Efficient Estimation in Infinite Dimensional GMM

JIN SEO CHO
Yonsei University
Email: jinseocho@yonsei.ac.kr

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

May 2026

Abstract

In GMM estimation it is well known that if the number of moment conditions grows with the sample size, GMM asymptotics differ from the standard case with moment size fixed as the sample size tends to infinity. The present work explores infinite dimensional GMM estimation under various conditions on the moment conditions and the weight matrix. Our approach employs a partial sum process formed by the moment conditions to represent high dimensional moments and an invariance principle to capture the infinite dimensional asymptotics as the moment size grows. Next, the GMM weight matrix is assumed to converge to one of two kernels at the limit: a continuous kernel or the Dirac delta function. Combining these different conditions enables development of a large sample theory for most efficient GMM estimation. The effects of permuting the moment conditions on GMM efficiency are also explored. The resulting theory is applied to weak instrumental variable estimation and the Angrist and Krueger (1991) data are re-analyzed in an empirical application of the new methods.

Key Words: Infinite dimensional GMM; Invariance principle; Neumann's series expansion; Stochastic integral; Weak IV, 2SLS.

Subject Classification: C13, C18, C36, C55, E24.

Acknowledgements: Min Seog Kim provided excellent and invaluable research assistance for our simulation work. Cho acknowledges support from Chang-am Fellowship and R. K. Cho Memorial Fellowship, and Phillips acknowledges support from the Kelly Fund at the University of Auckland and an LKC Fellowship at Singapore Management University.

1 Introduction

Structural estimation using many moment conditions has long been popular in several areas of applied economic research. For example, Angrist and Krueger (1991) in a now classic paper of labor economics estimated the monetary return to education using the instruments formed by an individual's birth year and quarter coupled with their interactions involving other relevant variables. In empirical finance, Jagannathan, Skoulakis, and Wang (2002) employed generalized method of moments (GMM) estimation formed by many moment conditions in estimating asset pricing models. In such applications, efficient estimation of the structural parameters is a useful primary goal. It is critical, for instance, in the determination of an education policy to estimate the returns to education as precisely as possible in learning the role of education in the labor market. Likewise, it is a central interest of financial analysts and investors to better understand the mechanisms of asset pricing and the differing returns from different assets.

In spite of this popularity, structural parameter estimation using standard GMM and instrumental variable (IV) techniques can be misleading as Phillips (1989); Staiger and Stock (1997); Chao and Swanson (2005) and much later research have pointed out. The strength, number (s_n) and stochastic properties of the moment conditions as well as the weight matrix used in defining the GMM distance all have roles in the limit behavior of standard GMM estimation. If even one of these components behaves irregularly, standard GMM asymptotics can fail. For example, if identically and independently distributed (iid) IVs are irrelevant or weak with numbers proportional to the sample size, two-stage least squares (2SLS) is asymptotically biased. The existing literature has provided various approaches to estimation in seeking to resolve the many issues associated with weakness and multiple moment conditions. For example, Nagar (1959) and Donald and Newey (2001) suggested removing bias and asymptotic bias by using *bias-corrected* 2SLS (bc-2SLS); and Anderson and Rubin (1949) and Anderson (2005) suggested limited information maximum-likelihood (LIML) estimation as an alternative approach to bias reduction. But their efficiency and mean squared error levels coupled with methods of improving these and allowing for infinite dimensional cases have not been systematically investigated in the literature.

The present study aims to provide in such irregular settings a systematic approach to infinite dimensional GMM (id-GMM). Just as in standard GMM analysis, efficient estimation remains an important element in id-GMM analysis. But as yet the literature has not tackled this aspect of the subject in depth. To achieve this goal and deal with the complexity of efficiency analysis in id-GMM we employ representations involving function space inner products associated with integral transforms that enable a systematic study of asymptotic efficiency. This approach which involves discretization in finite samples was developed recently in related work (Cho and Phillips, 2025) that dealt exclusively with Brownian motion and Brownian Bridge kernel covariance functions that apply in some particular cases.

The following discussion outlines the general approach taken in the present paper. We first assume that the moment conditions for GMM can be used to define a corresponding partial sum process. The number of moment conditions s_n is assumed to be proportional to the sample size n and the moment conditions are organized to satisfy an invariance principle with the partial sum process weakly converging by functional limit theory to a Wiener process (e.g., Phillips and Solo, 1992; de Jong and Davidson, 2000; Ibragimov and Phillips, 2008). This approach differs from that taken in Cho and Phillips (2025) who study GMM estimation with moment conditions that explicitly form a particular Gaussian process as s_n tends to infinity. We distinguish the two infinite dimensional GMM (id-GMM) estimations by naming the first *extended infinite dimensional* GMM (xid-GMM) and the latter *ordinary infinite dimensional* GMM (oid-GMM). The asymptotic theory turns out to be substantially different between xid-GMM and oid-GMM estimations.

The limit theory for xid-GMM estimation is established by supposing two general types of weight matrices for the moment conditions under certain conditions. Specifically, we suppose xid-GMM is conducted under either strong or weak moments, noting that many empirical applications employ moments that are considered to be weak, while classical GMM asymptotics assume strong moments. In addition, the weight matrix is assumed to converge as $s_n \to \infty$ either to a continuous kernel or to the Dirac delta function. In the first case, an arbitrary positive-definite continuous kernel limit function is assumed for the weight matrix, important practical examples of which include the Brownian motion kernel (BMK), the Brownian bridge kernel (BBK), segmented BMK or BBK kernels (Phillips and Jiang, 2025), and the squared exponential kernel (SEK). In the second case, the weight matrix is assumed to have the limiting form of the Dirac delta function, which is the kernel of the identity integral operator. The latter is motivated by the case where the instrumental variables (IVs) in GMM are iid. If s_n is fixed and finite, the optimal weight matrix in that case is the identity matrix, and the Dirac delta function is its infinite dimensional analogue.

The limit behavior of xid-GMM depends critically on the weight matrix. We select the two limiting forms of weight matrices because they are instrumental in determining GMM asymptotics and provide cornerstone extreme cases that enable the development of separate limit theory for xid-GMM relevant to empirical work. Specifically, the 2SLS or GMM environments employed in earlier work can be regarded as a variant of xid-GMM driven by the Dirac delta function. For example, the unweighted GMM examined by Han and Phillips (2006) is GMM with a weight matrix converging to the Dirac delta function; and Chao and Swanson (2005) and Newey and Windmeijer (2009) examine standard 2SLS and the continuous updating estimator (CUE) formed by homoskedasticity/heteroskedasticity consistent weight matrix under weak moment conditions. We can view the limits of the weight matrices as variants of the Dirac delta function in the sense that the limits are not continuous even for standard data. If the weight matrix converges to the Dirac delta function or its variants and the moment size grows proportional to the sample size, xid-GMM is not necessarily consistent for the unknown parameter as Chao and Swanson (2005), Han and Phillips (2006), and Newey and Windmeijer (2009) discuss. However, if the weight matrix converges to a continuous kernel, this outomce reverses. Even when the moment size grows proportional to the sample size, xid-GMM is asymptotically unbiased and normal under strong and weak moment conditions as we demonstrate below.

We further explore mechanisms to improve efficient estimation. Out of many weight matrices converging to continuous kernels, we aim at choosing a weight matrix that delivers an improved xid-GMM. Our approach notes that information matrix equality is attained if the kernel function used to compute the asymptotic covariance matrix of an xid-GMM is the Dirac delta function, although its discontinuity prevents practical implementation. Instead, we explore an xid-GMM that is driven by a continuous kernel which is asymptotically close to the Dirac delta function. Although the efficiency of the xid-GMM designed in this manner may not reach the efficiency level attained by the Dirac delta function, its efficiency improves as $s_n, n \to \infty$ and does so without inducing asymptotic bias. The environmental setting for this approach is discussed in the paper and a similar approach is explored for bc-2SLS. This technique seeks to take advantage of the good features of both approaches while reducing exposure to their respective disadvantages.

Further enhancement of xid-GMM estimation is possible to improve its efficiency. If the moment condition dimension s_n is large, the moments can be permuted in a vast number of ways, with each permutation producing different estimates even for the same weight matrix. Using these different estimates we may define another estimator, which we term *permuted xid*-GMM (pxid-GMM), and develop a procedure for which pxid-GMM is asymptotically more efficient than each individual xid-GMM. Our procedure also tackles the computational challenge involved in the inversion large matrices multiple times in xid-GMM estimation by using truncated Neumann series expansions which lead to a substantial improvement in computational efficiency.

The paper is organized as follows. Section 2 motivates xid-GMM and examines the limit theory under the two types of moment conditions and the two types of weight matrices employed. The pxid-GMM estimator is defined and analyzed in Section 3. Section 4 focuses on 2SLS with weak IVs and examines bc-2SLS. Section 5 discusses the use of the Neumann series expansion in computation and Section 6 reports the findings from simulations that explore the properties of the estimators defined in Sections 2-4. Section 7 applies the methodology to the empirical data provided by Angrist and Krueger (1991), and Section 8 concludes. An Online Supplement contains proofs, technical material, and additional empirical results.

For ease of reference we introduce some notation. For an arbitrary function $f(\cdot)$ and $j = 1, 2, \ldots$, we use $(d^j/dx^j)f(\bar{x})$ for $(d^j/dx^j)f(x)|_{x=\bar{x}}$. Euclidean distance is denoted $\|\cdot\|$, $L_p(A)$ is the Lebesgue L_p space over A, and $\mathbb{I}(\cdot)$ is the indicator function. The notations $\mathcal{C}^{(2)}(A)$ and $\mathcal{C}(A)$ denote the families of twice continuously differentiable functions and continuous functions defined on A. 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 $(A_i(\cdot), B_j(\cdot))$. Remaining notation is defined as it is introduced in the paper and for convenience is collected in a Glossary at the end of the Online Supplement in Section A.4.

2 Extended id-GMM Estimation

Standard finite dimensional GMM is an extremum estimator constructed by minimizing the GMM distance

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

where n is the sample size and there is assumed to be a unique vector $\theta_* \in \Theta$ satisfying the moment condition $\mathbb{E}[\bar{G}_n(\theta_*)] = 0$, where Θ is a convex and compact parameter space that is a subset of $\mathbb{R}^d$ ($d \in \mathbb{N}$). The sample moment vector $\bar{G}_n(\cdot)$ in (1) is given by

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

with $U_n(w_t, \cdot) : \Theta \mapsto \mathbb{R}^s$ continuously differentiable on Θ with probability 1, and where $\widehat{\Sigma}_n \in \mathbb{R}^{s\times s}$ is a symmetric, positive-definite matrix for each n. Here, $\{w_t \in \mathbb{R}^r : t = 1, 2, \ldots, n\}$ is a sequence of strictly stationary and ergodic random variables defined on a complete probability space. We let $\widehat{\theta}_n$ denote the GMM estimator, viz., $\widehat{\theta}_n := \arg\min_{\theta\in\Theta} \bar{q}_n(\theta)$, and call the dimension s of $G_n(\cdot)$ the *size* of the moment conditions.

Typically, the moment size s is invariant to the sample size n, although this is not always the case in practical work. Theory often provides a large number of possible moment conditions and practical implementations of GMM can reveal the potential for additional conditions in large datasets. Hansen (1982) and Bates and White (1985) among many others examine the asymptotic behavior of the GMM estimator for the case where $\widehat{\Sigma}_n$ is a consistent covariance estimator for Σ_*, say. If Σ_* is the covariance matrix of $U_{n,t}(\theta_*)$ and $H_*\Sigma_*^{-1}H_*'$ is positive definite, under standard conditions the GMM estimator is asymptotically normal and efficient with limit distribution $\sqrt{n}(\widehat{\theta}_n - \theta_*) \rightsquigarrow \mathcal{N}(0, (H_* \Sigma_*^{-1} H_*')^{-1})$, where $U_{n,t}(\cdot)$ abbreviates $U_n(w_t, \cdot)$ and $H_* := \mathbb{E}[\nabla_\theta U_{n,t}(\theta_*)]$.

The formulation in the present paper allows the moment size to depend on the sample size, so that $s = s_n$. Similarly indexed as being sample size dependent in dimension are $\Sigma_{n,*} = \mathrm{cov}(U_{n,t}(\theta_*))$ and $H_{n,*} = \mathbb{E}[\nabla_\theta U_{n,t}(\theta_*)]$. We assume that $s_n \to \infty$ diverges proportionally to n, so that $s_n = s_* n + o(n)$ for some $s_* > 0$. Estimation in this context is referred to as *infinite dimensional* GMM (id-GMM). A number of earlier studies examine econometric models under similar divergence conditions in addition to the examples below (e.g., Carrasco and Florens, 2000; Ledoit and Wolf, 2003; Shi, 2016). Our work is motivated particularly by the popular econometric approach 2SLS, but allows for weak IVs and IV dimension that tends to infinity, which we call *weak infinite dimensional* 2SLS (wid-2SLS).

Example (wid-2SLS)**.** In a standard linear structural model 2SLS has the following general form

$$\widehat{\theta}_n^{(s)} := (X'Z(\widehat{\Sigma}_n^{(s)})^{-1}Z'X)^{-1}X'Z(\widehat{\Sigma}_n^{(s)})^{-1}Z'y,$$

where y is an n-vector of observations of a dependent variable $y_t \in \mathbb{R}$, Z is an $n \times s$ matrix of IVs $z_t \in \mathbb{R}^s$, X is the $n \times d$ observation matrix of d explanatory variables $x_t \in \mathbb{R}^d$, and $\widehat{\Sigma}_n^{(s)}$ is a consistent estimator of a positive-definite matrix $\Sigma^{(s)}$. Let $u_t := y_t - x_t'\theta_*$ be the structural equation disturbance and suppose that $\{w_t := (x_t', z_t', u_t)' \in \mathbb{R}^{d+s+1} : t = 1, 2, \ldots, n\}$ is a sequence of iid observations. Under this condition and setting $\bar{G}_n(\theta) = n^{-1}Z'(y - X\theta)$ and $\widehat{\Sigma}_n = \widehat{\Sigma}_n^{(s)}$, 2SLS is in standard GMM form, minimizes $\bar{q}_n(\cdot)$, and satisfies $\widehat{\theta}_n^{(s)} \xrightarrow{\mathbb{P}} \theta_*$ with asymptotic distribution $\sqrt{n}(\widehat{\theta}_n^{(s)} - \theta_*) \rightsquigarrow \mathcal{N}[0, (P'(\Sigma^{(s)})^{-1}P)^{-1}]$, where $P := \mathbb{E}[z_t x_t']$.

The GMM framework for 2SLS estimation may be extended to accommodate high dimensional and weak instruments by setting $s = s_n$ with $s_n/n \to s_* > 0$ and using local-to-zero conditions on instrument relevance so that $\mathbb{E}[z_t x_t'] = O(n^{-\frac{1}{2}})$, the latter implying that the $s_n \times d$ matrix $\mathbb{E}[z_t x_t']$ has an increasing number of rows all of whose elements tend to zero at the special rate $O(n^{-\frac{1}{2}})$ as $n \to \infty$. Angrist and Krueger (1995), Donald and Newey (2001), and Hahn, Hausman, and Kuersteiner (2001) explore empirical settings and models with large numbers of instruments, and Staiger and Stock (1997) consider models with a finite number of weak instruments having the special setting of $O(n^{-\frac{1}{2}})$ relevance as $n \to \infty$. The latter setting is made primarily for asymptotic convenience and many other formulations are possible. In fact, weak IV with the specific moment matrix formulation of the order $O(n^{-\frac{1}{2}})$ is an example of a more general problem that arises in econometric modeling in the presence of a weak regression signal (Phillips, 2016).

A large literature has emerged following this early work on modified 2SLS and IV estimation under such extensions. Chao and Swanson (2005), Han and Phillips (2006), and Newey and Windmeijer (2009), among many others, explore the asymptotic properties of 2SLS in such circumstances. The primary lesson to emerge from these works is that the standard GMM estimator is typically inconsistent for the parameter of interest unless special conditions apply and that other procedures, conditions, or modifications to GMM are generally needed to achieve consistency. However, even in the absence of consistency when instrument relevance is of order $O(n^{-1/2})$, IV estimation still carries information about the parameter of interest and has limit theory that reflects what would in a Gaussian error setting be an exact finite sample distribution, as explained in (Phillips, 1989).[1]

As implied by the above research and as the following analysis shows, wid-2SLS entails quite different asymptotic behavior from standard GMM estimation. The analysis given below provides conditions for wid-2SLS to be consistent for the parameter of interest and to have an asymptotic normal distribution. Further, it provides an environment under which wid-2SLS is asymptotically most efficient. For this purpose the analysis provides a novel perspective on limit behavior by way of a stochastic integral representation. □

The limit behavior of id-GMM is influenced by the limiting properties of $\bar{G}_n(\theta_*)$. Many variants are possible and the present work focuses on the case, where $\sqrt{n}\bar{G}_n(\theta_*)$ is weakly correlated and its partial sum process converges to a stochastic integral. For such a case, more extensive analysis is required than that conducted by Cho and Phillips (2025), who suppose the moment conditions form a specific Gaussian process weakly. We refer to the current case as *extended* id-GMM (xid-GMM), whereas id-GMM in Cho and Phillips (2025) is referred to as *ordinary* id-GMM (oid-GMM).

Other approaches have been pursued in the literature that are relevant to parameter estimation which employs high dimensional moment equalities. Grenander (1981) developed a theory of abstract inference in a function space setting that relates to the current approach taken here. Carrasco and Florens (2000) work with a model that falls into the id-GMM estimation category. They note the associated covariance operator does not generally satisfy Picard's (1910) condition for the existence of a linear inverse operator at the limit, and they instead obtain limiting properties by way of Tikhonov regularization using spectral decomposition. Shi (2016) uses a different approach to parameter estimation by maximizing a constrained empirical likelihood and selecting informative moment conditions under circumstances that are similar to the current study. Donald and Newey (2001), Bai and Ng (2010), and Belloni, Chen, Chernozhukov, and Hansen (2012) also explore parameter estimation with high dimensional moment conditions with a focus on linear structural models. Donald and Newey (2001) assume $s_n^2 = o(n)$ and select the set of instruments by mean squared error. Bai and Ng (2010) assume $s_n = o(n)$ and estimate parameters by exploiting a specific data structure. Belloni et al. (2012) select 'informative' IVs by means of Lasso to estimate the unknown parameter by 2SLS under the condition that $\log(s_n) = o(n^{\frac{1}{2}})$.

The following assumption provides a data generating process (DGP) framework that governs the data properties and structures for xid-GMM estimation.

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;*

[1] In the unidentified structural equation case, the exact distribution involves a rational function of central Wishart distribution components, whereas in the weak $O(n^{-1/2})$ IV case, the exact distribution involves a rational function of non-central Wishart distribution components, for which the exact distribution is given in Phillips (1980). The central limit theory that delivers the weak IV asymptotics in the specific $O(n^{-1/2})$ setting then follows precisely as proved in (Phillips, 1989, Lemma 2.3), and as assumed in Staiger and Stock (1997).

(ii) *for each* $n \in \mathbb{N}$, $U_n : \mathbb{R}^p \times \Theta \mapsto \mathbb{R}^{s_n}$ *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, $\mathbb{E}[U_n(\theta_*)] = 0$*, and* $\widehat{\Sigma}_n^{-1}$ *is symmetric and positive definite uniformly in* n*, 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*

(v) $s_n/n = s_* + o(1)$ *for some fixed* $s_* > 0$. □

As in standard cases, the limit distribution of the xid-GMM estimator $\widehat{\theta}_n$ is determined by the limit behavior of the triplet $(\bar{G}_n(\theta_*), \widehat{\Sigma}_n, \nabla_\theta \bar{G}_n(\theta_*))$ by way of the linearized approximation of GMM by Taylor expansion as

$$\sqrt{n}(\widehat{\theta}_n - \theta_*) = -\left[\nabla_\theta \bar{G}_n(\theta_*)\widehat{\Sigma}_n^{-1}\nabla_\theta' \bar{G}_n(\theta_*)\right]^{-1}\left[\nabla_\theta \bar{G}_n(\theta_*)\widehat{\Sigma}_n^{-1}\widetilde{G}_n(\theta_*)\right] + o_{\mathbb{P}}(1), \tag{2}$$

where $\widetilde{G}_n(\theta_*) := \sqrt{n}\bar{G}_n(\theta_*)$. To characterize the limit behavior of xid-GMM, we translate the components of the triplet to a form in which they are stochastic processes defined on the unit interval or unit square.

We first translate the weight matrix $\widehat{\Sigma}_n^{-1}$ by mapping the elements into the unit square. Given that $\widehat{\Sigma}_n$ is an $s_n \times s_n$ matrix, it is possible to partition the unit square into s_n^2 small squares and place the values of $\widehat{\Sigma}_n$ at the crossing points so that a two-dimensional càdlàg function can be defined on $[0, 1]^2$ as follows:

$$\widehat{\sigma}_n(u_1, u_2) := \begin{cases} 0, & \text{if } u_1 \in [0, \frac{1}{s_n}] \text{ or } u_2 \in [0, \frac{1}{s_n}], \\ \widehat{\Sigma}_n^{(j,i)}, & \text{if } u_1 \in [\frac{j}{s_n}, \frac{j+1}{s_n}),\ u_2 \in [\frac{i}{s_n}, \frac{i+1}{s_n}), \text{ and } j, i = 1, 2, \ldots, s_n - 1, \\ \widehat{\Sigma}_n^{(s_n,i)}, & \text{if } u_1 = 1,\ u_2 \in [\frac{i}{s_n}, \frac{i+1}{s_n}), \text{ and } i = 1, 2, \ldots, s_n, \\ \widehat{\Sigma}_n^{(j,s_n)}, & \text{if } u_1 \in [\frac{j}{s_n}, \frac{i+1}{s_n}),\ u_2 = 1, \text{ and } j = 1, 2, \ldots, s_n, \text{ and} \\ \widehat{\Sigma}_n^{(s_n,s_n)}, & \text{if } u_1 = 1, \text{ and } u_2 = 1, \end{cases} \tag{3}$$

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 $i, j = 1, 2, \ldots, s_n$. In the same manner, we can also map the unit square to the values in $\widehat{\Sigma}_n^{-1}$ and assume that its limit shape converges to a kernel $\xi(\cdot, \circ)$ with convergence rate s_n^{-r} for some $r \geq 0$. More specifically, if we let $\xi_n(\cdot, \circ)$ be a function defined on the unit square obtained by mapping the unit square to the values in $\widehat{\Sigma}_n^{-1}$ and form $B_n := [b_n(\frac{1}{s_n}), b_n(\frac{2}{s_n}), \ldots, b_n(1)]'$ and $C_n := [c_n(\frac{1}{s_n}), c_n(\frac{2}{s_n}), \ldots, c_n(1)]'$, then for some $r \geq 0$ and a well-defined kernel $\xi(\cdot, \circ)$, we can create the following finite sample representation and limiting form

$$\frac{1}{s_n^{2+r}} B_n' \widehat{\Sigma}_n^{-1} C_n = \frac{1}{s_n^r}\int_0^1\int_0^1 b_n(u_1)\xi_n(u_1, u_2)c_n(u_2)du_1 du_2$$

$$= \int_0^1 \int_0^1 b(u_1)\xi(u_1,u_2)c(u_2)du_1du_2 + o_{\mathbb{P}}(1). \tag{4}$$

Here, we let $b_n(\cdot) := b(\lfloor n(\cdot)\rfloor/n)$ and $c_n(\cdot) := c(\lfloor n(\cdot)\rfloor/n)$ with $b(\cdot)$ and $c(\cdot)$ being continuous functions on $[0,1]$. This formulation handles the complexity involved in the inversion $\widehat{\Sigma}_n^{-1}$ and makes it more readily available for asymptotic analysis, as explained in Cho and Phillips (2025). For example, when $\widehat{\Sigma}_n$ is based on a Brownian motion kernel (BMK) or Brownian bridge kernel (BBK), Lemma 1 (i.a) of Cho and Phillips (2025) shows that $\xi(\cdot,\circ)$ is characterized by the generalized derivative $\delta''(\cdot-\circ)$, where $\delta(\cdot)$ is the Dirac delta function.[2] Furthermore, the double integral in (4) can be rewritten as an inner product involving an integral transform operator of $[0,1]^2$. That is, if we let Ξ_n and Ξ be the integral transform operators with kernels $\xi_n(\cdot,\circ)$ and $\xi(\cdot,\circ)$, respectively, viz., $\widehat{\Xi}_n c_n(\cdot) := \int_0^1 \xi_n(\cdot,u_2)c_n(u_2)du_2$ and $\Xi c(\cdot) := \int_0^1 \xi(\cdot,u_2)c(u_2)du_2$, then

$$\int_0^1 \int_0^1 b_n(u_1)\xi_n(u_1,u_2)c_n(u_2)du_1du_2 = (b_n(\cdot), \widehat{\Xi}_n c_n(\cdot)) \quad \text{and}$$

$$\int_0^1 \int_0^1 b(u_1)\xi(u_1,u_2)c(u_2)du_1du_2 = (b(\cdot), \Xi c(\cdot)).$$

With this notation, (4) above can be rewritten as $s_n^{-2-r}B_n'\widehat{\Sigma}_n^{-1}C_n = s_n^{-r}(b_n(\cdot), \widehat{\Xi}_n c_n(\cdot)) = (b(\cdot), \Xi c(\cdot)) + o_{\mathbb{P}}(1)$.

For the present study, we assume two extreme cases for the limiting kernel $\xi(\cdot,\circ)$. First, let $\xi(\cdot,\circ)$ be a continuous kernel on the unit square; and second, let $\xi(\cdot,\circ) = \delta(\cdot-\circ)$. These two limiting kernels are selected deliberately, as now explained.

Picard (1910) provided necessary and sufficient conditions for the existence of an inverse operator. But most kernels of interest do not satisfy the Picard conditions and regularization approaches are typically used to address this challenge. Carrasco and Florens (2000) and Kirsch (1996), for example, regularize $\widehat{\Sigma}_n^{-1}$ and examine the properties of the resulting estimators. The continuity condition here precludes the use of regularization. The second kernel is motivated by the assumption that the moment conditions are iid. In this framework, the covariance matrix of the associated moment conditions is diagonal for each n, and for finite s_n, the optimal GMM is defined by estimating using the identity matrix for the weight matrix. The infinite dimensional version of the identity matrix is the Dirac delta function, motivating $\delta(\cdot-\circ)$ as the kernel of the GMM distance, so that unweighted GMM can be viewed as GMM driven by the Dirac delta function. As we demonstrate below, the limit behavior of xid-GMM is substantially different under each weight matrix condition. Specifically, it turns out that xid-GMM is consistent and asymptotically normal if a continuous

[2] The intuition for this limiting kernel is that BMK is the covariance kernel of an integrated process, so that the inverse kernel is associated with an 'anti-integrated process', viz., a differential, c.f., Carrasco, Florens, and Renault (2007).

kernel is employed, whereas it is not necessarily consistent under the Dirac delta function but its asymptotic variance is the smallest. In other words, if a continuous kernel is employed for xid-GMM, its efficiency level turns out to be bounded from below by the asymptotic covariance matrix driven by the Dirac delta function, meaning that a substantial degree of efficiency can be expected by letting $\xi(\cdot, \circ)$ be a continuous function close to $\delta(\cdot - \circ)$. The two kernels are selected to exploit these features and are relevant in practical work as shown in the empirical application of Cho and Phillips (2026).

These conditions are collected in the following assumption.

Assumption 2. *(i)* $\widehat{\sigma}_n(\cdot, \circ)$ *is positive definite for each* n*; and*

(ii) for any càdlàg functions $b_n(\cdot)$ *and* $c_n(\cdot)$ *converging to* $b(\cdot) \in \mathcal{C}([0,1])$ *and* $c(\cdot) \in \mathcal{C}([0,1])$ *weakly or uniformly on* $[0,1]$*, there is an* $r \geq 0$ *such that* $s_n^{-r}(\widehat{\Xi}_n b_n(\cdot), c_n(\cdot)) = (\Xi b(\cdot), c(\cdot)) + o_{\mathbb{P}}(1)$*, with limiting operator* Ξ *satisfying either of the following conditions:*

(ii.a) an integral transform operator Ξ *with a continuous kernel* $\xi(\cdot, \circ)$*;*

(ii.b) an integral transform operator Ξ *with* $\xi(\cdot, \circ) = \delta(\cdot - \circ)$*.* □

We next transform $\bar{G}_n(\theta_*)$ into a càdlàg function defined on $[0,1]$. We define

$$g_n(u) := \begin{cases} 0, & \text{if } u \in [0, \frac{1}{s_n}), \\ \bar{G}_{n,j}(\theta_*), & \text{if } u \in [\frac{j}{s_n}, \frac{j+1}{s_n}),\ j = 1, 2, \ldots, s_n - 1; \text{ and} \\ \bar{G}_{n,s_n}(\theta_*), & \text{if } u = 1, \end{cases} \tag{5}$$

where $\bar{G}_{n,j}(\theta_*)$ is the j-th row element of $\bar{G}_n(\theta) \in \mathbb{R}^{s_n}$. Note that $\bar{G}_n(\theta_*)$ has s_n rows, and the above function $g_n(\cdot)$ is defined by translating $\bar{G}_n(\theta_*)$ into a function defined on the unit interval. If the moment condition $\sqrt{n}\bar{G}_{n,j}(\theta_*)$ is weakly correlated with $\sqrt{n}\bar{G}_{n,i}(\theta_*)$ $(j \neq i)$, $\sqrt{n}g_n(\cdot, \theta_*)$ may form a stochastic differential in the limit instead of a continuous Gaussian process as assumed in oid-GMM. Specifically, we suppose that $\sqrt{n}g_n(\cdot)$ is separated into two components with a central limit theorem (CLT) applying to one and a functional central limit theorem (FCLT) applying to the other. Thus, for some $\widetilde{u}_n$ and $\eta_{n,j} \in \mathbb{R}$, let

$$s_n(\cdot) := \frac{1}{\sqrt{s_n}} \sum_{j=1}^{\lfloor(\cdot)s_n\rfloor} \{\widetilde{G}_{n,j}(\theta_*) - \eta_{n,j}\widetilde{u}_n\} = \sqrt{s_n} \int_0^{(\cdot)} \{\widetilde{g}_n(u) - \eta_n(u)\widetilde{u}_n\} du,$$

and suppose that $\widetilde{u}_n$ satisfies a CLT and that $s_n(\cdot)$ converges weakly to a Brownian motion $\mathcal{B}_s(\cdot)$, with

$$\eta_n(u) := \begin{cases} 0, & \text{if } u \in [0, \frac{1}{s_n}), \\ \eta_{n,j}, & \text{if } u \in [\frac{j}{s_n}, \frac{j+1}{s_n}),\ j = 1, 2, \ldots, s_n - 1; \\ \eta_{s_n}, & \text{if } u = 1. \end{cases} \tag{6}$$

The formulation (6) can be obtained by noting that the zero moment condition $\mathbb{E}[U_{n,t}(\theta_*)] = 0$ may hold because different j's satisfy the moment condition separately, or because the same moment condition is shared by each j. For example, in the context of 2SLS, if we let the j-th IV of z_t be denoted $z_{t,j}$, it can be partitioned into two pieces: $z_{t,j} = \bar{z}_{t,j} + \mathbb{E}[z_{t,j}]$ by letting $\bar{z}_{t,j} := (z_{t,j} - \mathbb{E}[z_{t,j}])$, so that $\widetilde{G}_{n,j}(\theta_*) = n^{-\frac{1}{2}} \sum_{t=1}^{n} \bar{z}_{t,j} u_t + \mathbb{E}[z_{t,j}] n^{-\frac{1}{2}} \sum_{t=1}^{n} u_t$, and we let $\eta_{n,j}$ and $\widetilde{u}_n$ be $\mathbb{E}[z_{t,j}]$ and $n^{-\frac{1}{2}} \sum_{t=1}^{n} u_t$, respectively. For the second term on the right side, its asymptotic distribution can be obtained by application of a suitable CLT. But a CLT does not apply to the first term because the moment conditions differ for each j. We instead suppose that $s_n^{-\frac{1}{2}}\{\widetilde{g}(\cdot) - \eta_n(\cdot)\widetilde{u}_n\}$ forms a stochastic differential in the limit that corresponds to the differential of a Brownian motion, $d\mathcal{B}_s(\cdot)$. In view of this structure the analysis of xid-GMM differs from oid-GMM assuming that $\widetilde{g}_n(\cdot)$ asymptotically forms a continuous Gaussian stochastic process. As it turns out, the limit distribution of xid-GMM is determined by $s_n^{-\frac{1}{2}}\{\widetilde{g}(\cdot) - \eta_n(\cdot)\widetilde{u}_n\}$.

We now formulate the score function $\nabla_\theta \bar{G}_n(\theta_*)$ to facilitate derivation of the limit theory. As earlier, $\nabla_\theta \bar{G}_n(\theta_*)$ is written in terms of a set of functions defined on the unit interval. For $j = 1, 2, \ldots, s_n$ and $i = 1, 2, \ldots, d$, 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 for each $i = 1, 2, \ldots, d$, define

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

and let $H_n(\cdot) := [H_{n,1}(\cdot), H_{n,2}(\cdot), \ldots, H_{n,d}(\cdot)]'$. As in the construction of $g_n(\cdot)$, $H_{n,i}(\cdot)$ has a jump at each increment in j/s_n, where $j = 1, 2, \ldots, s_n$.

This formulation enables (2) to be represented in terms of Gramian matrices. Specifically, letting $\bar{A}_n := [\widehat{\boldsymbol{\Xi}}_n H_n(\cdot), H_n(\cdot)]$ and $\widetilde{d}_n := [\widehat{\boldsymbol{\Xi}}_n H_n(\cdot), \widetilde{g}_n(\cdot)]$, the xid-GMM can be rewritten as follows:[3]

$$\begin{aligned} \sqrt{n}(\widehat{\theta}_n - \theta_*) &= -\left[\nabla_\theta \bar{G}_n(\theta_*) \widehat{\Sigma}_n^{-1} \nabla'_\theta \bar{G}_n(\theta_*)\right]^{-1} \left[\nabla_\theta \bar{G}_n(\theta_*) \widehat{\Sigma}_n^{-1} \widetilde{G}_n(\theta_*)\right] + o_{\mathbb{P}}(1) \\ &= -[\widehat{\boldsymbol{\Xi}}_n H_n(\cdot), H_n(\cdot)]^{-1} [\widehat{\boldsymbol{\Xi}}_n H_n(\cdot), \widetilde{g}_n(\cdot)] + o_{\mathbb{P}}(1) = -\bar{A}_n^{-1} \widetilde{d}_n + o_{\mathbb{P}}(1), \end{aligned} \tag{7}$$

meaning that the limit distribution of $\widehat{\theta}_n$ is delivered through the limits of $\bar{A}_n$ and $\widetilde{d}_n$. From (7), asymptotics of xid-GMM are evidently affected by those of $H_n(\cdot)$. Correspondingly, the next step allows two forms of asymptotic behavior for $H_n(\cdot)$, according to whether strong or weak moment conditions apply.

[3]In deriving (7) it is implicitly assumed that the primary component of the Hessian matrix $\nabla^2_\theta \bar{q}_n(\theta_*)$ is determined by $\nabla_\theta \bar{G}_n(\theta_*) \widehat{\Sigma}_n^{-1} \nabla'_\theta \bar{G}_n(\theta_*)$, and the other term is asymptotically negligible. See Assumption 3 (ii) given below.

2.1 Strong Moment Conditions

Starting with a strong moment condition for $H_n(\cdot)$, the following representation of the centered and scaled score function is used in examining $\bar{A}_n := [\widehat{\boldsymbol{\Xi}}_n H_n(\cdot), H_n(\cdot)]$. For some $\nu \in [0, \frac{1}{2}]$, $\kappa \geq 0$, and $\mu_{n,j} \in \mathbb{R}^d$,

$$\frac{\sqrt{n}}{s_n^{\nu}} \left(\frac{1}{s_n^{\kappa}} \nabla_\theta \bar{G}_{n,j}(\theta_*) - \mu_{n,j} \right) = O_{\mathbb{P}}(1), \tag{8}$$

uniformly in $j = 1, 2, \ldots$. In case $\nabla_\theta \bar{G}_{n,j}(\theta_*)$ itself does not converge uniformly in j, (8) allows for the convergence rate to be adjusted by scaling with s_n^{κ} for some rate κ.

This assumption is effective in practice in various circumstances. For example, in the 2SLS analysis considered in Section 4, we suppose that the endogenous variable x_t is associated with an instrument z_t so that $x_t' = z_t'\phi_n + v_t'$ with $\phi_n = O(1)$. Then, if for each j, $\mathbb{E}[z_{t,j}] = 0$ without loss of generality, it follows that

$$\frac{1}{n}\sum_{t=1}^{n} z_{t,j}x_t - \sum_{\ell=1}^{s_n} \mathbb{E}[z_{t,j}z_{t,\ell}]\phi_{n,\ell} = \frac{1}{n}\sum_{t=1}^{n}\left(z_{t,j}\sum_{\ell=1}^{s_n} z_{t,\ell}\phi_{n,\ell} - \sum_{\ell=1}^{s_n}\mathbb{E}[z_{t,j}z_{t,\ell}]\phi_{n,\ell}\right) + \frac{1}{n}\sum_{t=1}^{n} z_{t,j}v_t,$$

where $\phi_{n,\ell}$ denotes the ℓ-th row element of ϕ_n. Here, for some κ and ν, if we suppose $\sum_{\ell=1}^{s_n}\mathbb{E}[z_{t,j}z_{t,\ell}]\phi_{n,\ell} = O(s_n^{\kappa})$ and the right side is $O_{\mathbb{P}}(s_n^{\nu+\kappa}n^{-\frac{1}{2}})$, then

$$\frac{\sqrt{n}}{s_n^{\nu}}\left(\frac{1}{s_n^{\kappa}}\left\{ \frac{1}{n}\sum_{t=1}^{n} z_{t,j}x_t \right\} - \frac{1}{s_n^{\kappa}}\sum_{\ell=1}^{s_n}\mathbb{E}[z_{t,j}z_{t,\ell}]\phi_{n,\ell}\right) = O_{\mathbb{P}}(1),$$

for each $j = 1, 2, \ldots$. From this result, (8) follows by letting $\nabla_\theta \bar{G}_{n,j}(\theta_*) := n^{-1}\sum_{t=1}^{n} z_{t,j}x_t$ and $\mu_{n,j} := s_n^{-\kappa}\sum_{\ell=1}^{s_n}\mathbb{E}[z_{t,j}z_{t,\ell}]\phi_{n,\ell}$. This approach effectively extends the moment condition between x_t and z_t when the moment size is fixed: that is, if $s_n = s$ and $\phi_n = \phi$, the condition can equivalently be rewritten as

$$\sqrt{n}\left(\frac{1}{n}\sum_{t=1}^{n} z_t x_t' - \mathbb{E}[z_t z_t']\phi \right) = O_{\mathbb{P}}(1),$$

which holds by virtue of a CLT under mild regularity conditions. Under our DGP condition we let s_n be proportional to n and scale the moment conditions appropriately by s_n^{ν} and s_n^{κ} to facilitate ransformation into a limiting stochastic process.

Now assume that the formulation in (8) weakly converges to a stochastic process by rewriting it using $H_n(\cdot)$. Specifically, let

$$a_n(\cdot) := \sqrt{s_n}\int_0^{(\cdot)} \frac{\sqrt{n}}{s_n^{\nu}}\left(\frac{1}{s_n^{\kappa}} H_n(u) - \mu_n(u) \right) du, \tag{9}$$

and suppose that $a_n(\cdot)$ weakly converges to a Brownian motion $\mathcal{B}_a(\cdot)$, where

$$\mu_n(u) := \begin{cases} 0, & \text{if } u \in [0, \frac{1}{s_n}), \\ \mu_{n,j}, & \text{if } u \in [\frac{j}{s_n}, \frac{j+1}{s_n}), j = 1, 2, \ldots, s_n - 1; \\ \mu_{n,s_n}, & \text{if } u = 1. \end{cases}$$

Different DGPs and model conditions imply different results here. If $\nu < \frac{1}{2}$, then (9) implies that $(s_n^{-\kappa} H_n(\cdot) - \mu_n(\cdot)) = o_{\mathbb{P}}(1)$ uniformly on $[0, 1]$, so that $s_n^{-\kappa} H_n(\cdot)$ uniformly converges to the limit of $\mu_n(\cdot)$, meaning that the limit of $\bar{A}_n$ can be delivered by continuous mapping. On the other hand, if $\nu = \frac{1}{2}$, $s_n^{-\kappa} H_n(\cdot)$ does not necessarily converge to the limit of $\mu_n(\cdot)$ from the fact that $s_n^{-\kappa} H_n(\cdot) - \mu_n(\cdot) = O_{\mathbb{P}}(1)$. Depending on the kernel of Ξ, the limit of $\bar{A}_n$ is also different. If $\xi(\cdot, \circ)$ is continuous, the limit of $\bar{A}_n$ can still be obtained by continuous mapping, but that is not the case for the limiting kernel $\delta(\cdot - \circ)$, as detailed below.

Next, we discuss regularity conditions for $\widetilde{d}_n := [\widehat{\Xi}_n H_n(\cdot), \widetilde{g}_n(\cdot)]$, whose asymptotic behavior is determined by the interactions between $H_n(\cdot)$ and $\widetilde{g}_n(\cdot)$ through $\widehat{\Xi}_n$. As for $\bar{A}_n$, if $\xi(\cdot, \circ)$ is continuous, the limit of $\widetilde{d}_n$ is delivered by continuous mapping, but more care is required for the Dirac delta function case. Before developing asymptotic behavior for the components $\bar{A}_n$ and $\widetilde{d}_n$, we formally state the assumptions employed so far.

Assumption 3. *(i) For some $\nu \in [0, \frac{1}{2}]$ and $\kappa \geq 0$, $\sqrt{n} s_n^{-\nu}(s_n^{-\kappa} H_n(\cdot) - \mu_n(\cdot)) = O_{\mathbb{P}}(1)$ uniformly on $[0, 1]$, where $\mu_n(\cdot) : [0, 1] \mapsto \mathbb{R}^d$, and for some nonzero $\mu(\cdot)$ and $\chi(\cdot) \in L_2([0, 1])$ and for some $\beta > 0$, $\mu_n(\cdot) = \mu(\cdot) + s_n^{-\beta} \chi(\cdot)$;*
(ii) $s_n^{-2} \nabla_\theta^2 \bar{q}_n(\theta_*) = \bar{A}_n + o_{\mathbb{P}}(s_n^{r+2\kappa})$; and
(iii) $[\Xi \mu(\cdot), \mu(\cdot)]$ is positive definite. □

Here, Assumption 3 (ii) supposes that the primary term of $\nabla_\theta^2 \bar{q}_n(\theta_*)$ is determined by $\bar{A}_n$ and the remainder is asymptotically negligible, so that the asymptotic approximation in (2) is effective. In addition, if $\xi(\cdot, \circ) = \delta(\cdot - \circ)$, then $[\Xi \mu(\cdot), \mu(\cdot)] = [\mu(\cdot), \mu(\cdot)]$ in Assumption 3 (iii).

Assumption 4. *Define the $(1 + d)$-vector stochastic process $\ell_n(\cdot) := (s_n(\cdot), a_n(\cdot)')'$.*
(i) For some $\rho(\cdot) \in L_2([0, 1])$ and for some $\alpha > \frac{1}{2}$, $\eta_n(\cdot) = s_n^{-\alpha} \rho(\cdot)$;
(ii) For a normal random variable $\mathcal{U} \sim \mathcal{N}(0, \sigma_u^2)$ and $(1 + d)$-vector standard Wiener process $\mathcal{W}_\ell(\cdot)$, there is a positive definite $\Sigma_\ell \in \mathbb{R}^{(1+d)\times(1+d)}$ such that

$$\begin{bmatrix} \ell_n(\cdot) \\ \widetilde{u}_n \end{bmatrix} \rightsquigarrow \begin{bmatrix} \int_0^{(\cdot)} d\mathcal{B}_\ell(u) \\ \mathcal{U} \end{bmatrix} := \begin{bmatrix} \Sigma_\ell \int_0^{(\cdot)} d\mathcal{W}_\ell(u) \\ \mathcal{U} \end{bmatrix},$$

and for some $\sigma_{\ell u} \in \mathbb{R}^{1+d}$, $\mathbb{E}[\int_0^{(\cdot)} d\mathcal{B}_\ell(u)\mathcal{U}] = \sigma_{\ell u} \int_0^{(\cdot)} du$. □

Prior literature provides primitive conditions for the weak convergence in Assumption 4 (ii) (e.g., Phillips and Solo, 1992; de Jong and Davidson, 2000). Here, if $\mathcal{U}$ and $\mathcal{W}_\ell(\cdot)$ are independent, we simply have

$\sigma_{\ell u} = 0$. For later purposes, partition $\mathcal{B}_\ell(\cdot)$ as $(\mathcal{B}_s(\cdot), \mathcal{B}_a(\cdot)')'$.

Assumptions 2, 3, and 4 help determine the large sample properties of the components constituting xid-GMM under strong moment conditions, which we call *strong* xid-GMM (sxid-GMM) in contrast to the xid-GMM formed by weak moment conditions as given below. We classify the Assumptions into two groups: **Group A** for a continuous kernel $\xi(\cdot, \circ)$; and **Group B** for the Dirac delta kernel $\xi(\cdot, \circ) = \delta(\cdot - \circ)$. Before proceeding with the asymptotic development of sxid-GMM, it is convenient to simplify notation by defining the inner products associated with the relevant stochastic differentials. If $\mathbf{\Psi}$ is an integral transform operator with a symmetric kernel $\psi(\cdot, \circ)$, write the associated inner product $(\mathbf{\Psi} d\mathcal{B}(\cdot), d\mathcal{B}(\cdot)) := \int_0^1 \int_0^1 \psi(u_1, u_2) d\mathcal{B}(u_1) d\mathcal{B}(u_2)$ and denote the j-th row element of $[\mathbf{\Psi} H(\cdot), d\mathcal{B}(\cdot)]$ by

$$\int_0^1 \int_0^1 H_j(u_1)\psi(u_1, u_2) du_1 d\mathcal{B}(u_2).$$

The following lemma provides the limit behavior of some key statistics under Groups A and B, notably the GMM distance function $q_n(\cdot)$ and the two Gramians $\widetilde{d}_n := [\widehat{\mathbf{\Xi}}_n H_n(\cdot), \widetilde{g}_n(\cdot)]$ and $\bar{A}_n = [\widehat{\mathbf{\Xi}}_n H_n(\cdot), H_n(\cdot)]$.

Lemma 1. *Given Assumptions 1, 2 (i), 3, and 4,*

(i) if Assumption 2 (ii.a) also holds,

- *(a)* $s_n^{-1-r} q_n(\theta_*) \rightsquigarrow (\mathbf{\Xi} d\mathcal{B}_s(\cdot), d\mathcal{B}_s(\cdot))$;
- *(b)* $s_n^{\frac{1}{2}-r-\kappa} \widetilde{d}_n \rightsquigarrow [\mathbf{\Xi}\mu(\cdot), d\mathcal{B}_s(\cdot)]$;
- *(c)* $s_n^{-r-2\kappa} \bar{A}_n \xrightarrow{\mathbb{P}} A_r := [\mathbf{\Xi}\mu(\cdot), \mu(\cdot)]$;

(ii) if Assumption 2 (ii.b) also holds,

- *(a)* $s_n^{-1-r} q_n(\theta_*) \xrightarrow{\mathbb{P}} \sigma_s^2 := R_1 \Sigma_\ell^2 R_1'$, *where* $R_1 := [1, 0_{1\times d}]$;
- *(b)* *(1) if* $\nu = 0$, *then* $s_n^{\frac{1}{2}-r-\kappa} \widetilde{d}_n \rightsquigarrow \tau_s + [\mu(\cdot), d\mathcal{B}_s(\cdot)]$, *where* $\tau_s := \sqrt{s_*}[d\mathcal{B}_a(\cdot), d\mathcal{B}_s(\cdot)]$;
  (2) if $\nu \in (0, \frac{1}{2}]$, *then* $s_n^{\frac{1}{2}-r-\nu-\kappa} \widetilde{d}_n \to \tau_s$;
- *(c)* *(1) if* $\nu \in [0, \frac{1}{2})$, *then* $s_n^{-r-2\kappa} \bar{A}_n \xrightarrow{\mathbb{P}} A_e := [\mu(\cdot), \mu(\cdot)]$;
  (2) if $\nu = \frac{1}{2}$, *then* $s_n^{-r-2\kappa} \bar{A}_n \xrightarrow{\mathbb{P}} A_f := \Upsilon_s + A_e$, *where* $\Upsilon_s := s_*[d\mathcal{B}_a(\cdot), d\mathcal{B}_a(\cdot)]$. □

Remark.

(i) $(\mathbf{\Xi} d\mathcal{B}_s(\cdot), d\mathcal{B}_s(\cdot))$ exists if and only if $\xi(\cdot, \circ) \in L^2([0,1]^2)$ (e.g., Kwapien and Woyczynski, 1987). Therefore, $(\mathbf{\Xi} d\mathcal{B}_s(\cdot), d\mathcal{B}_s(\cdot))$ exists as $\xi(\cdot, \circ) \in \mathcal{C}([0,1]^2)$ (e.g., Wang, 1975).

(ii) The limits in Lemma 1 depend on the support set dimension of $\xi(\cdot, \circ)$. For example, under Group A, the dimension is 2, whereas the dimension is 1 under Group B. For this reason $q_n(\theta_*)$ converges to a constant under Group B, but it is asymptotically central chi-squared under Group A since its limit is an inner product involving $d\mathcal{B}_s(\cdot)$. This feature implies that the J-test for overidentifying restrictions needs to be employed by incorporating the dimension of $\xi(\cdot, \circ)$. Likewise, $\widetilde{d}_n$ is asymptotically normal with nonzero location parameter or converges to a constant under Group B, but it is asymptotically normal with zero location parameter under Group A.

(iii) $\Upsilon_s + A_e$ is positive definite by Assumptions 3 and 4. □

We are now in a position to provide the limit distribution of the sxid-GMM for Groups A and B.

Theorem 1. *Given Assumptions 1, 2 (i), 3, and 4,*

(i) if Assumption 2 (ii.a) also holds, $s_n^{\frac{1}{2}+\kappa} n^{\frac{1}{2}}(\widehat{\theta}_n - \theta_*) \rightsquigarrow \mathcal{N}(0, A_r^{-1} B_r A_r^{-1})$*, where* $B_r := \sigma_s^2[\Xi\mu(\cdot), \Xi\mu(\cdot)]$*;*

(ii) if Assumption 2 (ii.b) also holds,

(a) if $\nu = 0$*, then* $s_n^{\frac{1}{2}+\kappa} n^{\frac{1}{2}}(\widehat{\theta}_n - \theta_*) \rightsquigarrow \mathcal{N}(\psi_s, \sigma_s^2 A_e^{-1})$*, where* $\psi_s := -A_e^{-1}\tau_s$*;*

(b) if $\nu \in (0, \frac{1}{2})$*, then* $s_n^{\frac{1}{2}-\nu+\kappa} n^{\frac{1}{2}}(\widehat{\theta}_n - \theta_*) \xrightarrow{\mathbb{P}} \psi_s$*; and*

(c) if $\nu = \frac{1}{2}$*, then* $s_n^{\kappa} n^{\frac{1}{2}}(\widehat{\theta}_n - \theta_*) \xrightarrow{\mathbb{P}} \psi_f := -A_f^{-1}\tau_s$*.* □

Remark.

(i) sxid-GMM is asymptotically normal with zero location under Group A, implying that the estimator is asymptotically unbiased. This result contrasts with those obtained using the Dirac delta function under Group B, wherein its limit distribution is asymptotically normal with nonzero mean or it converges to a constant. This property implies that using a continuous kernel is useful in obtaining an asymptotically unbiased sxid-GMM.

(ii) Given that $\kappa \geq 0$, sxid-GMM is super-consistent under Group A.

(iii) Theorem 1 (i) follows as a corollary of Lemma 1 (i), and we apply Itô isometry to obtain B_r.

(iv) In the context of 2SLS defined with weak IVs, Chao and Swanson (2005) estimated the kernel using a weight matrix that is optimal when a finite number of IVs are employed. They discovered the presence of an asymptotic bias in this case. They further examined the limit behavior of *bias corrected* 2SLS (bc-2SLS) and LIML to remove the bias. Theorem 1 (ii) demonstrates that a similar bias emerges even under the strong moment condition. From this property, estimating the kernel using the weight matrix assuming a finite number of IVs can be interpreted as use of a kernel similar to the Dirac delta function. Han and Phillips (2006) and Newey and Windmeijer (2009) also reported asymptotic bias in their work on GMM.

(v) Although sxid-GMM driven by the Dirac delta function is asymptotically biased, we can exploit the kernel to obtain a more efficient estimator without such bias. In particular, the asymptotic covariance matrix in Theorem 1 (ii.a) achieves the information matrix equality, which implies a formulation in which a more efficient unbiased estimator may be constructed by using a continuous kernel that is close in the limit to the Dirac delta function but still continuous; see Section 3.1. □

We next examine testing overidentifying restrictions using sxid-GMM under Group A conditions. From the results in Theorem 1, parameter estimation using a continuous kernel is more useful than one with a limiting Dirac delta function kernel. In developing methodology for testing over-identifying restrictions under Group A conditions we consider the following specific hypotheses: for each $t = 1, 2, \ldots$,

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

Our primary interest is in the J-test developed by Sargan (1958) and later extended to GMM by Hansen (1982), who examined the limit behavior of the GMM distance under each hypothesis. Lemma 1 (i.a) assumes the unknown parameter θ_* and provides the corresponding limit behavior of GMM distance in different environments. We now replace θ_* with $\widehat{\theta}_n$ and derive the limit behavior of the GMM distance.

First note that

$$q_n(\widehat{\theta}_n) = q_n(\theta_*) - \frac{1}{2}\sqrt{n}(\widehat{\theta}_n - \theta_*)'\nabla^2_\theta \bar{q}_n(\theta_*)\sqrt{n}(\widehat{\theta}_n - \theta_*) + o_{\mathbb{P}}(\|\widehat{\theta}_n - \theta_*\|^2), \tag{10}$$

and hence $q_n(\widehat{\theta}_n) = q_n(\theta_*) - \frac{1}{2}s_n^2\widetilde{d}_n'\bar{A}_n^{-1}\widetilde{d}_n + o_{\mathbb{P}}(\|\widehat{\theta}_n - \theta_*\|^2)$, since $s_n^{-2}\nabla^2_\theta\bar{q}_n(\theta_*) = \bar{A}_n + o_{\mathbb{P}}(s_n^{r+2\kappa})$; see Assumption 3 (ii). The limit behavior of each element on the right side is given in Lemma 1 (i), so the null limit behavior of the GMM distance follows straightforwardly. For this purpose, let $\mathbf{\Pi}_r$ be the integral operator with the following kernel: $\pi_r(\cdot,\circ) := \xi(\cdot,\circ) - \frac{1}{2}\lambda(\cdot)'A_r^{-1}\lambda(\circ)$, where $\lambda(\cdot) := \mathbf{\Xi}\mu(\cdot)$. The next result gives the null limit behavior of the J-test.

Theorem 2. *Given Assumptions 1, 2 (i, ii.a), 3, 4, and $\mathcal{H}_0$, $s_n^{-1-r}q_n(\widehat{\theta}_n) \rightsquigarrow (\mathbf{\Pi}_r d\mathcal{B}_s(\cdot), d\mathcal{B}_s(\cdot))$.* □

Remark.

(i) The null limit distribution of the J-test is asymptotically chi-squared with noncentrality determined by the continuous kernel of $\mathbf{\Pi}_r$. Here, $\mathbf{\Pi}_r$ differs from $\mathbf{\Xi}$ because the kernel is modified to allow parameter estimation error.

(ii) Given that the kernel $\mathbf{\Pi}_r$ is not straightforward to estimate, there may be difficulty using standard simulations to obtain the null distribution of the J-test. Instead, we can apply a resampling approach. For example, in the 2SLS framework, we can construct a resampled residual, say $\widehat{u}^b$, by multiplying iid standard normal random variates to the structural errors $\widehat{u} := y - X\widehat{\theta}_n$ and let $\widehat{y}^b := X\widehat{\theta}_n + \widehat{u}^b$. Using $(\widehat{y}^b, X, Z)$, the J-test can be computed, giving J_n^b, and the process repeated successively to obtain the null distribution of the J-test via the empirical distribution of J_n^b. If the J-test diverges under the alternative, the p-value of the J-test estimated by the percentage of the J_n^b greater than the original J-test decreases to zero as n tends to infinity.

(iii) Although not examined here, Donald, Imbens, and Newey (2003) explored another T-test for over-identifying restrictions when the GMM distance converges to a constant as in Lemma 1 (ii.a). If the GMM distance does not converge to a constant, it is not necessary to use the T-test.

(iv) Under the alternative, the J-test diverges and is not discussed further. □

2.2 Weak Moment Conditions

This section explores the large sample properties of xid-GMM under weak moment conditions for $\nabla_\theta\widetilde{G}_n(\theta_*)$. In this case we use the terminology *weak xid*-GMM (wxid-GMM). The asymptotic behavior of wxid-GMM is obtained in a similar manner to sxid-GMM, but allowing for weak moments and with other conditions as in sxid-GMM. The modification affects $H_n(\cdot)$ and using a similar formulation as in (9), we now have

$$\widetilde{a}_n(\cdot) := \sqrt{s_n}\int_0^{(\cdot)} \frac{\sqrt{n}}{s_n^\nu}\left(\frac{1}{s_n^\kappa}\widetilde{H}_n(u) - \mu_n(u)\right) du \tag{11}$$

converging weakly to a Brownian motion $\widetilde{\mathcal{B}}_a(\cdot)$, where $\widetilde{H}_n(\cdot) := \sqrt{n}H_n(\cdot)$. This condition is imposed by extending the fixed moment size condition as in the case of sxid-GMM. Thus, in the 2SLS setting below (8), if $s_n = s$ and $\phi_n = n^{-1/2}\phi$, the weak instrument condition is characterized as

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t x_t' - \mathbb{E}[z_t z_t']\phi = O_{\mathbb{P}}(1),$$

by applying a CLT to $n^{-\frac{1}{2}}\sum_{t=1}^{n} z_t v_t'$. We relax this condition by letting s_n be proportional to n and translate $-n^{-\frac{1}{2}}\sum_{t=1}^{n} z_{t,(\cdot)}x_t$ and $-\sum_{\ell=1}^{s_n}\mathbb{E}[z_{t,(\cdot)}z_{t,\ell}]\phi_{n,\ell}$ to $\widetilde{H}_n(\cdot)$ and $\mu_n(\cdot)$, respectively so that the weak moment conditions can be transformed to a stochastic process $\widetilde{a}_n(\cdot)$ by discounting them appropriately using s_n^{ν} and s_n^{κ}. This assumption parallels the formulation in (9) and, in view of the weak moments, we suppose an FCLT applies to the partial sum process driven by $\widetilde{H}_n(\cdot)$ instead of $H_n(\cdot)$. Otherwise the same regularity conditions are maintained as for sxid-GMM.

The limit behavior of wxid-GMM follows by expansion, similar to sxid-GMM. Specifically, from (7)

$$\widehat{\theta}_n - \theta_* = -[\widehat{\boldsymbol{\Xi}}_n\widetilde{H}_n(\cdot), \widetilde{H}_n(\cdot)]^{-1}[\widehat{\boldsymbol{\Xi}}_n\widetilde{H}_n(\cdot), \widetilde{g}_n(\cdot)] + o_{\mathbb{P}}(1) = -\widetilde{A}_n^{-1}\ddot{d}_n + o_{\mathbb{P}}(1),$$

where $\widetilde{A}_n := [\widehat{\boldsymbol{\Xi}}_n\widetilde{H}_n(\cdot), \widetilde{H}_n(\cdot)]$ and $\ddot{d}_n := [\widehat{\boldsymbol{\Xi}}_n\widetilde{H}_n(\cdot), \widetilde{g}_n(\cdot)]$ and the asymptotic behavior of these components follow as for $\bar{A}_n$ and $\widetilde{d}_n$. For a rigorous derivation of the limit distribution of wxid-GMM the following modified assumptions are employed.

Assumption 5. *(i) For some $\nu \in [0, \frac{1}{2}]$ and $\kappa \geq 0$, $\sqrt{n}s_n^{-\nu}(s_n^{-\kappa}\widetilde{H}_n(\cdot) - \mu_n(\cdot)) = O_{\mathbb{P}}(1)$ uniformly on $[0,1]$, where $\mu_n(\cdot) : [0,1] \mapsto \mathbb{R}^d$, and for some nonzero $\mu(\cdot)$ and $\chi(\cdot) \in L_2([0,1])$ and for some $\beta > 0$, $\mu_n(\cdot) = \mu(\cdot) + s_n^{-\beta}\chi(\cdot)$;*
(ii) $s_n^{-2}\nabla_\theta^2 q_n(\theta_*) = \widetilde{A}_n + o_{\mathbb{P}}(s_n^{r+2\kappa})$; and
(iii) $[\boldsymbol{\Xi}\mu(\cdot), \mu(\cdot)]$ is positive definite. □

Next, define $\widetilde{\ell}_n(\cdot) := (s_n(\cdot), \widetilde{a}_n(\cdot)')'$, which is a $(1+d)$-vector of component stochastic processes.

Assumption 6. *(i) For some $\rho(\cdot) \in L_2([0,1])$ and for some $\alpha > \frac{1}{2}$, $\eta_n(\cdot) = s_n^{-\alpha}\rho(\cdot)$*
(ii) for a normal random variable $\mathcal{U} \sim \mathcal{N}(0, \sigma_u^2)$ and $(1+d)$-vector standard Wiener process $\widetilde{\mathcal{W}}_\ell(\cdot)$, there is a positive definite $\widetilde{\Sigma}_\ell \in \mathbb{R}^{(1+d)\times(1+d)}$ such that

$$\begin{bmatrix} \widetilde{\ell}_n(\cdot) \\ \widetilde{u}_n \end{bmatrix} \rightsquigarrow \begin{bmatrix} \int_0^{(\cdot)} d\widetilde{\mathcal{B}}_\ell(u) \\ \mathcal{U} \end{bmatrix} := \begin{bmatrix} \widetilde{\Sigma}_\ell \int_0^{(\cdot)} d\widetilde{\mathcal{W}}_\ell(u) \\ \mathcal{U} \end{bmatrix},$$

and for some $\widetilde{\sigma}_{\ell u} \in \mathbb{R}^{1+d}$, $\mathbb{E}[\int_0^{(\cdot)} d\widetilde{\mathcal{B}}_\ell(u)\mathcal{U}] = \widetilde{\sigma}_{\ell u}\int_0^{(\cdot)} du$. □

Assumptions 5 and 6 replace the earlier Assumptions 3 and 4 and are influenced by the introduction of $\widetilde{H}_n(\cdot)$ in the definition of $\widetilde{a}_n$ in (11). Using these new conditions we obtain in an analogous way the following limit theory that applies under the Group A condition.

Corollary 1. *Given Assumptions 1, 2 (i, ii.a), 5, and 6, $s_n^{\frac{1}{2}+\kappa}(\widehat{\theta}_n - \theta_*) \rightsquigarrow \mathcal{N}(0, A_r^{-1}\widetilde{B}_r A_r^{-1})$, where $\widetilde{B}_r := \widetilde{\sigma}_s^2[\boldsymbol{\Xi}\mu(\cdot), \boldsymbol{\Xi}\mu(\cdot)]$ and $\widetilde{\sigma}_s^2 := R_1\widetilde{\Sigma}_\ell^2 R_1'$.* □

Remark.

(i) The major difference between Theorem 1 (i) and Corollary 1 is that the convergence rate of wxid-GMM is slower than that of sxid-GMM by $\sqrt{n}$.

(ii) The same approach to the limit theory applies to wxid-GMM when defined under the Group B condition and is therefore omitted.

(iii) The primary lesson of Corollary 1 matches that of Theorem 1 (i): viz., if wxid-GMM is defined using a continuous kernel, it is asymptotically unbiased and normally distributed, whereas if wxid-GMM is defined by the Dirac delta function it is asymptotically biased.

(iv) The asymptotics in Corollary 1 differ from the non-Gaussian asymptotics of weak 2SLS estimation when the moment size is finite. Staiger and Stock (1997) assumed a fixed moment size and gave asymptotics for 2SLS following Phillips's (1989) approach. In that work the limit theory is not normal and reflects the weak information by way of an internal CLT (in place of convergence in probability of sample moments of the endogenous variables and instruments) that leads to a replication of the exact finite sample theory of an IV or GMM procedure under Gaussianity in the underlying data (Phillips, 1980) . In contrast, Corollary 1 assumes the moment size tends to infinity, which enables consistent estimation of the integral transform of the population mean process of $n^{-1/2}\sum_{t=1}^{n} x_t z'_{t,(\cdot)}$, viz., $\mu(\cdot)$. Then asymptotics for $\widehat{\theta}_n$ are obtained quite differently from when the moment size is fixed. In particular, we have $s_n^{-2\kappa}\widetilde{A}_n = A_r + o_{\mathbb{P}}(1) = [\Xi\mu(\cdot), \mu(\cdot)] + o_{\mathbb{P}}(1)$ and $s_n^{\frac{1}{2}-\kappa}\ddot{d}_n \rightsquigarrow [\Xi\mu(\cdot), d\widetilde{\mathcal{B}}_s(\cdot)] \sim \mathcal{N}(0, \widetilde{B}_r)$, so that $s_n^{\frac{1}{2}+\kappa}(\widehat{\theta}_n - \theta_*) = -s_n^{\frac{1}{2}+\kappa}\widetilde{A}_n^{-1}\ddot{d}_n \rightsquigarrow \mathcal{N}(0, A_r^{-1}\widetilde{B}_r A_r^{-1})$, where $\widetilde{\mathcal{B}}_s(\cdot)$ is the first-row element of $\widetilde{\mathcal{B}}_\ell(\cdot)$.

(v) The result in Corollary 1 also differs from earlier studies with many weak moment conditions. Chao and Swanson (2005) compared standard wid-2SLS with bias-corrected wid-2SLS (wid-bc-2SLS) and LIML and showed that standard wid-bc-2SLS is consistent for the unknown parameter under weaker conditions. Han and Phillips (2006) explored limit theory of unweighted GMM driven by many weak moment conditions and demonstrated consistency and asymptotic normality when the moment size grows slower than the sample size. Along the same lines, Newey and Windmeijer (2009) studied CUE with a heteroskedasticity consistent weight matrix and established asymptotic normality under weak moment conditions, again with moment size growing slower than the sample size. Corollary 1 assumes a weight matrix converging to a continuous kernel in quite a different manner from these earlier studies to establish asymptotic normality: in this result the weak moment conditions are transformed to a stochastic process that follows functional limit theory and moment size is allowed to grow at the *same rate* as the sample size.

(vi) Section 4 provides further details, specifically examining wid-2SLS driven by the Dirac delta and continuous kernels which enables comparison with bias-corrected wid-2SLS estimation. □

We next discuss the limit behavior of the J-test under the Group A condition. For this purpose, using (10), we obtain that $q_n(\widehat{\theta}_n) = q_n(\theta_*) - \frac{1}{2}s_n^2\ddot{d}_n'\widetilde{A}_n^{-1}\ddot{d}_n + o_{\mathbb{P}}(\|\widehat{\theta}_n - \theta_*\|^2)$, similar to the strong moment case. The limit behavior of $\ddot{d}_n$ and $\widetilde{A}_n$ is identical to that of $\widetilde{d}_n$ and $\bar{A}_n$, and the asymptotic behavior of $q_n(\theta_*)$ is the same as for the strong moment case because it does not involve $\widetilde{H}_n(\cdot)$. Instead, it is defined only by $\widehat{\Xi}_n$ and $\widetilde{g}_n(\cdot)$. Hence, Theorem 2 can be applied even to the weak moment case.

3 Asymptotically Efficient xid-GMM

This section considers asymptotically most efficient xid-GMM using the results in Section 2.

3.1 Asymptotically Most Efficient Estimation

In view of the parallel relationship between sxid- and wxid-GMMs, we focus on the first without loss of generality in the following discussion. Primary interest lies in finding an asymptotically most efficient sxid-GMM driven by a continuous kernel such that the condition in Theorem 1 (i) holds. When employed with a continuous kernel sxid-GMM estimation is asymptotically unbiased and Theorem 1 (i) applies.

From Theorem 1 (i) the asymptotic covariance matrix of $s_n^{\frac{1}{2}+\kappa}\sqrt{n}(\widehat{\theta}_n - \theta_*)$ is $A_r^{-1}B_rA_r^{-1}$, and the information matrix equality holds for efficiency when for some $c > 0$, $B_r = cA_r$. Given that $H(\cdot)$ is arbitrary, if the continuous kernel of Ξ is close to $\delta(\cdot - \circ)$, B_r can be made arbitrarily close to cA_r. This property follows from Theorem 1 (ii.a). If the kernel function is the Dirac delta function, the asymptotic covariance matrix of the sxid-GMM satisfies the information matrix equality, but in that case it is asymptotically biased. That is, $\text{acov}[s_n^{\frac{1}{2}+\kappa}\sqrt{n}(\widehat{\theta}_n - \theta_*)] = A_e^{-1}B_eA_e^{-1}$ with $B_e = \sigma_s^2 A_e$. Thus, if sxid-GMM is estimated using a continuous kernel close to the Dirac delta function, $B_r \approx \sigma_s^2 A_r$ and $A_r \approx A_e$, meaning that the efficiency of the sxid-GMM with a continuous kernel improves although it is still bounded from below by $\sigma_s^2 A_e^{-1}$. There are many continuous kernels close to the Dirac delta function. For example, if we let $\xi(\cdot, \circ)$ be a ridge function such as the following squared exponential kernel (SEK):

$$\xi(\cdot, \circ; s^2) = \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(\cdot - \circ)^2}{2s^2}\right),$$

then $\xi(\cdot, \circ; s^2)$ approaches $\delta(\cdot - \circ)$ as s^2 reduces to zero.[4] Therefore, if s^2 is selected close to zero, we can gain efficiency from sxid-GMM while ensuring it is asymptotically unbiased. The above findings leads to the following formal result.

Corollary 2. *Given Assumptions 1 and 2 (i, ii.a),*

(i) if Assumptions 3 and 4 also hold and $\nu = 0$, then the asymptotic efficiency of sxid-GMM is bounded below by $\sigma_s^2 A_e^{-1}$; and

(ii) if Assumptions 5 and 6 also hold and $\nu = 0$, then the asymptotic efficiency of wxid-GMM is bounded below by $\widetilde{\sigma}_s^2 A_e^{-1}$. □

Remark.

(i) The proof of Corollary 2 (i) is in the Supplement, where it is shown that $A_r^{-1}B_rA_r^{-1} - \sigma_s^2 A_e^{-1}$ is positive semi-definite. Corollary 2 (ii) holds by analogy.

[4] In addition to SEK, there are other ridge functions that are flat along the diagonal. For example, the Matérn, exponential, γ-exponential, and rational quadratic kernels all belong to this class (see Rasmussen and Williams, 2005, Chapter 4).

(ii) Although Corollary 2 gives bounds on asymptotic efficiency, we cannot allow a continuous kernel to be too close to the Dirac delta function in finite samples. The reason is the existence of a trade off between the asymptotic variance and bias when n is finite. For example, as s^2 approaches zero for $\xi(\cdot, \circ; s^2)$, the magnitude of $A_r^{-1} B_r A_r^{-1}$ decreases, but the bias of sxid-GMM increases. The Supplement provides evidence of this linkage between efficiency and bias in simulation exercises.

(iii) We emphasize that Corollary 2 holds under the maintained conditions. In particular, Corollary 2 supposes the ordering of the moment conditions is fixed when computing xid-GMM. If the moment conditions are ordered differently, a different xid-GMM estimate is obtained and its efficiency level can differ from the original xid-GMM. The next section shows how this feature of GMM can be exploited to define another estimator with improved properties by permuting the moment conditions. □

3.2 Permuted xid-GMM

Corollary 2 shows that xid-GMM can attain good efficiency asymptotically. This section introduces another estimator called *permuted* xid-GMM (pxid-GMM) that is even more efficient. As in Section 3.1, we first restrict attention to sxid-GMM under the environment defined in Theorem 1 (i) and then extend results to wxid-GMM by analogy.

The xid-GMM estimator has a continuous kernel and this accommodates moment permutation. For a continuous kernel, permuting the moment conditions typically produces a different xid-GMM estimate. Specifically, suppose Q is an $s_n \times s_n$ permutation matrix and the moment conditions are permuted by application of Q to the moment conditions, thereby defining another GMM distance leading to

$$\bar{q}_n^p(\cdot) := \bar{G}_n(\cdot)' Q' \widehat{\Sigma}_n^{-1} Q \bar{G}_n(\cdot), \quad \text{with} \quad \widehat{\theta}_n^p := \arg\min_{\theta \in \Theta} \bar{q}_n^p(\cdot),$$

giving the new estimate $\widehat{\theta}_n^p$. In 2SLS for example, we have $\widehat{\theta}_n^p = (X'ZQ'\widehat{\Sigma}_n^{-1} QZ'X)^{-1} X'ZQ'\widehat{\Sigma}_n^{-1} QZ'y$. Then the number of permutations in total is $c_n := s_n!$, and $\widehat{\theta}_n^p$ typically differs with each permutation.

These differences can be used to define a more efficient estimator. To motivate the approach in a straightforward way, let $c_n = c$ be constant. Combine the c estimates linearly to define another estimator: $\bar{\theta}_n(\omega) := \sum_{p=1}^{c} \omega_p \widehat{\theta}_n^p$ with weights $\omega_p \geq 0$ satisfying $\sum_{p=1}^{c} \omega_p = 1$. Then, optimal weights are obtained by minimizing the asymptotic mean-square error (AMSE) (e.g., Taniguchi and Tresp, 1997; Hansen, 2009), giving $\omega_* := \arg\min_{\omega \geq 0} \omega' \Upsilon \omega$ subject to $\omega' \iota_c = 1$, where $\omega := (\omega_1, \omega_2, \ldots, \omega_c)'$ with ι_c being a c-vector of ones, and matrix

$$\Upsilon := \begin{bmatrix} \mathrm{tr}(\Omega_{\widehat{\theta}}^{(1,1)}) & \mathrm{tr}(\Omega_{\widehat{\theta}}^{(1,2)}) & \cdots & \mathrm{tr}(\Omega_{\widehat{\theta}}^{(1,c)}) \\ \mathrm{tr}(\Omega_{\widehat{\theta}}^{(2,1)}) & \mathrm{tr}(\Omega_{\widehat{\theta}}^{(2,2)}) & \cdots & \mathrm{tr}(\Omega_{\widehat{\theta}}^{(2,c)}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}(\Omega_{\widehat{\theta}}^{(c,1)}) & \mathrm{tr}(\Omega_{\widehat{\theta}}^{(c,2)}) & \cdots & \mathrm{tr}(\Omega_{\widehat{\theta}}^{(c,c)}) \end{bmatrix},$$

in which $\Omega_{\widehat{\theta}}^{(p,q)}$ is the p-th block-row and q-th block-column matrix of the $cd \times cd$ asymptotic covariance matrix $\Omega_{\widehat{\theta}} := \text{acov}[s_n^{\frac{1}{2}+\kappa} n^{\varpi}[(\widehat{\theta}_n^1 - \theta_*)', (\widehat{\theta}_n^2 - \theta_*)', \ldots, (\widehat{\theta}_n^c - \theta_*)']']$. In this formulation, $\varpi = \frac{1}{2}$ for sxid-GMM by Theorem 1 (i), and $\varpi = 0$ for wxid-GMM by Corollary 1. In addition, the objective function is the AMSE of $s_n^{\frac{1}{2}+\kappa} n^{\varpi}(\bar{\theta}_n(\omega) - \theta_*)$, viz., $\lim_{n\to\infty} s_n^{1+2\kappa} n^{2\varpi} \mathbb{E}[(\bar{\theta}_n(\omega) - \theta_*)'(\bar{\theta}_n(\omega) - \theta_*)]$. The following theorem provides the optimal weight estimator, the optimized AMSE, and the asymptotic covariance matrix estimator of $\bar{\theta}_n(\widehat{\omega}_n)$.

Theorem 3. *Given the conditions in Theorem 1 (i) or Corollary 1, if both $\Omega_{\widehat{\theta}}$ and Υ are positive definite,*

(i) $\omega_* = \Upsilon^{-1}\iota_c/(\iota_c'\Upsilon^{-1}\iota_c)$;

(ii) $\omega_*'\Upsilon\omega_* = (\iota_c'\Upsilon^{-1}\iota_c)^{-1}$; *and*

(iii) $\text{acov}[s_n^{\frac{1}{2}+\kappa} n^{\varpi}(\bar{\theta}_n(\omega_*) - \theta_*)] = (\iota_c'\Upsilon^{-1}\iota_c)^{-2}(\iota_c'\Upsilon_*^{-1} \otimes I_d)\Omega_{\widehat{\theta}}(\Upsilon^{-1}\iota_c \otimes I_d)$. □

Theorem 3 implies that $\bar{\theta}_n(\omega_*)$ may be more efficient than each individual xid-GMM estimator, taking advantage of the fact that the asymptotic covariance matrix of $\bar{\theta}_n(\omega_*)$ depends on c. When c is large, this asymptotic covariance matrix can be reduced relative to that of the original xid-GMM estimator, as indicated in the following corollary, where L_1 and U_1 are strictly lower and upper triangular matrices of ones.

Corollary 3. *Given the conditions in Theorem 3, if $\Omega_{\widehat{\theta}} = I_c \otimes \Psi + L_1 \otimes \Lambda + U_1 \otimes \Lambda'$ for a positive-definite and symmetric matrix $\Psi \in \mathbb{R}^{d\times d}$ and $\Lambda \in \mathbb{R}^{d\times d}$ such that $(\Lambda + \Lambda')$ is positive semi-definite,*

(i) $\omega_* = \frac{1}{c}\iota_c$;

(ii) $\omega_*'\Upsilon\omega_* = \frac{1}{c}\text{tr}(\Psi) + (1 - \frac{1}{c})\text{tr}(\Lambda)$;

(iii) $\text{acov}[s_n^{\frac{1}{2}+\kappa} n^{\varpi}(\bar{\theta}_n(\omega_*) - \theta_*)] = \frac{1}{c}\Psi + \frac{1}{2}(1 - \frac{1}{c})(\Lambda + \Lambda')$; *and*

(iv) $2\Psi - (\Lambda + \Lambda')$ *is positive definite.* □

Remark.

(i) The block-symmetric covariance matrix condition in Corollary 3 implicitly assumes that the moment conditions are iid with respect to the moment condition index. Specifically, Ψ denotes $\Omega_{\widehat{\theta}}^{(p,p)} := \text{avar}[s_n^{\frac{1}{2}+\kappa} n^{\varpi}(\widehat{\theta}_n^p - \theta_*)]$ for $p = 1, 2, \ldots, c$, and Λ denotes $\Omega_{\widehat{\theta}}^{(p,q)} := \text{avar}[s_n^{\frac{1}{2}+\kappa} n^{\varpi}(\widehat{\theta}_n^p - \theta_*), s_n^{\frac{1}{2}+\kappa} n^{\varpi} (\widehat{\theta}_n^q - \theta_*)]$ for $p, q = 1, 2, \ldots, c$ with $p \neq q$. Under the iid condition, $\Omega_{\widehat{\theta}}^{(p,q)} = \Lambda$ or Λ'. Using these, we represent $\Omega_{\widehat{\theta}}$ as $I_c \otimes \Psi + L_1 \otimes \Lambda + U_1 \otimes \Lambda'$.

(ii) Corollary 3 (i) implies that the optimal weight is equal for each permutation.

(iii) According to Corollary 3 (iii), the asymptotic covariance matrix of $\bar{\theta}_n(\omega_*)$ is determined by three factors: c, Λ, and Ψ. When c tends to infinity, the asymptotic covariance matrix of $\bar{\theta}_n(\omega_*)$ converges to $\frac{1}{2}(\Lambda + \Lambda')$, implying that $\bar{\theta}_n(\omega_*)$ becomes more efficient than each xid-GMM with the asymptotic covariance matrix Ψ. The implication is that the efficiency of $\bar{\theta}_n(\omega_*)$ is further improved by letting c be substantially large. □

Corollary 3 is now exploited to define a new pxid-GMM estimator more efficient than each individual xid-GMM. We first note that as n increases indefinitely, c_n also continues to rise, so that analyzing all potential xid-GMM estimates is extremely challenging. Even for s_n as small as 10, $s_n! = 3{,}628{,}800$, which is already too large to consider all permutations of the s_n moment conditions. We therefore analyze samples formed by independent draws of the xid-GMM estimates. Denote these samples $\{\widehat{\theta}_n^p : p = 1, 2, \ldots, m\}$, where m denotes a sufficiently large sample size selected by the researcher. Before applying a random sampling procedure, we note that each $\widehat{\theta}_n^p$ is asymptotically normally distributed around θ_* with an asymptotic covariance matrix Ψ as given by Theorem 1 (i) or Corollary 1. For example, Ψ generically denotes $A_r^{-1}B_rA_r^{-1}$ for the sxid-GMM. By letting $\xi(\cdot, \circ)$ be continuous and close to $\delta(\cdot - \circ)$, we can make the magnitude of Ψ moderately sized by virtue of Corollary 2. If the sampling procedure is further implemented in a manner to satisfy the condition in Corollary 3, we may then estimate θ_* more efficiently by the simple average

$$\bar{\theta}_n := \frac{1}{m}\sum_{p=1}^{m} \widehat{\theta}_n^p.$$

Here, the optimal weight is $\frac{1}{m}\iota_m$ from Corollary 3 (i). If m is sufficiently large, $\bar{\theta}_n$ is close to θ_* by the law of large numbers, and its asymptotic covariance matrix should be $\frac{1}{2}(\Lambda + \Lambda')$ by Corollary 3 (iii). We call the resulting estimate $\bar{\theta}_n$ the pxid-GMM.

4 Weak Infinite Dimensional and Bias-Corrected 2SLS Estimators

This section considers the limit behavior of wid-2SLS driven by the Dirac delta function and by a continuous kernel. The potential use of this procedure in applications is considerable and it is examined in relation to the earlier development in Section 2.2. Attention is also given to bc-2SLS and *permuted* wid-2SLS (pwid-2SLS) procedures.

4.1 wid-2SLS driven by the Dirac delta function

We first examine wid-2SLS estimation driven by the Dirac delta function and later compare it with wid-2SLS driven by a continuous kernel. To fix ideas the following DGP condition is assumed.

Assumption 7. *(i) For an iid sequence* $\{(y_t, x_t', z_t', u_t, v_t')' \in \mathbb{R}^{1+d+s_n+(1+d)} : t = 1, 2, \ldots, n\}$, $y_t = x_t'\theta_* + u_t$ *and* $x_t' = z_t'\pi_n + v_t'$, *where* $y_t \in \mathbb{R}$, $x_t \in \mathbb{R}^d$, $v_t \in \mathbb{R}^d$, *and* $\theta_* \in \mathbb{R}^d$;

(ii) $z_t := (z_{t,1}, \ldots, z_{t,s_n})' \in \mathbb{R}^{s_n}$ *is independent of* $(u_t, v_t')'$, *where* $\pi_n \in \mathbb{R}^{s_n \times d}$ *and* $\pi_n = n^{-\frac{1}{2}}\phi_n$ *such that* $\phi_n \in \mathbb{R}^{s_n \times d}$ *is* $O(1)$ *without being* $o(1)$; *and*

(iii) for some $s_* > 0$, $s_n/n = s_* + o(1)$. □

This DGP condition is typically assumed in wid-2SLS analysis (e.g., Chao and Swanson, 2005). The linear relationship in Assumption 7 (i) is written in observation form as

$$y = X\theta_* + u, \quad \text{and} \quad X = Z\pi_n + V,$$

where we let $Z := (z_1, \ldots, z_n)'$, $X := (x_1, \ldots, x_n)'$, $y := (y_1, \ldots, y_n)'$, $u := (u_1, \ldots, u_n)'$, and $V := (v_1, \ldots, v_n)'$.

The framework above relates closely to standard models for 2SLS estimation. But the DGP does not satisfy the regularity conditions for ordinary 2SLS analysis because the number of IVs tends to infinity as n increases. So standard 2SLS analysis and limit theory do not apply. Prior literature has tackled some aspects of 2SLS in this setting. First, Angrist and Krueger (1995), Donald and Newey (2001), and Hahn et al. (2001) among others have considered increasing numbers of IVs, although under strong moment conditions. Second, Phillips (1989) and Choi and Phillips (1992) studied IV estimation under lack of identification and partial identification. Staiger and Stock (1997) examined IV estimation under a local-to-zero condition as in Assumption 7 (ii), but with no allowance for the number of instruments to grow to infinity. Third, Chao and Swanson (2005), Han and Phillips (2006), Newey and Windmeijer (2009) assumed many weak IVs with dimension growing to infinity. Broadly speaking, they showed that GMM is not consistent for the structural parameter and/or suggested other procedures that are consistent or provide regularity conditions under which GMM is consistent for the parameter of interest. Anatolyev (2019) has surveyed recent developments of 2SLS with many weak IVs.

Under the framework of Assumption 7 we provide the limit distributions of the statistics relevant to estimation by wid-2SLS. Our analysis begins with the relevant stochastic processes. 2SLS is a special case of GMM, but it is convenient to define new notation using that of Section 2.2. First, let $\widehat{P}_{n,i}$ and $\widehat{w}_{n,i}$ be the i-th column elements of $\widehat{P}_n'$ and $\widehat{w}_n'$, where $\widehat{P}_n := n^{-1}Z'X \in \mathbb{R}^{s_n \times d}$ and $\widehat{w}_n := n^{-1}Z'u \in \mathbb{R}^{s_n \times 1}$. Next let $(\widetilde{P}_{n,i}', \widetilde{w}_{n,i})' := \sqrt{n}(\widehat{P}_{n,i}', \widehat{w}_{n,i})'$ and $\widehat{c}_n := n^{-1}Z'y \in \mathbb{R}^{s_n \times 1}$. Then

$$\widehat{\theta}_n = -(\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n)^{-1}(-\widehat{P}_n)'\widehat{\Sigma}_n^{-1}\widehat{c}_n = \theta_* - (\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n)^{-1}(-\widehat{P}_n)'\widehat{\Sigma}_n^{-1}\widehat{w}_n.$$

Second, we provide a useful stochastic process representation of 2SLS. Let $H_{n,j,i}$ be the j-th row and i-th column elements of $(-\widehat{P}_n)$ and $\mu_{n,j}$ be the j-th column of μ_n' where $\mu_n := -\mathbb{E}[z_t z_t']\phi_n$. With this notation, define the following stochastic processes:

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

$$g_n(u) \; := \; \begin{cases} 0, & \text{if } u \in [0, \frac{1}{s_n}) \text{ ;} \\ \widehat{w}_{n,j}, & \text{if } u \in [\frac{j}{s_n}, \frac{(j+1)}{s_n}), \, j = 1, 2, \ldots, s_n - 1\text{; and} \\ \widehat{w}_{n,n}, & \text{if } u = 1, \end{cases}$$

$$\mu_{n,i}(u) := \begin{cases} 0, & \text{if } u \in [0, \frac{1}{s_n}) \,; \\ \mu_{n,j,i}, & \text{if } u \in [\frac{j}{s_n}, \frac{(j+1)}{s_n}),\, j = 1, 2, \ldots, s_n - 1; \text{ and} \\ \mu_{n,s_n,i}, & \text{if } u = 1 \end{cases}$$

Next, let $H_n(\cdot) := [H_{n,1}(\cdot), \ldots, H_{n,d}(\cdot)]'$, $\mu_n(\cdot) := [\mu_{n,1}(\cdot), \ldots, \mu_{n,d}(\cdot)]'$, $[\widetilde{H}_n(\cdot)', \widetilde{g}_n(\cdot)]' := [\sqrt{n}H_n(\cdot)',$ $\sqrt{n}g_n(\cdot)]'$, and define $\widetilde{u}_n := n^{-\frac{1}{2}} \sum_{t=1}^n u_t$. Then construct the stochastic process

$$\eta_n(u) := \begin{cases} 0, & \text{if } u \in [0, \frac{1}{s_n}) \,; \\ \eta_j, & \text{if } u \in [\frac{j}{s_n}, \frac{(j+1)}{s_n}),\, j = 1, 2, \ldots, s_n - 1; \text{ and} \\ \eta_{s_n}, & \text{if } u = 1, \end{cases}$$

where $\eta_j := \mathbb{E}[z_{t,j}]$. It follows that in functional notation $\widehat{\theta}_n = \theta_* + [\widehat{\boldsymbol{\Xi}}_n \widetilde{H}_n(\cdot), \widetilde{H}_n(\cdot)]^{-1}[\widehat{\boldsymbol{\Xi}}_n \widetilde{H}_n(\cdot), \widetilde{g}_n(\cdot)]$.

These stochastic processes are key components in deriving the limit distribution of wid-2SLS. Our methodology is parallel to Section 2.2. Specifically, for some $\nu \in [0, \frac{1}{2}]$ and $\kappa \geq 0$, we let

$$\widetilde{\ell}_n(\cdot) := \begin{bmatrix} s_n(\cdot) \\ \widetilde{a}_n(\cdot) \end{bmatrix} := \frac{1}{\sqrt{s_n}} \sum_{j=1}^{\lfloor (\cdot) s_n \rfloor} \begin{bmatrix} \widetilde{w}_{n,j} - \eta_j \widetilde{u}_n \\ s_n^{-\nu} n^{\frac{1}{2}} (-s_n^{-\kappa} \widetilde{P}_{n,j} - \mu_{n,j}) \end{bmatrix}$$
$$= \sqrt{s_n} \int_0^{(\cdot)} \begin{bmatrix} \widetilde{g}_n(u) - \eta_n(u) \widetilde{u}_n \\ s_n^{-\nu} n^{\frac{1}{2}} (s_n^{-\kappa} \widetilde{H}_n(u) - \mu_n(u)) \end{bmatrix} du,$$

and suppose that for a positive-definite $\widetilde{\Sigma}_\ell \in \mathbb{R}^{(1+d)\times(1+d)}$ and $(1+d)$-variate Wiener process $\widetilde{\mathcal{W}}_\ell(\cdot) :=$ $(\widetilde{\mathcal{W}}_s(\cdot), \widetilde{\mathcal{W}}_a(\cdot)')' \in \mathbb{R}^{1+d}$,

$$\widetilde{\ell}_n(\cdot) \rightsquigarrow \int_0^{(\cdot)} d\widetilde{\mathcal{B}}_\ell(u) := \int_0^{(\cdot)} \begin{bmatrix} d\widetilde{\mathcal{B}}_s(u) \\ d\widetilde{\mathcal{B}}_a(u) \end{bmatrix} = \widetilde{\Sigma}_\ell \int_0^{(\cdot)} d\widetilde{\mathcal{W}}_\ell(u) := \widetilde{\Sigma}_\ell \int_0^{(\cdot)} \begin{bmatrix} d\widetilde{\mathcal{W}}_s(u) \\ d\widetilde{\mathcal{W}}_a(u) \end{bmatrix}.$$

In addition, we also suppose that $\widetilde{u}_n \rightsquigarrow \mathcal{U} \sim \mathcal{N}(0, \sigma_u^2)$ such that for some $\widetilde{\sigma}_{\ell u} \in \mathbb{R}^{1+d}$, $\mathbb{E}[\int_0^{(\cdot)} d\widetilde{\mathcal{B}}_\ell(u)\mathcal{U}] =$ $\widetilde{\sigma}_{\ell u} \int_0^{(\cdot)} du$ and that $[\mu(\cdot), \mu(\cdot)]$ is positive definite. If $\widetilde{\mathcal{B}}_\ell(\cdot)$ and $\mathcal{U}$ are independent, $\widetilde{\sigma}_{\ell u} = 0$.

These conditions parallel those in Assumptions 5 and 6. Specifically, Assumptions 5 (i) and 6 (ii) are explicitly assumed; and the other conditions in Assumptions 5 and 6 hold by the virtue of the linear model formulation of 2SLS. First, from the definitions of $\mu_n(\cdot)$, we can suppose that β in Assumption 3 is sufficiently large so that $\mu_n(\cdot)$ is almost identical to $\mu(\cdot)$. Second, the Hessian matrix $s_n^{-2}\nabla_\theta^2 q_n(\theta_*) = \widetilde{A}_n$ by virtue of the linear model condition.

We next make an assumption concerning the probability limit of $\widehat{\sigma}_n(\cdot, \circ)$. When $\widehat{\sigma}_n(\cdot, \circ)$ is defined by $\widehat{\Sigma}_n$ as in (3), we suppose that $\sup_{(u_1,u_2)} |\widehat{\xi}_n(u_1, u_2) - \mathbb{I}(u_1 = u_2)| \overset{\mathbb{P}}{\to} 0$. For example, Bickel and

Levina's (2008) regularized covariance matrix estimator achieves this consistently by letting their tuning parameter $\boldsymbol{\alpha}$ be proportional to the sample size. For this case, Assumption 2 (ii.b) holds by letting $r = 0$, so that $\widehat{\boldsymbol{\Xi}}_n$ converges to the identity operator. Then, for any $a(\cdot)$ and $b(\cdot) \in L^2([0,1])$, $(\widehat{\boldsymbol{\Xi}}_n a(\cdot), b(\cdot)) = (a(\cdot), b(\cdot)) + o_{\mathbb{P}}(1)$. This condition is assumed because 2SLS is constructed by inverting the covariance matrix of the moment condition. If the IVs here are iid with zero means and unit variances, their covariance matrix is the identity matrix for each finite n. The optimal 2SLS using a fixed number of IVs is therefore defined by estimating the identity matrix. The wid-2SLS driven by the Dirac delta function at the limit is motivated by this property. Although it is not assumed here that the IVs are necessarily iid with zero means and unit variances, estimating the Dirac delta function for iid IVs is parallel to estimating the identity matrix in the context of a finite number of IVs.

These conditions are collected together as follows.

Assumption 8.

(i) $\widehat{\sigma}_n(\cdot, \circ)$ *is positive definite for each* n*;*

(ii) $\sup_{(u_1,u_2)} |\widehat{\xi}_n(u_1, u_2) - \mathbb{I}(u_1 = u_2)| \xrightarrow{\mathbb{P}} 0$*;*

(iii) for some nonzero $\rho(\cdot)$, $\mu(\cdot)$ *and* $\chi(\cdot) \in L^2([0,1])$ *and for some* $\alpha > \frac{1}{2}$ *and* $\beta > 0$, $\eta_n(\cdot) = s_n^{-\alpha}\rho(\cdot)$ *and* $\mu_n(\cdot) = \mu(\cdot) + s_n^{-\beta}\chi(\cdot)$*;*

(iv) $A_e := [\mu(\cdot), \mu(\cdot)]$ *and* $\widetilde{A}_f := s_*[d\widetilde{\mathcal{B}}_a(\cdot), d\widetilde{\mathcal{B}}_a(\cdot)] + A_e$ *are positive definite;*

(v) for some $\gamma < \infty$, $s_n^{-1}\sum_{i=1}^{s_n}\mathbb{E}[z_{t,i}^2] \to \gamma$*; and*

(vi) for a positive definite $\widetilde{\Sigma}_\ell$, $\widetilde{\ell}_n(\cdot) \rightsquigarrow \int_0^{(\cdot)} d\widetilde{\mathcal{B}}_\ell(u) := \widetilde{\Sigma}_\ell \int_0^{(\cdot)} d\widetilde{\mathcal{W}}_\ell(u)$ *and* $\widetilde{u}_n \rightsquigarrow \mathcal{U} \sim \mathcal{N}(0, \sigma_u^2)$ *such that for some* $\widetilde{\sigma}_{\ell u} \in \mathbb{R}^{1+d}$, $\mathbb{E}[\int_0^{(\cdot)} d\widetilde{\mathcal{B}}_\ell(u)\mathcal{U}] = \widetilde{\sigma}_{\ell u}\int_0^{(\cdot)} du$*.* □

Assumption 8 (iv) is imposed to assure invertibility of A_e and $\widetilde{A}_f$, and $\widetilde{A}_f$ exists as a constant matrix by ergodicity. Assumption 8 (v) is imposed to assure regular behavior of the asymptotic bias. The asymptotic behavior of wid-2SLS follows by applying Theorem 1 (ii). We have the following result.

Corollary 4. *Given Assumptions 7 and 8,*

(i) if $\nu = 0$, $s_n^{\frac{1}{2}+\kappa}(\widehat{\theta}_n - \theta_*) \rightsquigarrow \mathcal{N}(\widetilde{\psi}_s, \widetilde{\sigma}_s^2 A_e^{-1})$*, where* $\widetilde{\psi}_s := -A_e^{-1}\widetilde{\tau}_s$ *and* $\widetilde{\sigma}_s^2 := R_1\widetilde{\Sigma}_\ell^2 R_1'$*, such that* $\widetilde{\tau}_s := \sqrt{s_*}[d\widetilde{\mathcal{B}}_a(\cdot), d\widetilde{\mathcal{B}}_s(\cdot)]$*;*

(ii) if $\nu \in (0, \frac{1}{2})$, $s_n^{\frac{1}{2}-\nu+\kappa}(\widehat{\theta}_n - \theta_*) \xrightarrow{\mathbb{P}} \widetilde{\psi}_s$*; and*

(iii) if $\nu = \frac{1}{2}$, $s_n^{\kappa}(\widehat{\theta}_n - \theta_*) \xrightarrow{\mathbb{P}} \widetilde{\psi}_f := -\widetilde{A}_f^{-1}\widetilde{\tau}_s$*.* □

Given the DGP and model conditions in Assumptions 7 and 8, we can apply Theorem 1 (ii) to obtain Corollary 4. The only difference lies in the fact that the convergence rate of the wid-2SLS is adjusted to accommodate the weak IV condition, making the convergence rate of wid-2SLS slower than that of sxid-GMM by $\sqrt{n}$. In view of the similarity between the two results we do not prove Corollary 4 in the Supplement.

There are differing viewpoints on asymptotic bias. First, we can view the asymptotic bias in terms of the joint distribution of $\widetilde{H}_n(\cdot)$ and $\widetilde{g}_n(\cdot)$. To illustrate, consider the condition in Corollary 4 (iii) with $\kappa = 0$. Note that $\widehat{\theta}_n - \theta_* = -(\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n)^{-1}(-\widehat{P}_n)'\widehat{\Sigma}_n^{-1}\widehat{w}_n$. Here, $s_n^{-1}\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n = [\widehat{\Xi}_n H_n(\cdot), H_n(\cdot)]$ and $\widehat{\Xi}_n H_n(\cdot) - H_n(\cdot) = o_{\mathbb{P}}(1)$ by Assumption 8 (ii), so that

$$\widetilde{A}_n := s_n^{-1} n\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n = [\widetilde{H}_n(\cdot), \widetilde{H}_n(\cdot)] + o_{\mathbb{P}}(1) \overset{\mathbb{P}}{\to} \widetilde{A}_f := s_*[d\widetilde{\mathcal{B}}_a(\cdot), d\widetilde{\mathcal{B}}_a(\cdot)] + [\mu(\cdot), \mu(\cdot)],$$

by applying Lemma 1 (ii.c.2). Similarly,

$$\ddot{d}_n := s_n^{-1} n(-\widehat{P}_n)'\widehat{\Sigma}_n^{-1}\widehat{w}_n = s_n^{-1}(-\widetilde{P}_n)'\widehat{\Sigma}_n^{-1}\widetilde{w}_n = [\widetilde{H}_n(\cdot), \widetilde{g}_n(\cdot)] + o_{\mathbb{P}}(1) \overset{\mathbb{P}}{\to} \widetilde{\tau}_s := \sqrt{s_*}[d\widetilde{\mathcal{B}}_a(\cdot), d\widetilde{\mathcal{B}}_s(\cdot)],$$

by applying Lemma 1 (ii.b.2), implying that $\widehat{\theta}_n - \theta_* = -\widetilde{A}_n^{-1}\ddot{d}_n \overset{\mathbb{P}}{\to} -\widetilde{A}_f^{-1}\widetilde{\tau}_s =: \widetilde{\psi}_f$. Unless $[d\widetilde{\mathcal{B}}_a(\cdot), d\widetilde{\mathcal{B}}_s(\cdot)] = 0$, asymptotic bias does not vanish and is traced back to the joint asymptotic distribution of $\widetilde{H}_n(\cdot)$ and $\widetilde{g}_n(\cdot)$. If the joint limit distribution is viewed in terms of the joint distribution of $d\widetilde{\mathcal{B}}_s(\cdot)$ and $d\widetilde{\mathcal{B}}_a(\cdot)$, then asymptotic independence occurs iff $[d\widetilde{\mathcal{B}}_a(\cdot), d\widetilde{\mathcal{B}}_s(\cdot)] = 0$.

To take a second viewpoint of the asymptotic bias, the probability limit of $[\widetilde{H}_n(\cdot), \widetilde{g}_n(\cdot)]$ is the probability limit of $[\widetilde{H}_n(\cdot) - \mu_n(\cdot), \widetilde{g}_n(\cdot)]$ because $[\mu_n(\cdot), \widetilde{g}_n(\cdot)] \overset{\mathbb{P}}{\to} 0$. In other words, $\widetilde{\tau}_s$ is related to the asymptotic covariance between $n^{-\frac{1}{2}}\sum_{t=1}^n(-x_t z_{t,i} - \mu_i)$ and $n^{-\frac{1}{2}}\sum_{t=1}^n z_{t,i}u_t$. Therefore, it follows that $\widetilde{\tau}_s = -(\lim_{n\to\infty} s_n^{-1}\sum_{i=1}^{s_n}\mathbb{E}[z_{t,i}^2])\mathbb{E}[u_t v_t] = -\gamma\mathbb{E}[u_t v_t]$ using Assumptions 7 (ii) and 8 (v) and the fact that $\pi_n = O(n^{-\frac{1}{2}})$ and $x_t = \pi_n' z_t + v_t$. Therefore, $\widetilde{\tau}_s = 0$ iff $\mathbb{E}[u_t v_t] = 0$. Hence, 2SLS has asymptotic bias when $\mathbb{E}[u_t v_t] \neq 0$.

4.2 wid-bc-2SLS

This section examines bc-2SLS in the framework of Section 4.1 and its performance is compared with wid-2SLS driven by a continuous kernel. Conditions under which bc-2SLS is asymptotically most efficient are also explored.

We first examine the weak infinite dimensional bias corrected 2SLS (wid-bc-2SLS) estimator.[5] In earlier work Donald and Newey (2001) and Chao and Swanson (2005) examined 2SLS and bc-2SLS under many weak IV conditions. Similarly, our primary interest is in removing the biases that are evident in Corollary 4 and obtaining the limit distribution of the wid-bc-2SLS. We define the wid-bc-2SLS estimator as

$$\widetilde{\theta}_n := -\left(\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n - \left(\frac{s_n^{\vartheta}}{n}\right)\gamma_n\widehat{\Omega}_{xx,n}\right)^{-1}\left((-\widehat{P}_n)'\widehat{\Sigma}_n^{-1}\widehat{c}_n - \left(\frac{s_n^{\vartheta}}{n}\right)\gamma_n(-\widehat{m}_n)\right),$$

[5] Nagar (1959) introduced bias corrected k-class and 2SLS ($k = 1$) estimators, exploring their finite sample properties by means of moment expansions.

where $\widehat{\Omega}_{xx,n} := n^{-1}X'X$, $\widehat{m}_n := n^{-1}X'y$, and $\gamma_n := (s_n n)^{-1}\sum_{i=1}^{s_n}\sum_{t=1}^{n} z_{t,i}^2$. Using $y = X\theta_* + u$, it follows that

$$\begin{aligned}(\widetilde{\theta}_n - \theta_*) &= -\left(\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n - \left(\frac{s_n^{\vartheta}}{n}\right)\gamma_n\widehat{\Omega}_{xx,n}\right)^{-1}\left((-\widehat{P}_n)'\widehat{\Sigma}_n^{-1}\widehat{w}_n - \left(\frac{s_n^{\vartheta}}{n}\right)\gamma_n(-\widehat{s}_n)\right)\\ &= -(\widetilde{A}_n - s_n^{\vartheta-1}\gamma_n\widehat{\Omega}_{xx,n})^{-1}(\ddot{d}_n - s_n^{\vartheta-1}\gamma_n(-\widehat{s}_n)),\end{aligned} \tag{12}$$

where $\widehat{s}_n := n^{-1}X'u$. This definition differs slightly from that in Chao and Swanson (2005). We remove the asymptotic bias by estimating γ by γ_n, and ϑ is not necessarily equal to unity. Different values for ϑ are used for each environment considered in Corollary 4 (i to iii).

Before proceeding, some regularity conditions are imposed for wid-bc-2SLS and the following notation is introduced. First, we apply the law of large numbers to $\widehat{\Omega}_{xx,n}$ and $\widehat{s}_n$. This application is standard for many data with finite moments. Further, we assume that γ_n is consistent for γ at a rate greater than $\sqrt{n}$. Next, we let $\phi_{n,j,i}$ be the j-th row and i-th column element of $\phi_n \in \mathbb{R}^{s_n\times d}$ and extend this to define the following function:

$$\phi_{n,i}(u) := \begin{cases} 0, & \text{if } u \in [0, \frac{1}{s_n})\ ;\\ \phi_{n,j,i}, & \text{if } u \in [\frac{j}{s_n}, \frac{j+1}{s_n}),\ j = 1, 2, \ldots, s_n - 1;\text{ and}\\ \phi_{n,s_n,i}, & \text{if } u = 1.\end{cases}$$

We further let $\phi_n(\cdot) := [\phi_{n,1}(\cdot), \ldots, \phi_{n,d}(\cdot)]'$ and suppose, for some nonzero $\phi(\cdot)$ and $\varphi(\cdot) \in L^2([0,1])$ and $\beta' > 0$, that $\phi_n(\cdot) = \phi(\cdot) + s_n^{-\beta'}\varphi(\cdot)$. This condition is imposed because $\phi(\cdot)$ affects the limit distribution of wid-2SLS. Finally, set $k_n := n^{-\frac{1}{2}}\sum_{t=1}^{n}(u_t v_t - \sigma_{uv})$, where $\sigma_{uv} := \mathbb{E}[u_t v_t]$, and assume that $k_n \rightsquigarrow \mathcal{K} \sim_d \mathcal{N}(0, \Sigma_k^2)$ and so asymptotically normal. In addition, for $\nu \in [0, \frac{1}{2}]$ and $\kappa \geq 0$, define

$$\begin{aligned}\ddot{\ell}_n(\cdot) := \begin{bmatrix} s_n(\cdot)\\ \ddot{a}_n(\cdot)\\ \ddot{c}_n(\cdot)\end{bmatrix} &:= \frac{1}{\sqrt{s_n}}\sum_{j=1}^{\lfloor(\cdot)s_n\rfloor}\begin{bmatrix} \widetilde{w}_{n,j} - \eta_{n,j}\widetilde{u}_n\\ s_n^{-\nu}n^{\frac{1}{2}}(-s_n^{-\kappa}\widetilde{P}_{n,j} - \mu_{n,j})\\ s_n^{-\nu}n^{\frac{1}{2}}(-s_n^{-\kappa}\widetilde{P}_{n,j} - \mu_{n,j})(\widetilde{w}_{n,j} - \eta_{n,j}\widetilde{u}_n) - \widetilde{\tau}_s/\sqrt{s_*}\end{bmatrix}\\ &= \sqrt{s_n}\int_0^{(\cdot)}\begin{bmatrix} \widetilde{g}_n(u) - \eta_n(u)\widetilde{u}_n\\ s_n^{-\nu}n^{\frac{1}{2}}(s_n^{-\kappa}\widetilde{H}_n(u) - \mu_n(u))\\ s_n^{-\nu}n^{\frac{1}{2}}(s_n^{-\kappa}\widetilde{H}_n(u) - \mu_n(u))(\widetilde{g}_n(u) - \eta_n(u)\widetilde{u}_n) - \widetilde{\tau}_s/\sqrt{s_*}\end{bmatrix}du,\end{aligned}$$

and suppose that for a positive-definite $\ddot{\Sigma}_\ell \in \mathbb{R}^{(1+2d)\times(1+2d)}$ and $(1+2d)$-variate Wiener process $\ddot{\mathcal{W}}_\ell(\cdot) \in \mathbb{R}^{1+2d}$, $\ddot{\ell}_n(\cdot) \rightsquigarrow \int_0^{(\cdot)} d\ddot{\mathcal{B}}_\ell(u) := \ddot{\Sigma}_\ell\int_0^{(\cdot)} d\ddot{\mathcal{W}}_\ell(u)$ and for some $\ddot{\Sigma}_{\ell h} \in \mathbb{R}^{(1+2d)\times(1+d)}$, $\mathbb{E}[\int_0^{(\cdot)} d\ddot{\mathcal{B}}_\ell(u)\mathcal{H}'] = \ddot{\Sigma}_{\ell h}\int_0^{(\cdot)} du$, where $\mathcal{H} := (\mathcal{U}, \mathcal{K}')'$. This supposition corresponds to Assumption 8 (vi), but additionally requires that $\sum_{j=1}^{\lfloor(\cdot)s_n\rfloor}\{s_n^{-\nu}n^{\frac{1}{2}}(-s_n^{-\kappa}\widetilde{P}_{n,j}-\mu_{n,j})(\widetilde{w}_{n,j}-\eta_j\widetilde{u}_n)-\widetilde{\tau}_s/\sqrt{s_*}\}$ is asymptotically normal jointly with the other components. For later purposes, $\ddot{\mathcal{B}}_\ell(\cdot)$ is partitioned as $\ddot{\mathcal{B}}_\ell(\cdot)\equiv(\ddot{\mathcal{B}}_s(\cdot),\ddot{\mathcal{B}}_a(\cdot)',\ddot{\mathcal{B}}_c(\cdot)')'\in\mathbb{R}^{1+2d}$. These conditions are collected in the following assumption.

Assumption 9.

(i) $\widehat{\Omega}_{xx,n}:=n^{-1}X'X\overset{\mathbb{P}}{\to}\Omega_{xx}:=\mathbb{E}[x_tx_t']$ *and* $\widehat{s}_n:=n^{-1}X'u\overset{\mathbb{P}}{\to}\mathbb{E}[u_tx_t]$, *and for some* $\gamma>0$, $\gamma_n:=(s_nn)^{-1}\sum_{i=1}^{s_n}\sum_{t=1}^{n}z_{t,i}^2=\gamma+o_{\mathbb{P}}(n^{-\frac{1}{2}})$;

(ii) $\ddot{A}_g:=(\widetilde{A}_f-\gamma\Omega_{xx})$ *is positive definite;*

(iii) for some nonzero $\phi(\cdot)$ *and* $\varphi(\cdot)\in L^2([0,1])$ *and* $\beta'>0$, $\phi_n(\cdot)=\phi(\cdot)+s_n^{-\beta'}\varphi(\cdot)$; *and*

(iv) for positive-definite matrices $\ddot{\Sigma}_\ell\in\mathbb{R}^{(1+2d)\times(1+2d)}$ *and* $\Sigma_h^2\in\mathbb{R}^{(1+d)\times(1+d)}$, $\ddot{\ell}_n(\cdot)\rightsquigarrow\int_0^{(\cdot)}d\ddot{\mathcal{B}}_\ell(u):=\ddot{\Sigma}_\ell\int_0^{(\cdot)}d\ddot{\mathcal{W}}(u)$ *and* $h_n:=(\widetilde{u}_n,k_n')'\rightsquigarrow\mathcal{H}:=(\mathcal{U},\mathcal{K}')'\sim_d\mathcal{N}(0,\Sigma_h^2)$ *such that for some* $\ddot{\Sigma}_{\ell h}\in\mathbb{R}^{(1+2d)\times(1+d)}$, $\mathbb{E}[\int_0^{(\cdot)}d\ddot{\mathcal{B}}_\ell(u)\mathcal{H}']=\ddot{\Sigma}_{\ell h}\int_0^{(\cdot)}du$. □

The limit theory for wid-bc-2SLS is now given according to the settings of Corollary 4. First, partition $\ddot{\Sigma}_{\ell h}$ and Σ_h^2 as follows

$$\underset{(1+2d)\times(1+d)}{\ddot{\Sigma}_{\ell h}}=\begin{bmatrix}\ddot{\sigma}_{su\ (1\times1)} & \ddot{\sigma}_{sk\ (1\times d)}\\ \ddot{\sigma}_{au\ (d\times1)} & \ddot{\Sigma}_{ak\ (d\times d)}\\ \ddot{\sigma}_{cu\ (d\times1)} & \ddot{\Sigma}_{ck\ (d\times d)}\end{bmatrix}\quad\text{and}\quad\underset{(1+d)\times(1+d)}{\Sigma_h^2}=\begin{bmatrix}\sigma_u^2\ {}_{(1\times1)} & \sigma_{uk}\ {}_{(1\times d)}\\ \sigma_{ku}\ {}_{(d\times1)} & \Sigma_k^2\ {}_{(d\times d)}\end{bmatrix},$$

and define $\ddot{\Sigma}_s^2:=\ddot{\sigma}_s^2[b(\cdot),b(\cdot)]+\sqrt{s_*}\varrho\ddot{R}_1\ddot{\Sigma}_\ell^2\ddot{R}_3'+\sqrt{s_*}\ddot{R}_3\ddot{\Sigma}_\ell^2\ddot{R}_1'\varrho'+s_*\ddot{R}_3\ddot{\Sigma}_\ell^2\ddot{R}_3'$, where $\ddot{\sigma}_s^2:=\ddot{R}_1\ddot{\Sigma}_\ell^2\ddot{R}_1'$, $b(\cdot):=\mu(\cdot)+s_*\gamma\phi(\cdot)$, $\varrho:=[b(\cdot),1]$, $\ddot{R}_1:=[1,0_{1\times2d}]$, and $\ddot{R}_3:=[0_{d\times(1+d)},I_{d\times d}]$. The following result collects the asymptotic theory under various conditions characterizing the respective DGPs.

Theorem 4. *Given Assumptions 7, 8 (i to v),*

(i) if Assumptions 8 (vi) and 9 (i) also hold, $\nu=0$, *and* $\vartheta=\frac{1}{2}+\kappa$, *then* $s_n^{\frac{1}{2}+\kappa}(\widetilde{\theta}_n-\theta_*)\rightsquigarrow\mathcal{N}(0,\widetilde{\sigma}_s^2A_e^{-1})$;

(ii) if Assumption 9 also holds, $\nu\in(0,\frac{1}{2})$, *and* $\vartheta=\frac{1}{2}+\nu+\kappa$, *then* $s_n^{\frac{1}{2}+\kappa}(\widetilde{\theta}_n-\theta_*)\rightsquigarrow\mathcal{N}(0,\ddot{\sigma}_s^2A_e^{-1})$;

(iii) if Assumption 9 also holds, $\nu=\frac{1}{2}$, *and* $\vartheta=1+\kappa$, *and*

  (a) if $\kappa=0$, $s_n^{\frac{1}{2}}(\widetilde{\theta}_n-\theta_*)\rightsquigarrow\mathcal{N}(0,\ddot{A}_g^{-1}H_h\ddot{A}_g^{-1})$;

  (b) if $\kappa>0$, $s_n^{\frac{1}{2}+\kappa}(\widetilde{\theta}_n-\theta_*)\rightsquigarrow\mathcal{N}(0,\widetilde{A}_f^{-1}H_h\widetilde{A}_f^{-1})$, *where* $H_h:=\ddot{\Sigma}_s^2+\sqrt{s_*}\gamma(\varrho\ddot{\sigma}_{sk}+\ddot{\sigma}_{sk}'\varrho')+s_*\gamma(\ddot{\Sigma}_{ck}+\ddot{\Sigma}_{ck}')+s_*\gamma^2\Sigma_k^2$. □

Remark.

(i) Under Theorem 4 (i and ii), wid-bc-2SLS satisfies the information matrix equality and is therefore asymptotically most efficient as implied by Corollary 2.

(ii) Under Theorem 4 (iii), the limit distribution of the wid-bc-2SLS depends on the parameters γ and s_*. Although consistent for θ_*, it is not necessarily more efficient than other consistent estimators, as confirmed by simulation evidence in Section 6.

(iii) Theorem 4 shows that wid-bc-2SLS is asymptotically normal, which is established by deriving the weak limits of the components that comprise wid-bc-2SLS. For example, under Theorem 4 (iii.a),

$$s_n^{\frac{1}{2}}(\widetilde{\theta}_n-\theta_*)$$

$$= -(ns_n^{-1}\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n - \gamma_n\widehat{\Omega}_{xx,n})^{-1}(s_n^{\frac{1}{2}}\{ns_n^{-1}(-\widehat{P}_n)'\widehat{\Sigma}_n^{-1}\widehat{w}_n - \widetilde{\tau}_s\} - s_n^{\frac{1}{2}}\{\gamma_n(-\widehat{s}_n) - \widetilde{\tau}_s\})$$
$$= -(\widetilde{A}_n - \gamma_n\widehat{\Omega}_{xx,n})^{-1}(s_n^{\frac{1}{2}}(\ddot{d}_n - \widetilde{\tau}_s) - s_n^{\frac{1}{2}}(\gamma_n(-\widehat{s}_n) - \widetilde{\tau}_s)).$$

We show that $\widetilde{A}_n - \gamma_n\widehat{\Omega}_{xx,n} \xrightarrow{\mathbb{P}} \ddot{A}_g$, $s_n^{\frac{1}{2}}(\ddot{d}_n - \widetilde{\tau}_s) \rightsquigarrow [\mu(\cdot), d\ddot{\mathcal{B}}_s(\cdot)] + \sqrt{s_*}\ddot{\mathcal{B}}_c(1)$, and $s_n^{\frac{1}{2}}(\gamma_n(-\widehat{s}_n) - \widetilde{\tau}_s) \rightsquigarrow \gamma\{s_*[\phi(\cdot), d\ddot{\mathcal{B}}_s(\cdot)] + \sqrt{s_*}\mathcal{K}\}$. Combining these limits delivers the limit distribution. □

4.3 wid-2SLS with a Continuous Kernel and Permuted wid-2SLS

This section examines pwid-2SLS estimation in the framework of Section 4.1. We start by considering wid-2SLS with a continuous kernel and use the following conditions.

Assumption 10.

(i) $\widehat{\sigma}_n(\cdot,\circ)$ *is positive definite uniformly in* n*;*
(ii) for a continuous kernel $\xi(\cdot,\circ)$, $\sup_{(u_1,u_2)}|\widehat{\xi}_n(u_1,u_2) - \xi(u_1,u_2)| \xrightarrow{\mathbb{P}} 0$*;*
(iii) for some nonzero $\rho(\cdot)$, $\mu(\cdot)$ *and* $\chi(\cdot) \in L^2([0,1])$ *and for some* $\alpha > \frac{1}{2}$ *and* $\beta > 0$, $\eta_n(\cdot) = s_n^{-\alpha}\rho(\cdot)$ *and* $\mu_n(\cdot) = \mu(\cdot) + s_n^{-\beta}\chi(\cdot)$*;*
(iv) $A_r := [\Xi\mu(\cdot), \mu(\cdot)]$ *is positive definite; and*
(v) for a positive definite $\widetilde{\Sigma}_\ell$, $\widetilde{\ell}_n(\cdot) \rightsquigarrow \widetilde{\Sigma}_\ell\int_0^{(\cdot)} d\widetilde{\mathcal{W}}_\ell(u)$ *and* $\widetilde{u}_n \rightsquigarrow \mathcal{U} \sim \mathcal{N}(0,\sigma_u^2)$. □

Assumption 10 adjusts the conditions in Assumption 8 to reflect use of a continuous kernel. Assumptions 10 (i, ii, iii, iv and v) correspond to Assumptions 2 (i, ii.a), 3 (i, iii), and 4 (i, ii), respectively. The other conditions in Assumptions 3 and 4 are straightforwardly satisfied by the DGP and model conditions. Assumption 10 (ii) also implies that $r = 0$.

In view of this correspondence, the limit theory of wid-2SLS with a continuous kernel is obtained from Corollary 1. Let $\ddot{\theta}_n$ be the wid-2SLS estimator, viz., $\ddot{\theta}_n := -(\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n)^{-1}(-\widehat{P}_n)'\widehat{\Sigma}_n^{-1}\widehat{c}_n$, so that $\ddot{\theta}_n = \theta_* - (\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n)^{-1}(-\widehat{P}_n)'\widehat{\Sigma}_n^{-1}\widehat{w}_n$. We use different notation here from $\widehat{\theta}_n$ in Section 4.1 to emphasize that $\ddot{\theta}_n$ is driven by a continuous kernel. Note also that

$$s_n^{-2-2\kappa}n\widehat{P}_n'\widehat{\Sigma}_n\widehat{P}_n = s_n^{-2\kappa}n[\widehat{\Xi}_nH_n(\cdot),H_n(\cdot)] = s_n^{-2\kappa}[\widehat{\Xi}_n\widetilde{H}_n(\cdot),\widetilde{H}_n(\cdot)] = s_n^{-2\kappa}\widetilde{A}_n = A_r + o_{\mathbb{P}}(1),$$

by Lemma 1 (i.c). In addition, $s_n^{-2-\kappa}n\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{w}_n = s_n^{-\kappa}[\widehat{\Xi}_n\widetilde{H}_n(\cdot),\widetilde{g}_n(\cdot)] = s_n^{-\kappa}\ddot{d}_n$, and Lemma 1 (i.b) implies that $s_n^{\frac{1}{2}-\kappa}\ddot{d}_n \rightsquigarrow [\Xi\mu(\cdot), d\widetilde{\mathcal{B}}_s(\cdot)]$. Using these results the limit distribution of wid-2SLS is as follows.

Corollary 5. *Given Assumptions 7 and 10,* $s_n^{\frac{1}{2}+\kappa}(\ddot{\theta}_n - \theta_*) \rightsquigarrow \mathcal{N}(0, A_r^{-1}\widetilde{B}_rA_r^{-1})$*, where* $\widetilde{B}_r := \widetilde{\sigma}_s^2[\Xi\mu(\cdot), \Xi\mu(\cdot)]$ *and* $\widetilde{\sigma}_s^2 := R_1\widetilde{\Sigma}_\ell^2R_1'$. □

Corollary 5 follows directly from Corollary 1. Further, Corollary 2 (ii) implies that if $\xi(\cdot,\circ)$ is a continuous kernel close to $\delta(\cdot-\circ)$, the efficiency of $\ddot{\theta}_n$ improves, and its level is bounded from below by $\widetilde{\sigma}_s^2[\mu(\cdot),\mu(\cdot)]^{-1}$.

We next examine pwid-2SLS estimation. Our main interest is in constructing a more efficient estimator using the wid-2SLS procedure driven by a continuous kernel when its limit distribution is non-degenerate, viz., under the environment of Corollary 5. We let m be the permutation size and suppose that each permutation is conducted by sampling the IVs with replacement. That is, given the set of the instruments $\{z_{t,j} : j = 1, 2, \ldots, s_n\}$, we sample the IVs from $\{z_{t,j} : j = 1, 2, \ldots, s_n\}$ with replacement thereby leading to a new permuted instrument $z_t^p := (z_{t,1}^p, \ldots, z_{t,s_n}^p)'$, in which the component $z_{t,j}^p$ is the j-th instrument that is drawn in the permutation. We permute the same set of instruments, so that only the ordering of the instruments is changed. The same permuted order of the instruments is applied to every individual $t = 1, 2, \ldots n$, defining $Z^p := (z_1^p, \ldots, z_n^p)'$, $\widehat{P}_n^p := n^{-1}X'Z^p$, and $\widehat{c}_n^p := n^{-1}Z^{p\prime}y$. With this permutation each new wid-2SLS is given as $\ddot{\theta}_n^p := -(\widehat{P}_n^{p\prime}\widehat{\Sigma}_n^{-1}\widehat{P}_n^p)^{-1}(-\widehat{P}_n^p)'\widehat{\Sigma}_n^{-1}\widehat{c}_n^p$. Proceeding in this way, independently permuting the instruments, computing the estimates $\ddot{\theta}_n^p$ for $p = 1, 2, \ldots, m$, and averaging, leads to the pwid-2SLS estimator given by

$$\bar{\theta}_n := \frac{1}{m}\sum_{p=1}^{m} \ddot{\theta}_n^p.$$

In this scheme, the same permutation order may replicate because of random resampling with replacement to produce permutations. The procedure and conditions are formally represented as follows.

Assumption 11.

(i) For each t, $\{z_{t,j} : j = 1, 2, \ldots, s_n\}$ is iid with respect to j;

(ii) the IVs are randomly sampled from $\{z_{t,j} : j = 1, 2, \ldots, s_n\}$ with replacement to define $z_t^p := (z_{t,(1)}, \ldots, z_{t,(s_n)})'$;

(iii) the IV matrix is defined as $Z^p := (z_1^p, \ldots, z_n^p)'$ by applying the same instrumental order as given in (ii) to every $t = 1, 2, \ldots, n$; and

(iv) for p and $q = 1, 2, \ldots, m$ ($p \neq q$), the random sampling procedure to define $Z^p := (z_1^p, \ldots, z_n^p)'$ is independent of the random sampling procedure to define $Z^q := (z_1^q, \ldots, z_n^q)'$. □

We now apply Corollary 3 to show that pwid-2SLS estimation is more efficient than wid-2SLS under the maintained assumptions. Let $\sigma_q^2 := \mathbb{E}[u_t^2 z_{t,j}^2]$ and $\zeta_1 := [\Xi\mu(\cdot), 1]$. The limit theory for pwid-2SLS is given in the following result.

Corollary 6. *Given Assumptions 7, 10, and 11,*

(i) $s_n^{\frac{1}{2}+\kappa}(\bar{\theta}_n - \theta_*) \rightsquigarrow \mathcal{N}(0, \frac{1}{m}A_r^{-1}\widetilde{B}_r A_r^{-1} + (1 - \frac{1}{m})\sigma_q^2 A_r^{-1}\zeta_1\zeta_1' A_r^{-1})$; and

(ii) $(A_r^{-1}\widetilde{B}_r A_r^{-1} - \sigma_q^2 A_r^{-1}\zeta_1\zeta_1' A_r^{-1})$ is positive definite. □

Corollary 6 follows by verifying the conditions in Corollary 3. In particular, we show that Ψ and Λ in Corollary 3 are $A_r^{-1}\widetilde{B}_r A_r^{-1}$ and $\sigma_q^2 A_r^{-1}\zeta_1\zeta_1' A_r^{-1}$, respectively, which in turn implies that if m is sufficiently large then pwid-2SLS is more efficient than wid-2SLS, as Corollary 6 (ii) assures.

5 Neumann Series Expansion

This section examines a Neumann series expansion applied to implement xid-GMM. When the number of moment conditions s_n rises and becomes large, computation of the inverse matrix $\widehat{\Sigma}_n^{-1}$ can become challenging. In such cases a Neumann series expansion can be useful.

First, note that if s_n is large and the maximum eigenvalue of $s_n^{-1}\widehat{\Sigma}_n$ is less than unity, then we have

$$\left(s_n^{-1}\widehat{\Sigma}_n\right)^{-1} = \left(I - \left(I - s_n^{-1}\widehat{\Sigma}_n\right)\right)^{-1} = \sum_{h=0}^{\infty}\left(I_{s_n} - s_n^{-1}\widehat{\Sigma}_n\right)^h \tag{13}$$

by Neumann expansion. Rewriting the binomial on the right side of (13) for each h gives

$$\left(I_{s_n} - s_n^{-1}\widehat{\Sigma}_n\right)^h = \sum_{\ell=0}^{h}(-1)^{\ell}\binom{h}{\ell}\left(s_n^{-1}\widehat{\Sigma}_n\right)^{\ell}, \tag{14}$$

where $(s_n^{-1}\widehat{\Sigma}_n)^0 := I_{s_n}$. Next, we represent the associated statistics in integral transform form. For this purpose let $\widehat{\Sigma}_n^{(\cdot,j_1)}$ and $\widehat{\Sigma}_n^{(j_{\ell-1},\circ)}$ be the j_1-th column and $j_{\ell-1}$-th row, respectively, so that we can let $\widehat{\sigma}_n(\cdot,\circ) := \widehat{\Sigma}_n^{(\cdot,\circ)}$. With this formulation, we can construct a recursive system in which for $\ell \in \{1,2,3,\ldots\}$,

$$\widehat{\sigma}_{n,\ell}(\cdot,\circ) = \int_0^1 \widehat{\sigma}_{n,\ell-1}(\cdot,u)\widehat{\sigma}_n(u,\circ)du,$$

such that $\widehat{\sigma}_{n,0}(\cdot,\circ) := s_n^{-1}\mathbb{I}(\cdot-\circ)$. If so, it follows that for $\ell \in \{1,2,3,\ldots\}$, $\widehat{\sigma}_{n,\ell}(\cdot,\circ) := s_n^{-(\ell+1)}\widehat{\Sigma}_n^{\ell}$.

Using this recursive system the GMM distance can be obtained without computing $\widehat{\Sigma}_n^{-1}$ directly. For each $\ell \in \{1,2,3,\ldots\}$, $\widehat{\sigma}_{n,\ell}(\cdot,\circ)$ is a kernel of an integral transformation operator. We let $\widehat{\boldsymbol{\Sigma}}_{n,\ell}$ be the integral transformation equipped with this kernel. Then, for the same B_n and C_n as defined in Section 4.1 and for $\ell \in \{1,2,3,\ldots\}$,

$$\frac{1}{s_n^2}B_n'\frac{1}{s_n^{\ell+1}}\widehat{\Sigma}_n^{\ell}C_n = \int_0^1\int_0^1 b_n(u_1)\widehat{\sigma}_{n,\ell}(u_1,u_2)c_n(u_2)du_1du_2 = (\widehat{\boldsymbol{\Sigma}}_{n,\ell}b_n(\cdot),c_n(\cdot)),$$

implying that

$$\frac{1}{s_n^2}B_n's_n^{-1}\left(s_n^{-1}\widehat{\Sigma}_n\right)^{-1}C_n = \sum_{h=0}^{\infty}\sum_{\ell=0}^{h}\binom{h}{\ell}\frac{(-1)^{\ell}}{s_n^{\ell+3}}B_n'\widehat{\Sigma}_n^{\ell}C_n = \sum_{h=0}^{\infty}\sum_{\ell=0}^{h}(-1)^{\ell}\binom{h}{\ell}(\widehat{\boldsymbol{\Sigma}}_{n,\ell}b_n(\cdot),c_n(\cdot)), \tag{15}$$

where the second equality follows from (13) and (14). From the definition of $\widehat{\boldsymbol{\Xi}}_n$, viz., $s_n^{-2}B_n'\widehat{\Sigma}_n^{-1}C_n = (\widehat{\boldsymbol{\Xi}}_nb_n(\cdot),c_n(\cdot))$, the kernel of $\widehat{\boldsymbol{\Xi}}_n$ must be $\sum_{h=0}^{\infty}\sum_{\ell=0}^{h}(-1)^{\ell}\binom{h}{\ell}\widehat{\sigma}_{n,\ell}(\cdot,\circ)$. Therefore, for $h=0,1,\ldots$, if we let $\widehat{\boldsymbol{\Xi}}_{n,h}$ be an integral transform operator with kernel $\widehat{\xi}_{n,h}(\cdot,\circ) := \sum_{\ell=0}^{h}(-1)^{\ell}\binom{h}{\ell}\widehat{\sigma}_{n,\ell}(\cdot,\circ)$, then $\widehat{\boldsymbol{\Xi}}_n$ becomes an integral transform operator with kernel $\widehat{\xi}_n(\cdot,\circ) = \sum_{h=0}^{\infty}\widehat{\xi}_{n,h}(\cdot,\circ)$ such that $s_n^{-2}B_n'\widehat{\Sigma}_n^{-1}C_n =$

$(\widehat{\Xi}_n b_n(\cdot), c_n(\cdot)) = \sum_{h=0}^{\infty}(\widehat{\Xi}_{n,h} b_n(\cdot), c_n(\cdot))$.

We now formally state the following additional condition to incorporate this procedure into xid-GMM estimation with the next theorem giving the operator representation in consequence.

Assumption 12. *The maximum eigenvalue of $s_n^{-1}\widehat{\Sigma}_n$ is less than unity uniformly in n.* □

Theorem 5. *Given Assumptions 1, 2, and 12,*

(i) $q_n(\theta_*) = \sum_{h=0}^{\infty}(\widehat{\Xi}_{n,h}\widetilde{g}_n(\cdot), \widetilde{g}_n(\cdot))$;

(ii) $\bar{A}_n = \sum_{h=0}^{\infty}[\widehat{\Xi}_{n,h} H_n(\cdot), H_n(\cdot)]$ *in probability; and*

(iii) $\widetilde{d}_n = \sum_{h=0}^{\infty}[\widehat{\Xi}_{n,h} H_n(\cdot), \widetilde{g}_n(\cdot)]$. □

6 Simulations

Our simulations have three main goals: (i) to corroborate the limit theory for wxid-GMM by studying findings for wid-2SLS and wid-bc-2SLS; (ii) to compare wid-bc-2SLS with wid-2SLS driven by a continuous kernel, with a key focus on their respective efficiency levels; and (iii) to examine the use of a Neumann series expansion as an approximation in the practical implementation of wid-2SLS.

The following DGP framework is employed. We let $d = 1$, $\phi_{n,*} = \iota_n := (1, 1, \ldots, 1)'$, and $v_t = u_t$ for every t, so that

$$y_t = x_{n,t}\theta_* + u_t \quad \text{and} \quad x_{n,t} := \frac{1}{\sqrt{n}}\sum_{j=1}^{n} z_{t,j} + u_t,$$

where $(z_t', u_t)' \sim \mathcal{N}(0, I_{n+1})$, $z_t := (z_{t,1}, \ldots, z_{t,n})'$, and so for $t \neq \tau$ and $j \neq i$, $(z_{t,j}, z_{\tau,i})' \sim \mathcal{N}(0, I_2)$. Then, $\alpha = \infty$ for an arbitrary continuous function $\rho(\cdot)$. Further, $\gamma := \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\mathbb{E}[z_{t,i}^2] = 1$. Here, $n^{-\frac{1}{2}}\sum_{i=1}^{n} z_{t,i}$ has unit variance and is defined by a local-to-zero parameter with convergence rate $n^{-\frac{1}{2}}$, so that $\pi_{n,*} = n^{-\frac{1}{2}}\iota_n$. By this condition, $\mu_n = -\iota_n$, and Assumption 8 (iii) holds by letting $\mu(\cdot) \equiv -1$ on $[0, 1]$ with $\beta = \infty$ for an arbitrary continuous function $\chi(\cdot)$. We also set $\theta_* = 1$ in the simulations. Note that $v_t \equiv u_t$ in the DGP, so that $\mathbb{E}[u_t v_t] = 1$, and $s_n = n$, which implies $s_* = 1$.

The wid-bc-2SLS is defined as

$$\widetilde{\theta}_n := -(\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n - \gamma_n\widehat{\Omega}_{xx,n})^{-1}((-\widehat{P}_n)'\widehat{\Sigma}_n^{-1}\widehat{c}_n - \gamma_n(-\widehat{m}_n))$$

by noting that $s_n n^{-1} = 1$, where $\widehat{P}_n := n^{-1}Z'X$, $\widehat{c}_n := n^{-1}Z'y$, $\widehat{\Omega}_{xx,n} := n^{-1}X'X$, and $\widehat{m}_n := n^{-1}X'y$. Under the current DGP, for each j, $\mathbb{E}[z_{t,j}^2] = 1$, so that $n^{-1}\sum_{j=1}^{n}\mathbb{E}[z_{t,j}^2] \equiv 1$ and $\gamma_n := (s_n n)^{-1}\sum_{j=1}^{n}\sum_{t=1}^{n} z_{t,j}^2$ is super-consistent for $\gamma = 1$. In addition, we let $\widehat{P}_{n,j}$ and $\widehat{w}_{n,j}$ be the j-th row element of $\widehat{P}_n$ and $\widehat{w}_n := n^{-1}Z'u$, respectively and further let $(\widetilde{P}_{n,j}, \widetilde{w}_{n,j})' := \sqrt{n}(\widehat{P}_{n,j}, \widehat{w}_{n,j})'$. Further, $\widetilde{\theta}_n = \theta_* - (\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n - \gamma_n\widehat{\Omega}_{xx,n})^{-1}((-\widehat{P}_n)'\widehat{\Sigma}_n^{-1}\widehat{w}_n - \gamma_n(-\widehat{s}_n))$, $\widehat{s}_n := n^{-1}X'u$, $\widehat{\Omega}_{xx,n} \stackrel{\mathbb{P}}{\to} \Omega_{xx} = 2$, and $\widehat{s}_n \stackrel{\mathbb{P}}{\to} 1$ by ergodicity and since $\mathbb{E}[x_{n,t}^2] = 2$ and $\mathbb{E}[x_{n,t}u_t] = 1$. Therefore, $\widetilde{\theta}_n \stackrel{\mathbb{P}}{\to} \theta_*$.

We start by considering the asymptotic distribution of wid-bc-2SLS. Let $k_n := n^{-\frac{1}{2}} \sum_{t=1}^{n} (u_t^2 - 1) \rightsquigarrow \mathcal{K} \sim \mathcal{N}(0, 2)$, and following some lengthy algebra find that

$$\ddot{\ell}_n(\cdot) := s_n^{-\frac{1}{2}} \sum_{j=1}^{\lfloor (\cdot) s_n \rfloor} \begin{bmatrix} \widetilde{w}_{n,j} \\ (-\widetilde{P}_{n,j} + 1) \\ (-\widetilde{P}_{n,j} + 1)\widetilde{w}_{n,j} + 1 \end{bmatrix} \rightsquigarrow \int_0^{(\cdot)} d\ddot{\mathcal{B}}_\ell(u) := \begin{bmatrix} 1 & -1 & -1 \\ -1 & 3 & 1 \\ -1 & 1 & 6 \end{bmatrix}^{\frac{1}{2}} \int_0^{(\cdot)} \begin{bmatrix} d\ddot{\mathcal{W}}_1(u) \\ d\ddot{\mathcal{W}}_2(u) \\ d\ddot{\mathcal{W}}_3(u) \end{bmatrix},$$

with $\mathbb{E}[\int_0^{(\cdot)} d\ddot{\mathcal{B}}_\ell(u)\mathcal{K}] = [0, 0, -2]' \int_0^{(\cdot)} du$. From this weak limit it follows that $\nu = \frac{1}{2}$ and $\kappa = 0$. We then apply Theorem 4 (iii.a) to obtain the limit distribution of $\widetilde{\theta}_n$. First, $\mu(\cdot) \equiv -1$ and $\phi(\cdot) \equiv 1$, so that $b(\cdot) := \mu(\cdot) + s_* \gamma \phi(\cdot) \equiv 0$ as $s_* = 1$ and $\gamma = 1$, implying that $\varrho = 0$ and $\ddot{\Sigma}_s^2 = 6$. Next, $\ddot{\Sigma}_{ck} = -2$ and together we have $H_h = \ddot{\Sigma}_s^2 + \sqrt{s_*}\gamma(\varrho \ddot{\sigma}_{sk} + \ddot{\sigma}'_{sk}\varrho') + s_*\gamma(\ddot{\Sigma}_{ck} + \ddot{\Sigma}'_{ck}) + s_*\gamma^2\Sigma_k^2 = 4$. Further, let $\widehat{\Sigma}_n$ be Bickel and Levina's (2008) regularized covariance matrix estimator obtained by setting their parameter $\boldsymbol{\alpha}$ to be proportional to n. This implies that Assumption 8 (i) holds and that $\widehat{\boldsymbol{\Xi}}_n$ asymptotically converges to an identity integral transform operator. With this setting it follows that $\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n = n^{-1}\widetilde{P}_n'\widehat{\Sigma}_n^{-1}\widetilde{P}_n \overset{\mathbb{P}}{\to} 3$, so that $\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n - \gamma_n\widehat{\Omega}_{xx,n} \overset{\mathbb{P}}{\to} 1$ as $\gamma_n\widehat{\Omega}_{xx,n} \overset{\mathbb{P}}{\to} \gamma\Omega_{xx} = 1 \cdot 2 = 2$. Therefore, $\ddot{A}_g = 1$ and $\sqrt{s_n}(\widetilde{\theta}_n - 1) \rightsquigarrow \mathcal{N}(0, 4)$.

Now consider estimating the unknown parameter θ_* by applying Neumann series expansion. If the inverse operator is approximated by a Neumann series, another level of bias is introduced. If k is finite, the finite sample behavior of the wid-bc-2SLS can be poor; but we can compensate by adjusting the bias correction term in $\widetilde{\theta}_n$ as follows

$$\widetilde{\theta}_{n,k} := \left(\sum_{h=0}^{k} (\widehat{\boldsymbol{\Xi}}_{n,h} H_n(\cdot), H_n(\cdot)) - \gamma_{n,k}\widehat{\Omega}_{xx,n} \right)^{-1} \left(\sum_{h=0}^{k} (\widehat{\boldsymbol{\Xi}}_{n,h} H_n(\cdot), c_n(\cdot)) - \gamma_{n,k}\widehat{m}_n \right),$$

and $\gamma_{n,k} := \gamma_n(1 - (1 - 1/n)^{k+1})$. If $k \to \infty$, $\gamma_{n,k} \to \gamma_n$, so that $\widetilde{\theta}_{n,k}$ approaches $\widetilde{\theta}_n$ and we have

$$\widetilde{\theta}_{n,k} = \theta_* + \left(\sum_{h=0}^{k} (\widehat{\boldsymbol{\Xi}}_{n,h} H_n(\cdot), H_n(\cdot)) - \gamma_{n,k}\widehat{\Omega}_{xx,n} \right)^{-1} \left(\sum_{h=0}^{k} (\widehat{\boldsymbol{\Xi}}_{n,h} H_n(\cdot), g_n(\cdot)) - \gamma_{n,k}\widehat{s}_n \right).$$

Under the current DGP, $\sum_{h=0}^{k} (\widehat{\boldsymbol{\Xi}}_{n,h} H_n(\cdot), H_n(\cdot)) = 3\gamma_{n,k} + o_{\mathbb{P}}(1)$, $\widehat{\Omega}_{xx,n} \overset{\mathbb{P}}{\to} 2$, and $\sum_{h=0}^{k} (\widehat{\boldsymbol{\Xi}}_{n,h} H_n(\cdot), g_n(\cdot)) = \gamma_{n,k} + o_{\mathbb{P}}(1)$, so that the finite sample bias introduced by Neumann series expansion is asymptotically removed. Here, $\sum_{h=0}^{k} (\widehat{\boldsymbol{\Xi}}_{n,h} H_n(\cdot), H_n(\cdot)) - \gamma_{n,k}\widehat{\Omega}_{xx,n} = \gamma_{n,k} + o_{\mathbb{P}}(1)$, implying that the inverse converges to unity in probability as n tends to infinity.

In addition to wid-bc-2SLS, we consider another wid-2SLS driven by a continuous kernel. If the kernel $\xi_n(\cdot, \circ)$ converges to a continuous kernel, Corollary 5 implies that wid-2SLS is asymptotically unbiased and normal. We compare its finite sample performance with wid-bc-2SLS. For this purpose, we let $\ddot{\theta}_n :=$

$-(\widehat{P}_n'\widehat{B}_n\widehat{P}_n)^{-1}(-\widehat{P}_n)'\widehat{B}_n\widehat{c}_n$, where $\widehat{B}_n$ is an $n \times n$ matrix such that its i-th row and j-th column element is identical to $\min[\frac{i}{n}, \frac{j}{n}]$. This BMK kernel converges to the continuous function $\min[\cdot, \circ]$ on the unit square as n tends to infinity.

The asymptotic variance of $\ddot{\theta}_n$ is obtained from Corollary 5. That is, $\text{acov}[\sqrt{n}(\ddot{\theta}_n - \theta_*)] = 1.20$ from the fact that $A_r = \frac{1}{3} = \int_0^1 \int_0^1 \min[x, y]dxdy$, $B_r = \frac{2}{15}\sigma^2 = \sigma^2 \int_0^1 (\int_0^1 \min[x, y]dx)^2 dy$, and $\sigma^2 = 1$, which is obtained from the square root of the global covariance matrix of $\widetilde{\ell}_n(1)$. This level of asymptotic variance ensures that the wid-2SLS is more efficient than the wid-bc-2SLS. As is affirmed in the Supplement by simulation, the asymptotic variance further decreases by employing SEK with s^2 close to zero. Although we cannot directly apply Corollary 2 (ii) to the current DGP from the fact that ν differs from 0, this behavior reveals that if the kernel function is continuous and close to the Dirac delta function, a more efficient wid-2SLS is obtained.

Finally, we estimate the unknown parameter by permuting the moment conditions. Given BMK, $\ddot{\theta}_n$ is not independent of the moment order. There are a total of $n!$ possible permutations, which is too large for practical implementation. Instead, $m = 5,000$ independent permutations of the moment orders are employed leading to the sample average $\bar{\theta}_n = m^{-1}\sum_{p=1}^m \ddot{\theta}_n^p$. Since the moment order choices are iid the optimal weight becomes $\frac{1}{m}$ for each permutation by Corollary 3 (i), justifying the use of the mean $\bar{\theta}_n$.

Under the current DGP the asymptotic variance of the pwid-2SLS estimate can be computed analytically. We note that $\text{avar}[\sqrt{n}(\ddot{\theta}_n^p - \theta_*)] = 1.20$; and for $p \neq q$, $\text{acov}[\sqrt{n}(\ddot{\theta}_n^p - \theta_*), \sqrt{n}(\ddot{\theta}_n^q - \theta_*)] = 1.00$. This result follows by noting that

$$\begin{aligned}\text{acov}[\sqrt{n}(\ddot{\theta}_n^p - \theta_*), \sqrt{n}(\ddot{\theta}_n^q - \theta_*)] &= A_r^{-1}\int_0^1 \cdots \int_0^1 \min[u, v]\min[u', v']\mathbb{E}[d\mathcal{W}^p(v)d\mathcal{W}^q(v')]dudu'A_r^{-1} \\ &= 9\left(\int_0^1 \min[u, v]dudv\right)^2 = 1, \end{aligned} \tag{16}$$

where the last equality follows from the fact that $A_r = \frac{1}{3}$, $\sigma^2 = 1$, and $\mathbb{E}[d\widetilde{\mathcal{B}}^p(v)d\widetilde{\mathcal{B}}^q(v')] = dvdv'$ under the current DGP. Applying Corollary 6 (i) gives $\text{acov}[\sqrt{n}(\bar{\theta}_n - \theta_*)] = \frac{1}{m}1.20 + (1 - \frac{1}{m})1.00 = 1.00004$ with $m = 5,000$. If m further increases to infinity, $\text{avar}[\sqrt{n}(\bar{\theta}_n - \theta_*)] \to 1$.

For the present simulations we used $n \in \{200, 500, 1,000, 1,500\}$ and $k \in \{10, 50, 100, 200, 500, \infty\}$ for $\widetilde{\theta}_{n,k}$ and for $k = \infty$, $\widetilde{\theta}_{n,k}$ is defined as $\widetilde{\theta}_n$. The number of replications were 3,000 and 15,000 for $\widetilde{\theta}_{n,k}$ and $(\widetilde{\theta}_n, \ddot{\theta}_n, \bar{\theta}_n)$, respectively. The simulation results are reported in Tables 1 and 2. Table 1 reports the sample averages of $\sqrt{n}(\widetilde{\theta}_n - \theta_*)$, $\sqrt{n}(\ddot{\theta}_n - \theta_*)$, $\sqrt{n}(\bar{\theta}_n - \theta_*)$, and $\sqrt{n}(\widetilde{\theta}_{n,k} - \theta_*)$ for each k, and Table 2 reports their sample variances. The results are summarized as follows

(i) As n increases, the finite sample distributions of $\sqrt{n}(\widetilde{\theta}_n - \theta_*)$ approach $\mathcal{N}(0, 4)$, the sample variance of $\sqrt{n}(\widetilde{\theta}_n - \theta_*)$ approaches 4, and bias approaches zero. But when n is small, the finite sample bias can be relatively large and the variance can be quite different from the asymptotic variance.

(ii) For $k \in \{10, 50, 100, 200, 500\}$, the overall distribution of $\sqrt{n}(\widetilde{\theta}_{n,k} - \theta_*)$ differs from that of $\sqrt{n}(\widetilde{\theta}_n - \theta_*)$, and the variance varies depending on k.

(iii) More efficient estimation is achieved by selecting k carefully when n is small. For example, when $n = 200$, the sample variance of $\sqrt{n}(\widetilde{\theta}_{n,k} - \theta_*)$ is smallest when $k = 100$, although its bias is slightly greater than that with $n = 500$. Hence, the efficiency of the wid-bc-2SLS depends on k when n is relatively small. On the other hand, if n is large, the sample variance of $\sqrt{n}(\widetilde{\theta}_{n,k} - \theta_*)$ decreases and converges to that of $\sqrt{n}(\widetilde{\theta}_n - \theta_*)$ as k increases. Consequently, the latter is more efficient than the former. The sample variance of $\sqrt{n}(\widetilde{\theta}_{n,k} - \theta_*)$ is always greater than that of $\sqrt{n}(\widetilde{\theta}_n - \theta_*)$ for $n = 500$, $1,000$, and $1,500$; and, further, the finite sample bias approaches zero as k increases, meaning that $\gamma_{n,k}$ removes the finite sample bias. This feature is also apparent in distributional shape. Figure 1 shows the empirical distributions of $\sqrt{n}(\widetilde{\theta}_{n,k} - \theta_*)$, revealing how it approaches that of $\sqrt{n}(\widetilde{\theta}_n - \theta_*)$ as k increases. We let $n = 500$ and increase k from 50 to 100, 200, and 500. With $k = 500$ the empirical distribution of $\sqrt{n}(\widetilde{\theta}_{n,k} - \theta_*)$ is almost identical to that of $\sqrt{n}(\widetilde{\theta}_n - \theta_*)$ and also almost identical to that of $\mathcal{N}(0, 4)$.

(iv) If k is small, the bias becomes large as n increases. For example, when $k = 10$, the sample average of $\sqrt{n}(\widetilde{\theta}_{n,k} - \theta_*)$ with $n = 500$ is more negatively valued than that with $n = 300$. This behavior arises from the observation that $\gamma_{n,k}$ converges to zero as $n \to \infty$ for fixed k, so that the asymptotic bias cannot be removed by increasing only n. The asymptotic bias of the wid-2SLS estimator is more easily passed on to $\widetilde{\theta}_{n,k}$ when k is small, meaning that for large n, k also needs to be large for improved performance.

(v) We also compare $\sqrt{n}(\ddot{\theta}_n - \theta_*)$ with $\sqrt{n}(\widetilde{\theta}_n - \theta_*)$. From the last third and second columns of Table 1, both estimators are distributed around zero, but the sample mean of $\sqrt{n}(\ddot{\theta}_n - \theta_*)$ is closer to zero than that of $\sqrt{n}(\widetilde{\theta}_n - \theta_*)$, implying that the finite sample bias of $\sqrt{n}(\ddot{\theta}_n - \theta_*)$ disappears more rapidly. Furthermore, the sample variance of $\sqrt{n}(\ddot{\theta}_n - \theta_*)$ is always smaller than that of $\sqrt{n}(\widetilde{\theta}_n - \theta_*)$. The asymptotic variance of $\sqrt{n}(\ddot{\theta}_n - \theta_*)$ is 1.20, which is close to the values in the second last column of Table 2. As the sample size increases, the sample variance approaches 1.20. This fact implies that the wid-2SLS driven by a continuous kernel is more efficient than the wid-bc-2SLS. In the Supplement, we also conduct simulations using SEK with various s^2's and observe that the sample variance approaches 1.00 as s^2 gets close to zero.

(vi) We compare the pwid-2SLS ($\bar{\theta}_n$) with the other procedures. Most of all, the sample average of

$\sqrt{n}(\bar{\theta}_n - \theta_*)$ is close to zero and converges to zero as n increases, implying that the pwid-2SLS is asymptotically unbiased. Further, the magnitude of the bias is smaller than the other estimators for large n. In addition, the sample variance of $\sqrt{n}(\bar{\theta}_n - \theta_*)$ is noticeably smaller than the others. This behavior corroborates Corollary 3 (iv). □

Although they are not reported here, simulations with other kernels were conducted similar to those in the present section. Simulation results are reported in the Supplement for BMK, BBK, and SEKs kernels with various s^2's and the Dirac delta function. The findings from these simulation exercises together confirm the properties and limit theory of wid-2SLS and pwid-2SLS estimation.

7 Empirical Application

This section reports an application of xid-GMM and wid-2SLS estimation to the empirical data in the study of Angrist and Krueger (1991), providing further empirical evidence on the returns to human capital in education. The literature on human capital and the relationship between education and its monetary return has been a popular applied research topic with the links between education and wage being a central issue. For example, Mincer (1958) provides a theory that describes a linear relationship between log wage and education; and Angrist and Krueger (1991) estimate a structural relationship between the variables using 2SLS. Here we report findings from the use of wid-2SLS estimation to the same empirical data.

We briefly review the well-known work of Angrist and Krueger (1991) to make the discussion here self contained. Their empirical investigation uses 5% samples of the US census data and notes that variation in education is created by compulsory school attendance laws because individuals born in the first quarter of year have shorter overall school attendance. This variation is used to construct instruments in the estimation of a monetary return equation by 2SLS. Specifically, they specify the following return equation based on Mincer (1958)

$$w_t = \xi_{0*} + \xi_{1*} x_t + d_t' \boldsymbol{\eta}_* + U_t, \tag{17}$$

where w_t denotes the weekly log wage (*wage*); x_t is schooling years (*educ*); d_t denotes other exogenous covariates, viz., $d_t = (\textit{birth}_{t,1}, \ldots, \textit{birth}_{t,9})'$; and $\textit{birth}_{t,c}$ denotes the birth year dummy for the t^{th}-individual; and $c = 1, 2, \ldots, 9$. There are 10 birth years in each cohort, so 9 dummies are used. There are 11 covariates in total on the right side of (17) including the constant (*const*), and x_t is the only endogenous variable, so that its coefficient ξ_{1*} measures the return to education. In addition, they construct the IV z_t using the exogenous variables (*const*, d_t')$'$ and the season of birth (*seas*), viz., $z_t = (1, d_t', \textit{seas}_{t,1}, \ldots, \textit{seas}_{t,30})'$, where $\textit{seas}_{t,j}$ is defined as $\textit{birth}_{t,j} \times \textit{qrtr}_{t,i}$ with $j = 1, 2, \ldots, 10$, and $\textit{qrtr}_{t,i}$ denotes the quarter of birth of the

t^{th}-individual with $i = 1, 2,$ and 3. Overall there are 40 IVs including the exogenous variables. We call this monetary return equation the basic model.

Using the three data sets constructed by the individuals belonging to 1920–1929, 1930–1939, and 1940–1949 cohorts, Angrist and Krueger (1991) estimated the basic model by ordinary least squares (OLS) and 2SLS and compared the estimation results, finding that individuals compelled to attend school for a shorter time than others earn lower wages as a result of their education shortfall. They also find that the returns to education estimated by 2SLS are statistically indistinguishable from those by OLS.[6] Angrist and Krueger (1991) estimated another model in parallel to the basic model by letting the regressor d_t include race dummies, a dummy for residence in an SMSA, a marital status dummy, and eight region-of-residence dummies. We call the model with these exogenous variables the extended model. Using this extended model, the authors obtained empirical results similar to those of the basic model.

The empirical findings of Angrist and Krueger (1991) motivated other research. For example, Staiger and Stock (1997) pointed out that their results were obtained using weak IVs, so that estimating the structural coefficient by 2SLS can be misleading; and instead used LIML estimation with the same data (see Anderson and Rubin, 1949; Anderson, 2005) in view of its improved properties under weak instrumentation. In other work, Donald and Newey (2001) and Chao and Swanson (2005) employed wid-bc-2SLS estimation to allow for the use of many weak IVs in estimation.

For our empirical illustration, we apply pwid-2SLS to the same data. The findings are contained in Table 3 based on estimation obtained as follows. First, OLS and 2SLS replicate Angrist and Krueger (1991): for the 3 cohorts and both basic and extended models, the structural coefficient was estimated by OLS and 2SLS; and 90% and 95% confidence intervals (CIs) given below the estimates were obtained by applying the residual bootstrap (e.g., Efron and Tibshirani, 1993), using 5,000 bootstrap iterations. Second, LIML and wid-bc-2SLS are obtained following Staiger and Stock (1997) and Chao and Swanson (2005). Again, 90% and 95% CIs are provided by the residual bootstrap. Third, we provide the pwid-2SLS along with its 90% and 95% CIs, and in estimation we employ the two continuous kernels BMK and BBK. Specifically, we randomly draw 5,000 different permutation orders with replacement and obtain 5,000 wid-2SLS estimates from each kernel. We compute the pwid-2SLS by averaging the 5,000 estimates, as discussed earlier in the paper. In addition, 90% and 95% CIs are obtained by bootstrapping 5,000 samples for each kernel.

The findings are summarized as follows.

(i) For the 1920-1929 cohort, we observe that the OLS, 2SLS, LIML, wid-bc-2SLS, and two pwid-2SLS

[6] Angrist and Krueger (1991) provide more details on the data in the following URL: http://economics.mit.edu/faculty/angrist/data1/data/angkru1991.

estimators produce similar estimates. In addition, the 90% and 95% CIs of both pwid-2SLS estimators are narrower than those of LIML and wid-bc-2SLS. Further, the 90% and 95% ranges of both pwid-2SLS estimators are narrower than those of the LIML and wid-bc-2SLS.

(ii) For the 1930-1939 cohort, both pwid-2SLS estimators produce estimates greater than OLS, 2SLS, LIML, and wid-bc-2SLS; and they are also similar to each other. In particular, both pwid-2SLS estimates are greater than OLS by 35.20% for the basic model and by 44.44% for the extended model; and the CIs of both pwid-2SLS estimators are narrower than those of LIML and wid-bc-2SLS. In addition, the 90% and 95% ranges of both pwid-2SLS estimators are narrower than those of LIML and wid-bc-2SLS.

(iii) For 1940-1949 cohort, different results are observed between the basic and extended models. For the basic model, the OLS, 2SLS, LIML, and wid-bc-2SLS produce similar estimation results, but both pwid-2SLS estimators produce smaller coefficients than the others. In addition, both pwid-2SLS estimators produce narrower CIs than LIML and wid-bc-2SLS. For the extended model, OLS differs from the others, and 2SLS and wid-bc-2SLS produce similar results smaller than OLS. LIML is much smaller than OLS, and both pwid-2SLS estimators produce the smallest estimates. Further, the size of the CIs and ranges are in the order of wid-bc-2SLS, the two wid-2SLSs, and LIML from the smallest to the largest.

(iv) Although we report estimation results for the 1940-1949 cohort in Table 3, the estimates need some care in interpretation. First, Staiger and Stock (1997) rejected the overidentification hypothesis using the Basmann-LIML test and concluded that the model specifications are unrealistic for the data, so that the estimates of the basic and extended models cannot provide an accurate indication of the structural relationship between the variables in this case. Second, Table 3 demonstrates that the estimates are indeed affected by the asymptotic bias induced by the model misspecification. The 2SLS, LIML, and wid-bc-2SLS estimates are smaller than OLS by at least 25% for the extended model, while estimates from the basic model are not so different from OLS although they are still smaller than OLS. Finally, the asymptotic bias is evident from both pwid-2SLS estimates: when compared to OLS, the return to education is estimated to be smaller than OLS by more than 63% for both basic and extended models. These findings help to reinforce doubts concerning the validity of the models for the 1940-1949 cohort. □

8 Conclusion

If the number of moment conditions grows to infinity proportional to the sample size, GMM asymptotics differ from the standard case where the number of moment conditions is constant. In particular, the new limit depends on the stochastic properties of the moment conditions and the weight matrix used to define the GMM distance. The present work provides new GMM asymptotics to address high dimensionality in which partial sums formed by the moment conditions satisfy an invariance principle. The moment conditions may be strong or weak and the GMM weight matrix may converge either to a continuous kernel or the Dirac delta function. The limit distribution of xid-GMM is obtained in this setting and shown to be normal and unbiased if the weight matrix converges to a continuous kernel.

A further innovation enables improvement in xid-GMM asymptotic efficiency. In the case where xid-GMM estimation is driven by a continuous kernel in the limit, a new estimator is developed called pxid-GMM obtained by permuting the moment conditions and averaging the resulting xid-GMM estimates. If the different estimates are based on randomly sampling and permuting the moment order, the asymptotic efficiency of the resulting estimator improves as the random sample size increases. In the 2SLS framework, this property is confirmed in simulations. For computing the inverse matrices of high dimension that arise in practical work, a Neumann series expansion is found to be a useful device to simplify calculations. All these methods are illustrated in an empirical application estimating the monetary return to education using the US census data sets in Angrist and Krueger (1991), showing the effect of efficient GMM estimation in applied work with many weak moment conditions.

References

ABADIR, K. M. AND J. R. MAGNUS (2012): *Matrix Algebra*, New York: Cambridge University Press. 8

ANATOLYEV, S. (2019): "Many Instruments and/or Regressors: A Friendly Guide," *Journal of Economic Surveys*, 33, 689–726. 23

ANDERSON, T. (2005): "Origins of the Limited Information Maximum Likelihood and Two-Stage Least Squares Estimators," *Journal of Econometrics*, 127, 1–16. 2, 37

ANDERSON, T. AND H. RUBIN (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63. 2, 37

ANGRIST, J. D. AND A. B. KRUEGER (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 106, 979–1014. 1, 2, 4, 36, 37, 39, 44

——— (1995): "Split-Sample Instrumental Variables the Return to Schooling," *Journal of Business & Economic Statistics*, 13, 225–235. 6, 23

BAI, J. AND S. NG (2010): "Instrumental Variable Estimation in a Data Rich Environment," *Econometric Theory*, 26, 1577–1606. 7

BATES, C. AND H. WHITE (1985): "A Unified Theory of Consistent Esetimation for Parametric Model," *Econometric Theory*, 1, 151–178. 5

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. 7

BICKEL, P. AND E. LEVINA (2008): "Regularized Estimation of Large Covariance Matrices," *Annals of Statitics*, 36, 199–227. 24, 25, 33, 16

CARRASCO, M. AND J.-P. FLORENS (2000): "Generalization of GMM to a Continuum of Moment Conditions," *Econometric Theory*, 16, 797–834. 6, 7, 9

CARRASCO, M., J.-P. FLORENS, AND E. RENAULT (2007): "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," *Handbook of Econometrics*, 6, 5633–5751. 9

CHAO, J. AND N. SWANSON (2005): "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica*, 73, 1673–1692. 2, 3, 4, 6, 15, 18, 22, 23, 26, 27, 37

CHO, J. S. AND P. C. B. PHILLIPS (2025): "GMM Estimation with Brownian Kernels Applied to Income Inequality Measurement," *Journal of Econometrics*, 252, 106110. 3, 7, 9

——— (2026): "Efficient Estimation in Infinite Dimensional GMM," Tech. rep., Yonsei Economic Research Institute, Yonsei University. 10

CHOI, I. AND P. C. B. PHILLIPS (1992): "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations," *Journal of Econometrics*, 51, 113–150. 23

DE JONG, R. AND J. DAVIDSON (2000): "The Functional Central Limit Theorem and Weak Convergence to Stochastic Integrals I," *Econometric Theory*, 16, 621–642. 3, 13

DONALD, S., G. IMBENS, AND N. NEWEY (2003): "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions," *Econometrica*, 69, 1161–1191. 16

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

1191. 2, 6, 7, 23, 26, 37

EFRON, B. AND R. J. TIBSHIRANI (1993): *An Introduction to the Bootstrap*, New York: Chapman & Hall/CRC. 37

GRENANDER, U. (1981): *Abstract Inference*, New York: John Wiley & Sons. 7

HAHN, J., J. HAUSMAN, AND G. KUERSTEINER (2001): "Accuracy of Higher Order Bias and Approximations," *Econometrics Journal*, 7, 272–306. 6, 23

HAN, C. AND P. C. B. PHILLIPS (2006): "GMM with Many Moment Conditions," *Econometrica*, 74, 147–192. 3, 4, 6, 15, 18, 23

HANSEN, B. E. (2009): "Averaging Estimators for Regressions with a Possible Structural Break," *Econometric Theory*, 25, 1498–1514. 20

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054. 5, 15

IBRAGIMOV, R. AND P. C. B. PHILLIPS (2008): "Regression Asymptotics Using Martingale Convergence Methods," *Econometric Theory*, 24, 888–947. 3

JAGANNATHAN, R., G. SKOULAKIS, AND Z. WANG (2002): "Generalized Method of Moments: Applications in Finance," *Journal of Business and Economic Statistics*, 20, 470–481. 2

KIRSCH, A. (1996): *An Introduction to the Mathematical Theory of Inverse Problems*, Applied Mathematical Sciences, Springer New York. 9

KWAPIEN, S. AND W. WOYCZYNSKI (1987): "Double Stochastic Integrals, Random Quadratic Forms and Random Series in Orlicz Spaces," *Annals of Probability*, 15, 1072–1096. 14

LEDOIT, O. AND M. WOLF (2003): "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrics," *Journal of Multivariate Analysis*, 88, 365–411. 6

MINCER, J. (1958): "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66, 281–302. 36

NAGAR, A. (1959): "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27, 575–595. 2, 26

NEWEY, W. AND F. WINDMEIJER (2009): "Generalized Method of Moments with Many Weak Moment Conditions," *Econometrica*, 77, 687–719. 3, 4, 6, 15, 18, 23

Phillips, P. C. B. (1980): "The Exact Distribution of Instrumental Variable Estimators in an Equation Containing $n+1$ Endogenous Variables," *Econometrica*, 861–878. 7, 18

——— (1989): "Partially Identified Econometric Models," *Econometric Theory*, 5, 181–240. 2, 7, 18, 23

——— (2016): "Inference in Near-Singular Regression," *Advances in Econometrics*, 36, 461–486. 6

Phillips, P. C. B. and L. Jiang (2025): "Cross Section Curve Data Autoregression," Discussion Paper 2439, Cowles Foundation, Yale University. 3

Phillips, P. C. B. and V. Solo (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 972–1001. 3, 13

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. 7, 9

Rasmussen, C. E. and C. K. I. Williams (2005): *Gaussian Processes for Machine Learning*, The MIT Press. 19

Sargan, J. D. (1958): "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393–415. 15

Shi, Z. (2016): "Econometric Estimation with High-Dimensional Moment Equalities," *Journal of Econometrics*, 195, 104–119. 6, 7

Staiger, D. and J. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586. 2, 6, 7, 18, 23, 37, 38

Taniguchi, M. and V. Tresp (1997): "Averaging Regularized Estimators," *Neural Computation*, 9, 1163–1178. 20

Wang, Y. (1975): "Characterization of Some Stochastic Processes," *Annals of Probability*, 3, 1038–1045. 14

$n\backslash$ Estimators	$k=10$	$k=50$	$k=100$	$k=200$	$k=500$	$k=\infty$	$\sqrt{n}(\ddot{\theta}_n-\theta_*)$	$\sqrt{n}(\bar{\theta}_n-\theta_*)$
200	-2.9476	-0.9008	-0.6599	-0.7224	-0.1682	-0.4551	-0.0701	-0.0751
500	-3.3566	-0.7509	-0.5252	-0.3614	-0.2968	-0.2389	-0.0282	-0.0369
1,000	-4.4116	-0.9275	-0.5499	-0.3709	-0.1959	-0.2370	-0.0348	-0.0246
1,500	-5.0503	-0.9960	-0.5207	-0.3834	-0.2093	-0.1449	-0.0346	-0.0233

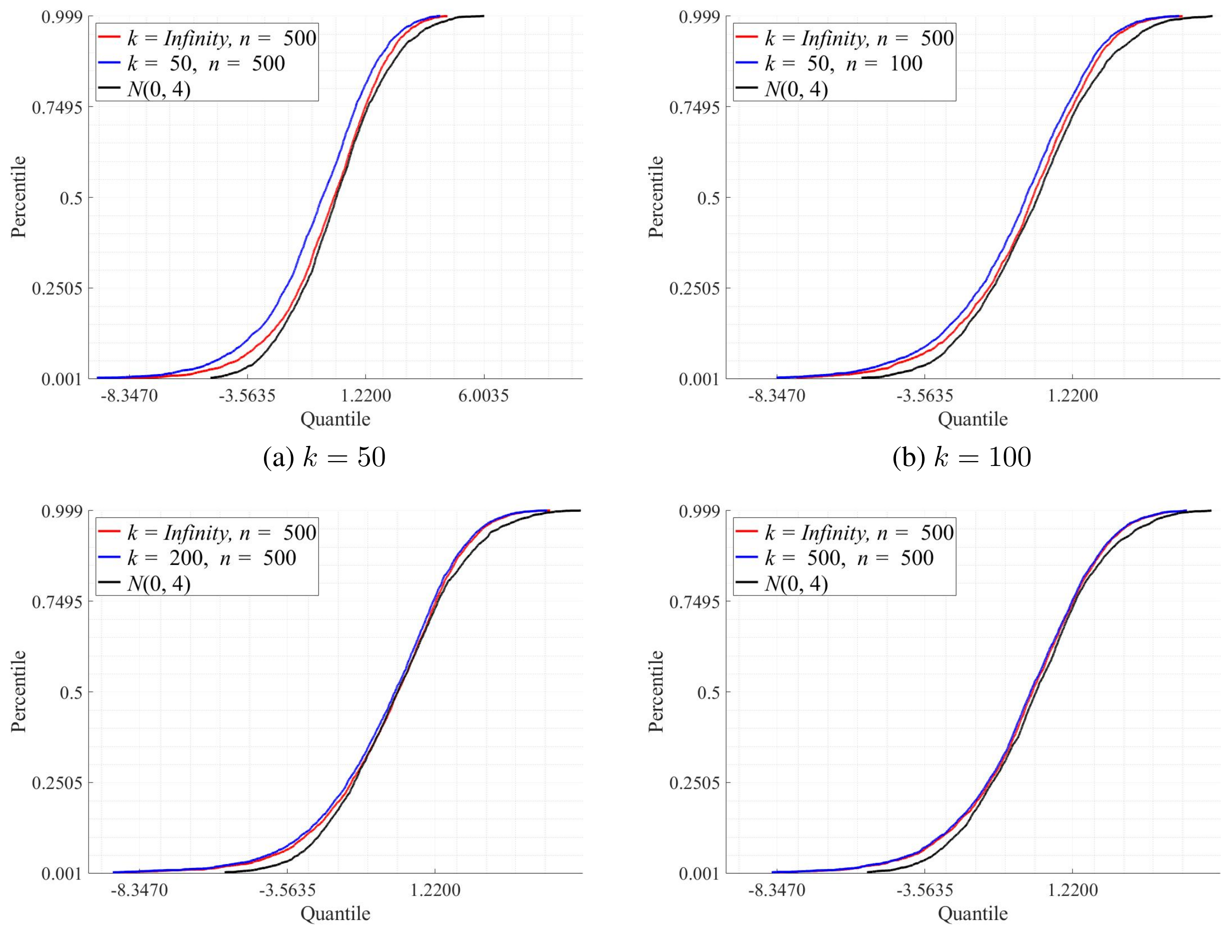
Table 1: SAMPLE BIASES OF WID-BC-2SLS, WID-2SLS, AND PERMUTED WID-2SLS. Each cell reports the finite sample bias of $\sqrt{n}(\widetilde{\theta}_{n,k}-\theta_*)$ that is obtained by implementing 3,000 independent experiments. If $k=\infty$, the cell reports the sample variance of $\sqrt{n}(\widetilde{\theta}_n-\theta_*)$ that is obtained by implementing 15,000 independent experiments. Here, k is the degree of Neumann's series expansion, and if $k=\infty$, the inverse matrix of $\widehat{\Sigma}_n$ is directly used as the weight matrix. Finally, the cells in the last two columns report the sample averages of $\sqrt{n}(\ddot{\theta}_n-\theta_*)$ and $\sqrt{n}(\bar{\theta}_n-\theta_*)$ obtained by 15,000 independent experiments, respectively, where $\ddot{\theta}_n$ and $\bar{\theta}_n$ are the wid-2SLS and pwid-2SLS driven by BMK.

$n\backslash$ Estimators	$k=10$	$k=50$	$k=100$	$k=200$	$k=500$	$k=\infty$	$\sqrt{n}(\ddot{\theta}_n-\theta_*)$	$\sqrt{n}(\bar{\theta}_n-\theta_*)$
200	45.442	10.438	6.6285	14.117	375.16	59.092	1.2556	1.0500
500	12.434	5.2559	4.9139	5.0342	4.6732	4.6000	1.2158	1.0097
1,000	9.7545	5.0122	4.5818	4.5868	4.5088	4.3848	1.2118	1.0205
1,500	8.9147	4.7059	4.2674	4.1417	4.1790	4.0493	1.1946	1.0070

Table 2: SAMPLE VARIANCE OF WID-BC-2SLS, WID-2SLS, AND PERMUTED WID-2SLS. Each cell reports the finite sample bias of $\sqrt{n}(\widetilde{\theta}_{n,k}-\theta_*)$ that is obtained by implementing 3,000 independent experiments. If $k=\infty$, the cell reports the sample variance of $\sqrt{n}(\widetilde{\theta}_n-\theta_*)$ that is obtained by implementing 15,000 independent experiments. Here, k is the degree of Neumann's series expansion, and if $k=\infty$, the inverse matrix of $\widehat{\Sigma}_n$ is directly used as the weight matrix. Finally, the cells in the last two columns report the sample variances of $\sqrt{n}(\ddot{\theta}_n-\theta_*)$ and $\sqrt{n}(\bar{\theta}_n-\theta_*)$ obtained by 15,000 independent experiments, respectively, where $\ddot{\theta}_n$ and $\bar{\theta}_n$ are the wid-2SLS and pwid-2SLS driven by BMK.

Data	1920-1929 Cohort		1930-1939 Cohort		1940-1949 Cohort	
Estimations \ Models	Basic	Extended	Basic	Extended	Basic	Extended
OLS	0.080	0.070	0.071	0.063	0.057	0.052
90% CI	[0.080, 0.081]	[0.069, 0.071]	[0.070, 0.072]	[0.063, 0.064]	[0.057, 0.058]	[0.057, 0.058]
95% CI	[0.079, 0.081]	[0.069, 0.071]	[0.070, 0.072]	[0.062, 0.064]	[0.051, 0.053]	[0.051, 0.053]
2SLS	0.077	0.067	0.089	0.081	0.055	0.039
90% CI	[0.053, 0.103]	[0.042, 0.093]	[0.061, 0.113]	[0.056, 0.119]	[0.031, 0.081]	[0.014, 0.067]
95% CI	[0.047, 0.108]	[0.037, 0.098]	[0.052, 0.104]	[0.046, 0.110]	[0.025, 0.086]	[0.009, 0.072]
LIML	0.076	0.066	0.093	0.084	0.054	0.029
90% CI	[0.033, 0.119]	[0.023, 0.108]	[0.055, 0.133]	[0.046, 0.123]	[0.000, 0.107]	[-0.032,0.086]
95% CI	[0.020, 0.128]	[0.014, 0.118]	[0.047, 0.142]	[0.038, 0.131]	[-0.014,0.119]	[-0.047,0.098]
wid-bc-2SLS	0.076	0.066	0.094	0.085	0.055	0.037
90% CI	[0.046, 0.108]	[0.036, 0.097]	[0.058, 0.121]	[0.049, 0.115]	[0.027, 0.084]	[0.008, 0.070]
95% CI	[0.040, 0.113]	[0.029, 0.103]	[0.052, 0.128]	[0.043, 0.122]	[0.021, 0.264]	[0.002, 0.076]
pwid-2SLS (BMK)	0.073	0.071	0.096	0.091	0.010	0.019
90% CI	[0.073, 0.074]	[0.071, 0.072]	[0.096, 0.097]	[0.091, 0.091]	[0.009, 0.011]	[0.018, 0.019]
95% CI	[0.073, 0.074]	[0.071, 0.072]	[0.096, 0.097]	[0.091, 0.092]	[0.009, 0.011]	[0.018, 0.019]
pwid-2SLS (BBK)	0.074	0.071	0.097	0.091	0.012	0.019
90% CI	[0.073, 0.074]	[0.071, 0.072]	[0.096, 0.097]	[0.091, 0.091]	[0.011, 0.013]	[0.019, 0.020]
95% CI	[0.073, 0.074]	[0.071, 0.072]	[0.096, 0.097]	[0.091, 0.092]	[0.011, 0.013]	[0.019, 0.020]

Table 3: STRUCTURAL PARAMETER ESTIMATION. This table shows the empirical estimation results from the basic and extended models. We use the same data as used by Angrist and Krueger (1991) and report the estimated coefficient of the schooling years in the model along with their confidence intervals. Six estimators are reported: the OLS, 2SLS, LIML, wid-bc-2SLS and two pwid-2SLS estimators driven by BMK and BBK.



(a) $k = 50$ (b) $k = 100$

(c) $k = 200$ (d) $k = 500$

Figure 1: EMPIRICAL DISTRIBUTIONS OF WID-2SLS. For $k = 50, 100, 200, 500$, each figure shows the empirical distribution of $\sqrt{n}(\widetilde{\theta}_{n,k} - \theta_*)$ that is obtained by implementing 3,000 independent experiments and letting n be 500. For comparison purpose, the distribution function of $\mathcal{N}(0, 4)$ is also drawn. Here, k is the degree of Neumann's series expansion, and if $k = \infty$, the inverse matrix of $\widehat{\Sigma}_n$ is directly used as the weight matrix.

Online Supplement for

'Efficient Estimation in Infinite-Dimensional GMM'

by

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

[a]Yonsei University

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

This Online Supplement is an Appendix that provides proofs of the results in the paper, including the lemmas. Additional simulation results and empirical findings are reported here; and notation used and defined throughout the main paper is collected in a Glossary at the end of this supplement in Section A.4.

A Appendix

A.1 Proofs of the Main Claims

We let $i_n := i/s_n$ and $j_n := j/s_n$ for notational simplicity.

Proof of Lemma 1: (i.a) First note that $s_n^{-2-r}q_n(\theta_*) = s_n^{-r}(\widehat{\boldsymbol{\Xi}}_n\widetilde{g}_n(\cdot), \widetilde{g}_n(\cdot)) = (\boldsymbol{\Xi}\widetilde{g}_n(\cdot), \widetilde{g}_n(\cdot)) + o_{\mathbb{P}}(1)$ by Assumption 2 (ii.a). Next observe that

$$
\begin{aligned}
&(\boldsymbol{\Xi}\widetilde{g}_n(\cdot), \widetilde{g}_n(\cdot))\\
=&\frac{1}{s_n^2}\sum_{i=1}^{s_n}\sum_{j=1}^{s_n}\xi\left(i_n, j_n\right)\left\{\widetilde{g}_n\left(i_n\right) - \eta_n\left(i_n\right)\widetilde{u}_n + \eta_n\left(i_n\right)\widetilde{u}_n\right\}\left\{\widetilde{g}_n\left(j_n\right) - \eta_n\left(j_n\right)\widetilde{u}_n + \eta_n\left(j_n\right)\widetilde{u}_n\right\}\\
=&\frac{1}{s_n^2}\sum_{i=1}^{s_n}\sum_{j=1}^{s_n}\xi\left(i_n, j_n\right)\left\{\widetilde{g}_n\left(i_n\right) - \eta_n\left(i_n\right)\widetilde{u}_n\right\}\left\{\widetilde{g}_n\left(j_n\right) - \eta_n\left(j_n\right)\widetilde{u}_n\right\}\\
&+\frac{2}{s_n^2}\sum_{i=1}^{s_n}\sum_{j=1}^{s_n}\xi\left(i_n, j_n\right)\left\{\widetilde{g}_n\left(i_n\right) - \eta_n\left(i_n\right)\widetilde{u}_n\right\}\eta_n\left(j_n\right)\widetilde{u}_n + \frac{1}{s_n^2}\sum_{i=1}^{s_n}\sum_{j=1}^{s_n}\xi\left(i_n, j_n\right)\eta_n\left(i_n\right)\eta_n\left(j_n\right)\widetilde{u}_n^2\\
&=: A_n + B_n + C_n.
\end{aligned}
$$

The limit behavior of A_n, B_n, and C_n is examined in turn. First,

$$
\begin{aligned}
s_nA_n &= \sum_{i=1}^{s_n}\sum_{j=1}^{s_n}\xi\left(i_n, j_n\right)s_n^{-\frac{1}{2}}\left\{\widetilde{g}_n\left(i_n\right) - \eta_n\left(i_n\right)\widetilde{u}_n\right\}s_n^{-\frac{1}{2}}\left\{\widetilde{g}_n\left(j_n\right) - \eta_n\left(j_n\right)\widetilde{u}_n\right\}\\
&\rightsquigarrow \int_0^1\int_0^1\xi(u,v)d\mathcal{B}_s(u)d\mathcal{B}_s(v) = (\boldsymbol{\Xi}d\mathcal{B}_s(\cdot), d\mathcal{B}_s(\cdot)),
\end{aligned}
$$

by Assumption 4. This fact implies that $A_n = O_{\mathbb{P}}(s_n^{-1})$.

Second, if for some $\alpha > \frac{1}{2}$, $\eta_n(\cdot) = s_n^{-\alpha}\rho(\cdot)$, then

$$s_n^{\frac{1}{2}+\alpha} B_n = \frac{2}{s_n} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi\left(i_n, j_n\right) s_n^{-\frac{1}{2}} \left\{\widetilde{g}_n\left(i_n\right) - \eta_n\left(i_n\right)\widetilde{u}_n\right\} s_n^{\alpha}\eta_n(j_n)\widetilde{u}_n$$
$$\rightsquigarrow 2\int_0^1 \int_0^1 \xi(u,v)\rho(v)d\mathcal{B}_s(u)dv\,\mathcal{U} = 2(\boldsymbol{\Xi}\rho(\cdot)\mathcal{U}, d\mathcal{B}_s(\cdot)), \quad \text{and}$$

$$s_n^{2\alpha} C_n = \frac{s_n^{2\alpha}}{s_n^2} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi\left(i_n, j_n\right) \eta_n(i_n)\eta_n(j_n)\widetilde{u}_n^2 \rightsquigarrow \int_0^1 \int_0^1 \xi(u,v)\rho(u)\rho(v)dudv\,\mathcal{U}^2 = (\boldsymbol{\Xi}\rho(\cdot)\mathcal{U}, \rho(\cdot)\mathcal{U}),$$

implying that $B_n = O_{\mathbb{P}}(s_n^{-\frac{1}{2}-\alpha})$ and $C_n = O_{\mathbb{P}}(s_n^{-2\alpha})$. Thus, if $\alpha > \frac{1}{2}$, $s_n^{-1-r} q_n(\theta_*) = s_n A_n + o_{\mathbb{P}}(1) \rightsquigarrow (\boldsymbol{\Xi} d\mathcal{B}_s(\cdot), d\mathcal{B}_s(\cdot))$. This proves (i.a).

(i.b) First, $s_n^{-r-\kappa}\widetilde{d}_n = s_n^{-\kappa}[\boldsymbol{\Xi} H_n(\cdot), \widetilde{g}_n(\cdot)] + o_{\mathbb{P}}(1)$ by Assumption 2 (ii.a), and

$$\begin{aligned}
&s_n^{-\kappa}[\boldsymbol{\Xi} H_n(\cdot), \widetilde{g}_n(\cdot)] \\
=&\frac{1}{s_n^2} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi\left(i_n, j_n\right) \left\{s_n^{-\kappa} H_n(i_n) - \mu_n(i_n) + \mu_n(i_n)\right\} \left\{\widetilde{g}_n\left(j_n\right) - \eta_n\left(j_n\right)\widetilde{u}_n + \eta_n\left(j_n\right)\widetilde{u}_n\right\} \\
=&\frac{1}{s_n^2} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi\left(i_n, j_n\right) \left\{s_n^{-\kappa} H_n(i_n) - \mu_n(i_n)\right\} \left\{\widetilde{g}_n\left(j_n\right) - \eta_n\left(j_n\right)\widetilde{u}_n\right\} \\
&+ \frac{1}{s_n^2} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi\left(i_n, j_n\right) \mu_n(i_n) \left\{\widetilde{g}_n\left(j_n\right) - \eta_n\left(j_n\right)\widetilde{u}_n\right\} \\
&+ \frac{1}{s_n^2} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi\left(i_n, j_n\right) \left\{s_n^{-\kappa} H_n(i_n) - \mu_n(i_n)\right\}\eta_n\left(j_n\right)\widetilde{u}_n + \frac{1}{s_n^2} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi\left(i_n, j_n\right) \mu_n(i_n)\eta_n\left(j_n\right)\widetilde{u}_n \\
&=: A_n + B_n + C_n + D_n.
\end{aligned}$$

We examine the limit behavior of A_n, B_n, C_n, and D_n in turn. First,

$$\frac{\sqrt{n}}{s_n^{-1+\nu}} A_n = \frac{1}{s_n} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi\left(i_n, j_n\right) \frac{\sqrt{n}}{s_n^{\nu}} \{s_n^{-\kappa} H_n(i_n) - \mu_n(i_n)\} \left\{\widetilde{g}_n\left(j_n\right) - \eta_n\left(j_n\right)\widetilde{u}_n\right\} = O_{\mathbb{P}}(1),$$

by Assumptions 3 and 4, which implies that $A_n = O_{\mathbb{P}}(s_n^{-1+\nu} n^{-\frac{1}{2}}) = o_{\mathbb{P}}(1)$ from the fact that $\nu \in [0, \frac{1}{2}]$. Second,

$$s_n^{\frac{1}{2}} B_n = \frac{1}{s_n^{3/2}} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi\left(i_n, j_n\right) \mu_n(i_n) \left\{\widetilde{g}_n\left(j_n\right) - \eta_n\left(j_n\right)\widetilde{u}_n\right\}$$
$$\rightsquigarrow \int_0^1 \int_0^1 \xi(u,v)\mu(u)d\mathcal{B}_s(v) = [\boldsymbol{\Xi}\mu(\cdot), d\mathcal{B}_s(\cdot)],$$

implying that $B_n = O_{\mathbb{P}}(s_n^{-\frac{1}{2}}) = o_{\mathbb{P}}(1)$. Third, from the assumption $\eta_n(\cdot) = s_n^{-\alpha}\rho(\cdot)$,

$$\frac{\sqrt{n}}{s_n^{-\frac{1}{2}+\nu-\alpha}} C_n = \frac{1}{s_n^{3/2}} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi\left(i_n, j_n\right) \frac{\sqrt{n}}{s_n^{\nu}} \{s_n^{-\kappa} H_n(i_n) - \mu_n(i_n)\} s_n^{\alpha} \eta_n\left(j_n\right) \widetilde{u}_n$$
$$\rightsquigarrow \int_0^1 \int_0^1 \xi(u,v)\rho(v) d\mathcal{B}_a(u),$$

by Assumptions 3 and 4, implying that $C_n = O_{\mathbb{P}}(s_n^{-\frac{1}{2}+\nu-\alpha} n^{-\frac{1}{2}})$, which is $o_{\mathbb{P}}(1)$ from the fact that $\nu \in [0, \frac{1}{2}]$. Furthermore, Assumptions 3 and 4 imply that

$$s_n^{\alpha} D_n = \frac{1}{s_n^2} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi\left(i_n, j_n\right) \mu_n(i_n) s_n^{\alpha} \eta_n\left(j_n\right) \widetilde{u}_n \rightsquigarrow \int_0^1 \int_0^1 \xi(u,v)\mu(u)\rho(v) du dv \, \mathcal{U} = [\boldsymbol{\Xi}\mu(\cdot), \rho(\cdot)\mathcal{U}],$$

implying that $D_n = O_{\mathbb{P}}(s_n^{-\alpha})$. From this, if $\alpha > \frac{1}{2}$, $s_n^{\frac{1}{2}-r-\kappa} \widetilde{d}_n = s_n^{\alpha} B_n + o_{\mathbb{P}}(1) \rightsquigarrow [\boldsymbol{\Xi}\mu(\cdot), d\mathcal{B}_s(\cdot)]$. This proves (i.b).

(i.c) We also note that $s_n^{-r-2\kappa} \bar{A}_n = s_n^{-r-2\kappa} [\widehat{\boldsymbol{\Xi}}_n H_n(\cdot), H_n(\cdot)] = s_n^{-2\kappa} [\boldsymbol{\Xi} H_n(\cdot), H_n(\cdot)] + o_{\mathbb{P}}(1)$ by Assumption 2 (ii.a). Focusing on the right side, we have

$$\begin{aligned}
& s_n^{-2\kappa} [\boldsymbol{\Xi} H_n(\cdot), H_n(\cdot)] \\
&= \frac{1}{s_n^2} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi(i_n, j_n) \{s_n^{-\kappa} H_n(i_n) - \mu_n(i_n) + \mu_n(i_n)\} \{s_n^{-\kappa} H_n(j_n) - \mu_n(j_n) + \mu_n(j_n)\}' \\
&= \frac{1}{s_n^2} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi(i_n, j_n) \{s_n^{-\kappa} H_n(i_n) - \mu_n(i_n)\} \{s_n^{-\kappa} H_n(j_n) - \mu_n(j_n)\}' \\
&\quad + \frac{1}{s_n^2} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi(i_n, j_n) \mu_n(i_n) \{s_n^{-\kappa} H_n(j_n) - \mu_n(j_n)\}' \\
&\quad + \frac{1}{s_n^2} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi(i_n, j_n) \{s_n^{-\kappa} H_n(i_n) - \mu_n(i_n)\} \mu_n(j_n)' + \frac{1}{s_n^2} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi(i_n, j_n) \mu_n(i_n) \mu_n(j_n)' \\
&\quad = A_n + B_n + C_n + D_n,
\end{aligned}$$

and examine the limit behavior of each component in turn. First,

$$n s_n^{1-2\nu} A_n = \frac{1}{s_n} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi(i_n, j_n) \frac{\sqrt{n}}{s_n^{\nu}} \{s_n^{-\kappa} H_n(i_n) - \mu_n(i_n)\} \frac{\sqrt{n}}{s_n^{\nu}} \{s_n^{-\kappa} H_n(j_n) - \mu_n(j_n)\}'$$
$$\rightsquigarrow \int_0^1 \int_0^1 \xi(u,v) d\mathcal{B}_a(u) d\mathcal{B}_a(v)'$$

by Assumption 4. Therefore, $A_n = O_{\mathbb{P}}(n^{-1}s_n^{2\nu-1})$. Given that $\nu \in [0, \frac{1}{2}]$, $A_n = o_{\mathbb{P}}(1)$. Second,

$$\frac{\sqrt{n}}{s_n^{\nu-\frac{1}{2}}} B_n = \frac{2}{s_n^{3/2}} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi(i_n, j_n) \frac{\sqrt{n}}{s_n^{\nu}} \mu_n(i_n) \{s_n^{-\kappa} H_n(j_n) - \mu_n(j_n)\}' \rightsquigarrow 2 \int_0^1 \int_0^1 \xi(u, v)\mu(u) d\mathcal{B}_a(v)' du,$$

by Assumptions 3, implying that $B_n = O_{\mathbb{P}}(n^{-\frac{1}{2}} s_n^{\nu-\frac{1}{2}})$. As $\nu \in [0, \frac{1}{2}]$, $B_n = o_{\mathbb{P}}(1)$. The same proof also holds for C_n by symmetry. Finally,

$$D_n = \frac{1}{s_n^2} \sum_{i=1}^{s_n} \sum_{j=1}^{s_n} \xi(i_n, j_n)\mu_n(i_n)\mu_n(j_n)' = \int_0^1 \int_0^1 \xi(u, v)\mu_n(u)\mu_n(v)' du dv = [\boldsymbol{\Xi}\mu(\cdot), \mu(\cdot)] + o_{\mathbb{P}}(1),$$

implying that $D_n = [\boldsymbol{\Xi}\mu(\cdot), \mu(\cdot)] + o_{\mathbb{P}}(1)$. It follows that $s_n^{-r-2\kappa} \bar{A}_n = s_n^{-2\kappa}[\boldsymbol{\Xi} H_n(\cdot), H_n(\cdot)] + o_{\mathbb{P}}(1) = [\boldsymbol{\Xi}\mu(\cdot), \mu(\cdot)] + o_{\mathbb{P}}(1)$. This proves (i.c).

(ii.a) Note that $s_n^{-1-r} q_n(\theta_*) = (\boldsymbol{\Xi}\widetilde{g}_n(\cdot), \widetilde{g}_n(\cdot)) + o_{\mathbb{P}}(1)$ by Assumption 2 (ii.a) and the fact that $\xi(i_n, j_n) = \mathbb{I}(i_n = j_n)$. Next,

$$\begin{aligned}
(\boldsymbol{\Xi}\widetilde{g}_n(\cdot), \widetilde{g}_n(\cdot)) &= \frac{1}{s_n} \sum_{i=1}^{s_n} \{\widetilde{g}_n(i_n) - \eta_n(i_n)\widetilde{u}_n + \eta_n(i_n)\widetilde{u}_n\}^2 \\
&= \frac{1}{s_n} \sum_{i=1}^{s_n} \{\widetilde{g}_n(i_n) - \eta_n(i_n)\widetilde{u}_n\}^2 + \frac{2}{s_n} \sum_{i=1}^{s_n} \eta_n(i_n) \{\widetilde{g}_n(i_n) - \eta_n(i_n)\widetilde{u}_n\} \widetilde{u}_n + \frac{1}{s_n} \sum_{i=1}^{s_n} \eta_n^2(i_n)\widetilde{u}_n^2 \\
&= A_n + B_n + C_n,
\end{aligned}$$

and examine the limit behavior of each component in turn. First,

$$A_n = \frac{1}{s_n} \sum_{i=1}^{s_n} \{\widetilde{g}_n(i_n) - \eta_n(i_n)\widetilde{u}_n\}^2 \rightsquigarrow \int_0^1 (d\mathcal{B}_s(u))^2 = (d\mathcal{B}_s(\cdot), d\mathcal{B}_s(\cdot))$$

by Assumption 4, so that $A_n = O_{\mathbb{P}}(1)$. Second, from the fact that $\eta_n(\cdot) = s_n^{-\alpha}\rho(\cdot)$, we have

$$s_n^{\frac{1}{2}+\alpha} B_n = \frac{2}{\sqrt{s_n}} \sum_{i=1}^{s_n} s_n^{\alpha} \eta_n(i_n)\widetilde{u}_n \{\widetilde{g}_n(i_n) - \eta_n(i_n)\widetilde{u}_n\} \rightsquigarrow 2 \int_0^1 \rho(u)\mathcal{U} d\mathcal{B}_s(u) = (d\mathcal{B}_s(\cdot), \rho(\cdot)\mathcal{U}),$$

which implies that $B_n = O_{\mathbb{P}}(s_n^{-\frac{1}{2}-\alpha})$. Next, note that

$$s_n^{2\alpha} C_n = s_n^{-1+2\alpha} \sum_{i=1}^{s_n} \eta_n^2(i_n)\widetilde{u}_n^2 \rightsquigarrow \int_0^1 \rho^2(u)\mathcal{U}^2 du = (\rho(\cdot)\mathcal{U}, \rho(\cdot)\mathcal{U}),$$

so that $C_n = O_{\mathbb{P}}(s_n^{-2\alpha})$. Hence, given $\alpha > 0$, it follows that $s_n^{-1-r} q_n(\theta_*) = A_n + o_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} \sigma_s^2$ as required for (ii.a).

(ii.b) We prove (ii.b.1) and (ii.b.2) together. Note that $s_n^{-r-\kappa} \widetilde{d}_n = s_n^{-\kappa}[\boldsymbol{\Xi} H_n(\cdot), \widetilde{g}_n(\cdot)] + o_{\mathbb{P}}(1)$ by

Assumption 2 (ii.b), and since $\xi(i_n, j_n) = \mathbb{I}(i_n = j_n)$ we have

$$s_n^{-\kappa}[\boldsymbol{\Xi} H_n(\cdot), \widetilde{g}_n(\cdot)] = \frac{1}{s_n}\sum_{i=1}^{s_n}\{s_n^{-\kappa}H_n(i_n) - \mu_n(i_n) + \mu_n(i_n)\}\{\widetilde{g}_n(i_n) - \eta_n(i_n)\widetilde{u}_n + \eta_n(i_n)\widetilde{u}_n\}$$

$$= \frac{1}{s_n}\sum_{i=1}^{s_n}\{s_n^{-\kappa}H_n(i_n) - \mu_n(i_n)\}\{\widetilde{g}_n(i_n) - \eta_n(i_n)\widetilde{u}_n\} + \frac{1}{s_n}\sum_{i=1}^{s_n}\mu_n(i_n)\{\widetilde{g}_n(i_n) - \eta_n(i_n)\widetilde{u}_n\}$$

$$+ \frac{1}{s_n}\sum_{i=1}^{s_n}\{s_n^{-\kappa}H_n(i_n) - \mu_n(i_n)\}\eta_n(i_n)\widetilde{u}_n + \frac{1}{s_n}\sum_{i=1}^{s_n}\mu_n(i_n)\eta_n(i_n)\widetilde{u}_n$$

$$= A_n + B_n + C_n + D_n,$$

and examine the limit behavior of each component in turn. First, by Assumptions 3 and 4

$$\frac{\sqrt{n}}{s_n^{\nu}}A_n = \frac{1}{s_n}\sum_{i=1}^{s_n}\frac{\sqrt{n}}{s_n^{\nu}}\{s_n^{-\kappa}H_n(i_n) - \mu_n(i_n)\}\{\widetilde{g}_n(i_n) - \eta_n(i_n)\widetilde{u}_n\} \rightsquigarrow [d\mathcal{B}_a(\cdot), d\mathcal{B}_s(\cdot)],$$

which implies $A_n = O_{\mathbb{P}}(s_n^{\nu}n^{-\frac{1}{2}}) = o_{\mathbb{P}}(1)$ if $\nu \in [0, \frac{1}{2})$. Second,

$$s_n^{\frac{1}{2}}B_n = \frac{1}{\sqrt{s_n}}\sum_{i=1}^{s_n}\mu_n(i_n)\{\widetilde{g}_n(i_n) - \eta_n(i_n)\widetilde{u}_n\} \rightsquigarrow [\mu(\cdot), d\mathcal{B}_s(\cdot)],$$

which implies $B_n = O_{\mathbb{P}}(s_n^{-\frac{1}{2}}) = o_{\mathbb{P}}(1)$. Third, from the fact that $\eta_n(\cdot) = s_n^{-\alpha}\rho(\cdot)$,

$$\frac{\sqrt{n}}{s_n^{-\frac{1}{2}+\nu-\alpha}}C_n = \frac{1}{\sqrt{s_n}}\sum_{i=1}^{s_n}\frac{\sqrt{n}}{s_n^{\nu}}\{s_n^{-\kappa}H_n(i_n) - \mu_n(i_n)\}s_n^{\alpha}\eta_n(i_n)\widetilde{u}_n \rightsquigarrow [d\mathcal{B}_a(\cdot), \rho(\cdot)\mathcal{U}]$$

by Assumptions 3 and 4, implying that $C_n = O_{\mathbb{P}}(s_n^{-\frac{1}{2}+\nu-\alpha}n^{-\frac{1}{2}}) = o_{\mathbb{P}}(1)$ if $\nu \in [0, \frac{1}{2})$. In addition, Assumptions 3 and 4 imply that

$$s_n^{\alpha}D_n = \frac{1}{s_n}\sum_{i=1}^{s_n}\mu_n(i_n)s_n^{\alpha}\eta_n(i_n)\widetilde{u}_n \rightsquigarrow [\mu(\cdot), \rho(\cdot)\mathcal{U}],$$

so that $D_n = O_{\mathbb{P}}(s_n^{-\alpha})$. Hence, if $\alpha > \frac{1}{2}$ and $\nu = 0$, $s_n^{\frac{1}{2}-r-\kappa}\widetilde{d}_n = \sqrt{s_n}(A_n + B_n) + o_{\mathbb{P}}(1) \rightsquigarrow \tau_s + [\mu(\cdot), d\mathcal{B}_s(\cdot)]$, proving (ii.b.1). If $\alpha > \frac{1}{2}$ and $\nu \in (0, \frac{1}{2})$, $s_n^{\frac{1}{2}-r-\nu-\kappa}\widetilde{d}_n = s_n^{\frac{1}{2}-\nu}A_n + o_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} \tau_s := \sqrt{s_*}[d\mathcal{B}_a(\cdot), d\mathcal{B}_s(\cdot)]$. When $\nu = \frac{1}{2}$, $A_n = O_{\mathbb{P}}(1)$, $B_n = O_{\mathbb{P}}(s_n^{-\frac{1}{2}})$, $C_n = O_{\mathbb{P}}(s_n^{-\alpha}n^{-\frac{1}{2}})$, and $D_n = O_{\mathbb{P}}(s_n^{-\alpha})$. It follows from $\alpha > \frac{1}{2}$ that $s_n^{-r-\kappa}\widetilde{d}_n = A_n + o_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} \tau_s$. Therefore, if $\nu \in (0, \frac{1}{2}]$, $s_n^{\frac{1}{2}-r-\nu-\kappa}\widetilde{d}_n \xrightarrow{\mathbb{P}} \tau_s$. This proves (ii.b.2).

(ii.c) We prove (ii.c.1) and (ii.c.2) together. Note that $s_n^{-r-2\kappa}\bar{A}_n = s_n^{-2\kappa}[\boldsymbol{\Xi} H_n(\cdot), H_n(\cdot)] + o_{\mathbb{P}}(1)$ by

Assumption 2 (ii.b) and focus on the right side. We have

$$
\begin{aligned}
s_n^{-2\kappa}[\boldsymbol{\Xi} H_n(\cdot), H_n(\cdot)] &= \frac{1}{s_n}\sum_{i=1}^{s_n}\{s_n^{-\kappa}H_n(i_n)-\mu_n(i_n)+\mu_n(i_n)\}\{s_n^{-\kappa}H_n(i_n)-\mu_n(i_n)+\mu_n(i_n)\}' \\
&= \frac{1}{s_n}\sum_{i=1}^{s_n}\{s_n^{-\kappa}H_n(i_n)-\mu_n(i_n)\}\{s_n^{-\kappa}H_n(i_n)-\mu_n(i_n)\}' + \frac{1}{s_n}\sum_{i=1}^{s_n}\mu_n(i_n)\{s_n^{-\kappa}H_n(i_n)-\mu_n(i_n)\}' \\
&\quad + \frac{1}{s_n}\sum_{i=1}^{s_n}\{s_n^{-\kappa}H_n(i_n)-\mu_n(i_n)\}\mu_n(i_n)' + \frac{1}{s_n}\sum_{i=1}^{s_n}\mu_n(i_n)\mu_n(i_n)' \\
&= A_n + B_n + C_n + D_n,
\end{aligned}
$$

and examine the limit behavior of each component in turn. First,

$$
\frac{n}{s_n^{2\nu}}A_n = \frac{1}{s_n}\sum_{i=1}^{s_n}\frac{\sqrt{n}}{s_n^{\nu}}\{s_n^{-\kappa}H_n(i_n)-\mu_n(i_n)\}\frac{\sqrt{n}}{s_n^{\nu}}\{s_n^{-\kappa}H_n(i_n)-\mu_n(i_n)\}' \rightsquigarrow [d\mathcal{B}_a(\cdot), d\mathcal{B}_a(\cdot)],
$$

by Assumption 4, so that $A_n = O_{\mathbb{P}}(n^{-1}s_n^{2\nu})$. If $\nu \in [0, \frac{1}{2})$, $A_n = o_{\mathbb{P}}(1)$; and if $\nu = \frac{1}{2}$, $A_n = O_{\mathbb{P}}(1)$. Second,

$$
\frac{\sqrt{n}}{s_n^{-\frac{1}{2}+\nu}}B_n = \frac{1}{\sqrt{s_n}}\sum_{i=1}^{s_n}\mu_n(j_n)\frac{\sqrt{n}}{s_n^{\nu}}\{s_n^{-\kappa}H_n(i_n)-\mu_n(i_n)\}' \rightsquigarrow \int_0^1 \mu(v)d\mathcal{B}_a(u)',
$$

by Assumptions 3, implying that $B_n = O_{\mathbb{P}}(n^{-\frac{1}{2}}s_n^{-\frac{1}{2}+\nu}) = o_{\mathbb{P}}(1)$, if $\nu \in [0, \frac{1}{2})$. The same proof holds for C_n by symmetry. Finally,

$$
D_n = \frac{1}{s_n}\sum_{i=1}^{s_n}\mu_n(i_n)\mu_n(i_n)' = \int_0^1 \mu_n(u)\mu_n(u)'du = [\mu(\cdot), \mu(\cdot)] + o_{\mathbb{P}}(1).
$$

Therefore, if $\nu \in [0, \frac{1}{2})$, it follows that $s_n^{-r-2\kappa}\bar{A}_n = s_n^{-2\kappa}[\boldsymbol{\Xi} H_n(\cdot), H_n(\cdot)] + o_{\mathbb{P}}(1) = [\mu(\cdot), \mu(\cdot)] + o_{\mathbb{P}}(1)$. This proves (ii..c.1). If $\nu = \frac{1}{2}$, $A_n = s_*[d\mathcal{B}_a(\cdot), d\mathcal{B}_s(\cdot)] + o_{\mathbb{P}}(1) = \Upsilon_s + o_{\mathbb{P}}(1)$ by noting that $s_n/n = s_* + o(1)$; $B_n = o_{\mathbb{P}}(1)$; $C_n = o_{\mathbb{P}}(1)$; and $D_n = [\mu(\cdot), \mu(\cdot)] + o_{\mathbb{P}}(1)$, so that $s_n^{-r-2\kappa}\bar{A}_n = A_n + D_n + o_{\mathbb{P}}(1) = \Upsilon_s + [\mu(\cdot), \mu(\cdot)] + o_{\mathbb{P}}(1)$, proving (ii.c.2). This completes the proof. ■

Proof of Theorem 1: (i) Using Lemma 1 (i.b and i.c), we have $s_n^{\frac{1}{2}+\kappa}n^{\frac{1}{2}}(\widehat{\theta}_n - \theta_*) \rightsquigarrow -A_r^{-1}[\boldsymbol{\Xi}\mu(\cdot), d\mathcal{B}_s(\cdot))]$. Here, $[\boldsymbol{\Xi}\mu(\cdot), d\mathcal{B}_s(\cdot)]$ is an integral transform of the increments of Brownian motion and is therefore normal. Using Itô isometry, it follows that $\mathbb{E}\,[[\boldsymbol{\Xi}\mu(\cdot), d\mathcal{B}_s(\cdot)][\boldsymbol{\Xi}\mu(\cdot), d\mathcal{B}_s(\cdot)]'] = \sigma_s^2\mathbb{E}[[\boldsymbol{\Xi}\mu(\cdot), \boldsymbol{\Xi}\mu(\cdot)]] =: B_r$, and so $[\boldsymbol{\Xi}\mu(\cdot), d\mathcal{B}_s(\cdot)] \sim \mathcal{N}(0, B_r)$. Therefore, $s_n^{\frac{1}{2}+\kappa}n^{\frac{1}{2}}(\widehat{\theta}_n - \theta_*) \rightsquigarrow \mathcal{N}(0, A_r^{-1}B_rA_r^{-1})$.

(ii.a) Using Lemmas 1 (ii.b.1) and (ii.c.1), note that $s_n^{\frac{1}{2}+\kappa}n^{\frac{1}{2}}(\widehat{\theta}_n - \theta_*) \rightsquigarrow -A_e^{-1}\{\tau_s + [\mu(\cdot), d\mathcal{B}_s(\cdot)]\} = \psi_s - A_e^{-1}[\mu(\cdot), d\mathcal{B}_s(\cdot)]$. Here, $[\mu(\cdot), d\mathcal{B}_s(\cdot)] \sim \mathcal{N}(0, \sigma_s^2A_e)$, so that $s_n^{\frac{1}{2}+\kappa}n^{\frac{1}{2}}(\widehat{\theta}_n - \theta_*) \rightsquigarrow \mathcal{N}(\psi_s, \sigma_s^2A_e^{-1})$.

(ii.b) Using Lemmas 1 (ii.b.2) and (ii.c.1) shows that $s_n^{\frac{1}{2}-\nu+\kappa}n^{\frac{1}{2}}(\widehat{\theta}_n - \theta_*) \xrightarrow{\mathbb{P}} \psi_s$.

(ii.c) Using Lemmas 1 (ii.b.2) and (ii.c.2) gives $s_n^{\kappa}n^{\frac{1}{2}}(\widehat{\theta}_n - \theta_*) \xrightarrow{\mathbb{P}} -A_f^{-1}\tau_s = \psi_f$, which completes the proof. ∎

Proof of Theorem 2: Apply a second-order Taylor expansion of $q_n(\cdot)$ around θ_* giving

$$q_n(\widehat{\theta}_n) = q_n(\theta_*) - \frac{1}{2}\sqrt{n}(\widehat{\theta}_n - \theta_*)'\nabla_\theta \bar{G}_n(\theta_*)\widehat{\Sigma}_n^{-1}\nabla_\theta' \bar{G}_n(\theta_*)\sqrt{n}(\widehat{\theta}_n - \theta_*) + o_{\mathbb{P}}(\|\widehat{\theta}_n - \theta_*\|^2). \tag{A.1}$$

Using (A.1), we have

$$s_n^{-1-r} q_n(\widehat{\theta}_n) = s_n^{-1-r} q_n(\theta_*) - \frac{1}{2} s_n^{\frac{1}{2}+\kappa}\sqrt{n}(\widehat{\theta}_n - \theta_*)' s_n^{-r-2\kappa}\bar{A}_n s_n^{\frac{1}{2}+\kappa}\sqrt{n}(\widehat{\theta}_n - \theta_*) + o_{\mathbb{P}}(1),$$

and Lemma 1 (i.a), Theorem 1, and Lemma 1 (i.c) imply that $s_n^{-1-r}q_n(\theta_*) \rightsquigarrow (\Xi d\mathcal{B}_s(\cdot), d\mathcal{B}_s(\cdot))$, $s_n^{\frac{1}{2}+\kappa}\sqrt{n}(\widehat{\theta}_n - \theta_*) \rightsquigarrow -A_r^{-1}[\Xi\mu(\cdot), d\mathcal{B}_s(\cdot)]$ and $s_n^{-r-2\kappa}\bar{A}_n \overset{\mathbb{P}}{\to} A_r$. Hence, $s_n^{-1-r}q_n(\widehat{\theta}_n) \rightsquigarrow (\Xi d\mathcal{B}_s(\cdot), d\mathcal{B}_s(\cdot)) - \frac{1}{2}[\Xi\mu(\cdot), d\mathcal{B}_s(\cdot)]'A_r^{-1}[\Xi\mu(\cdot), d\mathcal{B}_s(\cdot)]$, which is identical to $(\mathbf{\Pi}_r d\mathcal{B}_s(\cdot), d\mathcal{B}_s(\cdot))$. This completes the proof. ∎

Proof of Corollary 1: The proof is straightforward from Theorem 1 (i) and omitted. ∎

Proof of Corollary 2: (i) Given the conditions, we show that $Q = A_r^{-1}B_rA_r^{-1} - \sigma_s^2 A_e^{-1}$ is positive semi-definite. Note that $Q = A_r^{-1}(B_r - \sigma_s^2 A_r A_e^{-1}A_r)A_r^{-1}$ and first show that $B_r - \sigma_s^2 A_r A_e^{-1} A_r$ is positive semi-definite. By the definitions of B_r, A_r, and A_e,

$$\begin{aligned} B_r - \sigma_s^2 A_r A_e^{-1} A_r &= \sigma_s^2\left([\Xi\mu(\cdot), \Xi\mu(\cdot)] - [\Xi\mu(\cdot), \mu(\cdot)][\mu(\cdot), \mu(\cdot)]^{-1}[\mu(\cdot), \Xi\mu(\cdot)]\right) \\ &= \sigma_s^2\left([\lambda(\cdot), \lambda(\cdot)] - [\lambda(\cdot), \mu(\cdot)][\mu(\cdot), \mu(\cdot)]^{-1}[\mu(\cdot), \lambda(\cdot)]\right) = \sigma_s^2[\mathbf{\Upsilon}\lambda(\cdot), \lambda(\cdot)], \end{aligned}$$

where $\lambda(\cdot) := \Xi\mu(\cdot)$ and $\mathbf{\Upsilon}$ is the integral operator with kernel $\upsilon(\cdot, \circ) := \delta(\cdot, \circ) - \mu(\cdot)'[\mu(\cdot), \mu(\cdot)]^{-1}\mu(\circ)$. The last equality holds by noting that

$$\begin{aligned} \mathbf{\Upsilon}\lambda(\cdot) &= \int_0^1 \upsilon(\cdot, u)\lambda(u)du = \int_0^1 (\delta(\cdot, u) - \mu(\cdot)'[\mu(\cdot), \mu(\cdot)]^{-1}\mu(u))\lambda(u)du \\ &= \int_0^1 \delta(\cdot, u)\lambda(u)du - \int_0^1 \lambda(u)\mu(u)'du[\mu(\cdot), \mu(\cdot)]^{-1}\mu(\cdot) = \lambda(\cdot) - [\lambda(\cdot), \mu(\cdot)][\mu(\cdot), \mu(\cdot)]^{-1}\mu(\cdot), \end{aligned}$$

so that $[\mathbf{\Upsilon}\lambda(\cdot), \lambda(\cdot)] = [\lambda(\cdot), \lambda(\cdot)] - [\lambda(\cdot), \mu(\cdot)][\mu(\cdot), \mu(\cdot)]^{-1}[\mu(\cdot), \lambda(\cdot)]$. We also note that

$$\begin{aligned} [\mathbf{\Upsilon}\lambda(\cdot), \lambda(\cdot)] =& [\lambda(\cdot), \lambda(\cdot)] - [\lambda(\cdot), \mu(\cdot)][\mu(\cdot), \mu(\cdot)]^{-1}[\mu(\cdot), \lambda(\cdot)] \\ & - [\lambda(\cdot), \mu(\cdot)][\mu(\cdot), \mu(\cdot)]^{-1}[\mu(\cdot), \lambda(\cdot)] \\ & + [\lambda(\cdot), \mu(\cdot)][\mu(\cdot), \mu(\cdot)]^{-1}[\mu(\cdot), \mu(\cdot)][\mu(\cdot), \mu(\cdot)]^{-1}[\lambda(\cdot), \mu(\cdot)] = [\mathbf{\Upsilon}\lambda(\cdot), \mathbf{\Upsilon}\lambda(\cdot)] \end{aligned}$$

since $A_e := [\mu(\cdot), \mu(\cdot)]$ is positive definite. Further, $[\mathbf{\Upsilon}\lambda(\cdot), \mathbf{\Upsilon}\lambda(\cdot)]$ is positive semi-definite, so that

$[\mathbf{\Upsilon}\lambda(\cdot), \mathbf{\Upsilon}\lambda(\cdot)] = \mathbf{T}\mathbf{T}'$ for some matrix $\mathbf{T}$ (e.g., Abadir and Magnus, 2012, p. 219). It follows that

$$Q = \sigma_s^2 A_r^{-1}[\mathbf{\Upsilon}\lambda(\cdot), \lambda(\cdot)]A_r^{-1} = \sigma_s^2 A_r^{-1}[\mathbf{\Upsilon}\lambda(\cdot), \mathbf{\Upsilon}\lambda(\cdot)]A_r^{-1} = \sigma_s^2 A_r^{-1}\mathbf{T}\mathbf{T}'A_r^{-1} = \sigma_s^2(A_r^{-1}\mathbf{T})(A_r^{-1}\mathbf{T})',$$

as A_r^{-1} is symmetric. Hence, Q is positive semi-definite, which implies that the asymptotic efficiency of the sxid-GMM is bounded below by $\sigma_s^2 A_e^{-1}$.

(ii) The proof follows by analogy. ■

Proof of Theorem 3: The proof is straightforward and omitted. ■

Proof of Corollary 3: (i, ii, and iii) The proofs follow from Theorem 3 by noting that $\Upsilon^{-1}\iota_c = (\text{tr}(\Psi) + (c-1)\text{tr}(\Lambda))^{-1}\iota_c$.

(iv) By Assumption 9 $\Omega_{\widehat{\theta}} > 0$. Then, setting

$$D := \begin{bmatrix} \Psi & \Lambda' \\ \Lambda & \Psi \end{bmatrix},$$

D is positive definite, and so for any $x \in \mathbb{R}^{2d}$, $x'Dx > 0$. Partition $x = (x_1', -x_1')'$ such that $x_1 \in \mathbb{R}^d$. Then, $x'Dx = x_1'(2\Psi - (\Lambda + \Lambda'))x_1 > 0$, so that $2\Psi - (\Lambda + \Lambda') > 0$. This completes the proof. ■

Proof of Corollary 4: The proof directly follows from Theorem 1 (ii.a, ii.b, and ii.c). ■

Proof of Theorem 4: (i) Using (12), we have

$$\begin{aligned} s_n^{\frac{1}{2}+\kappa}(\widetilde{\theta}_n - \theta_*) &= -(s_n^{-2\kappa}\widetilde{A}_n - s_n^{\vartheta-1-2\kappa}\gamma_n\widehat{\Omega}_{xx,n})^{-1}(s_n^{\frac{1}{2}-\kappa}\ddot{d}_n - s_n^{\vartheta-\frac{1}{2}-\kappa}\gamma_n(-\widehat{s}_n)) \\ &= -(s_n^{-2\kappa}\widetilde{A}_n - s_n^{-\frac{1}{2}-\kappa}\gamma_n\widehat{\Omega}_{xx,n})^{-1}(s_n^{\frac{1}{2}-\kappa}\ddot{d}_n - \gamma_n(-\widehat{s}_n)), \end{aligned}$$

letting $\vartheta = \frac{1}{2} + \kappa$. Lemma 1 (ii.c.1) implies that $s_n^{-2\kappa}\widetilde{A}_n \overset{\mathbb{P}}{\to} A_e$, and Assumption 9 (i) implies that $\gamma_n\widehat{\Omega}_{xx,n} \overset{\mathbb{P}}{\to} \gamma\Omega_{xx}$. Therefore,

$$s_n^{-2\kappa}\widetilde{A}_n - s_n^{-\frac{1}{2}-\kappa}\gamma_n\widehat{\Omega}_{xx,n} \overset{\mathbb{P}}{\to} A_e. \tag{A.2}$$

Next, Lemma 1 (ii.b.1) implies that $s_n^{\frac{1}{2}-\kappa}\ddot{d}_n \rightsquigarrow \widetilde{\tau}_s + [\mu(\cdot), d\widetilde{\mathcal{B}}_s(\cdot)]$, and $\gamma_n(-\widehat{s}_n) \overset{\mathbb{P}}{\to} \widetilde{\tau}_s$ as shown in the proof of (ii), so that

$$s_n^{\frac{1}{2}-\kappa}\ddot{d}_n - \gamma_n(-\widehat{s}_n) \rightsquigarrow \widetilde{\tau}_s + [\mu(\cdot), d\widetilde{\mathcal{B}}_s(\cdot)] - \widetilde{\tau}_s = [\mu(\cdot), d\widetilde{\mathcal{B}}_s(\cdot)]. \tag{A.3}$$

Finally, combining (A.2) and (A.3) gives $s_n^{\frac{1}{2}+\kappa}(\widetilde{\theta}_n - \theta_*) = (s_n^{-2\kappa}\widetilde{A}_n - s_n^{-\frac{1}{2}-\kappa}\gamma_n\widehat{\Omega}_{xx,n})^{-1}(s_n^{\frac{1}{2}-\kappa}\ddot{d}_n - \gamma_n(-\widehat{s}_n)) \rightsquigarrow -A_e^{-1}[\mu(\cdot), d\widetilde{\mathcal{B}}_s(\cdot)]$; and since $[\mu(\cdot), d\widetilde{\mathcal{B}}_s(\cdot)] \sim \mathcal{N}(0, \widetilde{\sigma}_s^2 A_e)$, it follows that $s_n^{\frac{1}{2}+\kappa}(\widetilde{\theta}_n - \theta_*) \rightsquigarrow \mathcal{N}(0, \widetilde{\sigma}_s^2 A_e^{-1})$. This proves (i).

(ii) Using (12), note that

$$s_n^{\frac{1}{2}+\kappa}(\widetilde{\theta}_n - \theta_*) = -(s_n^{-2\kappa}\widetilde{A}_n - s_n^{\vartheta-1-2\kappa}\gamma_n\widehat{\Omega}_{xx,n})^{-1}s_n^{\nu}(s_n^{\frac{1}{2}-\nu-\kappa}\ddot{d}_n - s_n^{\vartheta-\frac{1}{2}-\nu-\kappa}\gamma_n(-\widehat{s}_n))$$
$$= -(s_n^{-2\kappa}\widetilde{A}_n - s_n^{-\frac{1}{2}+\nu-2\kappa}\gamma_n\widehat{\Omega}_{xx,n})^{-1}s_n^{\nu}(s_n^{\frac{1}{2}-\nu-\kappa}\ddot{d}_n - \gamma_n(-\widehat{s}_n)),$$

since $\vartheta = \frac{1}{2}+\nu+\kappa$. Here, Lemma 1 (ii.c.1) implies that $s_n^{-2\kappa}\widetilde{A}_n \overset{\mathbb{P}}{\to} A_e$ because $r=0$ by Assumption 8 (ii). Further, $s_n^{-\frac{1}{2}+\nu-2\kappa} = o(1)$ as $\nu \in (0,\frac{1}{2})$, and $\gamma_n\widehat{\Omega}_{xx,n} = O_{\mathbb{P}}(1)$ by Assumption 9. Therefore, it follows that

$$s_n^{-2\kappa}\widetilde{A}_n - s_n^{-\frac{1}{2}+\nu-2\kappa}\gamma_n\widehat{\Omega}_{xx,n} \overset{\mathbb{P}}{\to} A_e. \tag{A.4}$$

Now examine the asymptotic behavior of $s_n^{\nu}(s_n^{\frac{1}{2}-\nu-\kappa}\ddot{d}_n - \gamma_n(-\widehat{s}_n))$. From the definition of $\ddot{d}_n$, it follows that

$$s_n^{\nu}(s_n^{\frac{1}{2}-\nu-\kappa}\ddot{d}_n - \gamma_n(-\widehat{s}_n)) = s_n^{\nu}(s_n^{\frac{1}{2}-\nu-\kappa}\ddot{d}_n - \widetilde{\tau}_s) - s_n^{\nu}(\gamma_n(-\widehat{s}_n) - \widetilde{\tau}_s) =: A_n + B_n,$$

and the limit behavior of the components is examined in turn. First, consider A_n and note that

$$s_n^{-\kappa}\ddot{d}_n = [s_n^{-\kappa}\widetilde{H}_n(\cdot) - \mu_n(\cdot), \widetilde{g}_n(\cdot)] + [\mu_n(\cdot), \widetilde{g}_n(\cdot)] + o_{\mathbb{P}}(1) =: C_n + D_n,$$

Now $s_n^{-\nu}n^{\frac{1}{2}}C_n = [s_n^{-\nu}n^{\frac{1}{2}}(s_n^{-\kappa}\widetilde{H}_n(\cdot) - \mu_n(\cdot)), \widetilde{g}_n(\cdot)] \overset{\mathbb{P}}{\to} \widetilde{\tau}_s/\sqrt{s_*}$, and Assumption 9 (iv) implies that

$$\sqrt{s_n}\int_0^{(\cdot)}\left(\frac{\sqrt{n}}{s_n^{\nu}}\left\{\frac{\widetilde{H}_n(u)}{s_n^{\kappa}} - \mu_n(u)\right\}\widetilde{g}_n(u) - \widetilde{\tau}_s/\sqrt{s_*}\right)du \rightsquigarrow \int_0^{(\cdot)} d\ddot{\mathcal{B}}_c(u),$$

so that $s_n^{\frac{1}{2}}(s_n^{-\nu}n^{\frac{1}{2}}C_n - \widetilde{\tau}_s/\sqrt{s_*}) \rightsquigarrow \ddot{\mathcal{B}}_c(1)$. In other words, $C_n = s_n^{\nu}n^{-\frac{1}{2}}\widetilde{\tau}_s/\sqrt{s_*} + O_{\mathbb{P}}(s_n^{-\frac{1}{2}+\nu}n^{-\frac{1}{2}})$. Next, note that $s_n^{\frac{1}{2}}[\mu_n(\cdot), \widetilde{g}_n(\cdot)] \rightsquigarrow [\mu(\cdot), d\ddot{\mathcal{B}}_s(\cdot)]$, so that $D_n = O_{\mathbb{P}}(s_n^{-\frac{1}{2}})$. Combining these two factors,

$$A_n = s_n^{\nu}(s_n^{\frac{1}{2}-\nu-\kappa}\ddot{d}_n - \widetilde{\tau}_s) = s_n^{\frac{1}{2}-\nu}(C_n - s_n^{\nu}n^{-\frac{1}{2}}\widetilde{\tau}_s/\sqrt{s_*}) + s_n^{\frac{1}{2}}D_n = s_n^{\frac{1}{2}}D_n + o_{\mathbb{P}}(1), \tag{A.5}$$

using the assumption that $s_n/n = s_* + o(1)$ and $\nu \in (0,\frac{1}{2})$.

Next consider B_n. By Assumption 9 (i) we have $\widetilde{\tau}_s = -\gamma\sigma_{uv}$ and so

$$\gamma_n(-\widehat{s}_n) - \widetilde{\tau}_s = (\gamma_n - \gamma)((-\widehat{s}_n) - (-\sigma_{uv})) + (\gamma_n - \gamma)(-\sigma_{uv}) + \gamma((-\widehat{s}_n) - (-\sigma_{uv}))$$
$$= \gamma((-\widehat{s}_n) - (-\sigma_{uv})) + o_{\mathbb{P}}(n^{-\frac{1}{2}}).$$

Further,

$$s_n^{\frac{1}{2}}((-\widehat{s}_n) - (-\sigma_{uv})) = -(s_n/n)^{\frac{1}{2}}\phi_n'\widehat{w}_n - s_n^{\frac{1}{2}}(n^{-1}V'U - \sigma_{uv})$$

using that fact that $\widehat{s}_n = n^{-1}X'u$ and $X'U = n^{-\frac{1}{2}}\phi_n'Z'U + V'u$ by Assumption 8. Here, note that

$(s_n/n)^{\frac{1}{2}}\phi_n'\widehat{w}_n = (s_n/n)s_n^{\frac{1}{2}}[\phi_n(\cdot),\widetilde{g}_n(\cdot)] \rightsquigarrow s_*[\phi(\cdot),d\ddot{\mathcal{B}}_s(\cdot)]$ and $s_n^{\frac{1}{2}}(n^{-1}V'U-\sigma_{uv}) = (s_n/n)^{\frac{1}{2}}\sqrt{n}(n^{-1}V'U-\sigma_{uv}) \rightsquigarrow \sqrt{s_*}\mathcal{K}$. That is,

$$s_n^{\frac{1}{2}}((-\widehat{s}_n)-(-\sigma_{uv})) \rightsquigarrow -s_*[\phi(\cdot),d\ddot{\mathcal{B}}_s(\cdot)]-\sqrt{s_*}\mathcal{K}, \tag{A.6}$$

implying that $((-\widehat{s}_n)-(-\sigma_{uv})) = O_{\mathbb{P}}(s_n^{-\frac{1}{2}})$, so that $B_n = -s_n^{\nu}(\gamma_n(-\widehat{s}_n)-\gamma(-\sigma_{uv})) = O_{\mathbb{P}}(s_n^{\nu-\frac{1}{2}}) = o_{\mathbb{P}}(1)$ because $\nu\in(0,\frac{1}{2})$. It follows that

$$s_n^{\nu}(s_n^{\frac{1}{2}-\nu-\kappa}\ddot{d}_n-\gamma_n\widehat{s}_n) = s_n^{\nu}(s_n^{\frac{1}{2}-\nu-\kappa}\ddot{d}_n-\widetilde{\tau}_s)-s_n^{\nu}(\gamma_n(-\widehat{s}_n)-\widetilde{\tau}_s) = s_n^{\frac{1}{2}}D_n+o_{\mathbb{P}}(1) \rightsquigarrow [\mu(\cdot),d\ddot{\mathcal{B}}_s(\cdot)]. \tag{A.7}$$

Finally, combining (A.4) and (A.7) gives $s_n^{\frac{1}{2}+\kappa}(\widetilde{\theta}_n-\theta_*) \rightsquigarrow -A_e^{-1}[\mu(\cdot),d\ddot{\mathcal{B}}_s(\cdot)]$; and since $[\mu(\cdot),d\ddot{\mathcal{B}}_s(\cdot)] \sim \mathcal{N}(0,\ddot{\sigma}_s^2A_e)$, it follows that $s_n^{\frac{1}{2}+\kappa}(\widetilde{\theta}_n-\theta_*) \rightsquigarrow \mathcal{N}(0,\ddot{\sigma}_s^2A_e^{-1})$, giving the desired result.

(iii) We prove (iii.a) and (iii.b) together. Using (12), note that

$$\begin{aligned} s_n^{\frac{1}{2}+\kappa}(\widetilde{\theta}_n-\theta_*) &= (s_n^{-2\kappa}\widetilde{A}_n - s_n^{\vartheta-1-2\kappa}\gamma_n\widehat{\Omega}_{xx,n})^{-1}(s_n^{\frac{1}{2}-\kappa}\ddot{d}_n - s_n^{\vartheta-\frac{1}{2}-\kappa}\gamma_n(-\widehat{s}_n)) \\ &= (s_n^{-2\kappa}\widetilde{A}_n - s_n^{-\kappa}\gamma_n\widehat{\Omega}_{xx,n})^{-1}(s_n^{\frac{1}{2}-\kappa}\ddot{d}_n - s_n^{\frac{1}{2}}\gamma_n(-\widehat{s}_n)) \end{aligned}$$

letting $\vartheta = 1+\kappa$. First, Lemma 1 (ii.c.2) implies that $s_n^{-2\kappa}\widetilde{A}_n \xrightarrow{\mathbb{P}} \widetilde{A}_f$, and Assumption 9 (i) implies that $\gamma_n\widehat{\Omega}_{xx,n} \xrightarrow{\mathbb{P}} \gamma\Omega_{xx}$. Therefore, if $\kappa=0$,

$$s_n^{-2\kappa}\widetilde{A}_n - s_n^{-\kappa}\gamma_n\widehat{\Omega}_{xx,n} \xrightarrow{\mathbb{P}} \widetilde{A}_f-\gamma\Omega_{xx} =: \ddot{A}_g. \tag{A.8}$$

If $\kappa>0$,

$$s_n^{-2\kappa}\widetilde{A}_n - s_n^{-\kappa}\gamma_n\widehat{\Omega}_{xx,n} \xrightarrow{\mathbb{P}} \widetilde{A}_f. \tag{A.9}$$

Second, note that

$$s_n^{\frac{1}{2}-\kappa}\ddot{d}_n - s_n^{\frac{1}{2}}\gamma_n(-\widehat{s}_n) = s_n^{\frac{1}{2}}(s_n^{-\kappa}\ddot{d}_n-\widetilde{\tau}_s) - s_n^{\frac{1}{2}}(\gamma_n(-\widehat{s}_n)-\widetilde{\tau}_s) =: A_n+B_n.$$

Here, in (ii) asymptotic approximations of A_n and B_n were already derived without specifying the value of ν. Note from (A.5) that if $\nu=\frac{1}{2}$,

$$A_n := s_n^{\frac{1}{2}}(s_n^{-\kappa}\ddot{d}_n-\widetilde{\tau}_d) = (s_n/n)^{\frac{1}{2}}s_n^{\frac{1}{2}}(s_n^{-\frac{1}{2}}n^{\frac{1}{2}}C_n-\widetilde{\tau}_s/\sqrt{s_*}) + s_n^{\frac{1}{2}}D_n \rightsquigarrow \sqrt{s_*}\ddot{\mathcal{B}}_c(1)+[\mu(\cdot),d\ddot{\mathcal{B}}_s(\cdot)].$$

Furthermore, Assumption 9 (i) implies that $\gamma_n(-\widehat{s}_n)-\widetilde{\tau}_s = \gamma((-\widehat{s}_n)-(-\sigma_{uv}))+o_{\mathbb{P}}(n^{-\frac{1}{2}})$, and (A.6) implies that $B_n := \sqrt{s_n}(\gamma_n(-\widehat{s}_n) - \widetilde{\tau}_s) \rightsquigarrow -s_*\gamma[\phi(\cdot), d\ddot{\mathcal{B}}_s(\cdot)] - \sqrt{s_*}\gamma\mathcal{K}$. Therefore,

$$\begin{aligned} s_n^{\frac{1}{2}-\kappa}\ddot{d}_n - s_n^{\frac{1}{2}}\gamma_n(-\widehat{s}_n) &\rightsquigarrow \sqrt{s_*}\ddot{\mathcal{B}}_c(1) + [\mu(\cdot), d\ddot{\mathcal{B}}_s(\cdot)] + \gamma s_*[\phi(\cdot), d\ddot{\mathcal{B}}_s(\cdot)] + \gamma\sqrt{s_*}\mathcal{K} \\ &= [\mu(\cdot) + s_*\gamma\phi(\cdot), d\ddot{\mathcal{B}}_s(\cdot)] + \sqrt{s_*}(\ddot{\mathcal{B}}_c(1) + \gamma\mathcal{K}) \\ &= [b(\cdot), d\ddot{\mathcal{B}}_s(\cdot)] + \sqrt{s_*}\ddot{\mathcal{B}}_c(1) + \sqrt{s_*}\gamma\mathcal{K} \end{aligned} \tag{A.10}$$

by the definition of $b(\cdot)$. Finally, combine (A.8), (A.9), and (A.10) to complete the proof. If $\kappa = 0$, (A.8) and (A.10) imply that $s_n^{\frac{1}{2}}(\widetilde{\theta}_n - \theta_*) \rightsquigarrow -\ddot{A}_g^{-1}\{[b(\cdot), d\ddot{\mathcal{B}}_s(\cdot)] + \sqrt{s_*}\ddot{\mathcal{B}}_c(1) + \sqrt{s_*}\gamma\mathcal{K}\}$. In addition, note that $[b(\cdot), d\ddot{\mathcal{B}}_s(\cdot)] \sim \mathcal{N}(0, \ddot{\sigma}_s^2[b(\cdot), b(\cdot)])$ using Itô isometry, $\ddot{\mathcal{B}}_c(1) \sim \mathcal{N}(0, \ddot{R}_3\ddot{\Sigma}_\ell^2\ddot{R}_3')$, and $\mathrm{cov}[[b(\cdot), d\ddot{\mathcal{B}}_s(\cdot)], \ddot{\mathcal{B}}_c(1)] = [b(\cdot), 1]\ \ddot{R}_1\ddot{\Sigma}_\ell^2\ddot{R}_3'$, because

$$\begin{aligned} \mathrm{cov}[[b(\cdot),&d\ddot{\mathcal{B}}_s(\cdot)], \ddot{\mathcal{B}}_c(1)] = \mathbb{E}[[b(\cdot), d\ddot{\mathcal{B}}_s(\cdot)]\ddot{\mathcal{B}}_c(1)] = \mathbb{E}[[b(\cdot), d\ddot{\mathcal{B}}_s(\cdot)], [1, d\ddot{\mathcal{B}}_c(\cdot)]] \\ &= \mathbb{E}\left[\int_0^1 b(u)d\ddot{\mathcal{B}}_s(u)\int_0^1 d\ddot{\mathcal{B}}_c(u)'\right] = \int_0^1 b(u)\mathbb{E}[d\ddot{\mathcal{B}}_s(u)d\ddot{\mathcal{B}}_c(u)'] = \int_0^1 b(u)du\ddot{R}_1\ddot{\Sigma}_\ell^2\ddot{R}_3'. \end{aligned}$$

Therefore, $\mathrm{cov}[[b(\cdot), d\ddot{\mathcal{B}}_s(\cdot)] + \sqrt{s_*}\ddot{\mathcal{B}}_c(1)] = \ddot{\sigma}_s^2[b(\cdot), b(\cdot)] + \sqrt{s_*}\varrho\ddot{R}_1\ddot{\Sigma}_\ell^2\ddot{R}_3' + \sqrt{s_*}\ddot{R}_3\ddot{\Sigma}_\ell^2\ddot{R}_1'\varrho' + s_*\ddot{R}_3\ddot{\Sigma}_\ell^2\ddot{R}_3'$, which is defined as $\ddot{\Sigma}_s^2$. Next, derive the covariance between $[b(\cdot), d\ddot{\mathcal{B}}_s(\cdot)] + \sqrt{s_*}\ddot{\mathcal{B}}_c(1)$ and $\sqrt{s_*}\gamma\mathcal{K}'$

$$\begin{aligned} \mathbb{E}[([b(\cdot), d\ddot{\mathcal{B}}_s(\cdot)] + \sqrt{s_*}\ddot{\mathcal{B}}_c(1))(\sqrt{s_*}\gamma\mathcal{K}')] &= \sqrt{s_*}\gamma\int_0^1 b(u)\mathbb{E}[d\ddot{\mathcal{B}}_s(u)\mathcal{K}'] + s_*\gamma\int_0^1 \mathbb{E}[d\ddot{\mathcal{B}}_c(u)\mathcal{K}'] \\ &= \sqrt{s_*}\gamma\int_0^1 b(u)du\ddot{\sigma}_{sk} + s_*\gamma\ddot{\Sigma}_{sk} = \sqrt{s_*}\gamma\varrho\ddot{\sigma}_{sk} + s_*\gamma\ddot{\Sigma}_{sk}. \end{aligned}$$

By analogy, we have $\mathbb{E}[(\sqrt{s_*}\gamma\mathcal{K})([b(\cdot), d\ddot{\mathcal{B}}_s(\cdot)] + \sqrt{s_*}\ddot{\mathcal{B}}_c(1))'] = \sqrt{s_*}\gamma\ddot{\sigma}_{sk}'\varrho' + s_*\gamma\ddot{\Sigma}_{sk}'$, and it follows that

$$[b(\cdot), d\ddot{\mathcal{B}}_s(\cdot)] + \sqrt{s_*}\ddot{\mathcal{B}}_c(1) + \sqrt{s_*}\gamma\mathcal{K} \sim \mathcal{N}(0, \ddot{\Sigma}_s^2 + \sqrt{s_*}\gamma(\varrho\ddot{\sigma}_{sk} + \ddot{\sigma}_{sk}'\varrho') + s_*\gamma(\ddot{\Sigma}_{sk} + \ddot{\Sigma}_{sk}') + s_*\gamma^2\Sigma_k^2). \tag{A.11}$$

Since the asymptotic covariance matrix is defined as H_h, we have $\sqrt{s_n}(\widetilde{\theta}_n - \theta_*) \rightsquigarrow \mathcal{N}(0, \ddot{A}_g^{-1}H_h\ddot{A}_g^{-1})$. If $\kappa > 0$, (A.9), (A.10) and (A.11) imply that

$$\sqrt{s_n}(\widetilde{\theta}_n - \theta_*) \rightsquigarrow -\widetilde{A}_f^{-1}\{[b(\cdot), d\ddot{\mathcal{B}}_s(\cdot)] + \sqrt{s_*}\ddot{\mathcal{B}}_c(1) + \sqrt{s_*}\gamma\mathcal{K}\} \sim \mathcal{N}(0, \widetilde{A}_f^{-1}H_h\widetilde{A}_f^{-1}),$$

which completes the proof. ■

Proof of Corollary 4: As indicated in the paper, Theorem 1 (ii) can be used to obtain the required result after adjusting convergence rates. So the proof is omitted. ■

Proof of Corollary 5: The proof follows directly from Corollary 1 and is therefore omitted. ■

Proof of Corollary 6: We prove (i) and (ii) together. First let $\Omega_{\ddot{\theta}} := \mathrm{acov}[(s_n^{\frac{1}{2}+\kappa}(\ddot{\theta}_n^1 - \theta_*)', \ldots, s_n^{\frac{1}{2}+\kappa}(\ddot{\theta}_n^m -$

$\theta_*)')']$ and show that $\Omega_{\ddot{\theta}}$ satisfies the conditions in Corollary 4. Note that if $p = q$,

$$\text{acov}[(s_n^{\frac{1}{2}+\kappa}(\ddot{\theta}_n^p - \theta_*)', s_n^{\frac{1}{2}+\kappa}(\ddot{\theta}_n^q - \theta_*)')'] = A_r^{-1}\widetilde{B}_r A_r^{-1} \tag{A.12}$$

by Corollary 5. On the other hand, if $p \neq q$, we have

$$\text{acov}[(s_n^{\frac{1}{2}+\kappa}(\ddot{\theta}_n^p - \theta_*)', s_n^{\frac{1}{2}+\kappa}(\ddot{\theta}_n^q - \theta_*)')'] = \sigma_q^2 A_r^{-1}\zeta_1\zeta_1' A_r^{-1}, \tag{A.13}$$

as is now shown. First note that

$$\ddot{\theta}_n^p - \theta_* = -(\widehat{P}_n^{p\prime}\widehat{\Sigma}_n^{-1}\widehat{P}_n^p)^{-1}(-\widehat{P}_n^p)'\widehat{\Sigma}_n^{-1}\widehat{s}_n^p = -[\widehat{\Xi}_n \widetilde{H}_n^p(\cdot), \widetilde{H}_n^p(\cdot)]^{-1}[\widehat{\Xi}_n \widetilde{H}_n^p(\cdot), \widetilde{g}_n^p(\cdot)] = -(\widetilde{A}_n^p)^{-1}\ddot{d}_n^p,$$

where $\widehat{s}_n^p := Z^{p\prime}u$, and $\widetilde{H}_n^p(\cdot)$ and $\widetilde{g}_n^p(\cdot)$ are the functions defined using the permuted IVs. In addition, let $\widetilde{A}_n^p := [\widehat{\Xi}_n \widetilde{H}_n^p(\cdot), \widetilde{H}_n^p(\cdot)]$ and $\ddot{d}_n^p := [\widehat{\Xi}_n \widetilde{H}_n^p(\cdot), \widetilde{g}_n^p(\cdot)]$. Applying Lemma 1 (i.c) and (i.b) we have $s_n^{-2\kappa}\widetilde{A}_n^p \overset{\mathbb{P}}{\to} A_r$ and $s_n^{\frac{1}{2}-\kappa}\ddot{d}_n^p \rightsquigarrow [\Xi\mu(\cdot), d\widetilde{\mathcal{B}}_s^p(\cdot)]$, so that $s_n^{\frac{1}{2}+\kappa}(\ddot{\theta}_n^p - \theta_*) \rightsquigarrow -A_r^{-1}[\Xi\mu(\cdot), d\widetilde{\mathcal{B}}_s^p(\cdot)] \sim \mathcal{N}[0, A_r^{-1}\widetilde{B}_r A_r^{-1}]$, where $\widetilde{\mathcal{B}}_s^p(\cdot)$ is the weak limit of $\int_0^{(\cdot)} \widetilde{g}_n^p(u)du$. Likewise,

$$s_n^{\frac{1}{2}+\kappa}(\ddot{\theta}_n^q - \theta_*) = -(s_n^{-2\kappa}\widetilde{A}_n^q)^{-1}s_n^{\frac{1}{2}-\kappa}\ddot{d}_n^q \rightsquigarrow -A_r^{-1}[\Xi\mu(\cdot), d\widetilde{\mathcal{B}}_s^q(\cdot)] \sim \mathcal{N}[0, A_r^{-1}\widetilde{B}_r A_r^{-1}],$$

where $\widetilde{\mathcal{B}}_s^q(\cdot)$ is the weak limit of $\int_0^{(\cdot)} \widetilde{g}_n^q(u)du$. This implies

$$\text{acov}[(s_n^{\frac{1}{2}+\kappa}(\ddot{\theta}_n^p - \theta_*)', s_n^{\frac{1}{2}+\kappa}(\ddot{\theta}_n^q - \theta_*)')'] = A_r^{-1}\mathbb{E}[[\Xi\mu(\cdot), d\widetilde{\mathcal{B}}_s^p(\cdot)][\Xi\mu(\cdot), d\widetilde{\mathcal{B}}_s^q(\cdot)]']A_r^{-1}, \quad \text{and}$$

$$\mathbb{E}[[\Xi\mu(\cdot), d\widetilde{\mathcal{B}}_s^p(\cdot)][\Xi\mu(\cdot), d\widetilde{\mathcal{B}}_s^q(\cdot)]'] = \int_0^1 \cdots \int_0^1 \xi(u,v)\mu(u)\mu(u')'\xi(u',v')\mathbb{E}[d\widetilde{\mathcal{B}}_s^p(v)d\widetilde{\mathcal{B}}_s^q(v')]dudu'.$$

We derive the last equation by focusing on $\mathbb{E}[d\widetilde{\mathcal{B}}_s^p(v)d\widetilde{\mathcal{B}}_s^q(v')]$. First note that if we let $z_j^p := (z_{1,j}^p, z_{2,j}^p, \ldots, z_{n,j}^p)$ and $z_i^q := (z_{1,i}^q, z_{2,i}^q, \ldots, z_{n,i}^q)$, then for each v and $v' \in (0,1]$,

$$\begin{aligned}\text{cov}\left(\frac{1}{\sqrt{s_n n}}\sum_{j=1}^{\lfloor v s_n \rfloor} z_j^p U, \frac{1}{\sqrt{s_n n}}\sum_{i=1}^{\lfloor v' s_n \rfloor} z_i^q U\right) &= \frac{1}{s_n n}\sum_{t=1}^n\sum_{\tau=1}^n \mathbb{E}\left[u_t u_\tau \sum_{j=1}^{\lfloor v s_n \rfloor}\sum_{i=1}^{\lfloor v' s_n \rfloor} z_{t,j}^p z_{\tau,i}^q\right] \\ &= \frac{1}{s_n n}\sum_{t=1}^n \sum_{j=1}^{\lfloor v s_n \rfloor}\sum_{i=1}^{\lfloor v' s_n \rfloor} \mathbb{E}[u_t^2 z_{t,j}^p z_{t,i}^q] = vv' s_n \mathbb{E}[u_t^2 z_{t,j}^p z_{t,i}^q],\end{aligned}$$

since $\{(u_t, z_t)\}$ is an iid sequence over t. Further, from the random sampling procedure in Assumption 11 (ii), $z_{t,j}^p = z_{t,i}^q$ with probability $1/s_n$; and $z_{t,j}^p \neq z_{t,i}^q$ with probability $1 - 1/s_n$, so that $\mathbb{E}[u_t^2 z_{t,j}^p z_{t,i}^q] = s_n^{-1}\mathbb{E}[u_t^2 z_{t,j}^2] = s_n^{-1}\sigma_q^2$ by applying Assumptions 7 (ii) and 11 (i) and the definition of σ_q^2. Hence,

$$\text{cov}\left(\frac{1}{\sqrt{s_n n}}\sum_{j=1}^{\lfloor v s_n \rfloor} Z_j^p U, \frac{1}{\sqrt{s_n n}}\sum_{i=1}^{\lfloor v' s_n \rfloor} Z_i^q U\right) = vv'\sigma_q^2.$$

It follows that $\mathbb{E}[d\widetilde{\mathcal{B}}_s^p(v)d\widetilde{\mathcal{B}}_s^q(v')] = \sigma_q^2 dvdv'$ by noting that $(s_n n)^{-\frac{1}{2}}\sum_{j=1}^{\lfloor(\cdot)s_n\rfloor} Z_j^p U = \sqrt{s_n}\int_0^{(\cdot)} \widetilde{g}_n^p(v)dv$ and $(s_n n)^{-\frac{1}{2}}\sum_{j=1}^{\lfloor(\cdot)s_n\rfloor} Z_j^q U = \sqrt{s_n}\int_0^{(\cdot)} \widetilde{g}_n^q(v')dv'$. We now combine these results to show that

$$\begin{aligned}\mathbb{E}[[\boldsymbol{\Xi}\mu(\cdot), d\widetilde{\mathcal{B}}_s^p(\cdot)][\boldsymbol{\Xi}\mu(\cdot), d\widetilde{\mathcal{B}}_s^q(\cdot)]'] &= \sigma_q^2 \int_0^1\int_0^1\int_0^1\int_0^1 \xi(u,v)\mu(u)\mu(u')'\xi(u',v')dudu'dvdv' \\ &= \sigma_q^2 \int_0^1\int_0^1 \xi(u,v)\mu(u)dudv \int_0^1\int_0^1 \xi(u',v')\mu(u')'du'dv' \\ &= \sigma_q^2[\boldsymbol{\Xi}\mu(\cdot),1][\boldsymbol{\Xi}\mu(\cdot),1]' = \sigma_q^2\zeta_1\zeta_1',\end{aligned}$$

giving (A.13) as required.

Using (A.12) and (A.13), we have $\Omega_{\ddot{\theta}} = I_m\otimes A_r^{-1}\widetilde{B}_r A_r^{-1} + L_1\otimes\sigma_q^2 A_r^{-1}\zeta_1\zeta_1' A_r^{-1} + U_1\otimes\sigma_q^2 A_r^{-1}\zeta_1\zeta_1' A_r^{-1}$, which satisfies the condition in Corollary 3 by letting $\Psi = A_r^{-1}\widetilde{B}_r A_r^{-1}$ and $\Lambda = \Lambda' = \sigma_q^2 A_r^{-1}\zeta_1\zeta_1' A_r^{-1}$. Therefore,

$$\text{acov}[s_n^{\frac{1}{2}+\kappa}(\bar{\theta}_n - \theta_*)] = \frac{1}{m}A_r^{-1}\widetilde{B}_r A_r^{-1} + \left(1-\frac{1}{m}\right)\sigma_q^2 A_r^{-1}\zeta_1\zeta_1' A_r^{-1},$$

by Corollary 3 (iii). Furthermore, for each p, $s_n^{\frac{1}{2}+\kappa}(\ddot{\theta}_n^p - \theta_*)$ is asymptotically normal and centered at zero by Corollary 5, implying that $s_n^{\frac{1}{2}+\kappa}(\bar{\theta}_n^p - \theta_*)$ is asymptotically normal and centered at zero. These establish Corollary 6 (i).

Finally, Corollary 3 (iv) implies that $(2\Psi - \Lambda' - \Lambda)$ is positive definite. Under the given conditions, $\Psi = A_r^{-1}\widetilde{B}_r A_r^{-1}$ and $\Lambda = \Lambda' = \sigma_q^2 A_r^{-1}\zeta_1\zeta_1' A_r^{-1}$, so that $(2\Psi-\Lambda'-\Lambda) = 2(A_r^{-1}\widetilde{B}_r A_r^{-1} - \sigma_q^2 A_r^{-1}\zeta_1\zeta_1' A_r^{-1})$ is positive definite. This proves Corollary 6 (ii). ■

Proof of Theorem 5: (i) Given that $q_n(\theta_*) = \widetilde{G}_n(\theta_*)'\widehat{\Sigma}_n^{-1}\widetilde{G}_n(\theta_*)$, let both B_n and C_n be $\widetilde{G}_n(\theta_*)$ in (15), which then gives $q_n(\theta_*) = \sum_{h=0}^{\infty}(\widehat{\boldsymbol{\Xi}}_{n,h}\widetilde{g}_n(\cdot), \widetilde{g}_n(\cdot))$ using the definition of $\widetilde{g}_n(\cdot)$.

(ii) Since $\bar{A}_n = \nabla_\theta\bar{G}_n(\theta)\widehat{\Sigma}_n^{-1}\nabla_\theta'\bar{G}_n(\theta_*)$, let both B_n and C_n in (15) be $\nabla_\theta'\bar{G}_n(\theta_*)$, and then $\bar{A}_n = \sum_{h=0}^{\infty}(\widehat{\boldsymbol{\Xi}}_{n,h}H_n(\cdot), H_n(\cdot))$ by the definition of $H_n(\cdot)$.

(iii) Given that $\widetilde{d}_n = \nabla_\theta\bar{G}_n(\theta)\widehat{\Sigma}_n^{-1}\widetilde{G}_n(\theta_*)$, if we let B_n and C_n in (15) be $\nabla_\theta\bar{G}_n(\theta)$ and $\widetilde{G}_n(\theta_*)$, respectively, then $\widetilde{d}_n = \sum_{h=0}^{\infty}(\widehat{\boldsymbol{\Xi}}_{n,h}H_n(\cdot), \widetilde{g}_n(\cdot))$. This complete the proof. ■

A.2 Supplementary Simulations

This section reports supplementary simulation findings. Under the DGP condition in Section 6 we first examine the performance of wid-2SLS and pwid-2SLS with various continuous kernels and then assess the performance of wid-2SLS with a Dirac delta kernel.

A.2.1 wid-2SLS and pwid-2SLS with Various Kernels

This section reports simulations to assess the performance of wid-2SLS and pwid-2SLS driven by various kernels. We use the same DGP and model condition as in Section 6 of the main paper and estimate the unknown structural parameter θ_* by wid-2SLS. Six kernels are employed. They are BMK, BBK, SEK with $s^2 = 0.1, 0.01, 0.001$, and the Dirac delta function by an approximating sequence as explained below. Figure A.1 shows the functional shapes of the first five kernels. All of them are symmetric around the diagonal and the SEKs are flat over the diagonal. As s^2 decreases, the maximum value of the SEK increases and its volume is more concentrated on the diagonal. This property ensures that SEK gets closer to the Dirac delta function $\delta(\cdot, \circ)$ as $s^2 \to 0$.

The sample bias and variance of the wid-2SLS estimator driven by each kernel are computed. The number of replications is 15,000, and we observe how the sample bias and variance evolve as n increases, using the sample average and sample variance of $\sqrt{n}(\ddot{\theta}_n - \theta_*)$, for $n \in \{200, 500, 1,000, 1,500\}$. The sample bias and sample variance are reported in Tables A.1 and A.2, respectively. The asymptotic variances are given in the bottom row of Table A.2 and obtained using the analytic asymptotic variance formula given in Corollary 5.

The findings are summarized as follows.

(i) For each continuous kernel in Table A.1, the sample bias reduces as n increases. On the other hand, the sample bias increases for the Dirac delta function, and the sample biases are approximately $\sqrt{n}/3$. This corroborates Corollaries 4 (iii) and 5. Given that $\nu = \frac{1}{2}$ and $\kappa = 0$ for the current DGP, applying Corollary 4 (iii) implies that $(\ddot{\theta}_n - \theta_*) \overset{\mathbb{P}}{\to} \widetilde{\psi}_f$, which is $\frac{1}{3}$ as shown in Section A.2.2. From this, the asymptotic bias of $\sqrt{n}(\ddot{\theta}_n - \theta_*)$ is close to $\sqrt{n}/3$.

(ii) For the SEKs, the sample bias from $s^2 = 0.001$ is overall greater than the others. This behavior follows from the fact that SEK approaches the Dirac delta function as $s^2 \to 0$. Therefore, for fixed n, if s^2 is too small, the sample bias increases, as discussed in the remark below Corollary 2.

(iii) For each continuous kernel in Table A.2, the sample variance converges to its limit as n increases. The limit variance is smallest when using the SEK with $s^2 = 0.001$, and it further decreases to 1.00 if we let s^2 converge to 0. So the asymptotic variance is no smaller than 1.00. This variance can be obtained by applying the Dirac delta function to the asymptotic covariance matrix formula.

(iv) From Table A.2, the sample variance is smallest when the Dirac delta function is used for wid-2SLS estimation. But in this case the estimator suffers from bias. Also, the sample variance differs from the limit variance, so that the asymptotic variance given in Corollary 5 does not apply for the Dirac delta function kernel. Corollary 4 (iii) does not provide the asymptotic distribution of $(\ddot{\theta}_n - \theta_* - \widetilde{\psi}_f)$ for $\nu = \frac{1}{2}$. In Section A.2.2, we analytically derive the limit variance for the current DGP, and the derivation shows that the asymptotic variance is 0.50 given in (A.16), which is close to the sample variance. □

We next estimate the unknown parameter by pwid-2SLS. As for wid-2SLS, the same six kernels are used. The sample bias and sample variance of the pwid-2SLS estimator driven by each kernel are reported in Tables A.3 and A.4. The number of experiments is 5,000, and we examine how the sample bias and sample variance evolve as the sample sizes (n and m) increase. Here, the sample bias and sample variance are those of $\sqrt{n}(\bar{\theta}_n - \theta_*)$. We let $m = 1,000$ or $2,000$ and consider $\{n \in 100, 200, 500, 1,000\}$. Since the IVs are iid with respect to the instrument index, the optimal weight of the pwid-2SLS is $1/m$. In addition, we provide the corresponding limit variances in the bottom row of Table A.4. These are obtained analytically by letting $n, m \to \infty$, similar to (16).

The findings are summarized as follows.

(i) For each continuous kernel in Table A.3, the sample bias decreases as n increases. On the other hand, the sample bias increases as n increases for the Dirac delta function, and the sample biases are approximately $\sqrt{n}/3$. These features match those for wid-2SLS.

(ii) Focusing on the SEKs, the sample bias from $s^2 = 0.001$ exceeds the others. As $s^2 \to 0$, the SEK approaches the Dirac delta function. So, for fixed n as s^2 reduces, the sample bias increases. This also matches the findings for wid-2SLS driven by the SEK with $s^2 = 0.001$.

(iii) Overall, the sample bias is smallest when employing the SEK with $s^2 = 0.10$. In addition, increasing m does not modify the results substantially, implying that letting $m = 1,000$ is already too large to obtain a marginally significant improvement.

(iv) For all the continuous kernels, the limit variances are unity. This level is the smallest variance level that can be achieved by wid-2SLS. For each continuous kernel in Table A.4, the sample variance converges to its limit as n increases.

(v) Overall, the sample variance from the continuous kernels is smallest when using SEK with $s^2 = 0.001$. In addition, increasing m does not modify the sample variance substantially.

(vi) From Table A.4, the sample variance is smallest when the Dirac delta function is used. Even so, this procedure is not useful in practice due to its bias and its sample variance is quite different from the limit variance. This finding is shared by wid-2SLS driven by the Dirac delta function. □

A.2.2 wid-2SLS with the Dirac Delta Kernel

This section reports simulations to assess the performance of wid-2SLS driven by the Dirac delta kernel under the same DGP and model condition as in Section 6 of the main paper. The DGP condition in Section 6 delivers the limit distribution of the statistics relevant to the wid-2SLS. As before, for each j, we let $\widehat{P}_{n,j}$ and $\widehat{w}_{n,j}$ be the j-th row element of $\widehat{P}_n := n^{-1}Z'X$ and $\widehat{w}_n := n^{-1}Z'u$, respectively and further let $(\widetilde{P}_{n,j}, \widetilde{w}_{n,j})' := \sqrt{n}(\widehat{P}_{n,j}, \widehat{w}_{n,j})'$. Then, the following weak convergence holds

$$\widetilde{\ell}_n(\cdot) := \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor (\cdot) n \rfloor} \begin{bmatrix} \widetilde{w}_{n,j} \\ (-\widetilde{P}_{n,j} + 1) \end{bmatrix} \rightsquigarrow \int_0^{(\cdot)} \begin{bmatrix} d\widetilde{\mathcal{B}}_s(u) \\ d\widetilde{\mathcal{B}}_a(u) \end{bmatrix} := \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}^{\frac{1}{2}} \int_0^{(\cdot)} \begin{bmatrix} d\widetilde{\mathcal{W}}_s(u) \\ d\widetilde{\mathcal{W}}_a(u) \end{bmatrix}, \tag{A.14}$$

where $\widetilde{\mathcal{W}}_s(\cdot)$ and $\widetilde{\mathcal{W}}_a(\cdot)$ are two independent Wiener processes, and the coefficient matrix of the stochastic integral of $[\widetilde{\mathcal{W}}_s(\cdot), \widetilde{\mathcal{W}}_a(\cdot)]'$ in (A.14) is the square root of the limiting global covariance matrix of $\widetilde{\ell}_n(1)$. This weak convergence also implies that $\nu = \frac{1}{2}$ and $\kappa = 0$ as before. The application of the invariance principle here uses the fact that for $i \neq j$, $(-\widetilde{P}_{n,i}, \widetilde{w}_{n,i})'$ is asymptotically independent of $(-\widetilde{P}_{n,j}, \widetilde{w}_{n,j})'$.

The asymptotic bias of wid-2SLS arises from the dependence of $d\widetilde{\mathcal{B}}_s(\cdot)$ and $d\widetilde{\mathcal{B}}_a(\cdot)$. Note that $\widehat{\theta}_n = \theta_* - (\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n)^{-1}(-\widehat{P}_n)'\widehat{\Sigma}_n^{-1}\widehat{w}_n$, wherein $\widehat{\Sigma}_n$ is Bickel and Levina's (2008) regularized covariance matrix estimator obtained by setting their parameter $\boldsymbol{\alpha}$ to be proportional to n. This implies that Assumption 8 (i) holds and that $\widehat{\boldsymbol{\Xi}}_n$ asymptotically converges to an identity integral transform operator. With this setting it follows that $\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n = (\widetilde{H}_n(\cdot), \widetilde{H}_n(\cdot)) + o_{\mathbb{P}}(1) \overset{\mathbb{P}}{\to} 3$ and, similarly, $(-\widehat{P}_n)'\widehat{\Sigma}_n^{-1}\widehat{w}_n = (\widetilde{H}_n(\cdot), \widetilde{g}_n(\cdot)) + o_{\mathbb{P}}(1) \overset{\mathbb{P}}{\to} -\frac{3}{5}\int_0^1 du - \frac{2}{5}\int_0^1 du = -1$. Thus, $\widehat{\theta}_n - \theta_* \overset{\mathbb{P}}{\to} 1/3$, giving the asymptotic bias of wid-2SLS as $1/3$. This asymptotic bias corresponds to $\widetilde{\psi}_f$ in Corollary 4 (iii).

We next explore the sampling distribution of wid-2SLS. Although the limit theory was not given in Corollary 4 (iii), it can be derived here specifically for the given DGP via the limit distribution of $\sqrt{n}(\widehat{\theta}_n - \theta_* - 1/3)$. To do so, we use a second-order approximation obtained by eliminating the correlation between $\widetilde{P}_{n,i}$ and $\widetilde{w}_{n,i}$ and considering the limit distribution of $\widetilde{P}_{n,i}^2$. From the DGP conditions we have

$$\widetilde{m}_n(\cdot) := \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor (\cdot) n \rfloor} \begin{bmatrix} (\widetilde{P}_{n,j}\widetilde{w}_{n,j} - 1) \\ (\widetilde{P}_{n,j}^2 - 3) \end{bmatrix} \rightsquigarrow \int_0^{(\cdot)} \begin{bmatrix} d\widetilde{\mathcal{B}}_{pw}(u) \\ d\widetilde{\mathcal{B}}_{pp}(u) \end{bmatrix} := \begin{bmatrix} 9 & 13.5 \\ 13.5 & 40.5 \end{bmatrix}^{\frac{1}{2}} \int_0^{(\cdot)} \begin{bmatrix} d\widetilde{\mathcal{W}}_1(u) \\ d\widetilde{\mathcal{W}}_2(u) \end{bmatrix}, \tag{A.15}$$

where $(\widetilde{\mathcal{W}}_1(\cdot), \widetilde{\mathcal{W}}_2(\cdot))'$ is a vector of two independent Wiener processes. The matrix factor of the Wiener integral in (A.15) is the square root of the global covariance matrix of $\widetilde{m}_n(1)$. After some algebra, we have for each j, $\mathbb{E}[(\widetilde{P}_{n,j}\widetilde{w}_{n,j} - 1)^2] = 4 + o(1)$; and if $i \neq j$, $\mathbb{E}[(\widetilde{P}_{n,i}\widetilde{w}_{n,i} - 1)(\widetilde{P}_{n,j}\widetilde{w}_{n,j} - 1)] = 5/n + o(n^{-1})$.

This yields $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}[(\widetilde{P}_{n,i}\widetilde{w}_{n,i}-1)(\widetilde{P}_{n,j}\widetilde{w}_{n,j}-1)] = 9+o(1)$; and, similarly, the asymptotic global variance of $n^{-\frac{1}{2}}\sum_{j=1}^{n}(\widetilde{P}_{n,j}^{2}-3)$ is found to be 40.5 and the global asymptotic covariance between $n^{-\frac{1}{2}}\sum_{i=1}^{n}(\widetilde{P}_{n,j}^{2}-3)$, and $n^{-\frac{1}{2}}\sum_{i=1}^{n}(\widetilde{P}_{n,j}\widetilde{w}_{n,j}-1)$ is found to be 13.5, giving (A.15). The asymptotic distribution of wid-2SLS is obtained by combining these two properties. Thus

$$\begin{aligned}\sqrt{n}(\widehat{\theta}_n-\theta_*-1/3) &= -(\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n)^{-1}\sqrt{n}((-\widehat{P}_n)'\widehat{\Sigma}_n^{-1}\widehat{w}_n+1)+\sqrt{n}((\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n)^{-1}-1/3)\\ &\rightsquigarrow \frac{1}{3}\widetilde{\mathcal{B}}_{pw}(1)-\frac{1}{9}\widetilde{\mathcal{B}}_{pp}(1)\sim\mathcal{N}\left(0,\frac{1}{2}\right),\end{aligned} \tag{A.16}$$

using the fact that $\mathrm{cov}(\widetilde{\mathcal{B}}_{pw}(1),\widetilde{\mathcal{B}}_{pp}(1))=27/2$. This asymptotic variance 1/2 is estimated by the sample variances of the wid-2SLS and pwid-2SLS estimators that are reported in Tables A.2 and A.4.

To corroborate the limit theory $\mathcal{N}(0,\frac{1}{2})$ in (A.16) by simulation we generated independent data sets for the given DGP conditions and computed $\sqrt{n}(\widehat{\theta}_n-\theta_*-1/3)$ using the formula for wid-2SLS and Neumann's series expansion in the computation. In addition, we computed

$$\widehat{\theta}_{n,k} := \left(\sum_{h=0}^{k}(\widehat{\boldsymbol{\Xi}}_{n,h}H_n(\cdot),H_n(\cdot))\right)^{-1}\left(\sum_{h=0}^{k}(\widehat{\boldsymbol{\Xi}}_{n,h}H_n(\cdot),c_n(\cdot))\right),$$

and examined the finite sample distribution of $\sqrt{n}(\widehat{\theta}_{n,k}-\theta_*-1/3)$ by simulation for different levels of n and k, where $c_n(\cdot)$ is the càdlàg process constructed from $\widehat{c}_n := n^{-1}Z'y$.

Simulations were conducted according to the following plan: (i) observations were generated and both $\sqrt{n}(\widehat{\theta}_n-\theta_*-1/3)$ and $\sqrt{n}(\widehat{\theta}_{n,k}-\theta_*-1/3)$ were calculated for sample sizes $n\in\{50,100,200,300\}$ with $k\in\{10,50,100,200\}$; and (ii) $3,000$ replications were used to obtain the empirical distribution of $\sqrt{n}(\widehat{\theta}_{n,k}-\theta_*-1/3)$ and $12,000$ were used for the distribution of $\sqrt{n}(\widehat{\theta}_n-\theta_*-1/3)$. The results are reported in Tables A.5 and A.6. Table A.5 reports the sample averages of $\sqrt{n}(\widehat{\theta}_n-\theta_*-1/3)$ and $\sqrt{n}(\widehat{\theta}_{n,k}-\theta_*-1/3)$ for each k, and Table A.6 reports their sample variance. The findings are as follows:

(i) For each $k=10,50,100$, and 200, the finite sample biases of $\sqrt{n}(\widehat{\theta}_n-\theta_*-1/3)$ and $\sqrt{n}(\widehat{\theta}_{n,k}-\theta_*-1/3)$ are very close to zero. In addition, the sample variances are also close to $\frac{1}{2}$.

(ii) Despite the numerical differences in Table A.5, the overall differences between the distributions and the $\mathcal{N}(0,\frac{1}{2})$ limit are negligible. Figure A.2 shows the distributions of $\sqrt{n}(\widehat{\theta}_{n,k}-\theta_*-1/3)$ for each n and $\mathcal{N}(0,\frac{1}{2})$, revealing that the finite sample distributions of $\sqrt{n}(\widehat{\theta}_{n,k}-\theta_*-1/3)$ differ little from the limit theory even when k and n are moderately sized. So the distribution of $\widehat{\theta}_n$ appears to be very well approximated by that of $\widehat{\theta}_{n,k}$. □

A.3 Supplementary Empirical Results

This section provides empirical analyses supplementing those reported in Section 7 of the main paper. As Table 3 in the paper reports, estimation results are similar for pwid-2SLS estimation driven by BMK and BBK kernels. We investigate this further here by considering the sampling distributions of the pwid-2SLS estimators. As discussed in Section 7 of the paper, the moment condition orders are permuted randomly 5,000 times with replacement and each permutation used to estimate by wid-2SLS. Figure A.3 shows the histogram of the distribution based on 5,000 estimates using the BMK kernel. For each cohort, the sample average of the estimates are reported in Table 3 along with 90% and 95% ranges. Figure A.4 shows the histogram of the distribution based on the 5,000 estimates using the BBK kernel. As evident in these figures, the distributions are bell-shaped and are similar to each other.

To explore this similarity further we proceed to test whether the distributions of the pwid-2SLS estimators driven by BMK and BBK are identical. For this purpose, we apply the two-sample Kolmogorov-Smirnov (KS) test. Table A.7 reports the findings. For each cohort and for each model, we compute the two-sample KS-test along with its p-value. The outcome of the test shows that the two empirical distributions are identical in each case. Only for the basic model combined with 1940-1949 cohort data is the p-value less than 1%. The other p-values exceed 5%, meaning that the hypothesis that the two distributions are identical cannot be rejected for the 1920-1929 and 1930-1939 cohorts. The rejection for 1940-1949 cohort does not mean that the two empirical distributions differ substantially. In fact, Figure A.5 shows the empirical distributions and the distributions are evidently close in each case. This outcome corroborates the finding that BMK and BBK kernels do not yield significantly different estimates.

$n\backslash$ Kernels	BMK	BBK	SEK ($s^2 = 0.10$)	SEK ($s^2 = 0.01$)	SEK ($s^2 = 0.001$)	Dirac delta function
200	-0.0701	-0.0994	-0.0271	0.1712	0.6955	4.7104
500	-0.0282	-0.0405	-0.0174	0.1029	0.4646	7.4373
$1,000$	-0.0348	-0.0444	-0.0091	0.0590	0.3366	10.538
$1,500$	-0.0346	-0.0379	-0.0063	0.0447	0.2792	12.908

Table A.1: SAMPLE BIAS OF WID-2SLS. Each cell reports the finite sample bias of $\sqrt{n}(\ddot{\theta}_n - \theta_*)$ obtained by 15,000 independent experiments, where $\ddot{\theta}_n$ is the wid-2SLS estimator driven by BMK, BBK, SEKs with $s^2 = 0.10$, $s^2 = 0.01$, $s^2 = 0.001$, and the Dirac delta function, $\delta(\cdot - \circ)$.

$n\backslash$ Kernels	BMK	BBK	SEK ($s^2 = 0.10$)	SEK ($s^2 = 0.01$)	SEK ($s^2 = 0.001$)	Dirac delta function
200	1.2556	1.4499	1.0751	1.0032	0.8723	0.5272
500	1.2168	1.3552	1.0229	1.0057	0.9532	0.4987
$1,000$	1.2118	1.3208	1.0388	1.0299	0.9739	0.5070
$1,500$	1.1946	1.3508	1.0353	1.0219	0.9603	0.5063
$\text{acov}[\sqrt{n}(\ddot{\theta}_n - \theta_*)]$	1.2000	1.2000	1.0251	1.0200	1.0071	1.0000

Table A.2: SAMPLE VARIANCE OF WID-2SLS. Each cell reports the sample variance of $\sqrt{n}(\ddot{\theta}_n - \theta_*)$ obtained by 15,000 independent experiments, where $\ddot{\theta}_n$ is the wid-2SLS estimator driven by BMK, BBK, SEKs with $s^2 = 0.10$, $s^2 = 0.01$, $s^2 = 0.001$, and the Dirac delta function, $\delta(\cdot - \circ)$. The bottom line contains the asymptotic variance analytically obtained by letting $n \to \infty$.

m	$n\backslash$ Kernels	BMK	BBK	SEK ($s^2 = 0.10$)	SEK ($s^2 = 0.01$)	SEK ($s^2 = 0.001$)	Dirac delta function
1,000	100	-0.0954	-0.1462	-0.0386	0.2273	0.9352	3.3235
	200	-0.0635	-0.0965	-0.0245	0.1631	0.7228	4.6990
	500	-0.0242	-0.0442	0.0000	0.1188	0.5001	7.4554
	$1,000$	-0.0476	-0.0613	-0.0307	0.0535	0.3363	10.521
2,000	100	-0.0783	-0.1296	-0.0212	0.2472	0.9377	3.3559
	200	-0.0563	-0.0894	-0.0174	0.1706	0.7057	4.7118
	500	-0.0936	-0.1137	-0.0694	0.0503	0.4772	7.4359
	$1,000$	0.0110	-0.0029	0.0281	0.1119	0.3287	10.549

Table A.3: SAMPLE BIAS OF PWID-2SLS. Each cell reports the finite sample bias of $\sqrt{n}(\bar{\theta}_n - \theta_*)$ obtained by 5,000 independent experiments, where $\bar{\theta}_n$ is the pwid-2SLS estimator. Six kernels are used for the simulation: BMK, BBK, SEKs with $s^2 = 1.00$, $s^2 = 0.10$, $s^2 = 0.01$, and the Dirac delta function, $\delta(\cdot - \circ)$.

m	$n\backslash$ Kernels	BMK	BBK	SEK ($s^2 = 0.10$)	SEK ($s^2 = 0.01$)	SEK ($s^2 = 0.001$)	Dirac delta function
1,000	100	1.0983	1.1370	1.0579	0.8995	0.6577	0.5245
	200	1.0073	1.0226	0.9898	0.9119	0.7495	0.5133
	500	0.9725	0.9780	0.9660	0.9352	0.8736	0.4939
	$1,000$	1.0043	1.0067	1.0018	0.9861	0.9493	0.5022
2,000	100	1.1229	1.1623	1.0823	0.9189	0.6393	0.5004
	200	1.0195	1.0350	1.0022	0.9234	0.7586	0.5169
	500	1.0354	1.0410	1.0287	0.9965	0.8868	0.4973
	$1,000$	1.0042	1.0070	1.0008	0.9849	0.9358	0.5028
$\mathrm{acov}[\sqrt{n}(\bar{\theta}_n - \theta_*)]$		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table A.4: SAMPLE VARIANCE OF PWID-2SLS. Each cell reports $m\times$ the sample variance of $\sqrt{n}(\bar{\theta}_n - \theta_*)$ obtained by 5,000 independent experiments, where $\bar{\theta}_n$ is the pwid-2SLS estimator. Six kernels are used for the simulation: BMK, BBK, SEKs with $s^2 = 1.00$, $s^2 = 0.10$, $s^2 = 0.01$, and the Dirac delta function, $\delta(\cdot - \circ)$. The bottom line contains the asymptotic variance analytically obtained by letting n and $m \to \infty$.

$n\backslash$ Estimators	$k=10$	$k=50$	$k=100$	$k=200$	$k=\infty$
50	0.0149	−0.0174	−0.0002	0.0212	0.0124
100	−0.0276	−0.0306	0.0196	0.0286	0.0052
200	−0.0041	−0.0187	0.0117	0.0094	0.0091
300	−0.0042	−0.0130	0.0212	−0.0077	0.0031

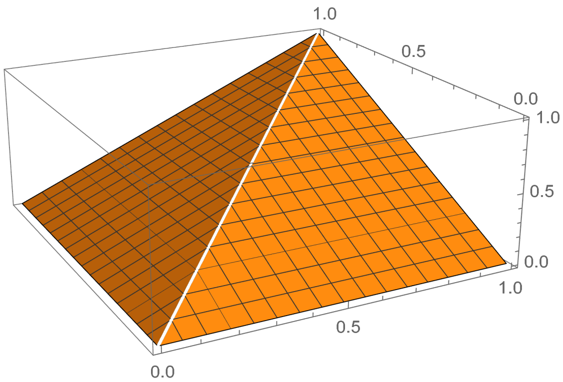
Table A.5: SAMPLE BIAS OF WID-2SLS. Each cell reports the finite sample mean of $\sqrt{n}(\widehat{\theta}_{n,k} - \theta_* - 1/3)$ that is obtained by implementing 3,000 independent experiments. If $k = \infty$, the cell reports the sample variance of $\sqrt{n}(\widehat{\theta}_n - \theta_* - 1/3)$ that is obtained by implementing 12,000 independent experiments. Here, k is the degree of Neumann's series expansion, and if $k = \infty$, the inverse matrix of $\widehat{\Sigma}_n$ is directly used as the weight matrix.

$n\backslash$ Estimators	$k=10$	$k=50$	$k=100$	$k=200$	$k=\infty$
50	0.5534	0.5191	0.5138	0.5299	0.5267
100	0.5396	0.5040	0.5196	0.5235	0.5206
200	0.5094	0.5139	0.5063	0.5438	0.5177
300	0.5089	0.4965	0.5252	0.5066	0.5088

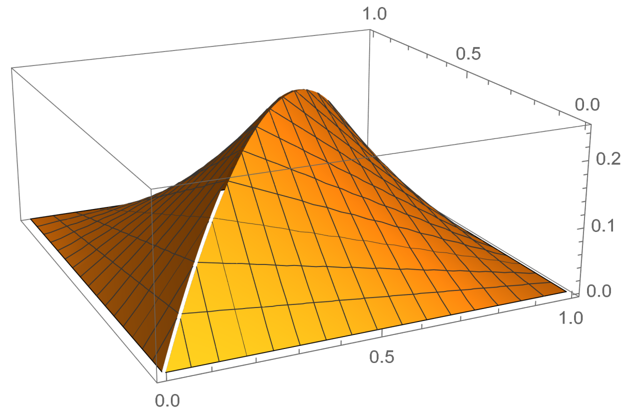
Table A.6: SAMPLE VARIANCE OF WID-2SLS. Each cell reports the sample variances of $\sqrt{n}(\widehat{\theta}_{n,k} - \theta_* - 1/3)$ that is obtained by implementing 3,000 independent experiments. If $k = \infty$, the cell reports the sample variance of $\sqrt{n}(\widehat{\theta}_n - \theta_* - 1/3)$ that is obtained by implementing 12,000 independent experiments. Here, k is the degree of Neumann's series expansion, and if $k = \infty$, the inverse matrix of $\widehat{\Sigma}_n$ is directly used as the weight matrix.

Data	1920-1929 Cohort		1930-1939 Cohort		1940-1949 Cohort	
Test $\backslash$ Models	Basic	Extended	Basic	Extended	Basic	Extended
KS-test	0.026	0.014	0.023	0.015	0.036	0.019
p-value (%)	6.700	70.84	12.18	57.38	0.240	27.82

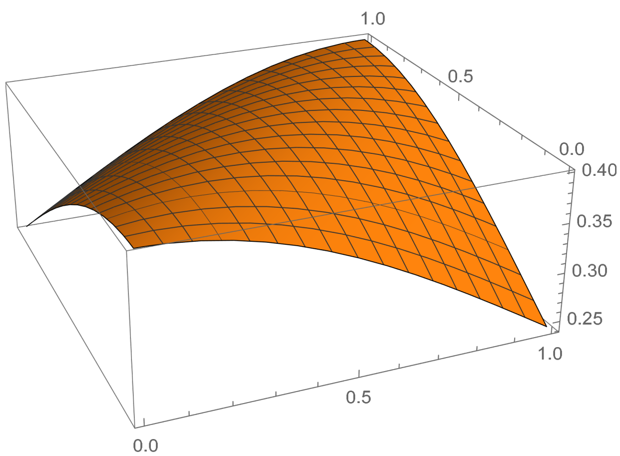
Table A.7: TWO-SAMPLE KS TESTS AND THEIR p-VALUES. This table shows the results obtained from the two-sample KS test applied to the sample distributions of the wid-2SLS estimators driven by BMK and BBK. For each cohort, we estimate the basic and extended models by wid-2SLS assuming BMK and BBK kernels. In total, the moment conditions are permuted 5,000 times with replacement to obtain the sampling distribution.



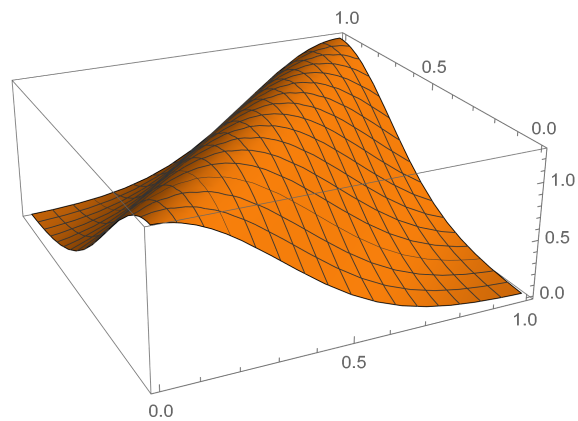
(a) Brownian motion kernel



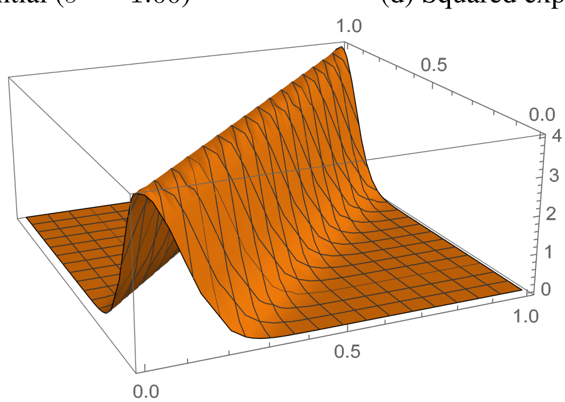
(b) Brownian bridge kernel



(c) Squared exponential ($s^2 = 1.00$)

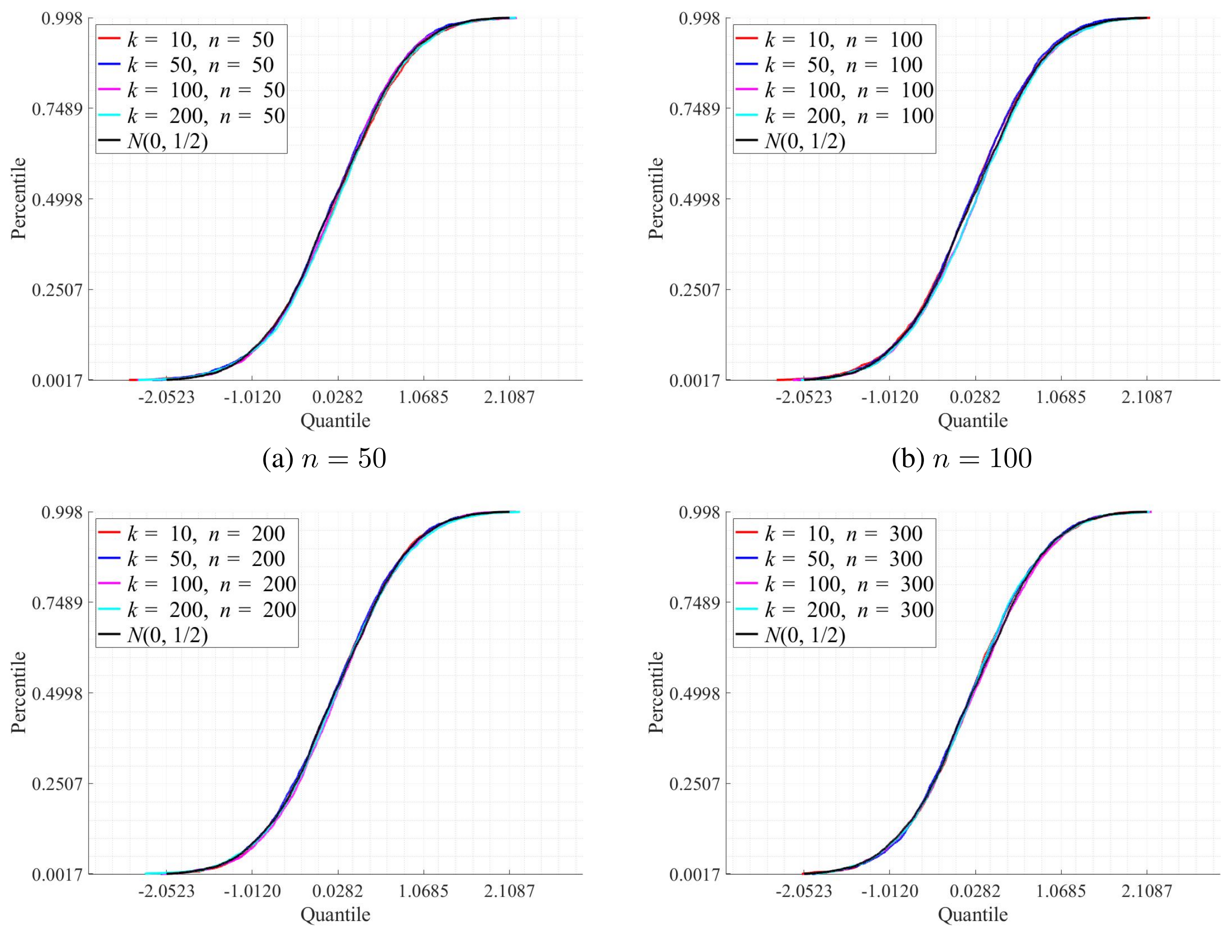


(d) Squared exponential ($s^2 = 0.10$)



(e) Squared exponential ($s^2 = 0.01$)

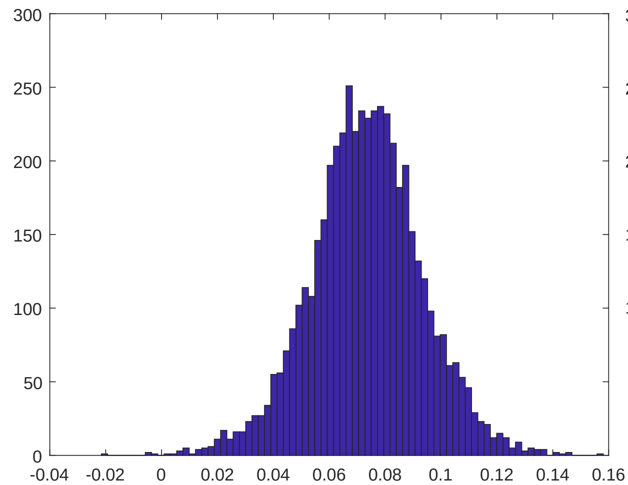
Figure A.1: KERNEL FUNCTIONS. These graphics show the functional shapes of the kernels used in the additional simulations.



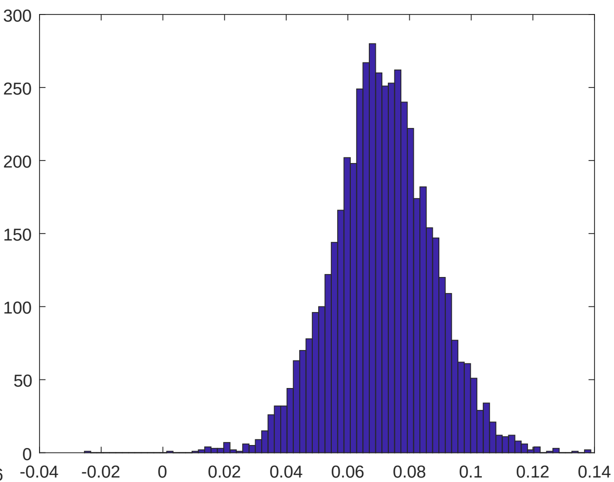
(a) $n = 50$ (b) $n = 100$

(c) $n = 200$ (d) $n = 300$

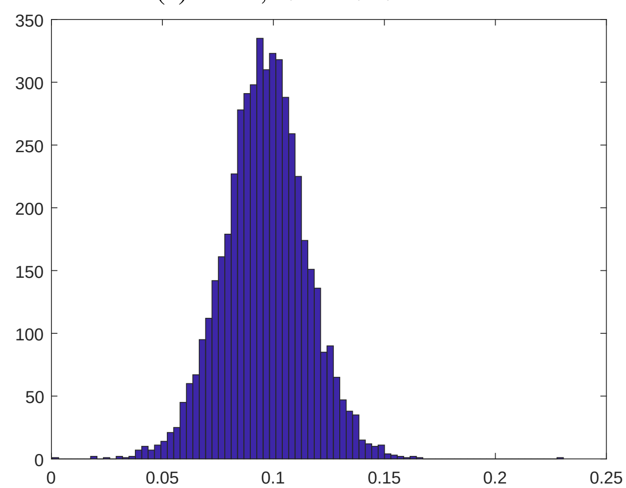
Figure A.2: EMPIRICAL DISTRIBUTIONS OF WID-2SLS. For $k = 50, 100, 200, 300$, each figure shows the empirical distribution of $\sqrt{n}(\widehat{\theta}_{n,k} - \theta_* - 1/3)$ obtained from 3,000 replications. For comparison, the distribution function of $\mathcal{N}(0, \frac{1}{2})$ is drawn. Here, k is the degree of Neumann's series expansion, and for $k = \infty$, the inverse matrix of $\widehat{\Sigma}_n$ is directly used as the weight matrix.



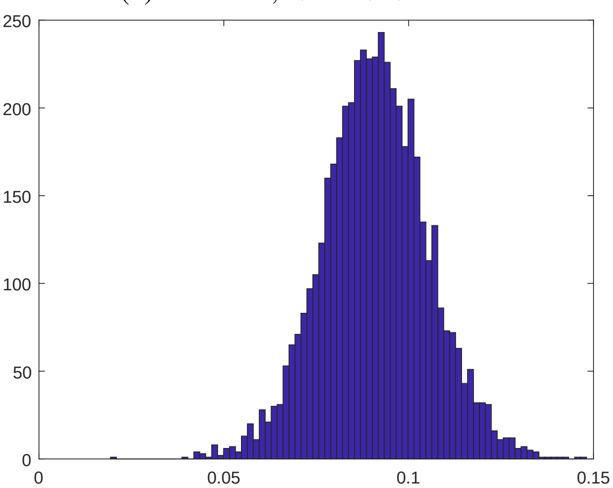
(a) Basic, 1920-1929 cohort



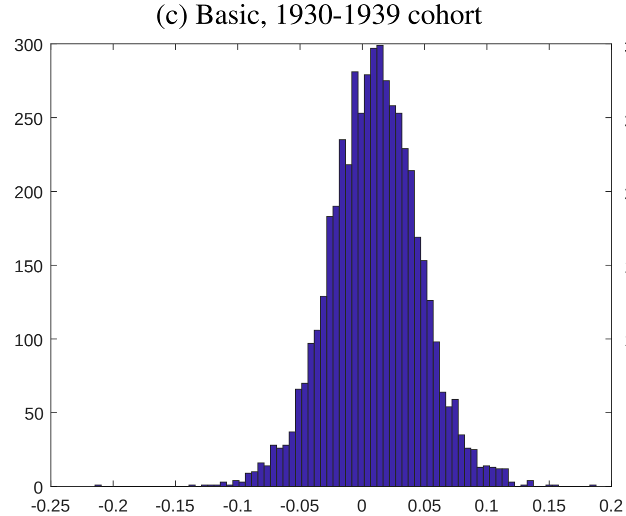
(b) Extended, 1920-1929 cohort



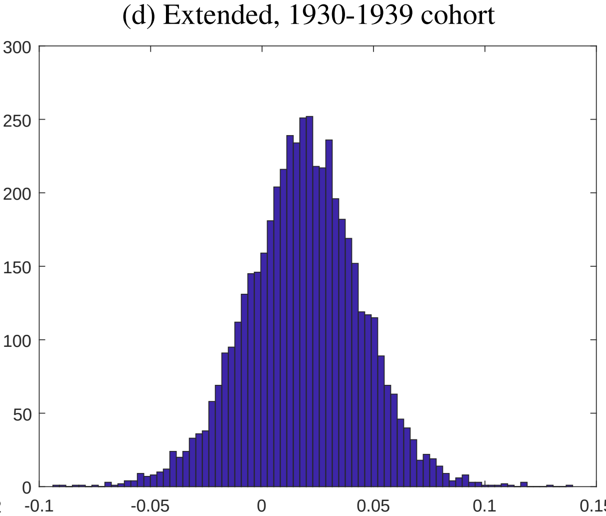
(c) Basic, 1930-1939 cohort



(d) Extended, 1930-1939 cohort

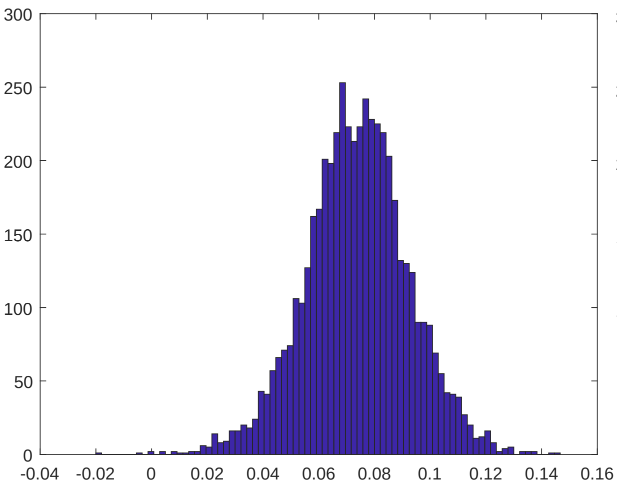


(e) Basic, 1940-1949 cohort

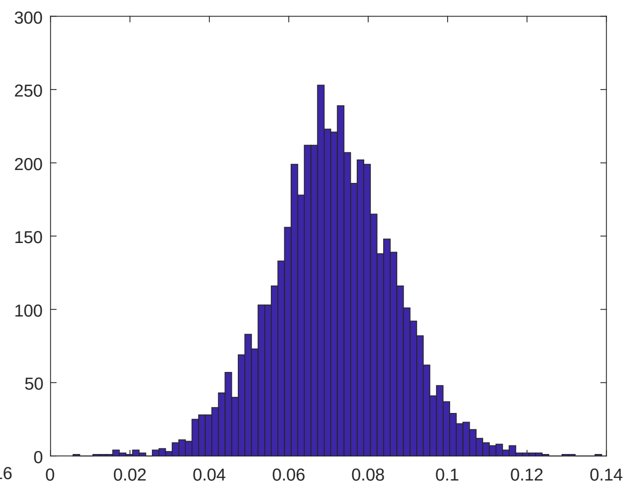


(f) Extended, 1940-1949 cohort

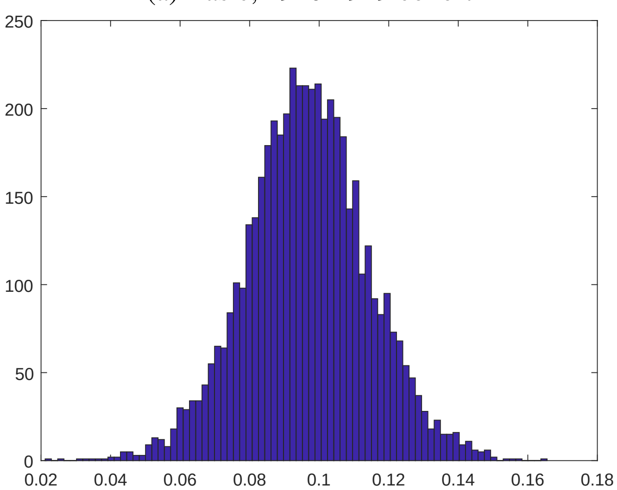
Figure A.3: DISTRIBUTIONS OF THE WID-2SLS ESTIMATOR DRIVEN BY BMK. This figure shows the histograms of the estimated coefficients of the schooling years. For each cohort, we estimated the basic and extended models by wid-2SLS driven by BMK. In total, the moment conditions are permuted 5,000 times with replacement to obtain the sampling distribution, which is drawn as a histogram using 80 bins.



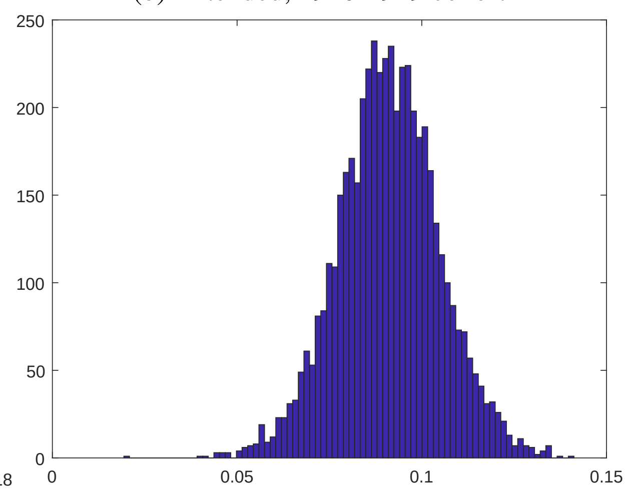
(a) Basic, 1920-1929 cohort



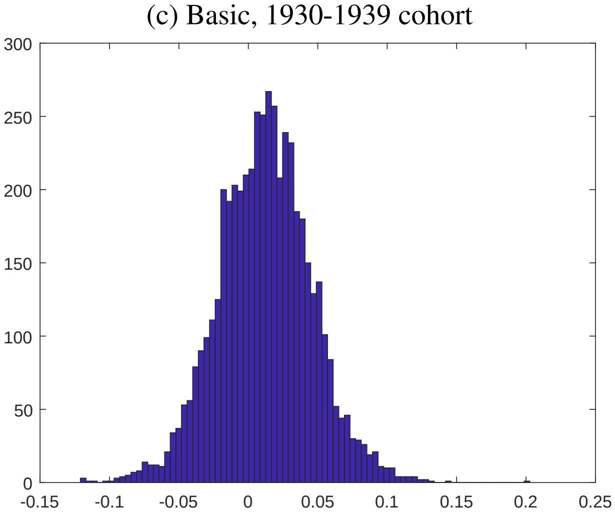
(b) Extended, 1920-1929 cohort



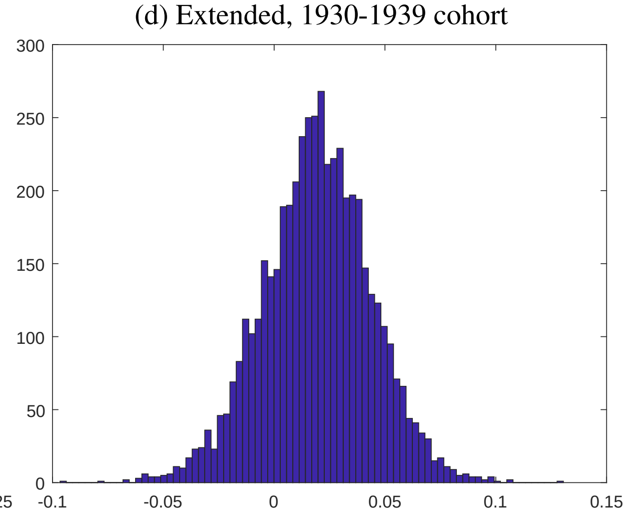
(c) Basic, 1930-1939 cohort



(d) Extended, 1930-1939 cohort

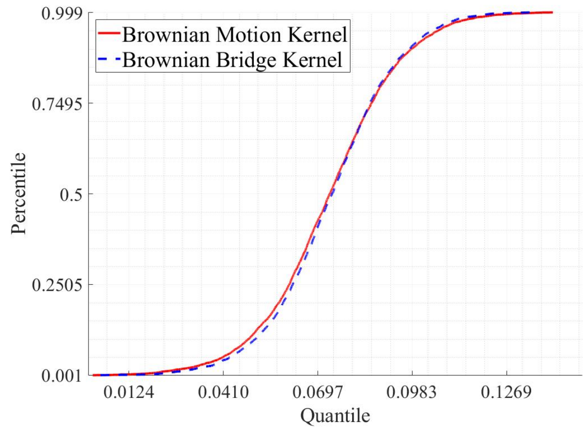


(e) Basic, 1940-1949 cohort

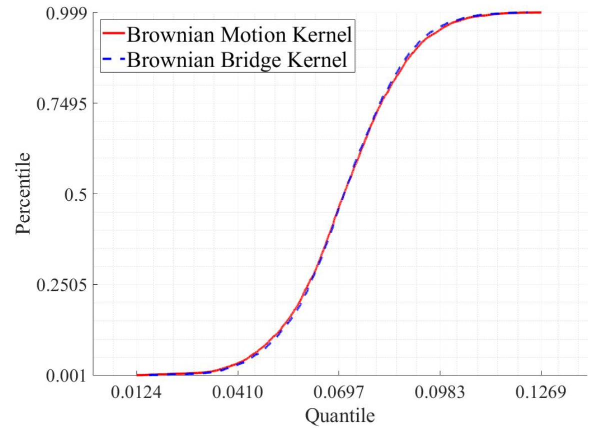


(f) Extended, 1940-1949 cohort

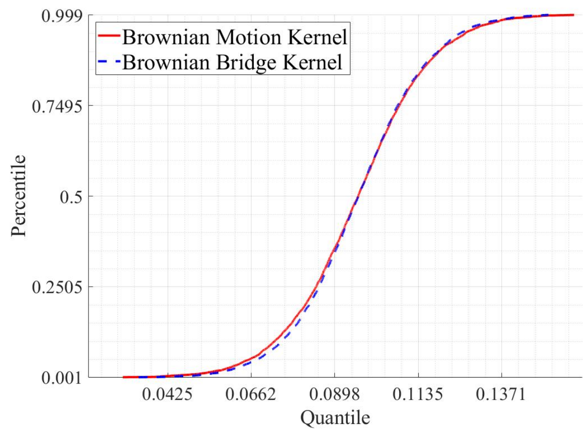
Figure A.4: DISTRIBUTIONS OF THE WID-2SLS ESTIMATOR DRIVEN BY BBK. This figure shows the histograms of the estimated coefficients of the schooling years. For each cohort, we estimate the basic and extended models by wid-2SLS driven by BBK. In total, the moment conditions are permuted 5,000 times with replacement to obtain the sampling distribution, which is drawn as a histogram using 80 bins.



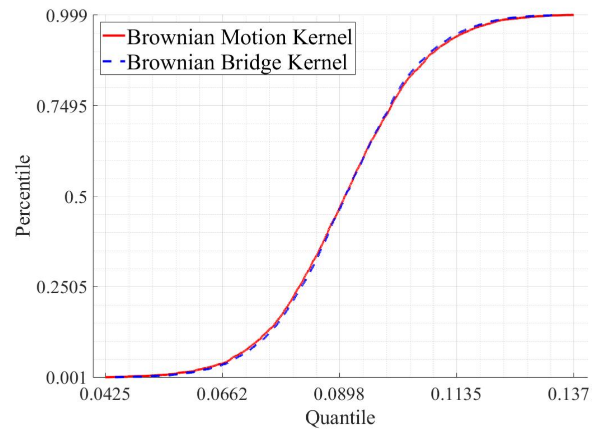
(a) Basic, 1920-1929 cohort



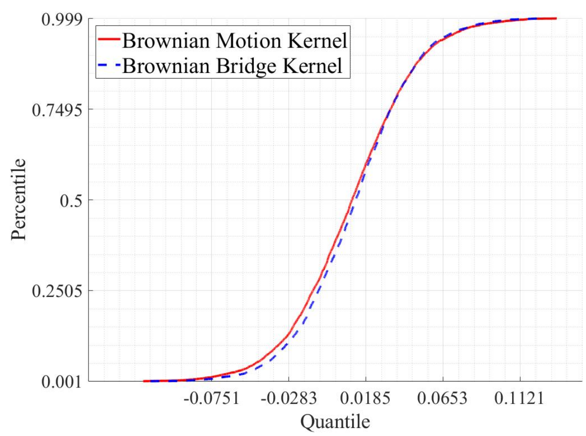
(b) Extended, 1920-1929 cohort



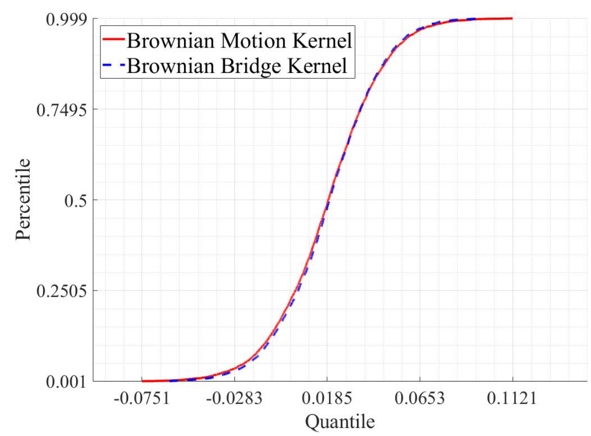
(c) Basic, 1930-1939 cohort



(d) Extended, 1930-1939 cohort



(e) Basic, 1940-1949 cohort



(f) Extended, 1940-1949 cohort

Figure A.5: EMPIRICAL DISTRIBUTIONS OF THE WID-2SLS ESTIMATORS DRIVEN BY BMK AND BBK. This figure shows the empirical distributions of the estimated coefficients of the schooling years. For each cohort, we estimated the basic and extended models by the wid-2SLS driven by BMK and BBK. In total, the moment condition is permuted 5,000 times with replacement to obtain the empirical distribution.

A.4 Glossary of Notation

Notation	Explanation	Notation	Explanation
$(\bar{q}_n(\cdot), q_n(\cdot))$	$(\bar{G}_n(\cdot)'\widehat{\Sigma}_n^{-1}\bar{G}_n(\cdot), n\bar{q}_n(\cdot))$	$(\bar{G}_n(\cdot), G_n(\cdot))$	$(n^{-1}\sum_{t=1}^n U_n(w_t,\cdot), n\bar{G}_n(\cdot)$
(n, s_n)	(sample size, moment size)	s_*	limit of s_n/n
$\widehat{\theta}_n$	$\arg\min_{\theta\in\Theta}\bar{q}_n(\theta)$	$\widetilde{G}_n(\cdot)$	$\sqrt{n}\bar{G}_n(\cdot)$
$\widehat{\Sigma}_n^{-1}$	weight matrix	$H_{n,*}$	$\mathbb{E}[\nabla_\theta U_n(w_t,\theta_*)]$
$\widehat{\sigma}_n(\cdot,\circ)$	càdlàg representation of $\widehat{\Sigma}_n$	$\xi_n(\cdot,\circ)$	càdlàg representation of $\widehat{\Sigma}_n^{-1}$
$\delta(\cdot)$	Dirac delta function	$\xi(\cdot,\circ)$	limit of $\xi_n(\cdot,\circ)$
$\widehat{\mathbf{\Xi}}_n$	integral transform operator with $\xi_n(\cdot,\circ)$	$\mathbf{\Xi}$	integral transform operator with $\xi(\cdot,\circ)$
$g_n(\cdot)$	càdlàg representation of $\bar{G}_n(\theta_*)$	$\widetilde{g}_n(\cdot)$	$\sqrt{n}g_n(\cdot)$
$\eta_n(\cdot)$	càdlàg representation of η_n	$\widetilde{u}_n$	a random variable satisfying CLT
$H_n(\cdot)$	càdlàg representation of $\nabla_\theta'\bar{G}_n(\theta_*)$	$(\bar{A}_n, \widetilde{d}_n)$	$([\widehat{\mathbf{\Xi}}_n H_n(\cdot), H_n(\cdot)], [\widehat{\mathbf{\Xi}}_n H_n(\cdot), \widetilde{g}_n(\cdot)])$
$\mu_n(\cdot)$	càdlàg representation of μ_n	$\mu(\cdot)$	limit of $\mu_n(\cdot)$
$\rho(\cdot)$	limit of $s_n^\alpha\eta_n(\cdot)$	$a_n(\cdot)$	$s_n^{\frac{1}{2}-\nu}n^{\frac{1}{2}}\int_0^{(\cdot)}(s_n^{-\kappa}H_n(u)-\mu_n(u))du$
$s_n(\cdot)$	$s_n^{\frac{1}{2}}\int_0^{(\cdot)}(\widetilde{g}_n(u)-\eta_n(u)\widetilde{u}_n)du$	$\ell_n(\cdot)$	$(s_n(\cdot), a_n(\cdot)')'$
$\mathcal{U}$	weak limit of $\widetilde{u}_n$	$\mathcal{B}_\ell(\cdot)$	BM and $\mathcal{B}_\ell(\cdot)=(\mathcal{B}_s(\cdot),\mathcal{B}_a(\cdot)')'$
$\mathcal{W}_\ell(\cdot)$	Wiener process	Σ_ℓ	global covariance matrix of $\ell_n(1)$
R_1	$[1, 0_{1\times d}]$	σ_s^2	$R_1\Sigma_\ell^2 R_1'$
τ_s	$\sqrt{s_*}[d\mathcal{B}_a(\cdot), d\mathcal{B}_s(\cdot)]$	(A_e, A_r, A_f)	$([\mu(\cdot),\mu(\cdot)], [\mathbf{\Xi}\mu(\cdot),\mu(\cdot)], \Upsilon_s + A_e)$
Υ_s	$s_*[d\mathcal{B}_a(\cdot), d\mathcal{B}_a(\cdot)]$	B_r	$\sigma_s^2[\mathbf{\Xi}\mu(\cdot), \mathbf{\Xi}\mu(\cdot)]$
(ψ_s, ψ_f)	$(-A_e^{-1}\tau_s, -A_f^{-1}\tau_s)$	$\pi_r(\cdot,\circ)$	$\xi(\cdot,\circ)-\frac{1}{2}\lambda(\cdot)'A_r^{-1}\lambda(\cdot)$
$\mathbf{\Pi}_r$	integral transform operator with $\pi_r(\cdot,\circ)$	$\lambda(\cdot)$	$\mathbf{\Xi}\mu(\cdot)$
$\widetilde{a}_n(\cdot)$	$s_n^{\frac{1}{2}-\nu}n^{\frac{1}{2}}\int_0^{(\cdot)}(s_n^{-\kappa}\widetilde{H}_n(u)-\mu_n(u))du$	$\widetilde{H}_n(\cdot)$	$\sqrt{n}H_n(\cdot)$
$(\widetilde{A}_n, \ddot{d}_n)$	$([\widehat{\mathbf{\Xi}}_n\widetilde{H}_n(\cdot), \widetilde{H}_n(\cdot)], [\widehat{\mathbf{\Xi}}_n\widetilde{H}_n(\cdot), \widetilde{g}_n(\cdot)])$	$\widetilde{\ell}_n(\cdot)$	$(s_n(\cdot), \widetilde{a}_n(\cdot)')'$
$\widetilde{\mathcal{W}}_\ell(\cdot)$	Wiener process	$\widetilde{\mathcal{B}}_\ell(\cdot)$	BM and $\widetilde{\mathcal{B}}_\ell(\cdot)=(\widetilde{\mathcal{B}}_s(\cdot),\widetilde{\mathcal{B}}_a(\cdot)')'$
$\widetilde{\Sigma}_\ell$	global covariance matrix of $\widetilde{\ell}_n(1)$	$\widetilde{\sigma}_s^2$	$R_1\widetilde{\Sigma}_\ell^2 R_1'$
$\widetilde{B}_r$	$\widetilde{\sigma}_s^2[\mathbf{\Xi}\mu(\cdot), \mathbf{\Xi}\mu(\cdot)]$	$\xi(\cdot,\circ,\sigma^2)$	squared exponential kernel
Q	permute matrix	c_n	$s_n!$
ω	vector of weights	$\bar{\theta}_n(\omega)$	$\sum_{p=1}^{c_n}\omega_p\widehat{\theta}_n^p$
$\widehat{\theta}_n^p$	permuted GMM	$(\widehat{P}_n, \widetilde{P}_n)$	$(n^{-1}Z'X, \sqrt{n}\widehat{P}_n)$
$(\widehat{c}_n, \widehat{m}_n)$	$(n^{-1}Z'y, n^{-1}X'y)$	$(\widehat{w}_n, \widehat{s}_n, \widetilde{w}_n)$	$(n^{-1}Z'u, n^{-1}X'u, \sqrt{n}\widehat{w}_n)$
$\widetilde{\tau}_s$	$\sqrt{s_*}[d\widetilde{\mathcal{B}}_a(\cdot), d\widetilde{\mathcal{B}}_s(\cdot)]$	$\widetilde{A}_f$	$s_*[d\widetilde{\mathcal{B}}_a(\cdot), d\widetilde{\mathcal{B}}_s(\cdot)] + A_e$
$(\widetilde{\psi}_s, \widetilde{\psi}_f)$	$(-A_e^{-1}\widetilde{\tau}_s, -\widetilde{A}_f^{-1}\widetilde{\tau}_s)$	γ_n	$(s_n n)^{-1}\sum_{i=1}^{s_n}\sum_{t=1}^n z_{t,i}^2$
$\phi_n(\cdot)$	càdlàg representation of $\phi_n(\cdot)$	$(\phi(\cdot), \gamma)$	limit of $(\phi_n(\cdot), s_n^{-1}\sum_{i=1}^{s_n}\mathbb{E}[z_{t,i}^2])$
$(\widehat{\Omega}_{xx,n}, \Omega_{xx})$	$(n^{-1}X'X, \mathbb{E}[x_t x_t'])$	$\ddot{a}_n(\cdot)$	$s_n^{\frac{1}{2}-\nu}n^{\frac{1}{2}}\int_0^{(\cdot)}(s_n^{-\kappa}\widetilde{H}_n(u)-\mu_n(u))du$
k_n	$n^{-\frac{1}{2}}\sum_{t=1}^n(u_t v_t - \mathbb{E}[u_t v_t])$	$\mathcal{K}$	weak limit of k_n
$\ddot{\ell}_n(\cdot)$	$(s_n(\cdot), \ddot{a}_n(\cdot)', \ddot{c}_n(\cdot)')'$	$\ddot{\mathcal{B}}_\ell(\cdot)$	BM and $\ddot{\mathcal{B}}_\ell(\cdot)=(\ddot{\mathcal{B}}_s(\cdot),\ddot{\mathcal{B}}_a(\cdot)',\ddot{\mathcal{B}}_c(\cdot)')'$
$\ddot{\Sigma}_\ell$	global covariance matrix of $\ddot{\ell}_n(1)$	$\ddot{\mathcal{W}}_\ell$	Wiener process
$\mathcal{H}$	$(\mathcal{U}, \mathcal{K}')'$	$\ddot{A}_g$	$\widetilde{A}_f - \gamma\Omega_{xx}$
$\ddot{\sigma}_s^2$	$\ddot{R}_1\ddot{\Sigma}_\ell^2\ddot{R}_1'$	$\ddot{R}_1$	$[1, 0_{1\times 2d}]$
$\ddot{R}_3$	$[0_{d\times(1+d)}, I_{d\times d}]$	$b(\cdot)$	$\mu(\cdot) + s_*\gamma\phi(\cdot)$
ϱ	$[b(\cdot), 1]$	$\ddot{\theta}_n$	$-(\widehat{P}_n'\widehat{\Sigma}_n^{-1}\widehat{P}_n)^{-1}(-\widehat{P}_n)'\widehat{\Sigma}_n^{-1}\widehat{c}_n$
$\ddot{\Sigma}_{\ell h}$	global covariance matrix between $\ddot{\ell}_n(1)$ and h_n	h_n	$(\widetilde{u}_n, k_n')'$
Σ_h^2	covariance matrix of $\mathcal{H}$	Σ_k^2	covariance matrix of $\mathcal{K}$
$\ddot{\sigma}_{sk}$	global covariance matrix between $s_n(1)$ and k_n	$\ddot{\Sigma}_{ck}$	global covariance matrix between $\ddot{c}_n(1)$ and k_n
σ_q^2	$\mathbb{E}[u_t^2 z_{t,j}^2]$	ζ_1	$[\mathbf{\Xi}\mu(\cdot), 1]$
$\ddot{c}_n(\cdot)$	$s_n^{\frac{1}{2}-\nu}n^{\frac{1}{2}}\int_0^{(\cdot)}(s_n^{-\kappa}\widetilde{H}_n(u)-\mu_n(u))(\widetilde{g}_n(u)-\eta_n(u)\widetilde{u}_n)-\widetilde{\tau}_s/s_8^{\mathbf{1}/2}du$		
$\ddot{\Sigma}_s^2$	$\ddot{\sigma}_s^2[b(\cdot), b(\cdot)] + \sqrt{s_*}\varrho\ddot{R}_1\ddot{\Sigma}_\ell^2\ddot{R}_3' + \sqrt{s_*}\ddot{R}_3\ddot{\Sigma}_\ell^2\ddot{R}_1'\varrho' + s_*\ddot{R}_3\ddot{\Sigma}_\ell^2\ddot{R}_3'$		
H_h	$\ddot{\Sigma}_s^2 + \sqrt{s_*}\gamma(\varrho\ddot{\sigma}_{sk} + \ddot{\sigma}_{sk}'\varrho') + s_*\gamma(\ddot{\Sigma}_{ck} + \ddot{\Sigma}_{ck}') + s_*\gamma^2\Sigma_k^2$		

Table A.8: NOTATION GLOSSARY. This table collects the notation used throughout the paper.