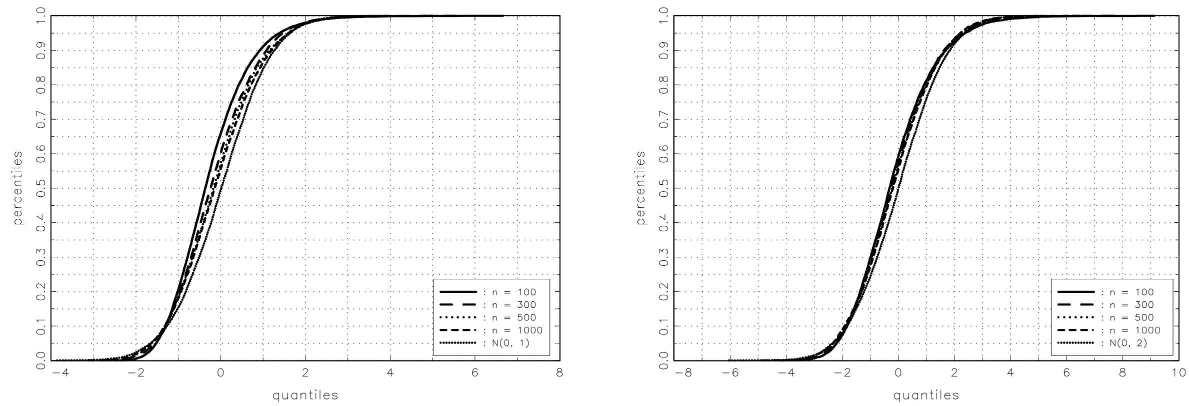
(c) Brownian bridge kernel ($n = 25$)



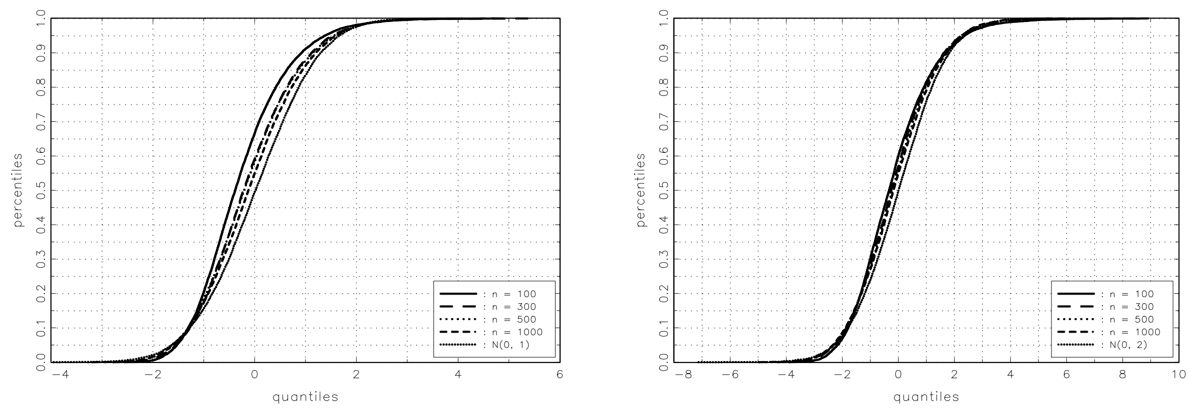
(d) Brownian bridge kernel ($n = 100$)

Figure 1: FUNCTIONAL SHAPES OF THE BROWNIAN MOTION AND BROWNIAN BRIDGE KERNELS. For $n = 25$ and 100, each figure shows the shapes of the Brownian motion and Brownian bridge kernel functions.

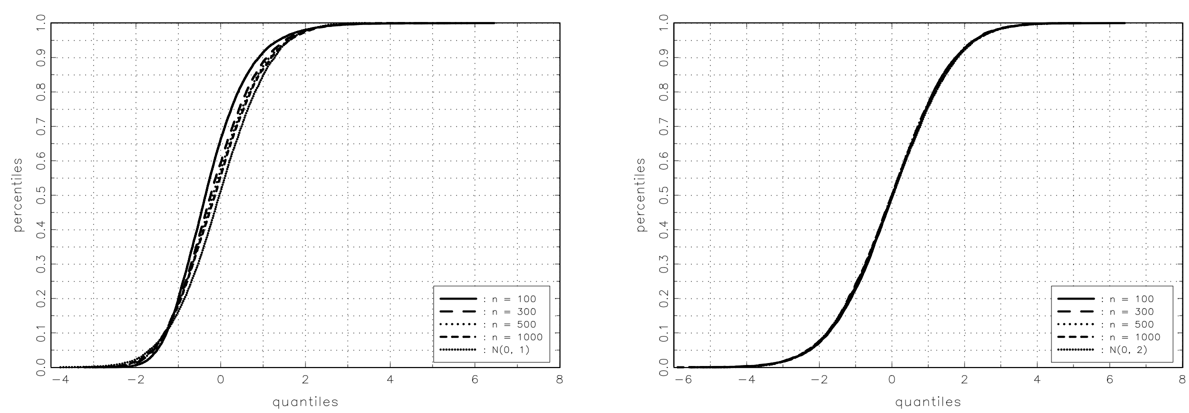
Distribution of the U-test Distribution of the MCMD estimator



(a) Exponential distribution case

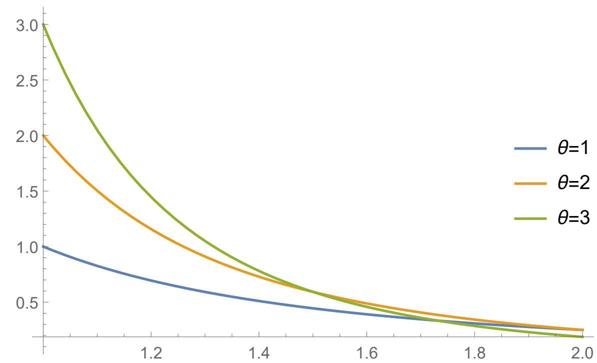


(b) Pareto distribution case

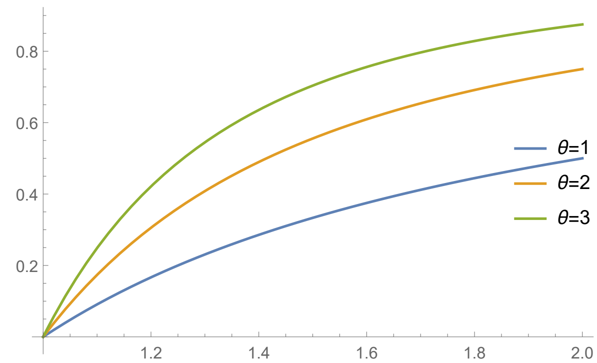


(c) Normal distribution case

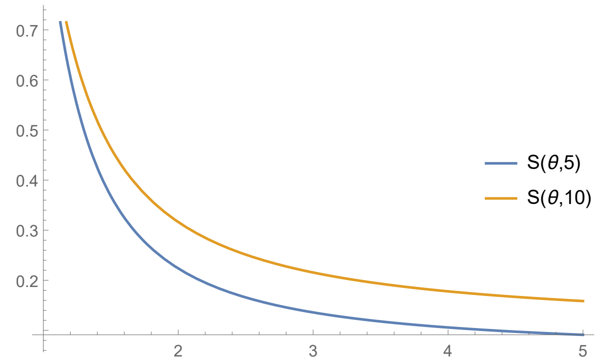
Figure 2: EMPIRICAL DISTRIBUTIONS OF THE U-TEST UNDER THE NULL AND THE MCMD ESTIMATOR. For $n = 100, 300, 500$, and $1,000$, each figure shows the null distributions of the U-test or the empirical distributions of the MCMD estimator. The distributions are obtained by repeating 10,000 independent experiments, and the limit distributions are drawn together for comparison purpose.



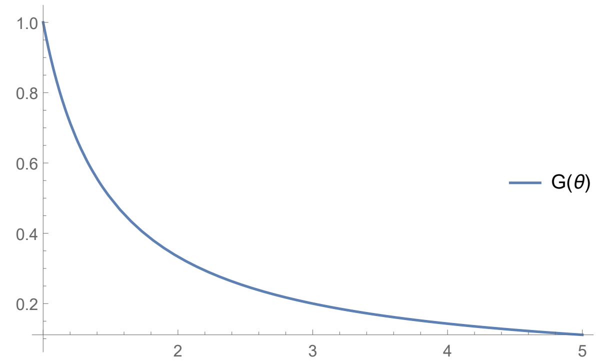
(a) Pareto PDFs



(b) Pareto CDFs

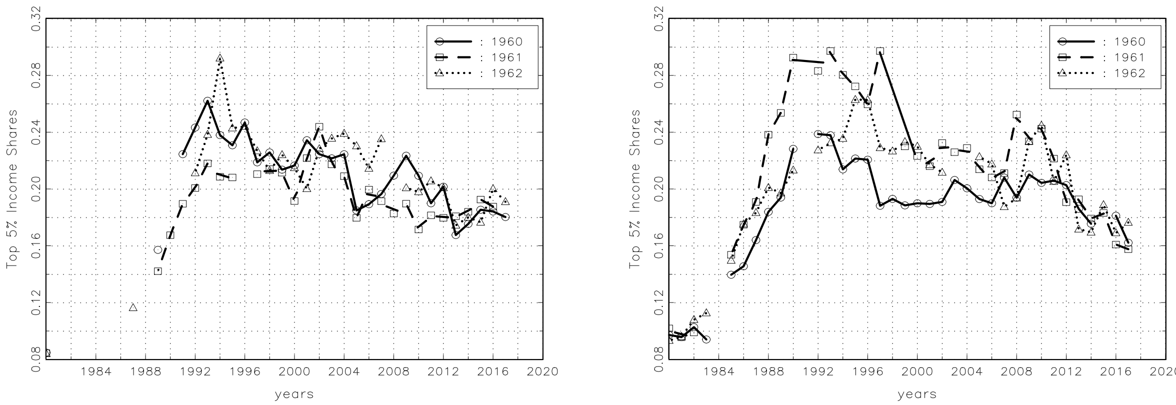


(c) Top income share function



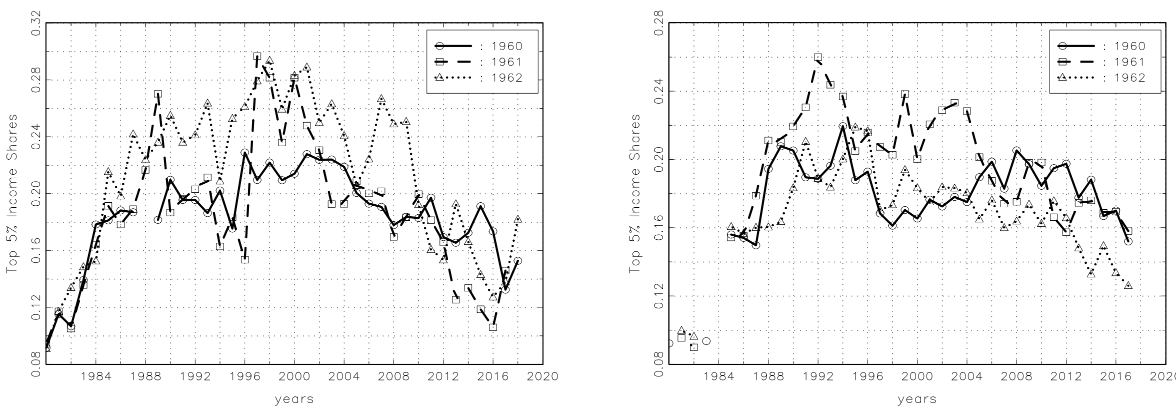
(d) Gini function

Figure 3: PDFs, CDFs, Top Income Share, and Gini Functions of the Pareto Random Variables. The figures in the upper panel show the shapes of the Pareto PDF and CDF for $\theta_* = 1\ 2$, and 3. The top income share function shows the functional shapes of the $q\%$ top income share coefficient as a function of θ_* for $q = 5$ and 10, and the Gini function shows the functional shape of the Gini coefficient as a function of θ_*.

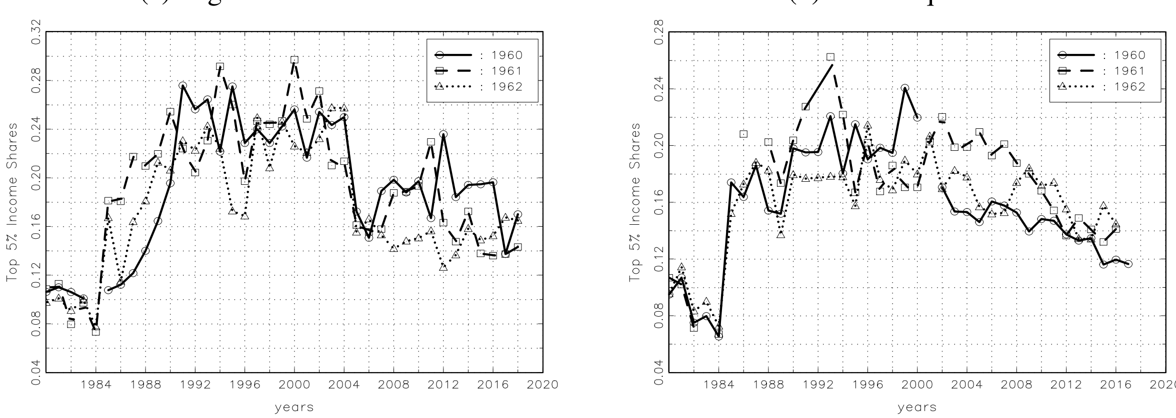


(a) Female (b) Male

Figure 4: TOP 5% INCOME SHARES OF FEMALE AND MALE COHORTS BETWEEN 1980 AND 2018. The figures show the top 5% income share coefficients of female and male cohorts estimated by imposing the Pareto distribution to the top 10% CWHS observations. Missing values signify that the p-value of the U-test is less than 1%.

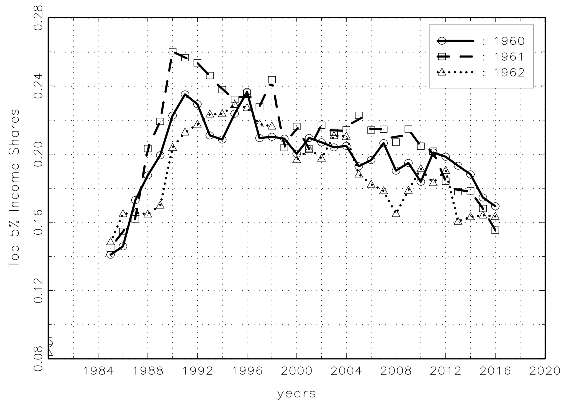


(a) High school or below (b) BA or equivalent

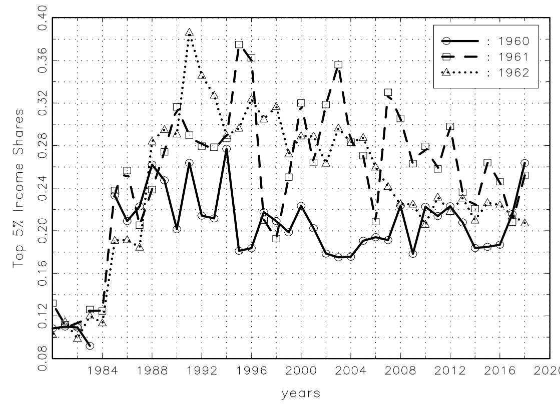


(c) MA or equivalent (d) Doctorate or equivalent

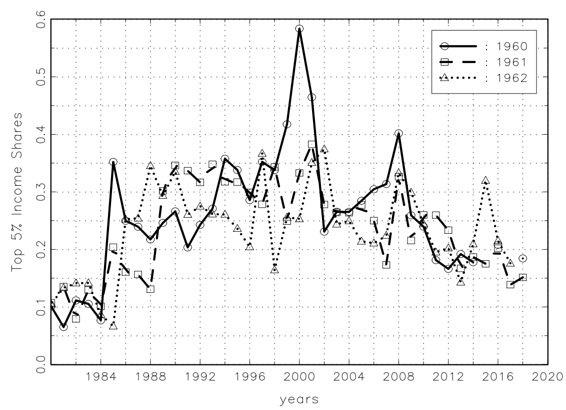
Figure 5: TOP 5% INCOME SHARES WITHIN THE SAME EDUCATION COHORTS BETWEEN 1980 AND 2018. The figures show the top 5% income share coefficients within the same education cohorts estimated by imposing the Pareto distribution to the top 10% CWHS observations. Missing values signify that the p-value of the U-test is less than 1%.



(a) White or Caucasian



(b) Black or African American



(c) Asian

Figure 6: TOP 5% INCOME SHARES WITHIN THE SAME RACE COHORTS BETWEEN 1980 AND 2018. The figures show the top 5% income share coefficients within the same race cohorts estimated by imposing the Pareto distribution to the top 10% CWHS observations. Missing values signify that the p-value of the U-test is less than 1%.

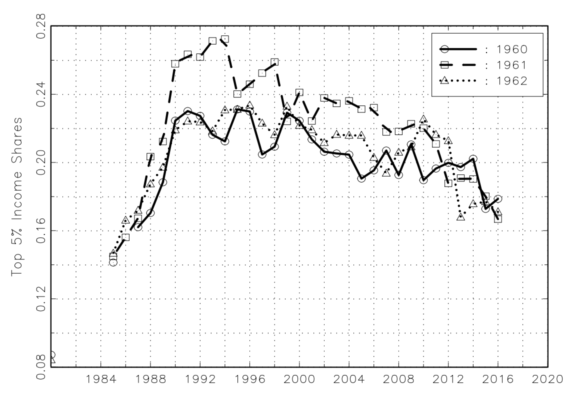


Figure 7: TOP 5% INCOME SHARES USING AGGREGATED OBSERVATIONS FOR EACH YEAR BETWEEN 1980 AND 2018. The figures show the top 5% income share coefficients of aggregated observations estimated by imposing the Pareto distribution to the top 10% CWHS observations. Missing values signify that the p-value of the U-test is less than 1%.

Online Supplement for
'GMM Estimation by the Brownian Kernels Applied to Measuring the Income Inequality'[1]

by

Jin Seo Cho[a] and Peter C. B. Phillips[b,c,d]

[a]Yonsei University

[b]University of Auckland, [c]Yale University, [d]Singapore Management University

This Online Supplement is an Appendix that provides various technical and empirical supplements to the main text.

A Appendix

The Appendix has five sections. Sections A.1 and A.2 explore the limit behavior of the quantities formed by transforming an Itô Process and a smooth Gaussian process. Section A.3 presents the moments of the statistics forming the infinite-dimensional MCMD estimator, and Section A.4 provides proofs of the main results. Section A.5 offers supplementary empirical studies to those in the paper.

A.1 Limit Difference of Transformed Itô Process

We derive the differential of a transformed Itô process. For this examination, suppose that a limit of a process is constructed as follows: $\bar{X}_n(\cdot) := \sqrt{n}(\widehat{X}_n(\cdot) - X(\cdot)) \Rightarrow \mathcal{G}(\cdot)$, where $\widehat{X}_n(\cdot)$ is a sample average of random processes defined on $[0, 1]$, $X(\cdot)$ is its population mean, and $\mathcal{G}(\cdot)$ is an Itô process satisfying Assumption 2 (ii) in the main paper. Here, we suppose that $X(\cdot)$ is differentiable on $[0, 1]$ and $\widehat{X}_n(\cdot)$ converges to $X(\cdot)$ uniformly on $[0, 1]$. For example, we can consider an empirical process as a specific example: $\widetilde{g}_n(\cdot) := \sqrt{n}\{\widehat{p}_n(\cdot) - (\cdot)\} \Rightarrow \mathcal{B}^0(\cdot)$, which determines the limit distribution of the infinite-dimensional MCMD estimator. For this case, $\widehat{X}_n(\cdot) = \widehat{p}_n(\cdot)$, $X(\cdot) = (\cdot)$, and $\mathcal{G}(\cdot) = \mathcal{B}^0(\cdot)$, such that $\mu(u, \mathcal{G}(u)) = -(1-u)^{-1}\mathcal{B}^0(\cdot)$ and $\sigma(u, \mathcal{G}(\cdot)) = 1$. As before, we let $\mu(\cdot)$ and $\sigma(\cdot)$ abbreviate $\mu(\cdot, \mathcal{G}(\cdot))$ and $\sigma(\cdot, \mathcal{G}(\cdot))$, respectively.

Given this, for a function $f : \mathbb{R} \mapsto \mathbb{R}$ in $\mathcal{C}^{(2)}([0, 1])$, we let $Q_n(\cdot) := f(\widehat{X}_n(\cdot))$ and derive the limits of the quantities associated with $\Delta Q_n(i_n)$ for $i = 0, 1, 2, \ldots, n$. Here, for each $t = 0, 1, 2, \ldots, n$ and a function $h : [0, 1] \mapsto \mathbb{R}$, we let $\Delta h(i_n) := h(i_n) - h(\frac{i-1}{n})$ for notational simplicity.

[1]Phillips acknowledges research support from the Kelly Fund at the University of Auckland and a KLC Fellowship at Singapore Management University.

A.1.1 Limit Behavior of $\sqrt{n}\Delta Q_n(\cdot)$

We obtain the limit behavior of $\sqrt{n}\Delta Q_n(\cdot)$ by applying Itô's lemma. Note that

$$\Delta Q_n(\cdot) = f'(\widehat{X}_n(\cdot))\Delta\widehat{X}_n(\cdot) + o_{\mathbb{P}}(1) \tag{A.1}$$

by Taylor expansion. Now $\Delta\widehat{X}_n(\cdot) = \Delta X(\cdot) + n^{-1/2}\bar{X}_n(\cdot)$ and $\Delta X(\cdot)$ can be approximated by $X'(\cdot)/n$. Therefore, we obtain that

$$\Delta\widehat{X}_n(\cdot) = \frac{1}{n}X'(\cdot) + \frac{1}{\sqrt{n}}\Delta\bar{X}_n(\cdot) + o_{\mathbb{P}}(1), \tag{A.2}$$

implying that

$$\sqrt{n}\Delta Q_n(\cdot) = f'(\widehat{X}_n(\cdot))\{n^{-1/2}X'(\cdot) + \Delta\bar{X}_n(\cdot)\} + o_{\mathbb{P}}(1) \tag{A.3}$$

by plugging (A.2) into (A.1). The stochastic differential equation of $\sqrt{n}\Delta Q_n(\cdot)$ is obtained from this limit. Note that $\widehat{X}_n(\cdot)$ converges to $X(\cdot)$ uniformly on $[0,1]$, and $\Delta\bar{X}_n(\cdot)$ is approximated by $d\mathcal{G}(\cdot)$. Therefore, if we let $d\mathcal{Q}(\cdot)$ denote the limit of $\sqrt{n}\Delta Q_n(\cdot)$, it follows that $d\mathcal{Q}(\cdot) = f'(X(\cdot))\mu(\cdot)du + f'(X(\cdot))\sigma(\cdot)d\mathcal{W}(\cdot)$.

For example, for the infinite-dimensional MCMD estimator we have $H_n(\cdot) = H(\widehat{p}_n(\cdot))$. This fact implies that

$$\sqrt{n}\Delta H_n(\cdot) = H'(\cdot)\{n^{-1/2} + \Delta\sqrt{n}\widehat{p}_n(\cdot)\} + o_{\mathbb{P}}(1) = n^{-1/2}H'(\cdot) + H'(\cdot)\Delta\widetilde{g}_n(\cdot) + o_{\mathbb{P}}(1) \tag{A.4}$$

by noting the definition of $\widetilde{g}_n(\cdot) := \sqrt{n}\{\widehat{p}_n(\cdot) - (\cdot)\}$ and that $\widehat{p}_n(\cdot) \to (\cdot)$ uniformly on $[0,1]$ with probability converging to 1. This fact can be related to Assumption 3 (ii) by noting that $C_1(\cdot) = C_2(\cdot) = H'(\cdot)$.

A.1.2 Limit Behavior of $n\sum_{t=1}^{n}\{\Delta Q_n(i_n)\}^2$

We examine the limit of $n\sum_{t=1}^{n}\{\Delta Q_n(i_n)\}^2$. From the first equality of (A.3), we note that $\sum_{i=1}^{n}\{\sqrt{n}\Delta Q_n(i_n)\}^2 = \sum_{i=1}^{n}\{f'(\widehat{X}_n(i_n))\}^2\{\frac{1}{n}(X'(i_n))^2 + (\Delta\bar{X}_n(i_n))^2\} + o_{\mathbb{P}}(1) = \frac{1}{n}\sum_{i=1}^{n}\{f'(X(i_n))\}^2\{(X'(i_n))^2 + \sigma^2(i_n)\} + o_{\mathbb{P}}(1)$, where the second equality holds by noting that $\widehat{X}_n(\cdot)$ converges to $X(\cdot)$ uniformly on $[0,1]$ and that $\sum_{i=1}^{n}\{f'(\widehat{X}_n(i_n))\}^2(\Delta\bar{X}_n(i_n))^2 = \sum_{i=1}^{n}\{f'(X(i_n))\}^2(\Delta\mathcal{G}(i_n))^2 + o_{\mathbb{P}}(1) = \frac{1}{n}\sum_{i=1}^{n}\{f'(X(i_n))\}^2\sigma^2(i_n) + o_{\mathbb{P}}(1)$. From this we can therefore derive that $n\sum_{i=1}^{n}\{\Delta Q_n(i_n)\}^2 \Rightarrow \int_0^1\{f'(X(u))\}^2\{(X'(u))^2 + \sigma^2(u,\mathcal{G}(u))\}du$. This result can also be generalized. If we let $\widetilde{Q}_n(\cdot) := h(\widehat{X}_n(\cdot))$ for a function $h : \mathbb{R} \mapsto \mathbb{R}$ in $\mathcal{C}^{(2)}([0,1])$, then we have $n\sum_{i=1}^{n}\{\Delta Q_n(i_n)\}\{\Delta\widetilde{Q}_n(i_n)\} \Rightarrow \int_0^1 f'(X(u))h'(X(u))\{X'(u) + \sigma^2(u,\mathcal{G}(u))\}du$.

For example, for the infinite-dimensional MCMD estimator, $H_n(\cdot) = H(\widehat{p}_n(\cdot))$ and $\widehat{p}_n(\cdot) \to (\cdot)$ uniformly on $[0,1]$ with probability converging to 1, and $X(\cdot) = (\cdot)$. This implies $n\sum_{i=1}^{n}\Delta H_n(i_n)\Delta H_n(i_n)'$ $\to 2\int_0^1 H'(u)H'(u)'du$ with probability converging to 1 by noting that $X'(\cdot) \equiv 1$ and $\sigma(\cdot) \equiv 1$ for the

Brownian bridge.

A.1.3 Limit Behavior of $n\sum_{t=1}^{n}\Delta Q_n(i_n)\Delta\bar{X}_n(i_n)$

Here we examine the limit behavior of $n\sum_{t=1}^{n}\Delta Q_n(i_n)\Delta\bar{X}_n(i_n)$. Note that (A.3) implies that

$$\begin{aligned} n\Delta Q_n(\cdot) =& f'(\widehat{X}_n(\cdot))\{X'(\cdot)+\sqrt{n}\Delta\bar{X}_n(\cdot)\} \\ &+\frac{1}{2}f''(\widehat{X}_n(\cdot))\left(\frac{1}{n}(X'(\cdot))^2+\frac{2}{\sqrt{n}}X'(\cdot)\Delta\bar{X}_n(\cdot)+(\Delta\bar{X}_n(\cdot))^2\right)+o_{\mathbb{P}}(1) \end{aligned} \tag{A.5}$$

by a second-order Taylor expansion. This expansion is obtained by using the following approximation: $(\Delta\widehat{X}(\cdot))^2=\frac{1}{n^2}(X'(\cdot))^2+\frac{2}{n\sqrt{n}}X'(\cdot)\Delta\bar{X}_n(\cdot)+\frac{1}{n}(\Delta\bar{X}_n(\cdot))^2+o_{\mathbb{P}}(1)$ based on (A.2). From this, we now obtain that

$$\begin{aligned} n\sum_{t=1}^{n}\left\{\Delta Q_n(i_n)\Delta\bar{X}_n(i_n)\right\} =& \sum_{i=1}^{n}f'\left(\widehat{X}_n(i_n)\right)X'(i_n)\Delta\bar{X}_n(i_n) \\ &+\sqrt{n}\sum_{i=1}^{n}f'\left(\widehat{X}_n(i_n)\right)\left(\Delta\bar{X}_n(i_n)\right)^2+o_{\mathbb{P}}(1), \end{aligned} \tag{A.6}$$

using the fact that $(\Delta\bar{X}_n(\cdot))^2=n^{-1}\sigma^2(\cdot)+o_{\mathbb{P}}(1)$. Note that the second-order term of (A.5) vanishes to 0 with probability converging to 1. This fact implies that the limit behavior of $n\sum_{t=1}^{n}\Delta Q_n(i_n)\Delta\bar{X}_n(i_n)$ by focusing on the first-order approximation of $n\Delta Q_n(\cdot)$. Therefore, we now obtain that

$$n\sum_{t=1}^{n}\left\{\Delta Q_n(i_n)\Delta\bar{X}_n(i_n)\right\}-\sqrt{n}\sum_{i=1}^{n}f'\left(\widehat{X}_n(i_n)\right)\left(\Delta\bar{X}_n(i_n)\right)^2\Rightarrow\int_0^1 f'(X(u))X'(u)d\mathcal{G}(u). \tag{A.7}$$

We derive the weak limit of $n\sum_{t=1}^{n}\{\Delta Q_n(i_n)\Delta\bar{X}_n(i_n)\}$ by elaborating the second term of the left side in (A.7). Note that $\sum_{i=1}^{n}f'(\widehat{X}_n(i_n))(\Delta\bar{X}_n(i_n))^2=\frac{1}{n}\sum_{i=1}^{n}f'(\widehat{X}_n(i_n))\sigma^2(i_n)+o_{\mathbb{P}}(1)\Rightarrow\int_0^1 f'(X(u))\sigma^2(u)du$. Therefore, if we further suppose that $\int_0^1 f'(X(u))\sigma^2(u)du=0$ and that $n^{-1/2}\sum_{i=1}^{n}f'(X(i_n))\sigma^2(i_n)\to 0$ with probability converging to 1, we can derive the weak limit of $n\sum_{t=1}^{n}\{\Delta Q_n(i_n)\Delta\bar{X}_n(i_n)\}$ more specifically. For this derivation, first note that

$$\begin{aligned} \sqrt{n}\sum_{i=1}^{n}f'\left(\widehat{X}_n(i_n)\right)\left(\Delta\bar{X}_n(i_n)\right)^2 =& \sqrt{n}\sum_{i=1}^{n}f'(X(i_n))\left(\Delta\bar{X}_n(i_n)\right)^2 \\ &+\sum_{i=1}^{n}f''(X(i_n))\sqrt{n}\left(\widehat{X}_n(i_n)-X(i_n)\right)\left(\Delta\bar{X}_n(i_n)\right)^2+o_{\mathbb{P}}(1), \end{aligned} \tag{A.8}$$

using the fact that

$$f'(\widehat{X}_n(\cdot))=f'(X(\cdot))+f''(X(\cdot))(\widehat{X}_n(\cdot)-X(\cdot))+o_{\mathbb{P}}(1). \tag{A.9}$$

Next, note that $\sqrt{n}\sum_{i=1}^{n} f'(X(i_n))(\Delta\bar{X}_n(i_n))^2 = \sqrt{n}\sum_{i=1}^{n} f'(X(i_n))\{(\Delta\bar{X}_n(i_n))^2 - \frac{1}{n}\sigma^2(i_n)\} + o_{\mathbb{P}}(1)$ from the supposition that $n^{-1/2}\sum_{i=1}^{n} f'(X(i_n))\sigma^2(i_n) = o_{\mathbb{P}}(1)$; and applying a CLT to the right side gives $\sqrt{n}\sum_{i=1}^{n} f'(X(i_n))(\Delta\bar{X}_n(i_n))^2 \Rightarrow \mathcal{Z} \sim \mathcal{N}(0, \Gamma)$, where $\Gamma := \lim_{n\to\infty} n\sum_{i=1}^{n}\sum_{j=1}^{n} f'(X(i_n))f'(X(j_n$ $))\mathbb{E}[\{(\Delta\bar{X}_n(i_n))^2 - \frac{1}{n}\sigma^2(i_n)\}\{(\Delta\bar{X}_n(j_n))^2 - \frac{1}{n}\sigma^2(j_n)\}]$. Third, note that $(\Delta\bar{X}_n(\cdot))^2 = n^{-1}\sigma^2(\cdot) + o_{\mathbb{P}}(1)$, and this implies that $\sum_{i=1}^{n} f''(X(i_n))\sqrt{n}(\widehat{X}_n(i_n) - X(i_n))(\Delta\bar{X}_n\,(i_n))^2 = \frac{1}{n}\sum_{i=1}^{n} f''(X(i_n))\sqrt{n}(\widehat{X}_n(i_n)$ $-X(i_n))\sigma^2(i_n) + o_{\mathbb{P}}(1) \Rightarrow \int_0^1 f''(X(u))\mathcal{G}(u)\sigma^2(u)du$ from the supposition that $\sqrt{n}(\widehat{X}_n(\cdot) - X(\cdot)) \Rightarrow \mathcal{G}(\cdot)$. Combining these two weak limits with (A.8) gives $\sqrt{n}\sum_{i=1}^{n} f'(\widehat{X}_n(i_n))(\Delta\bar{X}_n(i_n))^2 \Rightarrow \mathcal{Z} + \int_0^1 f''($ $X(u))\ \mathcal{G}(u)\sigma^2(u)du$, which further implies that $n\sum_{t=1}^{n}\{\Delta Q_n(i_n)\Delta\bar{X}_n(i_n)\} \Rightarrow \mathcal{Z} + \int_0^1 f''(X(u))\mathcal{G}(u)$ $\sigma^2(u)du + \int_0^1 f'(X(u))X'(u)d\mathcal{G}(u)$ by (A.7).

For example, if we consider the infinite-dimensional MCMD estimator, $H_n(\cdot) = H(\widehat{p}_n(\cdot))$ and $\widehat{p}_n(\cdot) \to (\cdot)$ uniformly on $[0,1]$ with probability converging to 1. If we further elaborate on (A.4) expanding it by using the fact that $H'(\widehat{p}_n(\cdot)) = H'(\cdot) + H''(\cdot)(\widehat{p}_n(\cdot) - (\cdot)) + o_{\mathbb{P}}(1)$, it follows that $\sqrt{n}\Delta H_n(\cdot) = n^{-1/2}H'(\cdot) + H'(\cdot)\Delta\widetilde{g}_n(\cdot) + n^{-1/2}H''(\cdot)\widetilde{g}_n(\cdot)\Delta\widetilde{g}_n(\cdot) + o_{\mathbb{P}}(n^{-1})$. Furthermore, $\widetilde{g}_n(\cdot) := \sqrt{n}\{\widehat{p}_n(\cdot) - (\cdot)\} \Rightarrow \mathcal{B}^0(\cdot)$, so that $\sigma^2(\cdot) \equiv 1$, and Section 2.4 shows that $\int_0^1 H'(u)du = 0$ and $n^{-1/2}\sum_{i=1}^{n} H'(i_n) = o(1)$ using theorem 1 (c) of Chui (1971). Hence, $n\sum_{t=1}^{n-1}\{\Delta H_n(i_n)\Delta\widetilde{g}_n(i_n)\} \Rightarrow \mathcal{Z} + \int_0^1 H''(u)\mathcal{B}^0(u)du + \int_0^1 H'(u)d\mathcal{B}^0(u)$, where $\mathcal{Z} \overset{\mathrm{A}}{\sim} \mathcal{N}(0, 8[H'(\cdot), H'(\cdot)])$, and $\int_0^1 H''(u)\mathcal{B}^0(u)du + \int_0^1 H'(u)d\mathcal{B}^0(u) = 0$ as Section 2.4 verifies.

A.2 Limit Differences of Smooth Gaussian Processes

This section derives the limit behavior of the same quantities examined in Section A.1 by supposing that $\mathcal{G}(\cdot)$ satisfies the condition in Assumption 2 (i). That is, $\mathcal{G}(\cdot)$ is differentiable with prob. 1 instead of being an Itô process. Note that the convergence rate of $\Delta\bar{X}_n(\cdot)$ is different from that of Section A.1. Specifically, Assumption 2 (i) implies that $n\Delta\bar{X}_n(\cdot) = \mathcal{G}'(\cdot) + o_{\mathbb{P}}(1)$. This different feature produces different limit behaviors for the quantities involved.

A.2.1 Limit Behavior of $\sum_{i=1}^{n}\Delta Q_n(i_n)$

We first examine the limit behavior of $\sqrt{n}\Delta Q_n(\cdot)$. If we combine (A.2), (A.3), and (A.9), we can derive the following:

$$\sqrt{n}\Delta Q_n(\cdot) = \frac{1}{\sqrt{n}}f'(\cdot)X'(\cdot) + f'(\cdot)\Delta\bar{X}_n(\cdot) + \frac{1}{n}X'(\cdot)\bar{X}_n(\cdot) + \frac{1}{\sqrt{n}}f''(\cdot)\bar{X}_n(\cdot)\Delta\bar{X}_n(\cdot) + o_{\mathbb{P}}(1), \quad \text{(A.10)}$$

so that $n\Delta Q_n(\cdot) = f'(\cdot)X'(\cdot) + o_{\mathbb{P}}(1)$. Therefore, it follows that $\sum_{i=1}^{n}\Delta Q_n(\cdot) = \frac{1}{n}\sum_{i=1}^{n} f'(i_n)X'(i_n) + o_{\mathbb{P}}(1) \to \int_0^1 f'(u)X'(u)du$ with probability converging to 1.

A.2.2 Limit Behavior of $n\sum_{t=1}^{n}\{\Delta Q_n(i_n)\}^2$

By (A.10), $n\sum_{i=1}^{n}\{\Delta Q_n(i_n)\}^2 = \frac{1}{n}\sum_{i=1}^{n}\{f'(i_n)X'(i_n)\}^2 + o_{\mathbb{P}}(1) \to \int_0^1\{f'(u)X'(u)\}^2du$ with probability converging to 1. The limit is identical to $(f'(\cdot)X'(\cdot), f'(\cdot)X'(\cdot))$.

A.2.3 Limit Behavior of $n\sum_{t=1}^{n}\Delta Q_n(i_n)\Delta\bar{X}_n(i_n)$

By (A.10) and the fact that $n\Delta\bar{X}_n(\cdot) = \mathcal{G}'(\cdot) + o_{\mathbb{P}}(1)$, it follows that $n\sum_{i=1}^{n}\Delta Q_n(i_n)\Delta\bar{X}_n(i_n) = \frac{1}{n}\sum_{i=1}^{n} f'(i_n)X'(i_n)n\Delta\bar{X}_n(i_n) + o_{\mathbb{P}}(1) \Rightarrow \int_0^1 f'(u)X'(u)\mathcal{G}'(u)du = (f'(\cdot)X'(\cdot), \mathcal{G}'(\cdot))$.

A.3 Asymptotic Behavior of Quantities involved in Infinite-Dimensional MCMD Estimation

This section explores the asymptotic variances of the quantities constituting the infinite-dimensional MCMD estimator. For this, we first note that $(\Delta\widehat{p}_n(\frac{1}{n}), \Delta\widehat{p}_n(\frac{2}{n}), \ldots, \Delta\widehat{p}_n(1))'$ follows a Dirichlet distribution with parameter ι_n. Using this condition, the following hold: for each i and $j = 1, 2, \ldots, n-1$ $(i \neq j)$,

$$\mathbb{E}\left[\Delta\widehat{p}_n\left(i_n\right)\right] = \frac{1}{n}, \tag{A.11}$$

$$\mathbb{E}\left[\left(\Delta\widehat{p}_n\left(i_n\right)\right)^2\right] = \frac{2}{n(n+1)}, \tag{A.12}$$

$$\mathbb{E}\left[\left(\Delta\widehat{p}_n\left(i_n\right)\right)^3\right] = \frac{6}{n(n+1)(n+2)}, \tag{A.13}$$

$$\mathbb{E}\left[\left(\Delta\widehat{p}_n\left(i_n\right)\right)^4\right] = \frac{24}{n(n+1)(n+2)(n+3)}, \tag{A.14}$$

$$\mathbb{E}\left[\left(\Delta\widehat{p}_n\left(i_n\right)\right)^2\Delta\widehat{p}_n\left(j_n\right)\right] = \frac{2}{n(n+1)(n+2)}, \tag{A.15}$$

$$\mathbb{E}\left[\left(\Delta\widehat{p}_n\left(i_n\right)\right)^2\left(\Delta\widehat{p}_n\left(j_n\right)\right)^2\right] = \frac{4}{n(n+1)(n+2)(n+3)}. \tag{A.16}$$

A.3.1 The Variance of $\sqrt{n}\sum_{i=1}^{n-1}H_j'(i_n)\{n(\Delta\widehat{p}_n(i_n))^2 - \frac{2n}{n(n+1)}\}$

The asymptotic variance of $\sqrt{n}\sum_{i=1}^{n-1}H_j'(i_n)\{n(\Delta\widehat{p}_n(i_n))^2 - \frac{2n}{n(n+1)}\}$ is shown to be $20(H_j'(\cdot), H_j'(\cdot))$. For simplicity let $U_{n,i} := n(\Delta\widehat{p}_n(i_n))^2 - \frac{2n}{n(n+1)}$, then $\text{var}[\sqrt{n}\sum_{i=1}^{n-1}H_j'(i_n)\{n(\Delta\widehat{p}_n(i_n))^2 - \frac{2n}{n(n+1)}\}] = n\sum_{i=1}^{n-1}\sum_{\ell=1}^{n-1}H_j'(i_n)H_j'(\ell_n)\mathbb{E}[U_{n,i}U_{n,\ell}]$, where $\ell_n := \frac{\ell}{n}$. Here, note that $\mathbb{E}[U_{n,i}U_{n,\ell}] = n^2\mathbb{E}[(\Delta\widehat{p}_n(i_n))^2 (\Delta\widehat{p}_n(\ell_n))^2] - \frac{4n^2}{n(n+1)}\mathbb{E}[(\Delta\widehat{p}_n(i_n))^2] + \frac{4n^2}{n^2(n+1)^2}$. If $i \neq \ell$, $\mathbb{E}[U_{n,i}U_{n,\ell}] = -16n^{-3} + o(n^{-3})$ by (A.12) and (A.16); and if $i = \ell$, $\mathbb{E}[U_{n,i}U_{n,\ell}] = 20n^{-2} + o(n^{-2})$ by (A.12) and (A.14). Combining these two facts, it follows that $\text{var}[\sqrt{n}\sum_{i=1}^{n-1}H_j'(i_n)\{n(\Delta\widehat{p}_n(i_n))^2 - \frac{2n}{n(n+1)}\}] = \frac{20}{n}\sum_{i=1}^{n-1}(H_j'(i_n))^2 - \frac{16}{n^2}(\sum_{i=1}^{n-1}H_j'(i_n))^2 + o(1) \to 20\int_0^1(H_j'(u))^2du = 20(H_j'(\cdot), H_j'(\cdot))$ since $\sum_{i=1}^{n-1}H_j'(i_n) = o(n^{-1})$ by theorem 1 (c) of Chui

(1971).

A.3.2 The Variance of $\sum_{i=1}^{n-1}(\Delta\widetilde{g}_n(i_n))^2 - 1$

The asymptotic variance of $\sum_{i=1}^{n-1}(\Delta\widetilde{g}_n(i_n))^2 - 1$ is now shown to be $4n^{-1} + o(n^{-1})$. We first note that $\Delta\widetilde{g}_n(\cdot) = \sqrt{n}(\Delta\widehat{p}_n(\cdot) - \frac{1}{n})$, so that $(\Delta\widetilde{g}_n(\cdot))^2 = n(\Delta\widehat{p}_n(\cdot))^2 - 2\Delta\widehat{p}_n(\cdot) + n^{-1}$. This fact implies that $\sum_{i=1}^{n-1}(\Delta\widetilde{g}_n(i_n))^2 - 1 = n\sum_{i=1}^{n-1}(\Delta\widehat{p}_n(i_n))^2 - 2 + 2\Delta\widehat{p}_n(1) + \frac{1}{n}$ by noting that $\sum_{i=1}^{n-1}\Delta\widehat{p}_n(i_n) = 1 - \Delta\widehat{p}_n(1)$. Hence, it follows that $\mathbb{E}[(\Delta\widetilde{g}_n(i_n))^2] = \frac{n-1}{n(n+1)}$ from (A.11) and (8) and that $\mathrm{var}[\sum_{i=1}^{n-1}(\Delta\widetilde{g}_n(i_n))^2 -1] = n^2\sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\mathbb{E}[(\Delta\widehat{p}_n(i_n))^2(\Delta\widehat{p}_n(j_n))^2] - 4n\sum_{i=1}^{n-1}\mathbb{E}[(\Delta\widehat{p}_n(i_n))^2] + 4 + o(n^{-1})$, implying that $\mathrm{var}[\sum_{i=1}^{n-1}(\Delta\widetilde{g}_n(i_n))^2 - 1] = \frac{4n^3+20n^2+72n}{n(n+1)(n+2)(n+3)} + o(n^{-1}) = \frac{4}{n} + o(n^{-1})$ by (A.12), (A.14), and (A.16).

A.3.3 The Covariance between $\sum_{i=1}^{n-1} H_j'(i_n)\Delta\widetilde{g}_n(i_n)$ and $\sqrt{n}\sum_{i=1}^{n-1} H_j'(i_n)\{n(\Delta\widehat{p}_n(i_n))^2 - \frac{2n}{n(n+1)}\}$

We show that $4(H_j'(\cdot), H_j'(\cdot))$ is the asymptotic covariance between $\sum_{i=1}^{n-1} H_j'(i_n)\Delta\widetilde{g}_n(i_n)$ and $\sqrt{n}\sum_{i=1}^{n-1} H_j'(i_n)U_{n,i}$, where $U_{n,i} := n(\Delta\widehat{p}_n(i_n))^2 - \frac{2n}{n(n+1)}$ as in Section A.3.1. We note that $\mathrm{cov}[\sum_{i=1}^{n-1} H_j'(i_n)\Delta\widetilde{g}_n(i_n), \sqrt{n}\sum_{i=1}^{n-1} H_j'(i_n)U_{n,i}] = \sum_{i=1}^{n-1}\sum_{\ell=1}^{n-1} H_j'(i_n)H_j'(\ell_n)[n^2\mathbb{E}[(\Delta\widehat{p}_n(i_n))^2\Delta\widehat{p}_n(\ell_n)] - \frac{2n}{n+1}]$ since $\Delta\widetilde{g}_n(\cdot) = \sqrt{n}\widehat{p}_n(\cdot) - n^{-1/2}$ and using (A.11) and (A.12), where $\ell_n := \frac{\ell}{n}$. We further use the moment conditions in (A.13) and (A.15) to obtain that $\mathrm{cov}[\sum_{i=1}^{n-1} H_j'(i_n)\Delta\widetilde{g}_n(i_n), \sqrt{n}\sum_{i=1}^{n-1} H_j'(i_n)U_{n,i}] = \frac{4n}{(n+1)(n+2)}\sum_{i=1}^{n-1} H_j'(i_n)^2 + \frac{4n^2}{(n+1)(n+2)}(\frac{1}{n}\sum_{i=1}^{n-1} H_j'(i_n))^2 \to 4(H_j'(\cdot), H_j'(\cdot))$, because $\sum_{i=1}^{n-1} H_j'(i_n) = o(n^{-1})$ by theorem 1 (c) of Chui (1971).

A.4 Proofs of the Main Claims

In this section we prove the main claims in the paper. In these proofs it is convenient to use some basic properties of generalized functions, particularly the Dirac delta function $\delta(x)$ and its derivatives that play critically important roles for the proofs. We note the following useful properties for functionals involving the delta function.[1]

Note that for a function $f : [0, 1] \mapsto \mathbb{R}$ in $\mathcal{C}^{(2)}([0, 1])$,

$$\int_0^1 \frac{\delta(x - \cdot - n^{-1}) - \delta(x - \cdot)}{n^{-1}} f(x)dx \to \int_0^1 -\delta'(x - \cdot)f(x)dx = f'(\cdot) \qquad \text{(A.17)}$$

uniformly on $[0, 1]$. If we further let $\delta_n(u - v) := n\widetilde{\mathbb{I}}_n(u, v)$, it follows that $\delta_n(x) = n\mathbb{I}_n[x \in [0, \frac{1}{n})]$, whose limit is $\delta(x)$ as n tends to infinity. Furthermore, we note that $\delta_n(u - \frac{1}{n} - v) = n\widetilde{\mathbb{I}}_n(u - \frac{1}{n}, v) = n\widetilde{\mathbb{J}}_n(u, v)$ by noting that $\widetilde{\mathbb{I}}_n(\cdot - \frac{1}{n}, \circ) = \widetilde{\mathbb{J}}_n(\cdot, \circ)$, so that the first-order derivative of the Dirac delta generalized function

[1] Readers are referred to Lightill (1959) for further details.

is obtained as follows:

$$\delta_n'(u-v) := \frac{\delta_n\left(u-v-\frac{1}{n}\right)-\delta_n(u-v)}{-n^{-1}} = n^2\{\widetilde{\mathbb{I}}_n(u,v)-\widetilde{\mathbb{J}}_n(u,v)\} \to \delta'(u-v). \tag{A.18}$$

We also note that the first-order derivative of the Dirac delta generalized function satisfies the following property:

$$-\delta'(u) = \delta'(-u). \tag{A.19}$$

The second-order derivative of the Dirac delta generalized function is obtained similarly to the first-order derivative

$$\begin{aligned}\delta_n''(u-v) &:= \frac{\delta_n'\left(u-v+\frac{1}{n}\right)-\delta_n'(u-v)}{n^{-1}} \\ &= \frac{n^2\{[\widetilde{\mathbb{J}}_n(u,v)-\widetilde{\mathbb{I}}_n(u,v)]-[\widetilde{\mathbb{I}}_n(v,u)-\widetilde{\mathbb{J}}_n(v,u)]\}}{n^{-1}} \to \delta''(u-v).\end{aligned} \tag{A.20}$$

Proof of Lemma 1: (i) We prove each statement in turn.

(i.a) First note that for any $n > 2$,

$$\widetilde{\Sigma}_n^{-1} = n\begin{bmatrix} 2 & -1 & \cdots & 0 & 0 \\ -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & \cdots & -1 & 2 \end{bmatrix},$$

so that if we let $\Omega_n := -\Omega_{1n} + \Omega_{2n}$,

$$\Omega_{1n} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \quad \text{and} \quad \Omega_{2n} = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

it follows that $\widetilde{\Sigma}_n^{-1} = -n\Omega_n - n\Omega_n'$. We now let $B_n := [b_n(\frac{1}{n}), b_n(\frac{2}{n}), \ldots, b_n(\frac{n-1}{n})]'$, so that the first-row and final-row elements of B_n converge to zero from the fact that $b(0) = b(1) = 0$. We further note that

$$n\Omega_n B_n = \frac{(\Omega_{2n}-\Omega_{1n})}{n^{-1}}B_n = \frac{1}{n^{-1}}\begin{bmatrix} b_n(\frac{2}{n}) - b_n(\frac{1}{n}) \\ \vdots \\ b_n(1) - b_n(\frac{n-1}{n}) \end{bmatrix} \to b'(\cdot) \tag{A.21}$$

uniformly on $[0,1]$, that is the same consequence as given in (A.17). We here use the fact that $b_n(1)=0$.

We further note that this result can be associated with the Dirac delta generalized function as follows:

$$\Omega_{2n}B_n = \begin{bmatrix} b_n(\frac{2}{n}) \\ \vdots \\ b_n(1) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n-1} b_n(i_n)\widetilde{\mathbb{J}}_n(i_n,\frac{1}{n}) \\ \vdots \\ \sum_{i=1}^{n-1} b_n(i_n)\widetilde{\mathbb{J}}_n(i_n,\frac{n-1}{n}) \end{bmatrix} = \begin{bmatrix} n\int_0^1 b_n(u)\widetilde{\mathbb{J}}_n(u,\frac{1}{n})du \\ \vdots \\ n\int_0^1 b_n(u)\widetilde{\mathbb{J}}_n(u,\frac{n-1}{n})du \end{bmatrix}$$

and

$$\Omega_{1n}B_n = \begin{bmatrix} b_n(\frac{1}{n}) \\ \vdots \\ b_n(\frac{n-1}{n}) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n-1} b_n(i_n)\widetilde{\mathbb{I}}_n(i_n,\frac{1}{n}) \\ \vdots \\ \sum_{i=1}^{n-1} b_n(i_n)\widetilde{\mathbb{I}}_n(i_n,\frac{n-1}{n}) \end{bmatrix} = \begin{bmatrix} n\int_0^1 b_n(u)\widetilde{\mathbb{I}}_n(u,\frac{1}{n})du \\ \vdots \\ n\int_0^1 b_n(u)\widetilde{\mathbb{I}}_n(u,\frac{n-1}{n})du \end{bmatrix}.$$

Therefore, if we let $\delta_n'(u-v) := n^2\{\widetilde{\mathbb{I}}_n(u,v)-\widetilde{\mathbb{J}}_n(u,v)\}$, it follows that

$$n\Omega_n B_n = n(\Omega_{2n}-\Omega_{1n})B_n = \begin{bmatrix} n^2\int_0^1 b_n(u)\{\widetilde{\mathbb{J}}_n(u,\frac{1}{n})-\widetilde{\mathbb{I}}_n(u,\frac{1}{n})\}du \\ \vdots \\ n^2\int_0^1 b_n(u)\{\widetilde{\mathbb{J}}_n(u,\frac{n-1}{n})-\widetilde{\mathbb{I}}_n(u,\frac{n-1}{n})\}du \end{bmatrix}$$

$$= \begin{bmatrix} -\int_0^1 b_n(u)\delta_n'(u-\frac{1}{n})du \\ \vdots \\ -\int_0^1 b_n(u)\delta_n'(u-\frac{n-1}{n})du \end{bmatrix} \to -\int_0^1 b(u)\delta'(u-\cdot)du = b'(\cdot)$$

by applying (A.17) and (A.18). This result shows that $n\Omega_n B_n \to b'(\cdot)$.

In a similar manner, we obtain

$$n\Omega_n' B_n = \frac{(\Omega_{2n}'-\Omega_{1n})}{n^{-1}}B_n = \frac{1}{n^{-1}}\begin{bmatrix} b_n(0)-b_n(\frac{1}{n}) \\ \vdots \\ b_n(\frac{n-2}{n})-b_n(\frac{n-1}{n}) \end{bmatrix} \to -b'(\cdot). \tag{A.22}$$

We here use the fact that $b_n(0)=0$. As before, (A.22) can be associated with the Dirac delta generalized function, viz.,

$$\Omega_{2n}' B_n = \begin{bmatrix} 0 \\ \vdots \\ b_n(\frac{n-2}{n}) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n-1} b_n(i_n)\widetilde{\mathbb{J}}_n(\frac{1}{n},i_n) \\ \vdots \\ \sum_{i=1}^{n-1} b_n(i_n)\widetilde{\mathbb{J}}_n(\frac{n-1}{n},i_n) \end{bmatrix} = \begin{bmatrix} n\int_0^1 b_n(u)\widetilde{\mathbb{J}}_n(\frac{1}{n},u)du \\ \vdots \\ n\int_0^1 b_n(u)\widetilde{\mathbb{J}}_n(\frac{n-1}{n},u)du \end{bmatrix},$$

$$\Omega_{1n}B_n = \begin{bmatrix} b_n(\frac{1}{n}) \\ \vdots \\ b_n(\frac{n-1}{n}) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n-1} b_n(i_n)\widetilde{\mathbb{I}}_n(\frac{1}{n}, i_n) \\ \vdots \\ \sum_{i=1}^{n-1} b_n(i_n)\widetilde{\mathbb{I}}_n(\frac{n-1}{n}, i_n) \end{bmatrix} = \begin{bmatrix} n\int_0^1 b_n(u)\widetilde{\mathbb{I}}_n(\frac{1}{n}, u)du \\ \vdots \\ n\int_0^1 b_n(u)\widetilde{\mathbb{I}}_n(\frac{n-1}{n}, u)du \end{bmatrix},$$

and

$$n\Omega_n' B_n = n(\Omega_{2n}' - \Omega_{1n})B_n = \begin{bmatrix} n^2\int_0^1 b_n(u)\{\widetilde{\mathbb{J}}_n(\frac{1}{n}, u) - \widetilde{\mathbb{I}}_n(\frac{1}{n}, u)\}du \\ \vdots \\ n^2\int_0^1 b_n(u)\{\widetilde{\mathbb{J}}_n(\frac{n-1}{n}, u) - \widetilde{\mathbb{I}}_n(\frac{n-1}{n}, u)\}du \end{bmatrix}$$

$$= \begin{bmatrix} -\int_0^1 b_n(u)\delta_n'(\frac{1}{n} - u)du \\ \vdots \\ -\int_0^1 b_n(u)\delta_n'(\frac{n-1}{n} - u)du \end{bmatrix} \to -\int_0^1 b(u)\delta'(\cdot - u)du = \int_0^1 b(u)\delta'(u - \cdot)du = -b'(\cdot),$$

by applying (A.17), (A.18), and (A.19). This result shows that $n\Omega_n' B_n \to -b'(\cdot)$. We also note that

$$\begin{aligned} \delta_n'(j_n - u) &= n^2\left\{\widetilde{\mathbb{I}}_n(j_n, u) - \widetilde{\mathbb{J}}_n(j_n, u)\right\} \\ &= -n^2\left\{\widetilde{\mathbb{J}}_n\left(u, \frac{j-1}{n}\right) - \widetilde{\mathbb{I}}_n\left(u, \frac{j-1}{n}\right)\right\} = -\delta_n'\left(u - \frac{j-1}{n}\right) \end{aligned} \qquad \text{(A.23)}$$

using the definition of $\widetilde{\mathbb{I}}_n(\cdot, \circ)$ and $\widetilde{\mathbb{J}}_n(\cdot, \circ)$.

Therefore, combining the two results in (A.21) and (A.22), it follows that

$$n\widetilde{\Sigma}_n^{-1}B_n = -n^2(\Omega_n + \Omega_n')B_n = -\frac{1}{n^{-2}}\begin{bmatrix} \vdots \\ b_n(\frac{j+1}{n}) - 2b_n(j_n) + b_n(\frac{j-1}{n}) \\ \vdots \end{bmatrix} \to -b''(\cdot),$$

by noting that for $j = 1, 2, \ldots, n-1$, $\frac{1}{n^{-2}}[b_n(\frac{j+1}{n}) - 2b_n(j_n) + b_n(\frac{j-1}{n})] = \frac{1}{n^{-1}}[\frac{1}{n^{-1}}\{b_n(\frac{j+1}{n}) - b_n(j_n)\} - \frac{1}{n^{-1}}\{b_n(j_n) - b_n(\frac{j-1}{n})\}] \to b''(\cdot)$. This result can also be related to the Dirac delta generalized function as follows:

$$-n^2(\Omega_n+\Omega_n')B_n = -\frac{1}{n^{-1}}\begin{bmatrix} \vdots \\ n^2\int_0^1 b_n(u)\{[\widetilde{\mathbb{J}}_n(u,j_n)-\widetilde{\mathbb{I}}_n(u,j_n)]-[\widetilde{\mathbb{I}}_n(j_n,u)-\widetilde{\mathbb{J}}_n(j_n,u)]\}du \\ \vdots \end{bmatrix}$$

$$= -\frac{1}{n^{-1}}\begin{bmatrix} \vdots \\ \int_0^1 b_n(u)\{-\delta_n'(u-j_n)+\delta_n'(u-\frac{j-1}{n})\}du \\ \vdots \end{bmatrix} = -\begin{bmatrix} \vdots \\ \int_0^1 b_n(u)\delta_n''(u-j_n)du \\ \vdots \end{bmatrix}$$

$$\to -\int_0^1 b(u)\delta''(u-\cdot)du = -b''(\cdot),$$

where the second equality holds by (A.23), and the third equality holds by noting that $\frac{1}{n^{-1}}[-\delta_n'(u-j_n)+\delta_n'(u-\frac{j-1}{n})] = \frac{1}{n^{-1}}[\delta_n'(u-j_n+\frac{1}{n})-\delta_n'(u-j_n)] = \delta_n''(u-j_n)$ using (A.20). Furthermore we can see that $n^3\{[\widetilde{\mathbb{J}}_n(u,j_n)-\widetilde{\mathbb{I}}_n(u,j_n)]-[\widetilde{\mathbb{I}}_n(j_n,u)-\widetilde{\mathbb{J}}_n(j_n,u)]\} = -\widetilde{\xi}_n(u,j_n)$ by noting that $\widetilde{\mathbb{I}}_n(u,j_n)=\widetilde{\mathbb{I}}_n(j_n,u)$, so it also follows that

$$\widetilde{\xi}_n(u,j_n) = -\delta_n''(u-j_n)\,, \tag{A.24}$$

implying that

$$n\widetilde{\Sigma}_n^{-1}B_n = \widetilde{\mathbf{\Xi}}_n b_n(\cdot) = -b''(\cdot)+o(1). \tag{A.25}$$

(i.b) Next note that if we let $C_n := [c_n(\frac{1}{n}), c_n(\frac{2}{n}), \ldots, c_n(\frac{n-1}{n})]'$,

$$C_n'\widetilde{\Sigma}_n^{-1}B_n = \frac{1}{n}C_n' n\widetilde{\Sigma}_n^{-1}B_n = \frac{1}{n}\sum_{j=1}^{n-1} c_n(j_n)\int_0^1 \widetilde{\xi}_n(j_n,u)\,b_n(u)du = \int_0^1\int_0^1 \widetilde{\xi}_n(v,u)\,b_n(u)c_n(v)dudv$$

$$= -\int_0^1\int_0^2 \delta_n''(v,u)\,b_n(u)c_n(v)dudv \to -\int_0^1 c(u)b''(u)du = \int_0^1 c'(u)b'(u)du, \tag{A.26}$$

where the second and fourth equalities follow from (A.24) and (A.25), and the last equality holds since $c(1)b'(1)-c(0)b'(0) = \int_0^1 d\{c(u)b'(u)\} = \int_0^1 c'(u)b'(u)du + \int_0^1 c(u)b''(u)du$. Note that $c(0)=c(1)=0$, so that the left side is zero, leading to (A.26). Therefore, it follows that

$$(\widetilde{\mathbf{\Xi}}_n b_n(\cdot), c_n(\cdot)) = \int_0^1\int_0^1 \widetilde{\xi}_n(v,u)b_n(u)c_n(v)dudv = \int_0^1 c'(u)b'(u)du + o(1) = (b'(\cdot),c'(\cdot))+o(1).$$

In addition, note that

$$(b'(\cdot),c'(\cdot)) = \int_0^1\int_0^1 \delta(u-v)c'(u)b'(v)dudv, \tag{A.27}$$

where the last equality follows from the fact that $\int_0^1 \delta(u_1-u_2)f'(u_1)du_1 = f'(u_2)$.

(ii) Using the fact that $\widetilde{\Sigma}_n^{-1} = -n\Omega_n - n\Omega_n'$, we obtain

$$\frac{1}{n}C_n'\widetilde{\Sigma}_n^{-1}B_n = -\sum_{i=2}^{n-1} c_n\left(\frac{i-1}{n}\right)\left\{b_n\left(i_n\right) - b_n\left(\frac{i-1}{n}\right)\right\} \\ -\sum_{i=2}^{n-1} b_n\left(\frac{i-1}{n}\right)\left\{c_n\left(i_n\right) + c_n\left(\frac{i-1}{n}\right)\right\} + 2b_n\left(\frac{n-1}{n}\right)c_n\left(\frac{n-1}{n}\right). \quad \text{(A.28)}$$

Also note that

$$b_n\left(\frac{i-1}{n}\right) = \frac{1}{2}\left(\left\{b_n\left(i_n\right) + b_n\left(\frac{i-1}{n}\right)\right\} - \left\{b_n\left(i_n\right) - b_n\left(\frac{i-1}{n}\right)\right\}\right), \quad \text{(A.29)}$$

$$c_n\left(\frac{i-1}{n}\right) = \frac{1}{2}\left(\left\{c_n\left(i_n\right) + c_n\left(\frac{i-1}{n}\right)\right\} - \left\{c_n\left(i_n\right) - c_n\left(\frac{i-1}{n}\right)\right\}\right), \quad \text{(A.30)}$$

so that if we plug these two equations into (A.28), it follows that $\frac{1}{n}C_n'\widetilde{\Sigma}_n^{-1}B_n = b_n(\frac{1}{n})c_n(\frac{1}{n}) + \sum_{i=2}^{n-1}\{b_n(i_n) - b_n(\frac{i-1}{n})\}\{c_n(i_n) - c_n(\frac{i-1}{n})\} + b_n(\frac{n-1}{n})c_n(\frac{n-1}{n})$. We here note that $b_n(0) = c_n(0) = b_n(1) = c_n(1) = 0$, so that $b_n(\frac{1}{n})c_n(\frac{1}{n}) = (b_n(\frac{1}{n}) - b_n(0))(c_n(\frac{1}{n}) - c_n(0))$ and $b_n(\frac{n-1}{n})c_n(\frac{n-1}{n}) = (b_n(1) - b_n(\frac{n-1}{n}))(c_n(1) - c_n(\frac{n-1}{n}))$. This fact implies

$$\frac{1}{n}C_n'\widetilde{\Sigma}_n^{-1}B_n = \sum_{i=1}^{n}\left\{b_n\left(i_n\right) - b_n\left(\frac{i-1}{n}\right)\right\}\left\{c_n\left(i_n\right) - c_n\left(\frac{i-1}{n}\right)\right\} \to \int_0^1 db(u)dc(u), \quad \text{(A.31)}$$

which is $(db(\cdot), dc(\cdot))$. Furthermore, (A.26) implies that $n^{-1}C_n'\widetilde{\Sigma}_n^{-1}B_n = n^{-1}(\widetilde{\boldsymbol{\Xi}}_n b_n(\cdot), c_n(\cdot))$. Therefore, $n^{-1}(\widetilde{\boldsymbol{\Xi}}_n b_n(\cdot), c_n(\cdot)) \to (db(\cdot), dc(\cdot))$. ■

Proof of Lemma 2: (i) First note that for $n \geq 2$,

$$\ddot{\Sigma}_n^{-1} = \left[\begin{array}{cccc|c} & & & & 0 \\ & & \widetilde{\Sigma}_n^{-1} & & \vdots \\ & & & & 0 \\ & & & & -n \\ \hline 0 & \cdots & 0 & -n & n \end{array}\right],$$

so that $\ddot{\Sigma}_n^{-1}$ is almost identical to $\widetilde{\Sigma}_n^{-1}$ except that the n-th row and n-th column element of $\ddot{\Sigma}_n^{-1}$ is n, whereas the $(n-1)$-th row and $(n-1)$-th column element of $\widetilde{\Sigma}_n^{-1}$ is $2n$. Therefore, the limit kernel of $\ddot{\Sigma}_n^{-1}$ can be similarly obtained to that of $\widetilde{\Sigma}_n^{-1}$. We again prove the statements in turn.

(i.a) Let $\ddot{B}_n := [B_n, b_n(1)]'$ and note that

$$n\ddot{\Sigma}_n^{-1}\ddot{B}_n = \left[\begin{array}{c} n\widetilde{\Sigma}_n^{-1}B_n \\ \hdashline -\frac{1}{n^{-2}}[b_n(\frac{n-1}{n}) - b_n(1)] \end{array}\right] = \left[\begin{array}{c} \vdots \\ -\int_0^1 b_n(u)\delta_n''(u-j_n)du \\ \vdots \\ \hdashline -\frac{1}{n^{-2}}[b_n(\frac{n-1}{n}) - b_n(1)] \end{array}\right].$$

Here, Lemma 1 (i) shows that $-\int_0^1 b_n(u)\delta_n''(u-j_n)du \to -b''(\cdot)$, and we further note that $-\frac{1}{n^{-2}}\{b_n(\frac{n-1}{n})- b_n(1)\} \to -b''(1)$ because $b_n(1-\frac{1}{n}) = b_n(1) - b_n'(1)\frac{1}{n} + b_n''(1)\frac{1}{n^2} + o(1)$ such that $b'(1) = 0$ and $b(\cdot) \in \mathcal{C}^{(2)}([0,1])$. Therefore, even the last row element converges to the negative second-order derivative of $b(\cdot)$, and this implies that

$$n\ddot{\Sigma}_n^{-1}\ddot{B}_n \to -b''(\cdot). \tag{A.32}$$

Note further that $\frac{-1}{n^{-2}}[b_n(1-\frac{1}{n}) - b_n(1)] = \frac{-1}{n^{-3}}\int_0^1 b_n(u)\{\ddot{\mathbb{J}}_n(1,u) + \ddot{\mathbb{J}}(u,1) - \ddot{\mathbb{I}}_n(1,u)\}du = \int_0^1 b_n(u) \ddot{\xi}_n(u,1)du$ using the definition of $\ddot{\xi}_n(\cdot,\circ)$. Therefore, it follows that

$$\left[\begin{array}{c} \vdots \\ -\int_0^1 b_n(u)\delta_n''(u-j_n)du \\ \vdots \\ \hdashline -\frac{1}{n^{-2}}[b_n(\frac{n-1}{n}) - b_n(1)] \end{array}\right] = \left[\begin{array}{c} \vdots \\ \int_0^1 b_n(u)\widetilde{\xi}_n(u,j_n)du \\ \vdots \\ \hdashline \int_0^1 b_n(u)\ddot{\xi}_n(u,1)du \end{array}\right] = \left[\begin{array}{c} \vdots \\ \int_0^1 b_n(u)\ddot{\xi}_n(u,j_n)du \\ \vdots \end{array}\right] = \ddot{\mathbf{\Xi}}_n b_n(\cdot), \tag{A.33}$$

where the first equality holds by (A.24), and the second equality holds by the fact that for $j = 1, 2, \ldots, n-1$, $\widetilde{\xi}_n(\cdot, j_n) = \ddot{\xi}_n(\cdot, j_n)$. Combining (A.32) and (A.33), it follows that $n\ddot{\Sigma}_n^{-1}\ddot{B}_n = \ddot{\mathbf{\Xi}}_n b_n(\cdot) \to -b''(\cdot)$.

(i.b) Let $\ddot{C}_n := [C_n, c_n(1)]'$ and obtain

$$\begin{aligned} \ddot{C}_n'\ddot{\Sigma}_n^{-1}\ddot{B}_n &= \frac{1}{n}\ddot{C}_n' n\ddot{\Sigma}_n^{-1}\ddot{B}_n = \frac{1}{n}\sum_{j=1}^n c_n(j_n)\int_0^1 \ddot{\xi}_n(j_n,u)\, b_n(u)du \\ &= \int_0^1\int_0^1 \ddot{\xi}_n(v,u)\, b_n(u)c_n(v)dudv = -\int_0^1\int_0^2 \delta_n''(v,u)\, b_n(u)c_n(v)dudv \\ &\to -\int_0^1 c(u)b''(u)du = \int_0^1 c'(u)b'(u)du, \end{aligned} \tag{A.34}$$

where the last equality holds by noting that $c(1)b'(1) - c(0)b'(0) = \int_0^1 d\{c(u)b'(u)\} = \int_0^1 c'(u)b'(u)du + \int_0^1 c(u)b''(u)du$. Note that $c(0) = b'(1) = 0$, so that the left side is zero, leading to (A.34). Therefore, it follows that $(\ddot{\mathbf{\Xi}}_n b_n(\cdot), c_n(\cdot)) = \int_0^1\int_0^1 \ddot{\xi}_n(v,u)b_n(u)c_n(v)dudv = \int_0^1 c'(u)b'(u)du + o(1) = (b'(\cdot), c'(\cdot)) + o(1)$.

(ii) Note that

$$\frac{1}{n}\ddot{C}_n'\ddot{\Sigma}_n^{-1}\ddot{B}_n = -\sum_{i=2}^{n} c_n\left(\frac{i-1}{n}\right)\left\{b_n\left(i_n\right) - b_n\left(\frac{i-1}{n}\right)\right\}$$
$$-\sum_{i=2}^{n} b_n\left(\frac{i-1}{n}\right)\left\{c_n\left(i_n\right) + c_n\left(\frac{i-1}{n}\right)\right\} + b_n\left(1\right)c_n\left(1\right). \tag{A.35}$$

Here, we plug (A.29) and (A.30) in (A.35) to obtain $\frac{1}{n}\ddot{C}_n'\ddot{\Sigma}_n^{-1}\ddot{B}_n = b_n(\frac{1}{n})c_n(\frac{1}{n}) + \sum_{i=2}^{n}\{b_n(i_n) - b_n(\frac{i-1}{n})\}$ $\{c_n(i_n) - c_n(\frac{i-1}{n})\} = \sum_{i=1}^{n}\{b_n(i_n) - b_n(\frac{i-1}{n})\}\{c_n(i_n) - c_n(\frac{i-1}{n})\}$ by noting that $b_n(0) = c_n(0) = 0$. This fact implies that as n tends to infinity,

$$\frac{1}{n}\ddot{C}_n'\ddot{\Sigma}_n^{-1}\ddot{B}_n = \sum_{i=1}^{n}\left\{b_n\left(i_n\right) - b_n\left(\frac{i-1}{n}\right)\right\}\left\{c_n\left(i_n\right) - c_n\left(\frac{i-1}{n}\right)\right\} \to \int_0^1 (db(u)dc(u)) \tag{A.36}$$

that is identical to $(db(\cdot), dc(\cdot))$. Furthermore, (A.34) implies that $n^{-1}\ddot{C}_n'\ddot{\Sigma}_n^{-1}\ddot{B}_n = n^{-1}(\ddot{\Xi}_n b_n(\cdot), c_n(\cdot))$. Therefore, $n^{-1}(\ddot{\Xi}_n b_n(\cdot), c_n(\cdot)) \to (db(\cdot), dc(\cdot))$. This completes the proof. ■

Proof of Lemma 3: (i) Given the conditions, we use Lemmas 1 and 2 to prove the statements in turn.

(i.a) Using the definition of $q_n(\theta_*)$, note that $q_n(\theta_*) = \widetilde{G}_n(\theta_*)'\widetilde{\Sigma}_n^{-1}\widetilde{G}_n(\theta_*) = (\widehat{\Xi}_n\widetilde{g}_n(\cdot), \widetilde{g}_n(\cdot)) \Rightarrow (\Xi\mathcal{G}(\cdot), \mathcal{G}(\cdot))$ by Assumption 2 (i). Further, Lemmas 1 (i and ii) or 2 (i and ii) imply that $(\Xi\mathcal{G}(\cdot), \mathcal{G}(\cdot)) = -(\mathcal{G}''(\cdot), \mathcal{G}(\cdot)) = (\mathcal{G}'(\cdot), \mathcal{G}'(\cdot)) =: \mathcal{Q}_d$. Therefore, $q_n(\theta_*) \Rightarrow \mathcal{Q}_d$.

(i.b) Using the definition of $\bar{A}_n$, note that $\bar{A}_n = \nabla_\theta\bar{G}_n(\theta_*)\widehat{\Sigma}_n^{-1}\nabla_\theta'\bar{G}_n(\theta_*) = [\widehat{\Xi}_n H_n(\cdot), H_n(\cdot)] \to [\Xi H(\cdot), H(\cdot)]$ with probability converging to 1 by Assumption 3 (i). Further, Lemmas 1 (i and ii) or 2 (i and ii) imply that $[\Xi H(\cdot), H(\cdot)] = -[H''(\cdot), H(\cdot)] = [H'(\cdot), H'(\cdot)]$; and Assumption 3 implies that $n\Delta H_n(\cdot) = C_1(\cdot) + o_{\mathbb{P}}(1)$ and $n\Delta H_n(\cdot) = H'(\cdot) + o_{\mathbb{P}}(1)$. Therefore, $[H'(\cdot), H'(\cdot)] = [C_1(\cdot), C_1(\cdot)] =: A_d$. Hence, $\bar{A}_n \to A_d$.

(i.c) By definition $D_n = \nabla_\theta\bar{G}_n(\theta_*)\widehat{\Sigma}_n^{-1}\widetilde{G}_n(\theta_*) = [H_n(\cdot), \widehat{\Xi}_n\widetilde{g}_n(\cdot)] \Rightarrow [H(\cdot), \Xi\mathcal{G}(\cdot)]$ by Assumptions 2 (i) and 3 (i). Note that Lemmas 1 (i and ii) or 2 (i and ii) imply that $[H(\cdot), \Xi\mathcal{G}(\cdot)] = -[H(\cdot), \mathcal{G}''(\cdot)] = [H'(\cdot), \mathcal{G}'(\cdot)]$; and further $H'(\cdot) = C_1(\cdot)$ by Assumption 3. Therefore, $[H'(\cdot), \mathcal{G}'(\cdot)] = [C_1(\cdot), \mathcal{G}'(\cdot)] =: \mathcal{D}_d$ and so $D_n \Rightarrow \mathcal{D}_d$.

(ii) We prove each statement in turn.

(ii.a) Given the conditions, we apply the proof of Lemma 1 (ii) or 2 (ii). For this, we note that $\frac{1}{n}q_n(\theta_*) = \frac{1}{n}\widetilde{G}_n(\theta_*)'\widetilde{\Sigma}_n^{-1}\widetilde{G}_n(\theta_*) = \frac{1}{n}(\widehat{\Xi}_n\widetilde{g}_n(\cdot), \widetilde{g}_n(\cdot)) = \sum_{i=1}^{s_n}(\Delta\widetilde{g}_n(i_n))^2 \Rightarrow \int_0^1(d\mathcal{G}(u))^2$, where the third equality holds by (A.27) or (A.36), and the weak convergence follows from Assumption 2 (ii). Furthermore, the same condition implies that $\int_0^1(d\mathcal{G}(u))^2 = \int_0^1\sigma^2(u)(d\mathcal{W}(u))^2 = \int_0^1\sigma^2(u)du = (\sigma(\cdot), \sigma(\cdot))$, where the second last equality holds by noting that $\mathcal{G}(\cdot)$ is an Itô process.

(ii.b) Given the conditions, it follows from (A.27) or (A.36) that $\frac{1}{n}\bar{A}_n = \frac{1}{n}\nabla_\theta \bar{G}_n(\theta_*)\widehat{\Sigma}_n^{-1}\nabla'_\theta \bar{G}_n(\theta_*) = \sum_{i=1}^{s_n} \Delta H_n(i_n)\Delta H_n(i_n)'$. Therefore, if we combine this equation with Assumption 3 (ii), it further follows that $\bar{A}_n = \frac{1}{n}\sum_{i=1}^{s_n} C_1(i_n)C_1(i_n)' + \sum_{i=1}^{s_n} C_2(i_n)C_2(i_n)'(\Delta\widetilde{g}_n(i_n))^2 + \frac{1}{\sqrt{n}}[\sum_{i=1}^{s_n} C_1(i_n)C_2(i_n)'\Delta\widetilde{g}_n(i_n) + \sum_{i=1}^{s_n} C_2(i_n)C_1(i_n)'\Delta\widetilde{g}_n(i_n)] + o_{\mathbb{P}}(1)$. Next examine the asymptotic behavior of each element on the right side. First, note that $\frac{1}{n}\sum_{i=1}^{s_n} C_1(i_n)C_1(i_n)' \to \int_0^1 C_1(u)C_1(u)'du = [C_1(\cdot), C_1(\cdot)]$.

Second, $(d\mathcal{G}(u))^2 = \sigma^2(u)du$, so that $\sum_{i=1}^{s_n} C_2(i_n)C_2(i_n)'(\Delta\widetilde{g}_n(i_n))^2 = \frac{1}{n}\sum_{i=1}^{s_n} C_2(i_n)C_2(i_n)'\sigma^2(i_n)$ $+ o_{\mathbb{P}}(1) \to \int_0^1 \sigma^2(u)C_2(u)C_2(u)'du = [\sigma(\cdot)C_2(\cdot), \sigma(\cdot)C_2(\cdot)]$.

Finally, we examine the asymptotic behavior of $n^{-1/2}\sum_{i=1}^{s_n} C_2(i_n)C_1(i_n)'\Delta\widetilde{g}_n(i_n)$. Note that $\Delta\widetilde{g}_n(\cdot)$ can be approximated by $d\mathcal{G}(\cdot)$ if n is sufficiently large, so that it follows that $\frac{1}{\sqrt{n}}\sum_{i=1}^{s_n} C_1(i_n)C_2(i_n)'\Delta\widetilde{g}_n(i_n) = \frac{1}{\sqrt{n}}\int_0^1 C_1(u)C_2(u)'\mu(u)du + \frac{1}{\sqrt{n}}\int_0^1 C_1(u)C_2(u)'\sigma(u)d\mathcal{W}(u) + o_{\mathbb{P}}(1)$. Note that the first two terms on the right side are $o_{\mathbb{P}}(1)$, implying that $\bar{A}_n = \frac{1}{n}\sum_{i=1}^{s_n} C_1(i_n)C_1(i_n)' + \sum_{i=1}^{s_n} C_2(i_n)C_2(i_n)'(\Delta\widetilde{g}_n(i_n))^2 + o_{\mathbb{P}}(1) \to [C_1(\cdot), C_1(\cdot)] + [\sigma(\cdot)C_2(\cdot), \sigma(\cdot)C_2(\cdot)]$ with probability converging to 1.

(ii.c) Given the conditions, (A.27) or (A.36) implies that $D_n = n\sum_{i=1}^{s_n} \Delta H_n(i_n)\Delta\widetilde{g}_n(i_n)$, and Assumption 3 (ii) implies that $\Delta H_n(\cdot) = n^{-1}C_1(\cdot) + n^{-1/2}C_2(\cdot)\Delta\widetilde{g}_n(\cdot) + n^{-1}C_3(\cdot)\widetilde{g}_n(\cdot)\Delta\widetilde{g}_n(\cdot) + o_{\mathbb{P}}(n^{-1})$, so that $D_n - \sqrt{n}\sum_{i=1}^{s_n} C_2(i_n)(\Delta\widetilde{g}_n(i_n))^2 = \sum_{i=1}^{s_n} C_1(i_n)\Delta\widetilde{g}_n(i_n) + \sum_{i=1}^{s_n} C_3(i_n)\widetilde{g}_n(i_n)(\Delta\widetilde{g}_n(i_n))^2 + o_{\mathbb{P}}(1) \Rightarrow \int_0^1 C_1(u)d\mathcal{G}(u) + \int_0^1 \sigma^2(u)C_3(u)\mathcal{G}(u)du =: \mathcal{D}_u$ by noting that $(\Delta\widetilde{g}_n(\cdot))^2 = n^{-1}\sigma^2(\cdot) + o_{\mathbb{P}}(1)$ and $\widetilde{g}_n(\cdot) \Rightarrow \mathcal{G}(\cdot)$.

(iii.d) Given Assumption 2 (ii), it follows that $(d\mathcal{G})^2(u) = \sigma^2(u)du$. Approximate $\Delta\widetilde{g}_n(\cdot)$ by $d\mathcal{G}(\cdot)$ and then $\sum_{i=1}^{s_n} C_2(i_n)(\Delta\widetilde{g}_n(i_n))^2 = \int_0^1 C_2(u)\sigma^2(u)du + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1)$ from the condition that $\int_0^1 C_2(u)\sigma^2(u) du = 0$ with probability 1, implying that $n^{-1/2}\sum_{i=1}^{s_n} C_2(i_n)\sigma^2(i_n) = o_{\mathbb{P}}(1)$ by applying theorem 1 (c) of Chui (1971). This also implies that $n^{-1/2}\sum_{i=1}^{s_n} C_2(i_n)\mathbb{E}[\sigma^2(i_n)] = o(1)$ Therefore, $\sqrt{n}\sum_{i=1}^{s_n} C_2(i_n)(\Delta\widetilde{g}_n(i_n))^2 = \frac{1}{\sqrt{n}}\sum_{i=1}^{s_n} C_2(i_n)\{(\sqrt{n}\Delta\widetilde{g}_n(i_n))^2 - \mathbb{E}[\sigma^2(i_n)]\} + o_{\mathbb{P}}(1)$. We here note that $\{(\sqrt{n}\Delta\widetilde{g}_n(i_n))^2\}$ is a mixingale process of size -1 by Assumption 3 such that $\mathrm{var}[\frac{1}{\sqrt{n}}\sum_{i=1}^{s_n} C_2(i_n)\{(\sqrt{n}\ \Delta\widetilde{g}_n(i_n))^2 - \mathbb{E}[\sigma^2(i_n)]\}] = \frac{1}{n}\sum_{i=1}^{s_n} \gamma_1(i_n)C_2(i_n)C_2(i_n)' + \frac{1}{n^2}\sum_{i=1}^{s_n}\sum_{j=1_{i\neq j}}^{s_n} \gamma_2(i_n, j_n)C_2(i_n)C_2(j_n)' + o(1) \to \int_0^1 \gamma_1(u)C_2(u)C_2(u)'du + \int_0^1\int_0^1 \gamma_2(u,v)C_2(u)C_2(v)'dudv = \Gamma$, which is finite by Assumptions 2 (ii.c) and 3. Therefore, it follows from the mixingale CLT (e.g., White, 2001, theorem 5.16) that $\frac{1}{\sqrt{n}}\sum_{i=1}^{s_n} C_2(i_n) \{(\sqrt{n}\Delta\widetilde{g}_n(i_n))^2 - \sigma^2(i_n)\} \Rightarrow \mathcal{Z} \sim \mathcal{N}(0, \Gamma)$ by noting that Γ is positive definite. This implies $\sqrt{n}\sum_{i=1}^{s_n} C_2(i_n)(\Delta\widetilde{g}_n(i_n))^2 \Rightarrow \mathcal{Z}$. Combining this result with (ii.c) gives $D_n = \sqrt{n}\sum_{i=1}^{s_n} C_2(i_n)(\Delta\widetilde{g}_n(i_n))^2 + \sum_{i=1}^{s_n} C_1(i_n)\Delta\widetilde{g}_n(i_n) + \sum_{i=1}^{s_n} C_3(i_n)\widetilde{g}_n(i_n)(\Delta\widetilde{g}_n(i_n))^2 + o_{\mathbb{P}}(1) \Rightarrow \mathcal{Z} + \mathcal{D}_u =: \mathcal{D}_w$. This completes the proof. ■

Proof of Theorem 1: (i) Given Lemma 3 (i) and the fact that $\sqrt{n}(\widehat{\theta}_n - \theta_*) = -\bar{A}_n^{-1}D_n + o_{\mathbb{P}}(1)$, $\sqrt{n}(\widehat{\theta}_n -$

$\theta_*) \Rightarrow -A_d^{-1}\mathcal{D}_d$.

(ii) Given Lemma 3 (ii.d), the desired results follow from the fact that $\sqrt{n}(\widehat{\theta}_n - \theta_*) = -\bar{A}_n^{-1}D_n + o_{\mathbb{P}}(1) \Rightarrow -A_u^{-1}\mathcal{D}_w$. ■

Proof of Theorem 2: (i) We first apply a second-order Taylor expansion to $q_n(\cdot)$ around θ_* and obtain that

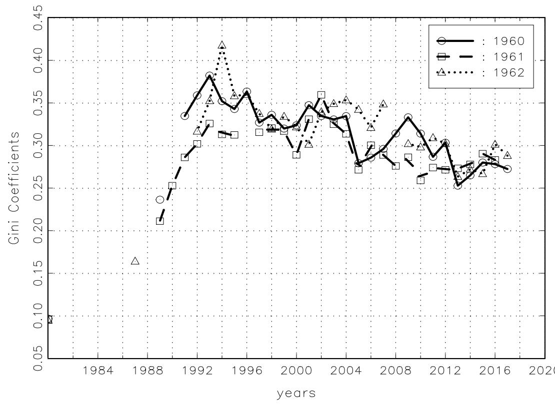
$$J_n := q_n(\widehat{\theta}_n) = q_n(\theta_*) - \sqrt{n}(\widehat{\theta}_n - \theta_*)'(\nabla_\theta \bar{G}_n(\theta_*)\widehat{\Sigma}_n^{-1}\nabla_\theta' \bar{G}_n(\theta_*))\sqrt{n}(\widehat{\theta}_n - \theta_*) + o_{\mathbb{P}}(1). \tag{A.37}$$

It now follows from (1) that $J_n = q_n(\theta_*) - \widetilde{G}_n(\theta_*)'\widehat{\Sigma}_n^{-1}\nabla_\theta'\bar{G}_n(\theta_*)(\nabla_\theta\bar{G}_n(\theta_*)\widehat{\Sigma}_n^{-1}\nabla_\theta'\bar{G}_n(\theta_*))^{-1}\nabla_\theta\bar{G}_n(\theta_*)$ $\widehat{\Sigma}_n^{-1}\widetilde{G}_n(\theta_*) + o_{\mathbb{P}}(1) = q_n(\theta_*) - D_n'\bar{A}_n^{-1}D_n + o_{\mathbb{P}}(1) \Rightarrow \mathcal{Q}_d - \mathcal{D}_d'A_d^{-1}\mathcal{D}_d$ by Lemma 3 (i). We further note that $\mathcal{Q}_d - \mathcal{D}_d'A_d^{-1}\mathcal{D}_d = (\boldsymbol{\Xi}\mathcal{G}(\cdot), \mathcal{G}(\cdot)) - [\lambda_d(\cdot)', \mathcal{G}(\cdot)]A_d^{-1}[\lambda_d(\cdot), \mathcal{G}(\cdot)] = (\boldsymbol{\Xi}\mathcal{G}(\cdot), \mathcal{G}(\cdot)) - [\lambda_d(\cdot)'A_d^{-1}[\lambda_d(\circ), \mathcal{G}(\circ)],$ $\mathcal{G}(\cdot)] = (\boldsymbol{\Xi}\mathcal{G}(\cdot), \mathcal{G}(\cdot)) - ((\lambda_d(\cdot)'A_d^{-1}\lambda_d(\circ), \mathcal{G}(\circ)), \mathcal{G}(\cdot)) = (\boldsymbol{\Pi}_d\mathcal{G}(\cdot), \mathcal{G}(\cdot)) =: \mathcal{J}_d$, where the last equality holds from the fact that $(\lambda_d(\cdot)'A^{-1}\lambda_d(\circ), \mathcal{G}(\circ)) = \lambda_d(\cdot)'A^{-1}\int_0^1 \lambda_d(u)\mathcal{G}(u)du = \boldsymbol{\Lambda}_d\mathcal{G}(\cdot)$, so that $(\boldsymbol{\Xi}\mathcal{G}(\cdot), \mathcal{G}(\cdot)) - ((\lambda_d(\cdot)'A^{-1}\lambda_d(\circ), \mathcal{G}(\circ)), \mathcal{G}(\cdot)) = (\boldsymbol{\Xi}\mathcal{G}(\cdot), \mathcal{G}(\cdot)) - (\boldsymbol{\Lambda}_d\mathcal{G}(\cdot), \mathcal{G}(\cdot)) = ((\boldsymbol{\Xi} - \boldsymbol{\Lambda}_d)\mathcal{G}(\cdot), \mathcal{G}(\cdot)) = (\boldsymbol{\Pi}_d\mathcal{G}(\cdot), \mathcal{G}(\cdot))$.

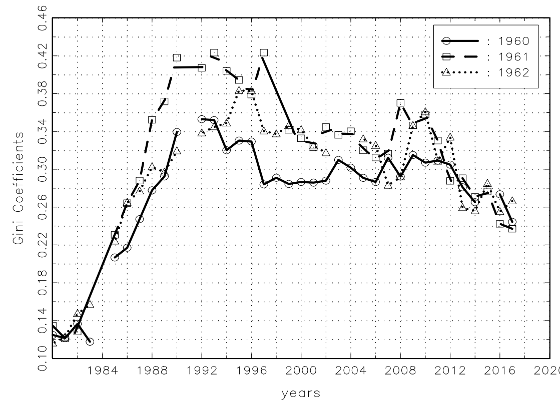
(ii) Given (A.37), it follows that $n^{-1}J_n = n^{-1}q_n(\theta_*) + o_{\mathbb{P}}(1)$. Furthermore, Lemma 3 (ii.a) implies that $n^{-1}q_n(\theta_*) \to q_u := \int_0^1 \sigma^2(u)du$ with probability converging to 1, which is identical to $\int_0^1 \mathbb{E}[\sigma^2(u)]du$. Now $n^{-1}q_n(\theta_*) = n^{-1}\sum_{i=1}^{s_n}(\sqrt{n}\Delta\widetilde{g}_n(i_n))^2$. It follows that $\sqrt{n}(\frac{1}{n}J_n - q_u) = \frac{1}{\sqrt{n}}\sum_{i=1}^{s_n}[(\sqrt{n}\Delta\widetilde{g}_n(i_n))^2 - \mathbb{E}[\sigma^2(i_n)]] + o_{\mathbb{P}}(1)$ by noting that $\left|\frac{1}{n}\sum_{i=1}^{s_n}\sigma^2(i_n) - q_u\right| = o_{\mathbb{P}}(n^{-1/2})$, which is implied by theorem 1 (c) of Chui (1971). Now $\{(\sqrt{n}\Delta\widetilde{g}_n(i_n))^2 - \mathbb{E}[\sigma^2(i_n)]\}$ is a mixingale of size -1, as assumed by Assumption 2 (ii). Therefore, $\text{var}[\frac{1}{\sqrt{n}}\sum_{i=1}^{s_n}\{(\sqrt{n}\Delta\widetilde{g}_n(i_n))^2 - \mathbb{E}[\sigma^2(i_n)]\}] = \frac{1}{n}\sum_{i=1}^{s_n}\gamma_1(i_n) + \frac{1}{n^2}\sum_{i=1}^{s_n}\sum_{j=1_{i\neq j}}^{s_n}\gamma_2(i_n, j_n) + o_{\mathbb{P}}(1) \to \int_0^1 \gamma_1(u)du + \int_0^1\int_0^1 \gamma_2(u,v)dudv =: v^2$, which is finite by Assumptions 2 (ii). Therefore, it follows from the mixingale CLT that $\frac{1}{\sqrt{n}}\sum_{i=1}^{s_n}\{(\sqrt{n}\Delta\widetilde{g}_n(i_n))^2 - \mathbb{E}[\sigma^2(i_n)]\} \overset{\text{A}}{\sim} \mathcal{N}(0, v^2)$, which also implies that $\sqrt{n}(\frac{1}{n}J_n - q_u) + o_{\mathbb{P}}(1) = \frac{(J_n - nq_u)}{\sqrt{n}} + o_{\mathbb{P}}(1) \overset{\text{A}}{\sim} \mathcal{N}(0, v^2)$. From this, we obtain that $U_n \overset{\text{A}}{\sim} \mathcal{N}(0,1)$ under $\mathcal{H}_0$, as required. ■

A.5 Empirical Supplements

In this section we provide the estimated Gini coefficients using annual data of each year. We draw the evolution of the Gini coefficients in parallel to Figures 4 to 7. For the female and male cohorts, Figure A.1 shows the estimated Gini coefficients as functions of time between 1980 and 2018. Likewise, Figures A.2 and A.3 show the estimated Gini coefficients from the data sets classified by education and race, respectively. Figure A.4 shows the Gini coefficients when data sets are not classified. As the implications of these figures are the same as those from Figures 4 to 7, for brevity they are not repeated.

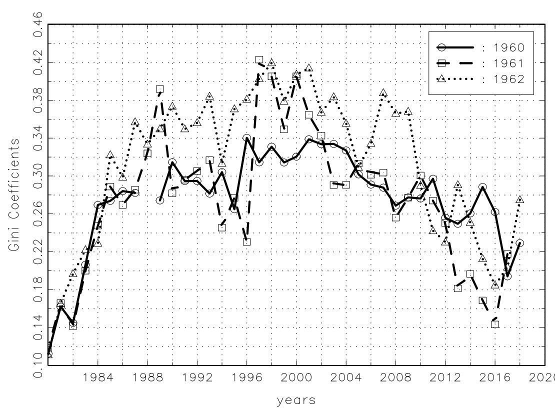


(a) Female

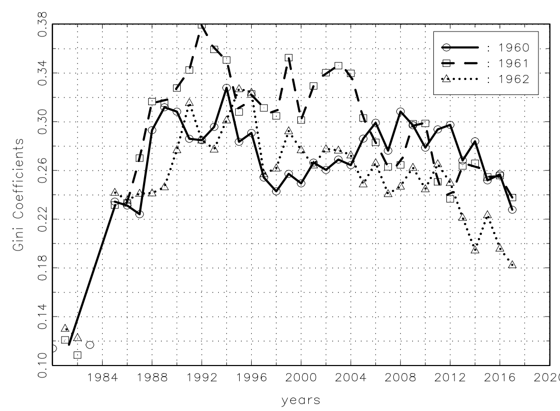


(b) Male

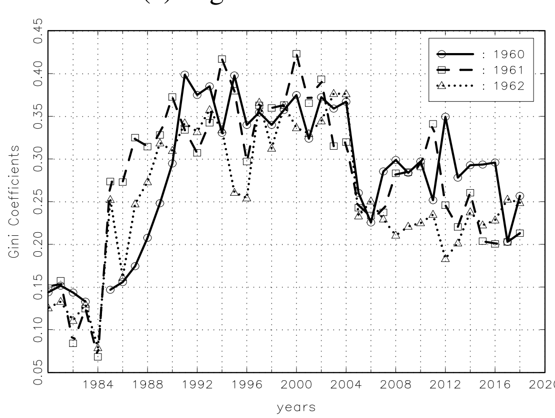
Figure A.1: GINI COEFFICIENTS OF FEMALE AND MALE COHORTS BETWEEN 1980 AND 2018. The figures show the Gini coefficients of female and male cohorts estimated by imposing the Pareto distribution to the top 10% CWHS observations. Missing values signify that the p-value of the U-test is less than 1%.



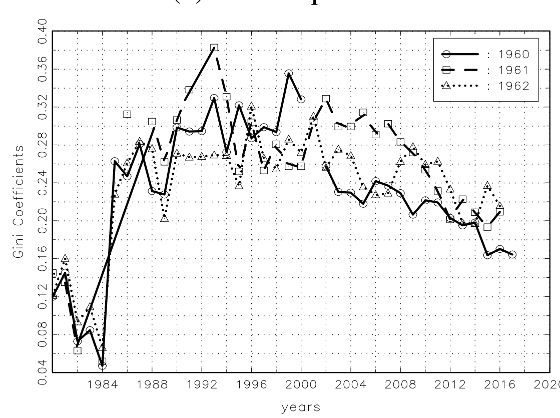
(a) High school or below



(b) BA or equivalent

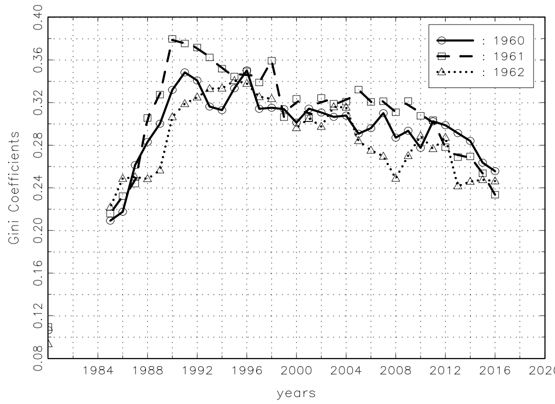


(c) MA or equivalent

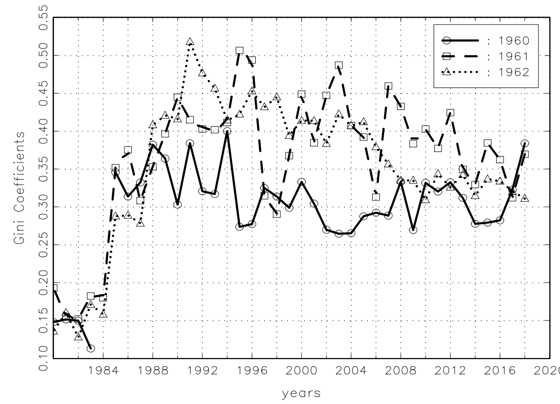


(d) Doctorate or equivalent

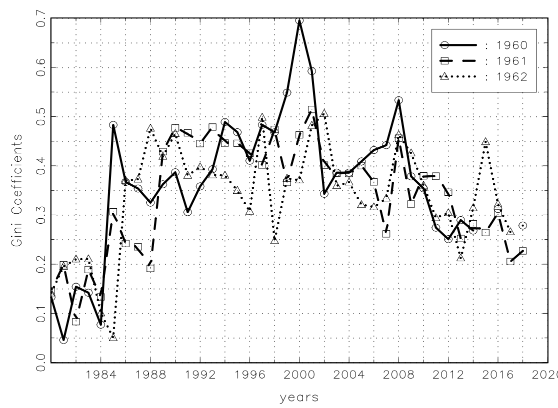
Figure A.2: GINI COEFFICIENTS WITHIN THE SAME EDUCATION COHORTS BETWEEN 1980 AND 2018. The figures show the Gini coefficients within the same education cohorts estimated by imposing the Pareto distribution to the top 10% CWHS observations. Missing values signify that the p-value of the U-test is less than 1%.



(a) White or Caucasian



(b) Black or African American



(c) Asian

Figure A.3: GINI COEFFICIENTS WITHIN THE SAME RACE COHORTS BETWEEN 1980 AND 2018. The figures show the Gini coefficients within the same race cohorts estimated by imposing the Pareto distribution to the top 10% CWHS observations. Missing values signify that the p-value of the U-test is less than 1%.

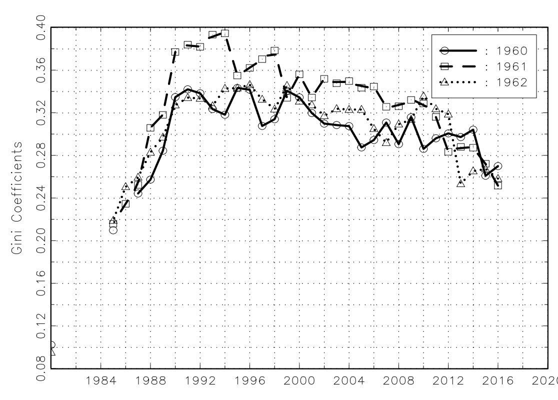


Figure A.4: GINI COEFFICIENTS USING AGGREGATED OBSERVATIONS FOR EACH YEAR BETWEEN 1980 AND 2018. The figures show the Gini coefficients of aggregated observations estimated by imposing the Pareto distribution to the top 10% CWHS observations. Missing values signify that the p-value of the U-test is less than 1%.