MONASH University

MONASH BUSINESS SCHOOL

DEPARTMENT OF ECONOMETRCS AND BUSINESS STATISTICS

ISSN 1440-771X

WORKING PAPER SERIES

Nonparametric Series Estimation of Cointegrated Systems with Time–Varying Coefficients

Yu Bai, **Jiti Gao and Hsein Kew**

Working Paper No. 09/26
October 2026

Nonparametric Series Estimation of Cointegrated Systems with Time–Varying Coefficients

Yu Bai
City University of Macau

Jiti Gao
Monash University

Hsein Kew
Monash University

October 2026

Abstract

In this paper, we propose a nonparametric series estimation framework for a cointegrated system with coefficients that evolve smoothly over time. We first establish an asymptotic theory for the series-based OLS (SOLS) estimation method under nonstationary (unconditional) volatility of unknown form. Then, we develop an asymptotic efficiency theory for a broad class of series estimators. Building on this development, we further propose a series-based dynamic OLS estimator (SDOLS) that attains the efficiency bound and is easy to implement. Monte Carlo simulations show that the SDOLS estimator has satisfactory finite-sample performance and typically performs better than the existing alternatives. An empirical application to the consumption-wealth ratio shows that allowing for time variation in the cointegrating relationship often improves equity premium forecast accuracy relative to both the conventional fixed-coefficient specification and the historical-mean benchmark.

Keywords: series estimation, asymptotic efficiency, SDOLS estimator, consumption–wealth ratio
JEL-Classification: C22, C32, C14, G17

1 — Introduction

The investigation of structural change in cointegration models has received considerable attention over the past several decades. This interest is closely related to the role of cointegration in characterizing long-run equilibrium relationships among nonstationary economic variables. Although this long-run interpretation may appear to imply permanence, such relationships are typically examined over extended sample periods during which the underlying economic environment may evolve. A time-varying cointegrating relationship may therefore be interpreted as a slowly evolving equilibrium relationship that prevails at each point in time, rather than as a single relationship that remains invariant over the entire sample period (see Koop et al., 2011, for a related interpretation). Under this interpretation, deviations from the prevailing relationship remain stable even though the coefficients characterizing that relationship evolve over time. Such evolution may reflect changes in policy regimes, institutions, technologies, or preferences.

Much of the earlier literature on structural change and cointegration has focused on developing procedures to detect structural breaks and/or to estimate their dates. Studies addressing these issues in single-equation settings include Hansen (1992), Andrews et al. (1996), and Kuo (1998), among others. Bai et al. (1998) develop procedures for testing and dating common breaks in multivariate time series. Related stability tests for cointegrated systems are proposed by Seo (1998), Hansen and Johansen (1999), and Quintos (1998). The parameter evolution considered in much of this literature is deterministic and abrupt, with coefficients modeled as piecewise constant and changing discontinuously at a finite number of break dates.

One strand of the literature takes an alternative approach and allows the cointegrating vector to evolve randomly over time. Evolution can be discrete, as in Markov Switching models (e.g. Hall et al., 1997) or threshold models (e.g. Balke and Fomby, 1997), or continuous. In turn, continuous parameter evolution can be driven either by observable variables, as in smooth transition models (e.g. Saikkonen and Choi, 2004) and functional coefficient cointegration models (e.g. Xiao, 2009; Phillips and Wang, 2023), or by unobservable shocks combined with time series models for the parameters, as in Bayesian time-varying cointegration models (e.g. Koop et al., 2011).

Yet another strand of the literature retains a deterministic specification of parameter variation but replaces finite abrupt shifts with smooth evolution. This modelling approach is nonparametric,

leaving the functional form of the coefficient paths unspecified apart from suitable smoothness restrictions. Two broad estimation approaches have emerged in this literature: series methods and kernel methods. The series estimation approach dates back to Park and Hahn (1999), who develop a series estimator for time-varying cointegrating coefficients in a single-equation framework. Bierens and Martins (2010) propose a time-varying vector error correction model (VECM) in which the cointegration relationship varies smoothly over time and is modelled using an orthonormal series expansion, with the series truncation order held fixed in their asymptotic analysis. Within a kernel-estimation framework, Phillips et al. (2017) show that conventional asymptotic theory for stationary regressors does not carry over to integrated regressors, because the kernel-weighted signal matrix becomes asymptotically singular whenever the dimension of the regressors exceeds one. They introduce a novel rotational decomposition and establish the corresponding limit theory. Li et al. (2020) further extend the work of Phillips et al. (2017) by allowing for cointegration among the regressors and a mixture of stochastic and deterministic trends. More recently, Gao et al. (2025) develop a structural break test for the cointegrating parameters in a VECM model, while allowing the adjustment coefficients and short-run dynamics to evolve smoothly over time.

A related literature allows for heteroskedastic innovations in cointegrated systems. In fixed-coefficient VECM, Cavaliere et al. (2010) develop cointegration testing under nonstationary volatility, and Boswijk et al. (2016) consider estimation and hypothesis testing on cointegrating coefficients under both conditional and unconditional volatility of a quite general and unknown form. Gao et al. (2025) also allow the error innovations to cover dependent and locally stationary errors with conditional heteroskedasticity of unknown form.

In this paper, we develop a nonparametric series estimation framework for cointegrated systems with coefficients that evolve smoothly over time. Building on the triangular representation of cointegrated systems as in Phillips (1991), our first contribution is to extend the series-based estimation framework of Park and Hahn (1999) to cointegrated system with nonstationary (unconditional) volatility of unknown form. We establish the corresponding limit theory for the series-based ordinary least squares (SOLS) estimator. A key finding is that the limiting signal matrix is not uniformly invertible as the truncation order increases, a feature that must be accommodated in the asymptotic analysis. Nevertheless, the limit theory is formulated directly for the original coefficient functions, without the rotation required by kernel-based methods.

A further contribution concerns the formal justification of asymptotic efficiency in time-varying cointegrating regressions. Although Park and Hahn (1999) describe their series-based canonical cointegrating regression (SCCR) estimator as asymptotically efficient, the asymptotic efficiency theory in the corresponding nonparametric setting has not been worked out. Establishing such an asymptotic efficiency theory requires additional analysis even under homoskedastic innovations, as the series truncation order increases with the sample size, so efficiency arguments developed for fixed-coefficient cointegrating regressions do not apply directly. We extend the concentration-based efficiency criterion of Saikkonen (1991) and derive an efficiency bound within a broad class of series-based estimators. This framework provides a formal justification for the efficiency of SCCR estimator. We further propose a series-based dynamic OLS (SDOLS) estimator that attains the same efficiency bound, which extends the dynamic OLS approach of Saikkonen (1991) and Stock and Watson (1993) to the time-varying cointegration setting. The estimator augments the regression with leads and lags of the first differences of the integrated regressors, and all coefficients can be estimated jointly by least squares, making the procedure straightforward to implement. Under suitable conditions, we show that SDOLS estimator has a mixed-normal limiting distribution.

We show the practical relevance and applicability of the proposed model and estimation method through an empirical application to the consumption-wealth ratio (cay) and equity premium forecasting. The results show that using cay measures constructed from the time-varying cointegrated system generally improves forecast accuracy relative to both their fixed-coefficient counterparts and the historical mean benchmark, although the forecasting gains depend on the implementational procedure and vary across business cycle states.

The finite sample performance of the proposed estimators is also evaluated using an extensive Monte Carlo study in Section F of the Online Supplement. The results show that SDOLS with a small to moderate number of leads and lags generally improves estimation accuracy relative to both the SOLS estimator and SCCR estimator. It also produces pointwise confidence intervals with coverage probabilities closer to the nominal level. These performance patterns are broadly robust across alternative coefficient dynamics.

The rest of the paper is organized as follows. Section 2 presents the model, the series estimation methodology, and establishes the baseline asymptotic theory. Section 3 develops the asymptotic

efficiency theory and proposes the series-based dynamic OLS estimator. Section 4 presents an empirical application to the consumption–wealth ratio and equity premium forecasting. Section 5 concludes. Details of the data construction are provided in the Appendix. The Online Supplement contains all mathematical proofs of the main results, supporting lemmas, and Monte Carlo simulations.

Notation. Bold symbols denote vectors or matrices, whereas nonbold symbols denote scalars. $\|\cdot\|$ denotes the Euclidean norm for vectors and the Frobenius norm for matrices, while $\|\cdot\|_{sp}$ denotes the spectral norm for matrices. We use $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ to denote the smallest and largest eigenvalues of $\mathbf{A}$, respectively. $\mathbf{I}_d$ and $\mathbf{0}_{d_1\times d_2}$ denote the identity and zero matrices of the indicated dimensions, respectively. The notation $\lfloor a \rfloor$ denotes the largest integer not exceeding a. For two positive sequences a_n and b_n, $a_n \asymp b_n$ means that $a_n/b_n = O(1)$ and $b_n/a_n = O(1)$ as $n \to \infty$. For a sufficiently smooth function $g(\cdot)$, $g^{(j)}(\cdot)$ denotes its jth derivative with respect to its argument. $\mathcal{D}_{\mathbb{R}^{m\times n}}[0,1]$ denotes the space of $m \times n$ matrices of Càdlàg functions on $[0,1]$. Finally, $\overset{a.s.}{\to}$, $\overset{p}{\to}$, and $\Rightarrow$ denote almost sure convergence, convergence in probability and convergence in distribution, respectively, while "$a.s.$" means almost surely.

2 — Model and estimation theory

2.1. Model setup

We consider the time-varying coefficients cointegrated system (cf. Phillips, 1991)

$$\mathbf{y}_t = \mathbf{A}_t \mathbf{X}_t + \mathbf{u}_{0t}, \tag{1}$$

$$\mathbf{X}_t = \mathbf{X}_{t-1} + \mathbf{u}_{xt}, \tag{2}$$

where $t = 1, 2, \cdots, n$. $\mathbf{y}_t$ denotes a d_y-dimensional vector, and $\mathbf{X}_t$ is a $d_x \times 1$ vector of integrated regressors. Here, $\mathbf{A}_t := \mathbf{A}(\tau_t)$ is the $d_y \times d_x$ matrix of time-varying cointegrating coefficients, where $\tau_t := t/n$ denotes the rescaled time point.[1] The process $\mathbf{X}_t$ is initialized at $\mathbf{X}_0 = O_p(1)$.

1. This specification differs from the functional coefficient cointegration models in which the index variable in the functional coefficients is random, rather than deterministic functions of scaled time point.

Remark 1. Let $\mathbf{Y}_t = \left(\mathbf{y}_t^\top, \mathbf{X}_t^\top\right)^\top$. Consider a time-varying VECM model of the form

$$\Delta \mathbf{Y}_t = \mathbf{\Pi}_t^{(\text{ecm})} \mathbf{Y}_{t-1} + \sum_{j=1}^{p-1} \mathbf{\Gamma}_{jt}^{(\text{ecm})} \Delta \mathbf{Y}_{t-j} + \mathbf{u}_t^{(\text{ecm})}.$$

where $\mathbf{\Pi}_t^{(\text{ecm})} = \boldsymbol{\alpha}_t^{(\text{ecm})} \boldsymbol{\beta}_t^{(\text{ecm})\top}$. However, without prior restrictions, the speed of adjustment $\boldsymbol{\alpha}_t^{(\text{ecm})}$ and time-varying cointegration vector $\boldsymbol{\beta}_t^{(\text{ecm})}$ cannot be separately identified. The implementations usually impose restrictions on which components are allowed to vary and how they vary. For instance, Bierens and Martins (2010) allow the cointegrating vectors to vary smoothly over time while keeping the speed of adjustment coefficients $\boldsymbol{\alpha}_t^{(\text{ecm})}$ fixed. In Gao et al. (2025), $\boldsymbol{\alpha}_t^{(\text{ecm})}$ varies smoothly over time, while the cointegration vector is kept fixed in the baseline specification. They further extend the framework to allow unknown structural breaks in $\boldsymbol{\beta}^{(\text{ecm})}$. Between consecutive break points, the speed of adjustment is allowed to vary smoothly over time.

The cointegrated system in (1)-(2) should be viewed as a regression-based formulation of the time-varying cointegrating relationship, rather than as a full time-varying VECM model. Here, we do not attempt to distinguish the sources of time variation within a full VECM, but instead provide a tractable regression-based framework for estimation and inference of the time-varying cointegrating relationship itself.

The innovations $\mathbf{u}_t = \left(\mathbf{u}_{0t}^\top, \mathbf{u}_{xt}^\top\right)^\top$ are assumed to satisfy the following linear process structure with heteroskedastic innovations:

$$\mathbf{u}_t = \mathbf{\Phi}(\mathcal{L})\boldsymbol{\varepsilon}_t = \sum_{j=0}^{\infty} \mathbf{\Phi}_j \boldsymbol{\varepsilon}_{t-j}, \tag{3}$$

$$\boldsymbol{\varepsilon}_t = \mathbf{H}_t \mathbf{e}_t, \quad \mathbf{e}_t \overset{i.i.d.}{\sim} \left(\mathbf{0}_{(d_y+d_x)\times 1}, \mathbf{I}_{d_y+d_x}\right). \tag{4}$$

Following Cavaliere and Taylor (2006) and Cavaliere et al. (2010), the sequence $\boldsymbol{\varepsilon}_t$ is formed as the product of two components, $\mathbf{H}_t$ and $\mathbf{e}_t$. Since $\mathbf{e}_t$ is a sequence of independent and identically distributed $(i.i.d.)$ random vectors, conditionally on $\mathbf{H}_t$, the error term $\boldsymbol{\varepsilon}_t$ has mean zero and time-varying covariance matrix $\mathbf{\Sigma}_{\varepsilon,t} := \mathbf{H}_t \mathbf{H}_t^\top$. We impose the following conditions.

Assumption 1.

(i) The coefficient matrices in (3) satisfy $\sum_{j=0}^{\infty} j^2\|\mathbf{\Phi}_j\| < \infty$ and $\mathbf{\Phi}(1) = \sum_{j=0}^{\infty} \mathbf{\Phi}_j$ is nonsingular.

(ii) $\max_{1\le t\le n} \mathbb{E}\|\mathbf{e}_t\|^p < \infty$, for some $p > 4$.

(iii) The matrix $\mathbf{H}_t$ is non-stochastic and satisfies $\mathbf{H}_t := \mathbf{H}(\tau_t)$ for all $t = 1, 2, \cdots, n$, where $\mathbf{H}(\cdot) \in \mathcal{D}_{\mathbb{R}^{(d_y+d_x)\times(d_y+d_x)}}[0,1]$. Moreover, the smallest singular value of $\mathbf{H}(u)$, denoted by $\sigma_{\min}(\mathbf{H}(u))$, satisfies $\inf_{u\in[0,1]} \sigma_{\min}(\mathbf{H}(u)) > 0$ for all $u \in [0,1]$, and $\sup_t \|\mathbf{H}_t\|_{sp} < \infty$ for $t \le 0$. In addition, the total variation of $\mathbf{H}(\cdot)$ satisfies

$$V_H[0,1] := \sup_{\{r_i\}_{i=0}^N \in \Pi} \left\{ \sum_{i=1}^{N} \|\mathbf{H}(r_i) - \mathbf{H}(r_{i-1})\| \right\} < \infty,$$

where Π is the collection of all partitions on the interval $[0,1]$, i.e. sequences $\{r_i\}_{i=0}^N$ such that $0 \le r_0 < r_1 < \cdots < r_N \le 1$ with $N \in \mathbb{N}$.

Remark 2. As explained in Remark 2.1 of Cavaliere et al. (2010), Assumption 1(iii) accommodates a broad class of the innovation variance dynamics, including single or multiple variance or covariance shifts, linear or broken trend, and smooth transition variance shifts. It also implies that $\mathbf{\Sigma}_{\varepsilon,t}$ is uniformly bounded: $\max_{1\le t\le n} \|\mathbf{\Sigma}_{\varepsilon,t}\| < \infty$. In addition, the requirement that $\mathbf{H}_t$ is non-stochastic can be relaxed to allow for cases where $\mathbf{e}_t$ and $\mathbf{H}_s$ are stochastically independent for all $t, s = 1, \cdots, n$. See, e.g., Remark 2.1 of Cavaliere and Taylor (2007) and Remark 2.2 of Cavaliere et al. (2010), for a more detailed discussion.

Under Assumption 1, the following functional central limit theory (FCLT) follows directly from Lemma 2 of Cavaliere et al. (2010):

$$\begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \boldsymbol{\varepsilon}_t \\ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{\varepsilon}_t \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} \boldsymbol{\varepsilon}_i \right)^\top \end{pmatrix} \Rightarrow \begin{pmatrix} \mathbf{B}_H(r) \\ \int_0^1 d\mathbf{B}_H(r) \mathbf{B}_H^\top(r) \end{pmatrix}, \tag{5}$$

where $\mathbf{B}_H(\cdot) := \int_0^{\cdot} \mathbf{H}(s) d\mathbf{B}(s)$ is a multivariate continuous local martingale and $\mathbf{B}(\cdot)$ is a standard Brownian motion vector. Interestingly, as pointed out in Remark 3 of Cavaliere and Taylor (2006), the stochastic differential representation of this process $d\mathbf{B}_H(r) = \mathbf{H}(r) d\mathbf{B}(r)$ is discussed by Barndorff-Nielsen and Shephard (2004), who name $\mathbf{B}_H(r)$ as a multivariate stochastic volatility

process with $\boldsymbol{\Sigma}_\varepsilon(\cdot)$ and $\int_0^{\cdot} \boldsymbol{\Sigma}_\varepsilon(r)dr$ being the corresponding spot covolatility process and integrated covariance process, respectively.

Remark 3. By the Skorohod representation theorem (see, e.g., section 23.3 of Davidson (2021)), there exists, in a richer probability space, a random vector $(\boldsymbol{\varepsilon}_1^*, \cdots, \boldsymbol{\varepsilon}_n^*)^\top$, satisfying $(\boldsymbol{\varepsilon}_1^*, \cdots, \boldsymbol{\varepsilon}_n^*)^\top \overset{d}{=} (\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n)^\top$, for which

$$\begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \boldsymbol{\varepsilon}_t^* \\ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{\varepsilon}_t^* \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t-1} \boldsymbol{\varepsilon}_i^* \right)^\top \end{pmatrix} \xrightarrow{a.s.} \begin{pmatrix} \mathbf{B}_H(r) \\ \int_0^1 d\mathbf{B}_H(r) \mathbf{B}_H^\top(r) \end{pmatrix}.$$

To simplify the notation, we write $(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n)^\top = (\boldsymbol{\varepsilon}_1^*, \cdots, \boldsymbol{\varepsilon}_n^*)^\top$ and all the subsequent proofs should be understood in this richer probability space.

By virtue of Assumption 1, $\mathbf{u}_t$ has time-varying covariance matrix $\boldsymbol{\Sigma}_t = \sum_{j=0}^{\infty} \boldsymbol{\Phi}_j \boldsymbol{\Sigma}_{\varepsilon,t-j} \boldsymbol{\Phi}_j^\top$, with $\max_{1 \le t \le n} \mathbb{E}\|\mathbf{u}_t\|^p < \infty$, for some $p > 4$. It follows immediately from (5) that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \mathbf{u}_t \Rightarrow \underline{\mathbf{B}}(r) := \boldsymbol{\Phi}(1) \mathbf{B}_H(r),$$

where $\underline{\mathbf{B}}(r) = \left(\underline{\mathbf{B}}_0^\top(r), \underline{\mathbf{B}}_X^\top(r)\right)^\top$ is a $(d_y + d_x)$-variate continuous martingale with spot covariance matrix $\boldsymbol{\Omega}(r) = \boldsymbol{\Phi}(1) \boldsymbol{\Sigma}_\varepsilon(r) \boldsymbol{\Phi}(1)^\top > 0$. Furthermore, for $r \in [0, 1]$, we define the local contemporaneous covariance matrix as $\boldsymbol{\Sigma}(r) := \sum_{j=0}^{\infty} \boldsymbol{\Phi}_j \boldsymbol{\Sigma}_\varepsilon(r) \boldsymbol{\Phi}_j^\top$. For ease of theoretical development, we partition $\boldsymbol{\Sigma}(r)$, $\boldsymbol{\Phi}(1)$, and $\boldsymbol{\Omega}(r)$ conformably with $\mathbf{u}_t$ as

$$\boldsymbol{\Sigma}(r) = \begin{bmatrix} \boldsymbol{\Sigma}_{00}(r) & \boldsymbol{\Sigma}_{0x}(r) \\ \boldsymbol{\Sigma}_{x0}(r) & \boldsymbol{\Sigma}_{xx}(r) \end{bmatrix}, \quad \boldsymbol{\Phi}(1) = \begin{bmatrix} \boldsymbol{\Phi}_0(1) \\ \boldsymbol{\Phi}_x(1) \end{bmatrix}, \quad \boldsymbol{\Omega}(r) = \begin{bmatrix} \boldsymbol{\Omega}_{00}(r) & \boldsymbol{\Omega}_{0x}(r) \\ \boldsymbol{\Omega}_{x0}(r) & \boldsymbol{\Omega}_{xx}(r) \end{bmatrix}, \tag{6}$$

where $\boldsymbol{\Phi}_0(1)$ and $\boldsymbol{\Phi}_x(1)$ are of dimension $d_y \times (d_y + d_x)$ and $d_x \times (d_y + d_x)$, respectively. Using the Beveridge–Nelson (BN) decomposition, we have the following decomposition for $\mathbf{u}_t$:

$$\mathbf{u}_t = \boldsymbol{\Phi}(1) \boldsymbol{\varepsilon}_t - \Delta \widetilde{\boldsymbol{u}}_t, \text{ for } \widetilde{\mathbf{u}}_t = \sum_{j=0}^{\infty} \widetilde{\boldsymbol{\Phi}}_j \boldsymbol{\varepsilon}_{t-j}, \quad \widetilde{\boldsymbol{\Phi}}_j = \sum_{k=j+1}^{\infty} \boldsymbol{\Phi}_k, \tag{7}$$

where Assumption 1(i) ensures $\sum_{j=0}^{\infty} \left\| \widetilde{\mathbf{\Phi}}_j \right\| < \infty$. Corresponding to the partition of $\mathbf{u}_t$, we write

$$\mathbf{u}_{0t} = \sum_{j=0}^{\infty} \mathbf{\Phi}_{0j} \boldsymbol{\varepsilon}_{t-j}, \quad \widetilde{\mathbf{u}}_{0t} = \sum_{j=0}^{\infty} \widetilde{\mathbf{\Phi}}_{0j} \boldsymbol{\varepsilon}_{t-j},$$

and also

$$\mathbf{u}_{xt} = \sum_{j=0}^{\infty} \mathbf{\Phi}_{xj} \boldsymbol{\varepsilon}_{t-j}, \quad \widetilde{\mathbf{u}}_{x,t} = \sum_{j=0}^{\infty} \widetilde{\mathbf{\Phi}}_{xj} \boldsymbol{\varepsilon}_{t-j}, \tag{8}$$

where $\mathbf{\Phi}_{0j}$, $\widetilde{\mathbf{\Phi}}_{0j}$ are of dimension $d_y \times (d_y + d_x)$ and $\mathbf{\Phi}_{xj}$, $\widetilde{\mathbf{\Phi}}_{xj}$ are of dimension $d_x \times (d_y + d_x)$, respectively. In addition, the limit theory involves the partitioned components of the local one-sided long-run covariance matrix, defined for $r \in [0, 1]$ by

$$\mathbf{\Lambda}(r) := \begin{bmatrix} \mathbf{\Lambda}_{00}(r) & \mathbf{\Lambda}_{0x}(r) \\ \mathbf{\Lambda}_{x0}(r) & \mathbf{\Lambda}_{xx}(r) \end{bmatrix},$$

where

$$\mathbf{\Lambda}(r) = \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} \mathbf{\Phi}_{k+j} \mathbf{\Sigma}_{\varepsilon}(r) \mathbf{\Phi}_k^{\top} = \sum_{k=0}^{\infty} \left(\sum_{j=1}^{\infty} \mathbf{\Phi}_{k+j} \right) \mathbf{\Sigma}_{\varepsilon}(r) \mathbf{\Phi}_k^{\top} = \sum_{k=0}^{\infty} \widetilde{\mathbf{\Phi}}_k \mathbf{\Sigma}_{\varepsilon}(r) \mathbf{\Phi}_k^{\top}.$$

For $r \in [0, 1]$, these definitions also imply that

$$\mathbf{\Omega}(r) = \mathbf{\Sigma}(r) + \mathbf{\Lambda}(r) + \mathbf{\Lambda}^{\top}(r).$$

2.2. Estimation procedure

We begin by constructing a series approximation to $\mathbf{A}(\cdot)$. Let $a(\cdot)$ be a generic (i, j)th element in $\mathbf{A}(\cdot)$, where $i = 1, 2, \cdots, d_y$, $j = 1, 2, \cdots, d_x$. Suppose that $a(\cdot) \in L^2[0, 1] = \{u(r) : \int_0^1 u^2(r) dr < \infty\}$, in which the inner product is given by $\langle u_1, u_2 \rangle = \int_0^1 u_1(r) u_2(r) dr$ and the induced norm $\|u\|^2 = \langle u, u \rangle$. Furthermore, let

$$\psi_0(r) = 1, \quad \psi_j(r) = \sqrt{2} \cos(j\pi r), \tag{9}$$

where $j = 1, 2, 3, \cdots$. Then, $\{\psi_j(r)\}$ is a complete orthonormal basis in the Hilbert space $L^2[0,1]$, where $\langle \psi_i(r), \psi_j(r) \rangle = \delta_{ij}$ is the Kronecker delta and completeness follows from that of the trigonometric system by even extension; see Section 4.23 of Rudin (1987).

The basis functions defined in (9) are commonly referred to as Chebyshev time polynomials. We adopt this basis because it has some attractive features to the relevant econometrics literature. They were first introduced into econometrics applications by Bierens (1997) for testing a unit root against a nonlinear trend-stationary alternative, and were later used by Bierens and Martins (2010) in their time-varying VECM model. As discussed in Dong and Gao (2025), the use of Chebyshev time polynomials is mainly a matter of convention, which can readily be replaced by some other orthonormal basis functions on $L^2[0,1]$.

Then, the basis $\{\psi_j(r)\}$ is used to expand the unknown function $a(r) \in L^2[0,1]$ into orthonormal series, that is,

$$a(r) = \sum_{i=0}^{\infty} b_{1,i} \psi_i(r), \tag{10}$$

where $b_{1,i} = \langle a, \psi_i \rangle$. For a positive integer k, we can define the truncated series

$$a_k(r) = \sum_{i=0}^{k-1} b_{1,i} \psi_i(r). \tag{11}$$

Completeness, as Theorem 4.18 (ii) of Rudin (1987), implies Parseval's identity (Theorem 4.18 (iv) of Rudin (1987)) and hence the convergence of $\sum_{i=0}^{\infty} b_{1,i}^2$. Therefore,

$$\int_0^1 |a(r) - a_k(r)|^2 \, dr = \sum_{j=k}^{\infty} b_{1,j}^2 \to 0,$$

as $k \to \infty$. However, L^2 convergence alone is insufficient for our purpose, as the subsequent asymptotic analysis requires an explicit rate of convergence. To obtain such a rate, we impose additional restrictions on $a(\cdot)$.

Assumption 2. The function $a(\cdot)$ is q–times continuously differentiable for some integer $q > 1$. Moreover, $a^{(2\ell-1)}(0) = a^{(2\ell-1)}(1) = 0$ for every positive integer ℓ satisfying $2\ell - 1 < q$.

In particular, the endpoint restrictions $a^{(2\ell-1)}(0) = a^{(2\ell-1)}(1) = 0$ are the compatibility

conditions for the even periodic extension underlying the Fourier cosine series, see Theorem 4 of Boyd (2001) and the subsequent discussions. They eliminate the relevant boundary terms arising from repeated integration by parts, thereby ensuring that the Fourier cosine coefficients of $a(\cdot)$ decay sufficiently fast to yield the following L^2 approximation rate.

Lemma 1. *Under Assumption 2, we have*

$$\int_0^1 |a(r) - a_k(r)|^2 \, dr = O\left(k^{-2q}\right). \tag{12}$$

Since the expansion (10) holds elementwise, the scalar coefficients $b_{1,i}$ can be collected into matrix coefficients $\mathbf{B}_i$, yielding the matrix series representation:

$$\mathbf{A}(r) = \sum_{i=0}^{\infty} \mathbf{B}_i \psi_i(r),$$

where $\mathbf{B}_i$ is of dimension $d_y \times d_x$. We can similarly define a matrix truncated series

$$\mathbf{A}_k(r) = \sum_{i=0}^{k-1} \mathbf{B}_i \psi_i(r) = \mathbf{B}^{(k)} \left(\boldsymbol{\psi}_{(k)}(r) \otimes \mathbf{I}_{d_x}\right), \tag{13}$$

where $\boldsymbol{\psi}_{(k)}(r) = (\psi_0(r), \psi_1(r), \cdots, \psi_{k-1}(r))^\top$ and $\underset{d_y \times k d_x}{\mathbf{B}^{(k)}} = (\mathbf{B}_0, \mathbf{B}_1, \cdots, \mathbf{B}_{k-1})$.

With (13), we can rewrite the model (1) as

$$\mathbf{y}_t = \sum_{i=0}^{k-1} \mathbf{B}_i \psi_i(\tau_t) \mathbf{X}_t + \mathbf{u}_{0t}^{(k)}, \tag{14}$$

where $\mathbf{u}_{0t}^{(k)} = \mathbf{u}_{0t} + (\mathbf{A}_t - \mathbf{A}_k(\tau_t))\, \mathbf{X}_t$. Define $\underset{k d_x \times 1}{\mathbf{X}_t^{(k)}} = \left(\mathbf{X}_t^\top, \psi_1(\tau_t)\mathbf{X}_t^\top, \cdots, \psi_{k-1}(\tau_t)\mathbf{X}_t^\top\right)^\top$, we can write (14) more conveniently as

$$\mathbf{y}_t = \mathbf{B}^{(k)} \mathbf{X}_t^{(k)} + \mathbf{u}_{0t}^{(k)}. \tag{15}$$

$\mathbf{B}^{(k)}$ can thus be estimated by LS (least squares):

$$\widehat{\mathbf{B}}_n^{(k)} = \left(\sum_{t=1}^{n} \mathbf{y}_t \mathbf{X}_t^{(k)\top}\right) \left(\sum_{t=1}^{n} \mathbf{X}_t^{(k)} \mathbf{X}_t^{(k)\top}\right)^{-1} \tag{16}$$

Then, for any $z_0 \in [0,1]$,

$$\widehat{\mathbf{A}}_n(z_0) = \widehat{\mathbf{B}}_n^{(k)} \mathbf{T}_{(k)}(z_0), \tag{17}$$

naturally defines the series-based OLS (SOLS) estimator of the unknown function $\mathbf{A}(\cdot)$, where $\mathbf{T}_{(k)}(z_0) = \boldsymbol{\psi}_{(k)}(z_0) \otimes \mathbf{I}_{d_x}$.

2.3. Asymptotic theory

We now establish the asymptotic properties of the SOLS estimator. From (17), for each fixed $z_0 \in [0,1]$, we have the decomposition

$$\widehat{\mathbf{A}}_n(z_0) - \mathbf{A}(z_0) = \left(\widehat{\mathbf{B}}_n^{(k)} - \mathbf{B}^{(k)}\right) \mathbf{T}_{(k)}(z_0) + \left(\mathbf{A}_k(z_0) - \mathbf{A}(z_0)\right). \tag{18}$$

where the second part $\mathbf{A}_k(z_0) - \mathbf{A}(z_0)$ is just the deterministic approximation error from the truncation, and the first term contributes to the stochastic behaviour of $\widehat{\mathbf{A}}_n(z_0)$.

It is shown in the proof of Theorem 1 that, under suitable restrictions on k, the deterministic approximation error is asymptotically of smaller order, so that the limiting distribution of the SOLS estimator is determined by the following estimation error component:

$$\left(\widehat{\mathbf{B}}_n^{(k)} - \mathbf{B}^{(k)}\right) \mathbf{T}_{(k)}(z_0),$$

Before proceeding, we present a lemma on the limit of the corresponding sample signal matrix.

Lemma 2. *Let $k = \lfloor c \cdot n^{\kappa} \rfloor$ for some $c > 0$, where $0 < \kappa < 1$. Under Assumption 1, we have, as $n \to \infty$*

$$\left\| \frac{1}{n^2} \sum_{t=1}^{n} \mathbf{X}_t^{(k)} \mathbf{X}_t^{(k)\top} - \boldsymbol{\Lambda}_{xx}^{(k)} \right\|_{sp} = O_p\left(\frac{k}{n^a}\right) + O_p\left(\frac{k^2}{n}\right),$$

where $a = \frac{p-2}{2p}$ and

$$\boldsymbol{\Lambda}_{xx}^{(k)} := \int_0^1 \left\{ \boldsymbol{\psi}_{(k)}(r) \boldsymbol{\psi}_{(k)}^{\top}(r) \otimes \left(\underline{\mathbf{B}}_X(r) \underline{\mathbf{B}}_X^{\top}(r) \right) \right\} dr.$$

Moreover,

(i) $\sup_{k \geq 1} \left\| \boldsymbol{\Lambda}_{xx}^{(k)} \right\|_{sp} < \infty$, *a.s.;*

(ii) for every fixed $k \geq 1$, $\left\| \left(\mathbf{\Lambda}_{xx}^{(k)} \right)^{-1} \right\|_{sp} < \infty \; a.s.$;

(iii) $\left\| \left(\mathbf{\Lambda}_{xx}^{(k)} \right)^{-1} \right\|_{sp} \to \infty \; a.s.$ *as* $k \to \infty$.

Although $\lambda_{\max}\left(\mathbf{\Lambda}_{xx}^{(k)} \right) = \left\| \mathbf{\Lambda}_{xx}^{(k)} \right\|_{sp}$ is uniformly bounded over k $a.s.$, $\lambda_{\min}\left(\mathbf{\Lambda}_{xx}^{(k)} \right)$, while strictly positive $a.s.$ for every fixed $k \geq 1$, converges to zero $a.s.$ as $k \to \infty$. Consequently, $\left\| \left(\mathbf{\Lambda}_{xx}^{(k)} \right)^{-1} \right\|_{sp} = \frac{1}{\lambda_{\min}\left(\mathbf{\Lambda}_{xx}^{(k)} \right)} \to \infty \; a.s.$. A formal argument is provided in the proof of Lemma 2.

Given the lack of uniform invertibility of $\mathbf{\Lambda}_{xx}^{(k)}$, we impose the following assumption.

Assumption 3. There exists a deterministic sequence ϱ_k, with $\varrho_k \to \infty$, such that, as $k \to \infty$, $\left\| \left(\mathbf{\Lambda}_{xx}^{(k)} \right)^{-1} \right\|_{sp} = O_p(\varrho_k)$.

Let $\mathbf{U}_n(r) = \mathbf{\Phi}_0(1) \frac{1}{\sqrt{n}} \sum_{s=1}^{\lfloor nr \rfloor} \boldsymbol{\varepsilon}_s$, $\mathbf{V}_n^x(r) = \mathbf{\Phi}_x(1) \frac{1}{\sqrt{n}} \sum_{s=1}^{\lfloor nr \rfloor} \boldsymbol{\varepsilon}_s$, and $\mathbf{V}_n^x(r-) = \lim_{u \uparrow r} \mathbf{V}_n^x(u)$. By (5) and the continuous mapping theorem, we obtain

$$\int_0^1 d\mathbf{U}_n(r) \mathbf{V}_n^{x\top}(r-) \Rightarrow \int_0^1 d\underline{\mathbf{B}}_0(r) \underline{\mathbf{B}}_X^\top(r).$$

However, such weak convergence result is not enough for our purposes because it does not provide an approximation rate. We therefore impose the following strong invariance principle. Primitive conditions to establish results of this type can be found in Theorem 2.1 of Kifer (2026) and Theorem 2.2 of Friz and Kifer (2024).

Assumption 4. For some $\delta > 0$,

$$\sup_{r \in [0,1]} \left\| \int_0^r d\mathbf{U}_n(s) \mathbf{V}_n^{x\top}(s-) - \int_0^r d\underline{\mathbf{B}}_0(s) \underline{\mathbf{B}}_X^\top(s) \right\| = O_p\left(\frac{1}{n^\delta} \right).$$

The following proposition establishes the leading term in $n\left(\widehat{\mathbf{B}}_n^{(k)} - \mathbf{B}^{(k)} \right)$.

Proposition 1. *Suppose that Assumptions 1-4 are satisfied. Let* $k = \lfloor c \cdot n^\kappa \rfloor$ *for some* $c > 0$, *where* $0 < \kappa < a$, $a = \frac{p-2}{2p}$, *and* $p > 4$ *is as in Assumption 1. Suppose further that*

$$\frac{\varrho_k^2 k^{3/2}}{n^a} \to 0, \quad \frac{\varrho_k k^{3/2}}{n^{\min(\delta, 1/2)}} \to 0, \quad \frac{\varrho_k n}{k^q} \to 0,$$

where ϱ_k is as in Assumption 3 and $\delta > 0$ is as in Assumption 4. Then, as $n \to \infty$,

$$\left\| n \cdot \left(\widehat{\mathbf{B}}_n^{(k)} - \mathbf{B}^{(k)} \right) - \mathbf{\Gamma}_{ux}^{(k)} \left(\mathbf{\Lambda}_{xx}^{(k)} \right)^{-1} \right\|_{sp} = o_p(1),$$

where $\mathbf{\Lambda}_{xx}^{(k)}$ is defined as in Lemma 2 and

$$\mathbf{\Gamma}_{ux}^{(k)} := \int_0^1 d\underline{\mathbf{B}}_0(r) \left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r) \right)^\top + \int_0^1 \boldsymbol{\psi}_{(k)}^\top(r) \otimes \left(\mathbf{\Lambda}_{0x}(r) + \mathbf{\Sigma}_{0x}(r) \right) dr.$$

Remark 4. Three conditions $\frac{\varrho_k^2 k^{3/2}}{n^a} \to 0$, $\frac{\varrho_k k^{3/2}}{n^{\min(\delta,1/2)}} \to 0$, and $\frac{\varrho_k n}{k^q} \to 0$ are imposed so that the leading term in $n \cdot \left(\widehat{\mathbf{B}}_n^{(k)} - \mathbf{B}^{(k)} \right)$ is given by $\mathbf{\Gamma}_{ux}^{(k)} \left(\mathbf{\Lambda}_{xx}^{(k)} \right)^{-1}$. If we assume that $\varrho_k \asymp k^\gamma$ for some $\gamma > 0$, these conditions imply that

$$\frac{1}{q - \gamma} < \kappa < \min \left\{ \frac{p-2}{p(4\gamma+3)}, \frac{2\min(\delta, 1/2)}{2\gamma+3} \right\}.$$

For the above interval to be nonempty, we need an additional restriction on q:

$$q > \gamma + \max \left\{ \frac{p(4\gamma+3)}{p-2}, \frac{2\gamma+3}{2\min(\delta, 1/2)} \right\}.$$

In practice, k can be selected using the generalized cross–validation (GCV) method discussed in Gao et al. (2002), which we implement in both our empirical application and simulations.

As explained following (18), under the rate conditions imposed in Proposition 1, the deterministic approximation error is asymptotically negligible relative to the estimation error. Therefore, the leading term of $n \cdot \left\{ \widehat{\mathbf{A}}_n(z_0) - \mathbf{A}(z_0) \right\}$ is determined by $n \cdot \left(\widehat{\mathbf{B}}_n^{(k)} - \mathbf{B}^{(k)} \right) \mathbf{T}_{(k)}(z_0)$. By Proposition 1 and the fact that $\|\mathbf{T}_{(k)}(z_0)\| = O(\sqrt{k})$, it follows that

$$\left\| n \cdot \left(\widehat{\mathbf{A}}_n(z_0) - \mathbf{A}(z_0) \right) - \mathbf{\Gamma}_{ux}^{(k)} \left(\mathbf{\Lambda}_{xx}^{(k)} \right)^{-1} \mathbf{T}_{(k)}(z_0) \right\|_{sp} = o_p(\sqrt{k}). \tag{19}$$

Note that, the limiting distribution of the estimator need not be mixed normal, as for any $r, r' \in [0,1]$, we have

$$\operatorname{cov}\left(\underline{\mathbf{B}}_0(r), \underline{\mathbf{B}}_X(r') \right) = \int_0^{r \wedge r'} \mathbf{\Omega}_{0x}(s) ds,$$

which may not be a zero matrix. Under additional restrictions, however, a mixed normal limiting

distribution can be obtained. The results are summarized in the next theorem.

Theorem 1. *Suppose that conditions in Proposition 1 are satisfied and for almost every* $r \in [0,1]$, $\mathbf{\Omega}_{0x}(r) = \mathbf{0}_{d_y \times d_x}$ *and* $\mathbf{\Lambda}_{0x}(r) + \mathbf{\Sigma}_{0x}(r) = \mathbf{0}_{d_y \times d_x}$. *Then,*

$$n \left(\mathbb{V}_{(k)}(z_0)\right)^{-1/2} \operatorname{vec}\left(\widehat{\mathbf{A}}_n(z_0) - \mathbf{A}(z_0)\right) \Rightarrow \mathcal{N}\left(\mathbf{0}_{d_y d_x \times 1}, \mathbf{I}_{d_y d_x}\right), \tag{20}$$

for any fixed $z_0 \in [0,1]$, *where*

$$\begin{aligned}
&\mathbb{V}_{(k)}(z_0) \\
&= \int_0^1 \left\{ \left(\mathbf{T}_{(k)}^{\top}(z_0) \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1} \left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r)\right) \right) \left(\mathbf{T}_{(k)}^{\top}(z_0) \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1} \left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r)\right) \right)^{\top} \right\} \otimes \mathbf{\Omega}_{00}(r)\, dr.
\end{aligned}$$

Here, the estimator has a mixed normal limiting distribution. In addition, as shown in the proof of Theorem 1, $\left\| \left(\mathbb{V}_{(k)}(z_0)\right)^{-1/2} \right\|_{sp} = O_p(1) k^{-1/2}$, the order involved in (20) is therefore $O_p(1)\frac{n}{\sqrt{k}}$. This is comparable with the kernel estimate in the literature. Theorem 3.1 in Phillips et al. (2017) shows that the *type I* normalization is given by $nh^{1/2}$, where h is the bandwidth parameter. Thinking of k^{-1} as equivalent to the bandwidth h, the two orders in the two situations are quite comparable. However, due to kernel estimation degeneracy issues and rotation matrix introduced in their approach, the kernel estimate also involves a *type II* normalization, which is given by nh.

The above inferential results become operational if we can construct an estimated version of $\mathbb{V}_{(k)}(z_0)$. Following Yan et al. (2025), we may estimate $\mathbb{V}_{(k)}(z_0)$ by

$$\widehat{\mathbb{V}}_{(k)}(z_0) = \widehat{\mathbf{\Xi}}_{n,0}^{(k)}(z_0) + \sum_{j=1}^{\ell_n} \nu\left(j/\ell_n\right) \left[\widehat{\mathbf{\Xi}}_{n,j}^{(k)}(z_0) + \widehat{\mathbf{\Xi}}_{n,j}^{(k)\top}(z_0) \right], \tag{21}$$

where $\nu(\cdot)$ is a kernel function and ℓ_n is the lag truncation number satisfying $\ell_n \to \infty$ and $\ell_n/n \to 0$ as $n \to \infty$, and

$$\widehat{\mathbf{\Xi}}_{n,j}^{(k)}(z_0) = \frac{1}{n} \sum_{t=j+1}^{n} \left[\left(\mathbf{T}_{(k)}^{\top}(z_0) \left(\mathbf{\Lambda}_{n,xx}^{(k)}\right)^{-1} \frac{\mathbf{X}_t^{(k)}}{\sqrt{n}} \right) \left(\mathbf{T}_{(k)}^{\top}(z_0) \left(\mathbf{\Lambda}_{n,xx}^{(k)}\right)^{-1} \frac{\mathbf{X}_{t-j}^{(k)}}{\sqrt{n}} \right)^{\top} \right] \otimes \left[\widehat{\mathbf{u}}_{0t} \widehat{\mathbf{u}}_{0,t-j}^{\top} \right].$$

in which $\widehat{\mathbf{u}}_{0t}$ is the LS residuals from the regression (15), and $\mathbf{\Lambda}_{n,xx}^{(k)} := \frac{1}{n^2} \sum_{t=1}^{n} \mathbf{X}_t^{(k)} \mathbf{X}_t^{(k)\top}$.

3 — Efficient estimation

In the previous section, we establish the asymptotic theory of the SOLS estimator under non-stationary volatility. We now turn to efficient estimation of the time-varying cointegrating coefficients. To be better connected with the existing literature, throughout this section, we specialize to the stationary case by assuming $\mathbf{H}_t = \mathbf{H}$ for all t. Consequently, $\mathbf{\Sigma}_\varepsilon(r) = \mathbf{\Sigma}_\varepsilon$ and $\mathbf{\Sigma}(r)$, $\mathbf{\Lambda}(r)$, $\mathbf{\Omega}(r)$ reduces to $\mathbf{\Sigma}$, $\mathbf{\Lambda}$, $\mathbf{\Omega}$, respectively. In addition, throughout this section, we denote $\mathbf{B}(r) = \left(\mathbf{B}_0^\top(r), \mathbf{B}_X^\top(r)\right)^\top$ as a $(d_y + d_x)$-dimensional vector Brownian motion with covariance matrix $\mathbf{\Omega}$ and $\mathbf{B}_{0\cdot x}(r) := \mathbf{B}_0(r) - \mathbf{\Omega}_{0x}\mathbf{\Omega}_{xx}^{-1}\mathbf{B}_X(r)$.

3.1. Preliminaries

In general, for any fixed $z_0 \in [0,1]$, the series estimator $\widehat{\mathbf{A}}_n(z_0)$ has a nonstandard limiting distribution. Hence, the conventional minimum variance approach to assess asymptotic efficiency is not directly applicable. In this section, we adapt the efficiency framework of Saikkonen (1991) for fixed-coefficient cointegrated system to the present time-varying setting. The framework is formulated in terms of the concentration, or peakedness, of the limiting distribution over an appropriate class of estimators.

For any fixed $z_0 \in [0,1]$, $\mathbf{A}_k(z_0) = \mathbf{B}^{(k)}\left(\boldsymbol{\psi}_{(k)}(z_0) \otimes \mathbf{I}_{d_x}\right)$. Accordingly, we define the relevant class in terms of the estimators of $\mathbf{B}^{(k)}$. Throughout this and the next subsection, k denotes a sequence of integers satisfying $k \to \infty$ as $n \to \infty$. It is further understood that k satisfies rate conditions required for the expansions of the estimators under consideration, like the ones imposed in Proposition 1.

Definition 1. For any fixed $z_0 \in [0,1]$, let $\widetilde{\mathbf{A}}_n(z_0) = \widetilde{\mathbf{B}}_n^{(k)}\mathbf{T}_{(k)}(z_0)$ be a generic series estimator of $\mathbf{A}(z_0)$ in (1). We say that $\widetilde{\mathbf{A}}_n(z_0)$ belongs to the class $\mathcal{S}(z_0)$ if, as $n \to \infty$, $\widetilde{\mathbf{B}}_n^{(k)}$ admits the following

$$\left\| n \cdot \left(\widetilde{\mathbf{B}}_n^{(k)} - \mathbf{B}^{(k)}\right) - \left[\left(\int_0^1 d\mathbf{B}_0(r)\mathbf{L}_{(k)}^\top(r)\right)\mathbf{V}_{(k)} + \mathbf{U}_{(k)}\right]\right\|_{sp} = o_p(1). \tag{22}$$

Here, $\mathbf{T}_{(k)}(z_0) = \boldsymbol{\psi}_{(k)}(z_0) \otimes \mathbf{I}_{d_x}$. $\mathbf{L}_{(k)}(r)$ is a vector-valued, $\mathcal{F}_X(r)$-adapted stochastic process such that the indicated stochastic integral is well defined, where $\mathcal{F}_X(r) = \sigma\left(\mathbf{B}_X(s) : 0 \le s \le r \le 1\right)$.

$\mathbf{V}_{(k)}$ and $\mathbf{U}_{(k)}$ are $\mathcal{F}_X(1)$-measurable random matrices of conformable dimensions. The subscript (k) in $\mathbf{L}_{(k)}(r)$, $\mathbf{V}_{(k)}$, and $\mathbf{U}_{(k)}$ indicate that they involve the basis vector $\boldsymbol{\psi}_{(k)}(r)$. Furthermore, for each fixed $z_0 \in [0,1]$,

$$\left(\int_0^1 \left(\boldsymbol{\psi}_{(k)}(r) \otimes \mathbf{B}_X(r)\right) \mathbf{L}_{(k)}^{\top}(r)\, dr\right) \mathbf{V}_{(k)} \mathbf{T}_{(k)}(z_0) = \mathbf{T}_{(k)}(z_0), \qquad a.s. \tag{23}$$

To motivate the class $\mathcal{S}(z_0)$ defined in Definition 1, consider an estimator of $\mathbf{B}^{(k)}$ which is of the general form

$$\widetilde{\mathbf{B}}_n^{(k)} = \left(\sum_{t=1}^{n} \mathbf{y}_t \mathbf{w}_t^{\top(k)}\right) \mathbf{V}_n^{(k)} + \mathbf{U}_n^{(k)},$$

where the dimensions of $\mathbf{w}_t^{(k)}$, $\mathbf{V}_n^{(k)}$, and $\mathbf{U}_n^{(k)}$ are the same as $\mathbf{L}_{(k)}(r)$, $\mathbf{V}_{(k)}$, and $\mathbf{U}_{(k)}$, respectively. Clearly, if we choose $\mathbf{w}_t^{(k)} = \mathbf{X}_t^{(k)}$, $\mathbf{U}_n^{(k)} = \mathbf{0}_{d_y \times k d_x}$, and $\mathbf{V}_n^{(k)} = \left(\sum_{t=1}^{n} \mathbf{X}_t^{(k)} \mathbf{X}_t^{(k)\top}\right)^{-1}$, we obtain the SOLS estimator defined in (16). Furthermore, let

$$\begin{aligned} \mathbf{y}_t^* &= \mathbf{y}_t - \mathbf{B}^{(k)} \left\{\boldsymbol{\psi}_{(k)}(\tau_t) \otimes \left(\mathbf{C}^{\top} \mathbf{u}_t\right)\right\} - \mathbf{D}^{\top} \mathbf{u}_t, \\ \mathbf{X}_t^* &= \mathbf{X}_t - \mathbf{C}^{\top} \mathbf{u}_t, \end{aligned} \tag{24}$$

where

$$\mathbf{D} = \left[\mathbf{0}_{d_y \times d_y}, \boldsymbol{\Omega}_{0x} \boldsymbol{\Omega}_{xx}^{-1}\right]^{\top}, \quad \mathbf{C} = \boldsymbol{\Sigma}^{-1} \begin{bmatrix} \boldsymbol{\Lambda}_{0x} + \boldsymbol{\Sigma}_{0x} \\ \boldsymbol{\Lambda}_{xx} + \boldsymbol{\Sigma}_{xx} \end{bmatrix}.$$

Then, if we choose

$$\begin{aligned} \mathbf{w}_t^{(k)} &= \boldsymbol{\psi}_{(k)}(\tau_t) \otimes \mathbf{X}_t^*, \;\; \mathbf{V}_n^{(k)} = \left(\sum_{t=1}^{n} \mathbf{w}_t^{(k)} \mathbf{w}_t^{(k)\top}\right)^{-1}, \\ \mathbf{U}_n^{(k)} &= -\left(\sum_{t=1}^{n} \left[\mathbf{B}^{(k)} \left\{\boldsymbol{\psi}_{(k)}(\tau_t) \otimes \left(\mathbf{C}^{\top} \mathbf{u}_t\right)\right\} + \mathbf{D}^{\top} \mathbf{u}_t\right] \mathbf{w}_t^{(k)\top}\right) \mathbf{V}_n^{(k)}. \end{aligned}$$

we obtain the multivariate extension of the series-based canonical cointegration regressions estimator (SCCR) considered in Park and Hahn (1999).

Based on the previous discussion, $\mathbf{w}_t^{(k)}$ should be highly correlated with $\mathbf{X}_t^{(k)}$. In addition, the formulation of SCCR estimator also suggests that $\mathbf{U}_n^{(k)}$ may also depend on $\mathbf{B}^{(k)}$. Suppose that the

dependence of $\mathbf{U}_n^{(k)}$ on $\mathbf{B}^{(k)}$ has the linear form

$$\mathbf{U}_n^{(k)} = \mathbf{B}^{(k)}\mathbf{R}_n^{(k)} + \overline{\mathbf{U}}_n^{(k)},$$

where $\mathbf{R}_n^{(k)}$ is a $kd_x \times kd_x$ random matrix and $\overline{\mathbf{U}}_n^{(k)}$ is a $d_y \times kd_x$ random matrix that does not depend on $\mathbf{B}^{(k)}$. Then,

$$\begin{aligned}\widetilde{\mathbf{A}}_n(z_0) - \mathbf{A}(z_0) = \mathbf{B}^{(k)} &\left[\left(\sum_{t=1}^n \mathbf{X}_t^{(k)}\mathbf{w}_t^{\top(k)}\right)\mathbf{V}_n^{(k)} + \mathbf{R}_n^{(k)} - \mathbf{I}_{kd_x}\right]\mathbf{T}_{(k)}(z_0) \\ &+ \left(\sum_{t=1}^n \mathbf{u}_{0t}^{(k)}\mathbf{w}_t^{\top(k)}\right)\mathbf{V}_n^{(k)}\mathbf{T}_{(k)}(z_0) + \overline{\mathbf{U}}_n^{(k)}\mathbf{T}_{(k)}(z_0) + O(k^{-q+1/2}).\end{aligned}$$

To avoid dealing with the dependence on $\mathbf{B}^{(k)}$, the estimator of $\mathbf{B}^{(k)}$ must be constructed such that

$$\left(\sum_{t=1}^n \mathbf{X}_t^{(k)}\mathbf{w}_t^{\top(k)}\right)\mathbf{V}_n^{(k)} + \mathbf{R}_n^{(k)} - \mathbf{I}_{kd_x} = \mathbf{0}_{kd_x\times kd_x}. \tag{25}$$

We can easily verify that both SOLS and SCCR estimators satisfy the condition (25). It is also worth mentioning (23) in Definition 1 is a normalization condition used in the characterization of asymptotic efficiency developed in the next subsection. It is straightforward to verify that this condition is satisfied by both the SOLS and SCCR estimators.

Then, under (25), the expansion for $\widetilde{\mathbf{B}}_n^{(k)}$ simplifies to

$$n\cdot\left(\widetilde{\mathbf{B}}_n^{(k)} - \mathbf{B}^{(k)}\right) = \left(\frac{1}{n}\sum_{t=1}^n \mathbf{u}_{0t}^{(k)}\mathbf{w}_t^{\top(k)}\right)\left(n^2\mathbf{V}_n^{(k)}\right) + n\,\overline{\mathbf{U}}_n^{(k)}.$$

Then, (22) is justified if

$$\begin{aligned}&\left\|\left(\frac{1}{n}\sum_{t=1}^n \mathbf{u}_{0t}^{(k)}\mathbf{w}_t^{\top(k)}\right)\left(n^2\mathbf{V}_n^{(k)}\right) - \left(\int_0^1 d\mathbf{B}_0(r)\mathbf{L}_{(k)}^{\top}(r)\right)\mathbf{V}_{(k)}\right\|_{sp} = o_p(1), \\ &\left\|n\overline{\mathbf{U}}_n^{(k)} - \mathbf{U}_{(k)}\right\|_{sp} = o_p(1).\end{aligned}$$

These results assumed above are known to hold under certain regularity conditions. For example, under the conditions imposed in Theorem 1, they are satisfied for the SOLS estimator.

Related results are also established by Park and Hahn (1999) for the SCCR estimator in single equation case ($d_y = 1$) under similar assumptions. Thus, the class $\mathcal{S}(z_0)$ is sufficiently broad to include the series-based estimators considered in the literature.

3.2. Characterization of asymptotic efficiency

The following efficiency criterion is adapted from Saikkonen (1991).

Definition 2. For any fixed $z_0 \in [0,1]$, let $\widetilde{\mathbf{A}}_n(z_0)$ and $\widetilde{\mathbf{A}}^*_n(z_0)$ be two series estimators of $\mathbf{A}(z_0)$ belonging to the class $\mathcal{S}(z_0)$. We say that $\widetilde{\mathbf{A}}_n(z_0)$ is asymptotically more efficient than $\widetilde{\mathbf{A}}^*_n(z_0)$ if, for any convex set $\mathcal{C} \subset \mathbb{R}^{d_y d_x}$, symmetric around the origin,

$$\limsup_{n\to\infty} \left[\mathbb{P}\left\{n \cdot \left(\mathbb{V}^e_{(k)}(z_0)\right)^{-1/2} \widetilde{\mathbf{R}}^*_n(z_0) \in \mathcal{C}\right\} - \mathbb{P}\left\{n \cdot \left(\mathbb{V}^e_{(k)}(z_0)\right)^{-1/2} \widetilde{\mathbf{R}}_n(z_0) \in \mathcal{C}\right\}\right] \leq 0,$$

where

$$\widetilde{\mathbf{R}}^*_n(z_0) = \operatorname{vec}\left(\widetilde{\mathbf{A}}^*_n(z_0) - \mathbf{A}(z_0)\right), \quad \widetilde{\mathbf{R}}_n(z_0) = \operatorname{vec}\left(\widetilde{\mathbf{A}}_n(z_0) - \mathbf{A}(z_0)\right).$$

Here,

$$\mathbb{V}^e_{(k)}(z_0) = \left\{\mathbf{T}^\top_{(k)}(z_0)\left(\mathbf{\Lambda}^{(k)}_{xx}\right)^{-1}\mathbf{T}_{(k)}(z_0)\right\} \otimes \mathbf{\Omega}_{0\cdot x},$$

with $\mathbf{\Omega}_{0\cdot x} = \mathbf{\Omega}_{00} - \mathbf{\Omega}_{0x}\mathbf{\Omega}^{-1}_{xx}\mathbf{\Omega}_{x0}$. If the inequality holds for all $\widetilde{\mathbf{A}}^*_n(z_0) \in \mathcal{S}(z_0)$, $\widetilde{\mathbf{A}}_n(z_0)$ is said to be asymptotically efficient within the class $\mathcal{S}(z_0)$.

In Definition 2, asymptotic efficiency between two estimators is measured by concentration, or peakedness, over a convex symmetric set $\mathcal{C}$. However, several issues are worth mentioning. First, by the construction of the estimators in $\mathcal{S}(z_0)$, the order of both $n \cdot \widetilde{\mathbf{R}}^*_n(z_0)$ and $n \cdot \widetilde{\mathbf{R}}_n(z_0)$ depends on k. Concentration probabilities over such bounded $\mathcal{C}$ may tend to zero, making the unnormalized comparison uninformative. We therefore normalize both terms by $\left(\mathbb{V}^e_{(k)}(z_0)\right)^{-1/2}$, which will be made clear later that this is the efficient random variance matrix. Second, by the construction of the estimator, the leading terms are k-dependent. This differs from the fixed coefficient cointegrated system considered in Saikkonen (1991), where it can be argued that the associated concentration probabilities can be compared through ordinary limits. We therefore formulate the efficiency comparison using the $\limsup$ of the difference between the two concentration probabilities. This

formulation does not require either probability sequence to have an ordinary limit and ensures that the two estimators are compared along the common sequence indexed by n, with $k := k_n$.

In the next theorem, we derive an upper bound for the lim sup concentration probabilities of arbitrary estimators in $\mathcal{S}(z_0)$ as defined in Definition 2. To state compactly the condition under which this bound is strict, define

$$\boldsymbol{\phi}_{(k)}(z_0) := \left(\mathbb{V}^e_{(k)}(z_0)\right)^{-1/2} \operatorname{vec}\left(\left[\int_0^1 d\mathbf{B}_{0\cdot x}(r)\left(\boldsymbol{\psi}_{(k)}(r) \otimes \mathbf{B}_X(r)\right)^\top\right]\left(\boldsymbol{\Lambda}^{(k)}_{xx}\right)^{-1}\mathbf{T}_{(k)}(z_0)\right).$$

Clearly, by construction, we have

$$\boldsymbol{\phi}_{(k)}(z_0)|\mathcal{F}_X(1) \equiv \mathcal{N}\left(\mathbf{0}_{d_y d_x \times 1}, \mathbf{I}_{d_y d_x}\right), \qquad a.s.$$

Theorem 2. *For any fixed $z_0 \in [0,1]$, let $\widetilde{\mathbf{A}}_n(z_0) \in \mathcal{S}(z_0)$ be an arbitrary estimator of $\mathbf{A}(z_0)$. Let $\boldsymbol{\xi} \sim \mathcal{N}\left(\mathbf{0}_{d_y d_x \times 1}, \mathbf{I}_{d_y d_x}\right)$ denote a generic standard Normal random vector.*

(i) Let $\mathcal{C}$,$\mathbb{V}^e_{(k)}(z_0)$, and $\widetilde{\mathbf{R}}_n(z_0)$ be defined as in Definition 2. Then,

$$\limsup_{n\to\infty} \mathbb{P}\left\{n\ \left(\mathbb{V}^e_{(k)}(z_0)\right)^{-1/2}\widetilde{\mathbf{R}}_n(z_0) \in \mathcal{C}\right\} \leq \mathbb{P}\left\{\boldsymbol{\xi} \in \mathcal{C}\right\}.$$

(ii) Suppose that there exists $\varepsilon_0 > 0$, such that

$$\liminf_{n\to\infty} \mathbb{P}\left\{\left\|n\ \left(\mathbb{V}^e_{(k)}(z_0)\right)^{-1/2}\widetilde{\mathbf{R}}_n(z_0) - \boldsymbol{\phi}_{(k)}(z_0)\right\| > \varepsilon_0\right\} > 0. \tag{26}$$

Then, for every bounded convex set $\mathcal{C}$ symmetrically around the origin with $\mathbb{P}\{\boldsymbol{\xi} \in \mathcal{C}\} > 0$,

$$\limsup_{n\to\infty} \mathbb{P}\left\{n\ \left(\mathbb{V}^e_{(k)}(z_0)\right)^{-1/2}\widetilde{\mathbf{R}}_n(z_0) \in \mathcal{C}\right\} < \mathbb{P}\left\{\boldsymbol{\xi} \in \mathcal{C}\right\}.$$

By Theorem 2, for any fixed $z_0 \in [0,1]$, if an estimator $\widetilde{\mathbf{A}}_n(z_0) \in \mathcal{S}(z_0)$ satisfies

$$n\ \left(\mathbb{V}^e_{(k)}(z_0)\right)^{-1/2} \operatorname{vec}\left(\widetilde{\mathbf{A}}_n(z_0) - \mathbf{A}(z_0)\right) \Rightarrow \boldsymbol{\xi}, \tag{27}$$

where $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}_{d_y d_x \times 1}, \mathbf{I}_{d_y d_x})$, then it is asymptotically efficient within $\mathcal{S}(z_0)$. Since Theorem 10 of Park and Hahn (1999) establishes (27) for their SCCR estimator in the single-equation case

$(d_y = 1)$, that estimator is asymptotically efficient within $\mathcal{S}(z_0)$. Moreover, under the conditions of Theorem 1, and the stationary innovations setting maintained in this section, for any fixed $z_0 \in [0, 1]$, SOLS estimator is also asymptotically efficient within $\mathcal{S}(z_0)$ in the pure cointegration case where $\mathbf{\Omega}_{0x} = \mathbf{0}_{d_y \times d_x}$ and $\mathbf{\Lambda}_{0x} + \mathbf{\Sigma}_{0x} = \mathbf{0}_{d_y \times d_x}$.

Although an efficient estimator already exists in the literature, two issues remain. First, the interpretation of the SCCR transformation in (24) is less direct than that of the conventional CCR transformation introduced by Park (1992). In the fixed-coefficient setting, when both $\mathbf{y}_t$ and $\mathbf{X}_t$ are integrated processes, transformations that differ from them only by stationary components preserve the same cointegrating relationship or, more precisely, the same cointegrating matrix $\left[\mathbf{I}_{d_y}, -\mathbf{A}\right]$. In the present setting, however, the cointegrating matrix varies over time, and $\mathbf{y}_t$ need not be integrated in the conventional sense.[2] Although the transformation in (24) is algebraically constructed so that the transformed regression retains the same series coefficient matrix $\mathbf{B}^{(k)}$, its interpretation as a canonical transformation of an integrated system is therefore less immediate. Second, (24) involves the unknown quantities $\mathbf{B}^{(k)}$, $\mathbf{\Omega}_{0x}$, $\mathbf{\Omega}_{xx}$, $\mathbf{\Lambda}_{0x}$, $\mathbf{\Sigma}_{0x}$, $\mathbf{\Lambda}_{xx}$, and $\mathbf{\Sigma}_{xx}$, all of which must be estimated before the transformation can be implemented. Thus, the SCCR estimator is inherently a two-step procedure, and the quality of the initial estimation may be crucial for its finite-sample performance.

In the next subsection, we develop a new series-based estimator that achieves asymptotic efficiency through a simple time domain correction, thereby avoiding both the transformation of $\mathbf{y}_t$ and $\mathbf{X}_t$ and the preliminary estimation of the model.

3.3. Series-based dynamic OLS estimator

To illustrate the idea, suppose that $d_y = d_x = 1$ and $\mathbf{u}_t = (u_{0t}, u_{xt})^\top$ is an $i.i.d.$ bivariate normal vector with a full covariance matrix. By the property of multivariate normal distribution, the linear projection of u_{0t} on u_{xt} is simply $u_{0t} = \lambda u_{xt} + v_t$, where v_t is uncorrelated with u_{xt}. The model can then be written as $y_t = A_t X_t + \lambda u_{xt} + v_t$. Since $\mathbf{u}_t$ is $i.i.d.$ over time, v_t is also

2. To see this, note that from (1),

$$\Delta \mathbf{y}_t = \mathbf{A}_t \mathbf{u}_{xt} + (\mathbf{A}_t - \mathbf{A}_{t-1}) \mathbf{X}_{t-1} + \Delta \mathbf{u}_{0t}.$$

The term $\Delta \mathbf{u}_{0t}$ is stationary under the maintained assumptions. Moreover, since $\mathbf{A}_t$ is sufficiently smooth, the term $(\mathbf{A}_t - \mathbf{A}_{t-1}) \mathbf{X}_{t-1}$ is asymptotically negligible. However, $\mathbf{A}_t \mathbf{u}_{xt}$ generally has time-varying second-order moments. Consequently, $\Delta \mathbf{y}_t$ need not be stationary, and $\mathbf{y}_t$ need not be integrated in the conventional sense.

uncorrelated with u_{xs} for all s, and therefore with X_t. We can simply run OLS estimation to obtain an efficient estimator of $A(z_0)$ for any fixed $z_0 \in [0,1]$. In the fixed-coefficient case, closely related augmentations have been used in the predictive regression literature to remove the so-called embedded endogeneity; see Amihud and Hurvich (2004); Amihud et al. (2008). Cai and Wang (2014) also use this idea to handle endogeneity in predictive regressions with nonstationary regressors, although their framework assumes that $\mathbf{u}_t$ and v_t are strictly stationary and strong mixing. In the present setting, however, $\mathbf{u}_t$ is generated by the linear process in (3). A contemporaneous augmentation by $u_{xt} = \Delta X_t$ alone is then insufficient, because u_{0t} may be correlated with $u_{x,t+j}$ for any j.

The above discussion motivates us to consider the general linear projection of the form

$$\mathbf{u}_{0t} = \sum_{j=-\infty}^{\infty} \mathbf{\Pi}_j \mathbf{u}_{x,t-j} + \mathbf{v}_t, \tag{28}$$

where $\sum_{j=-\infty}^{\infty} \|\mathbf{\Pi}_j\| < \infty$ and $\mathbf{v}_t$ is a stationary process satisfying $\mathbb{E}\left[\mathbf{u}_{xt}\mathbf{v}_{t+j}^\top\right] = \mathbf{0}_{d_x \times d_y}$ for all j. The representation in (28), under the maintained stationarity of $\mathbf{u}_t = \left(\mathbf{u}_{0t}^\top, \mathbf{u}_{xt}^\top\right)^\top$, follows from Theorem 8.3.1 of Brillinger (1975), which gives conditions under which the linear projection of one stationary process onto the closed linear span generated by another stationary process admits a two-sided linear filter representation. For the process generated by (3) under homoskedasticity, these conditions reduce to standard spectral regularity conditions, whose verification is postponed to the proof of Theorem 3.

By plugging (28) back to (1), we have

$$\mathbf{y}_t = \mathbf{A}_t \mathbf{X}_t + \sum_{j=-\infty}^{\infty} \mathbf{\Pi}_j \mathbf{u}_{x,t-j} + \mathbf{v}_t.$$

Given $\sum_{j=-\infty}^{\infty} \|\mathbf{\Pi}_j\| < \infty$, $\mathbf{\Pi}_j$ would be negligible for $|j| > K$ with sufficiently large K. This suggests that we can truncate the doubly infinite sum in (28) at $|j| = K$. Inserting this approximation, together with the series truncation (13), we obtain the regression

$$\mathbf{y}_t = \mathbf{B}^{(k)} \mathbf{X}_t^{(k)} + \sum_{j=-K}^{K} \mathbf{\Pi}_j \Delta \mathbf{X}_{t-j} + \widetilde{\mathbf{v}}_t^{(k)}, \tag{29}$$

where $\widetilde{\mathbf{v}}_t^{(k)} = \mathbf{v}_t + \sum_{|j|>K} \mathbf{\Pi}_j \mathbf{u}_{x,t-j} + (\mathbf{A}_t - \mathbf{A}_k(\tau_t))\,\mathbf{X}_t$. As in Section 2, the following strong invariance principle is needed.

Assumption 5. For some $\delta_* > 0$,

$$\sup_{r\in[0,1]} \left\| \int_0^r d\mathbf{U}_n^*(s)\mathbf{V}_n^{x\top}(s-) - \int_0^r d\mathbf{B}_v(s)\mathbf{B}_X^\top(s) \right\| = O_p\left(\frac{1}{n^{\delta_*}}\right),$$

where $\mathbf{U}_n^*(r) := \frac{1}{\sqrt{n}} \sum_{s=1}^{\lfloor nr \rfloor} \mathbf{v}_s$, $\mathbf{B}_v(r) := \mathbf{B}_0(r) - \mathbf{\Omega}_{0x}\mathbf{\Omega}_{xx}^{-1}\mathbf{B}_X(r)$, and $\mathbf{V}_n^{x\top}(s-)$ is defined as in Section 2.

Let $\mathbf{Z}_t^{(K)} = \left(\Delta\mathbf{X}_{t+K}^\top, \cdots, \Delta\mathbf{X}_{t-K}^\top\right)^\top$. $\mathbf{B}^{(k)}$ and $\mathbf{\Pi}^{(K)} = [\mathbf{\Pi}_{-K}, \cdots, \mathbf{\Pi}_K]$ can be estimated by LS:

$$\begin{bmatrix} \widetilde{\mathbf{B}}_n^{(k)} & \widehat{\mathbf{\Pi}}_n^{(K)} \end{bmatrix} = \begin{bmatrix} \sum_{t=K+1}^{n-K} \mathbf{y}_t \mathbf{X}_t^{(k)\top} & \sum_{t=K+1}^{n-K} \mathbf{y}_t \mathbf{Z}_t^{(K)\top} \end{bmatrix} \begin{bmatrix} \sum_{t=K+1}^{n-K} \mathbf{X}_t^{(k)} \mathbf{X}_t^{(k)\top} & \sum_{t=K+1}^{n-K} \mathbf{X}_t^{(k)} \mathbf{Z}_t^{(K)\top} \\ \sum_{t=K+1}^{n-K} \mathbf{Z}_t^{(K)} \mathbf{X}_t^{(k)\top} & \sum_{t=K+1}^{n-K} \mathbf{Z}_t^{(K)} \mathbf{Z}_t^{(K)\top} \end{bmatrix}^{-1}.$$

Then, for any $z_0 \in [0,1]$, $\widetilde{\mathbf{A}}_n(z_0) = \widetilde{\mathbf{B}}_n^{(k)} \mathbf{T}_{(k)}(z_0)$. Following the convention in the fixed-coefficient cointegration regression literature (Saikkonen, 1991; Stock and Watson, 1993), we refer to $\widetilde{\mathbf{A}}_n(z_0)$ as the series-based dynamic OLS (SDOLS) estimator. The limiting distribution of the SDOLS estimator is provided in the next theorem.

Theorem 3. *Suppose that Assumptions 1, 2, 3, and 5 are satisfied and the spectral density matrix* $\mathbf{f}_{uu}(\lambda)$[3] *satisfies* $\inf_{\lambda\in[-\pi,\pi]} \lambda_{\min}\left(\mathbf{f}_{uu}(\lambda)\right) > 0$. *Let* $k = \lfloor c \cdot n^\kappa \rfloor$ *for some* $c > 0$, *where* $0 < \kappa < a$, $a = \frac{p-2}{2p}$, *and* $p > 4$ *is as in Assumption 1. Suppose further that*

$$\frac{\varrho_k^2 k^{3/2}}{n^a} \to 0, \quad \frac{\varrho_k k^{3/2}}{n^{\delta_*}} \to 0, \quad \frac{\varrho_k n}{k^q} \to 0, \tag{30}$$

where ϱ_k *is as in Assumption 3 and* $\delta_* > 0$ *is as in Assumption 5. If, in addition,* $\varrho_k K\sqrt{\frac{k}{n}} \to 0$ *and* $\varrho_k\sqrt{kn}\sum_{|j|>K}\|\mathbf{\Pi}_j\| \to 0$. *Then, for any given* $z_0 \in [0,1]$, *we have, as* $n \to \infty$

$$n\,\left(\mathbb{V}_{(k)}^e(z_0)\right)^{-1/2} \mathrm{vec}\left(\widetilde{\mathbf{A}}_n(z_0) - \mathbf{A}(z_0)\right) \Rightarrow \mathcal{N}\left(\mathbf{0}_{d_y d_x \times 1}, \mathbf{I}_{d_y d_x}\right), \tag{31}$$

3. $\mathbf{f}_{uu}(\lambda)$ denotes the spectral density matrix of $\mathbf{u}_t$, partitioned conformably with $\mathbf{u}_t = \left(\mathbf{u}_{0t}^\top, \mathbf{u}_{xt}^\top\right)^\top$. For $a, b \in \{0, x\}$, let $\mathbf{f}_{u_a u_b}(\lambda)$ denote the corresponding (a,b) block.

where

$$\mathbb{V}^{e}_{(k)}(z_0) := \left\{ \mathbf{T}^{\top}_{(k)}(z_0) \left(\mathbf{\Lambda}^{(k)}_{xx} \right)^{-1} \mathbf{T}_{(k)}(z_0) \right\} \otimes \mathbf{\Omega}_{0\cdot x},$$

$\mathbf{\Omega}_{0\cdot x} = \mathbf{\Omega}_{00} - \mathbf{\Omega}_{0x}\mathbf{\Omega}^{-1}_{xx}\mathbf{\Omega}_{x0}$*, and* $\mathbf{\Lambda}^{(k)}_{xx}$ *is defined as in Lemma 2, with* $\underline{\mathbf{B}}_X(\cdot)$ *replaced by* $\mathbf{B}_X(\cdot)$*.*

Theorem 3 shows that the limiting distribution of the SDOLS estimator is mixed Normal. The estimator can be viewed as an extension of the dynamic OLS estimator of Saikkonen (1991) and Stock and Watson (1993) to the present time-varying cointegration setting. In view of Theorem 2, the SDOLS estimator is also asymptotically efficient just as the SCCR estimator. However, because the correction of the source of nonstandard inference and final computation of the estimator of $\mathbf{B}^{(k)}$ are accomplished simultaneously in the time domain, no initial estimation of the model is required.

There are two additional conditions imposed on K. The condition $\varrho_k K \sqrt{\frac{k}{n}} \to 0$ specifies an upper bound for the rate at which K is allowed to grow with n, whereas $\varrho_k \sqrt{kn} \sum_{|j|>K} \|\mathbf{\Pi}_j\| \to 0$ gives an analogous lower bound so that truncation has no impact on the estimator. The Monte Carlo evidence reported in Section F suggests that a moderate value of K is generally advisable in practice.

Remark 5. To illustrate the joint feasibility of these conditions imposed in Theorem 3, suppose, as in Remark 4, that $\varrho_k \asymp k^{\gamma}$ for some $\gamma > 0$. Suppose further that $K = \lfloor c_K n^{\zeta} \rfloor$ for some $c_K > 0$ and that, for some $\beta > 0$,

$$\sum_{|j|>K} \|\mathbf{\Pi}_j\| = O(K^{-\beta}).$$

Since $k \asymp n^{\kappa}$, the two conditions involving K are satisfied whenever

$$\frac{1 + \kappa(2\gamma + 1)}{2\beta} < \zeta < \frac{1 - \kappa(2\gamma + 1)}{2}.$$

Such a choice of ζ exists provided that $\beta > 1$ and

$$\kappa < \frac{\beta - 1}{(\beta + 1)(2\gamma + 1)}.$$

Combining this restriction with those in (30) gives

$$\frac{1}{q-\gamma} < \kappa < \min\left\{\frac{p-2}{p(4\gamma+3)}, \frac{2\delta_*}{2\gamma+3}, \frac{\beta-1}{(\beta+1)(2\gamma+1)}\right\}.$$

For the above interval to be nonempty, it suffices to have

$$q > \gamma + \max\left\{\frac{p(4\gamma+3)}{p-2}, \frac{2\gamma+3}{2\delta_*}, \frac{(\beta+1)(2\gamma+1)}{\beta-1}\right\}.$$

Finally, under the maintained assumption on stationary innovations, inference can be made operational in a similar way to Theorem 1. We can replace $\mathbf{\Lambda}_{xx}^{(k)}$ with the sample version $\frac{1}{n^2}\sum_{t=1}^{n}\mathbf{X}_t^{(k)}\mathbf{X}_t^{(k)\top}$ and find a consistent estimator of $\mathbf{\Omega}_{0\cdot x}$. Let $\widehat{\mathbf{v}}_t = \mathbf{y}_t - \widetilde{\mathbf{B}}_n^{(k)}\mathbf{X}_t^{(k)} - \widehat{\mathbf{\Pi}}_n^{(K)}\mathbf{Z}_t^{(K)}$ be the LS residuals from the regression (29), we may construct the following estimator:

$$\widehat{\mathbf{\Omega}}_{0\cdot x} = \frac{1}{n}\sum_{|j|\leq \ell_n} \nu\left(j/\ell_n\right) \sum_{t=\max\{K+1,K+j+1\}}^{\min\{n-K,n+j-K\}} \widehat{\mathbf{v}}_t\widehat{\mathbf{v}}_{t-j}^{\top}, \tag{32}$$

where $\nu(\cdot)$ is the same kernel function as in (21), and ℓ_n is the lag truncation parameter satisfying $\ell_n \to \infty$ as $n \to \infty$.

To complement the asymptotic results, Section F in the Online Supplement contains a Monte Carlo study to assess the finite-sample performance of the proposed estimators. Across a range of coefficient dynamics and specifications for the innovations, SDOLS with small to moderate augmentation orders generally improves estimation accuracy relative to SOLS and SCCR and produces pointwise confidence intervals with empirical coverage closer to the nominal level. Including leads of Δx_t often provides additional gains. These patterns remain broadly similar for coefficient paths that satisfy Assumption 2 and those that violate its endpoint compatibility conditions.

4 — An empirical application

We illustrate the proposed method in an empirical application to the consumption–wealth ratio (cay) and examine its predictive ability to the U.S. quarterly equity premium. The cay variable, originally proposed by Lettau and Ludvigson (2001), measures the deviation in the common trend

between consumption (c), aggregate wealth (a), and labor income (y):

$$\text{cay}_t = c_t - \omega a_t - (1-\omega) y_t, \tag{33}$$

where ω equals the average share of asset holdings in total wealth. Under certain conditions, it can be shown that

$$\text{cay}_t = \mathbb{E}_t \sum_{i=1}^{\infty} \rho_\omega^i \left\{ [\omega r_{a,t+i} + (1-\omega) r_{h,t+i}] - \Delta c_{t+i} \right\} + (1-\omega) z_t, \tag{34}$$

where $\mathbb{E}_t$ is the conditional expectation on information available at time t, ρ_ω is the steady-state ratio of new investment to total wealth, $r_{a,t}$ is the return on asset holdings, $r_{h,t}$ is the return on human capital, and z_t is a mean zero stationary random variable. (34) suggests that the consumption-wealth ratio cay_t may contain information about future asset returns $r_{a,t+i}$. In particular, its predictive content is stronger when future returns on human capital $r_{h,t+i}$ and consumption growth Δc_{t+i} exhibit limited volatility, or these variables are closely related to expected returns on assets.

An important first step in using cay to forecast asset returns is to estimate the cointegrating relationship among consumption, aggregate wealth, and labor income as specified in (33). Lettau and Ludvigson (2001) estimate a fixed coefficient version of this relationship using the dynamic ordinary least squares (DOLS) procedure of Stock and Watson (1993), which augments the cointegrating regression with leads and lags of the first differences of aggregate wealth and labor income. Given the resulting estimates $\widehat{\beta}_a$ and $\widehat{\beta}_y$, the conventional measure is constructed as

$$\widehat{\text{cay}}_t = c_t - \widehat{\beta}_a a_t - \widehat{\beta}_y y_t,$$

which measures the deviation of consumption from its estimated common trend with aggregate wealth and labor income.

Although Lettau and Ludvigson (2001) argue that the coefficients entering cay_t represent steady-state wealth shares and are therefore unlikely to vary over time, subsequent empirical evidence has questioned the stability of this equilibrium relationship. In particular, Dauber and Lawrenz (2025) estimate $(\beta_a, \beta_y)^\top$ over different sample periods and find substantial instability in

the estimated coefficients. Motivated by this evidence, we use the proposed SDOLS estimator to obtain time-varying estimates of the cointegrating coefficients and construct corresponding time-varying measures of cay. We then examine whether these measures improve forecast accuracy of the U.S. equity premium relative to the conventional cay measure based on a fixed coefficient cointegrating regression.

To avoid complications arising from the pandemic period, we use quarterly data spanning 1952Q1-2019Q4. Data on consumption and income are taken from the Bureau of Economic Analysis (BEA). For consumption, we consider two alternative measures. The first measure is based on the nondurables plus services minus clothing and footwear (NDS): c_t^{NDS}. Following a change of methodology outlined in Martin Lettau's website (https://sites.google.com/view/martinlettau/data), we also consider using total personal consumption expenditures (PCE): c_t^{PCE}, as an alternative proxy for aggregate consumption.

The wealth series are obtained from the Board of Governors of the Federal Reserve System (FED). The aggregated wealth and labour income series are all deflated using the PCE-type deflator and then transformed into per capita terms (in logarithm). Finally, we use CRSP value-weighted index return provided by the Center for Research in Security Prices (CRSP). The risk-free rate is the Treasury-bill rate. A more detailed description of how each variables is constructed can be found in section A of the Appendix.

Let $\mathbf{c}_t = \left(c_t^{\mathrm{PCE}}, c_t^{\mathrm{NDS}}\right)^\top$ and $\mathbf{X}_t = (a_t, y_t)^\top$. To estimate the cointegrating relationships jointly, we estimate the model

$$\mathbf{c}_t = \boldsymbol{\alpha} + \mathbf{A}_t \mathbf{X}_t + \sum_{j=-K}^{K} \boldsymbol{\Pi}_j \Delta \mathbf{X}_{t-j} + \boldsymbol{\epsilon}_t, \tag{35}$$

using the SDOLS estimator proposed in Section 3.3, where $\mathbf{A}_t = (\boldsymbol{\beta}_{a,t}, \boldsymbol{\beta}_{y,t})$, $\boldsymbol{\beta}_{a,t} = \left(\beta_{a,t}^{\mathrm{PCE}}, \beta_{a,t}^{\mathrm{NDS}}\right)^\top$, and $\boldsymbol{\beta}_{y,t} = \left(\beta_{y,t}^{\mathrm{PCE}}, \beta_{y,t}^{\mathrm{NDS}}\right)^\top$. It is worth noting that, as the two equations in (35) contain the same regressors and no cross-equation restrictions are imposed, the system SDOLS estimator is algebraically identical to estimating the two equations separately, provided that k and K are set to be the same across specifications. Nevertheless, the system formulation is convenient for treating the two consumption measures jointly to account for the cross-equation dependence of their disturbances.

As in Section F, we select the series truncation order k by GCV criterion over $\{1, 2, \ldots, 12\}$.

Motivated by the simulation results, we set the lead-lag order $K = 4$. We also consider a parsimonious specification by fixing $k = 2$. Finally, imposing $\mathbf{A}_t = \mathbf{A}$ in (35) yields the conventional fixed coefficient specification, which we also estimate with $K = 4$.

We now assess whether allowing the coefficients in cay to vary over time improves out-of-sample (OOS) forecasting performance of the quarterly U.S. equity premium. The equity premium is measured as the CRSP value-weighted index return minus the lagged Treasury-bill rate. Using the estimates from (35), we construct two time-varying cay measures:

$$\widehat{\mathrm{cay}}_t^{\mathrm{PCE}} = c_t^{\mathrm{PCE}} - \widehat{\beta}_{a,t}^{\mathrm{PCE}} a_t - \widehat{\beta}_{y,t}^{\mathrm{PCE}} y_t,$$
$$\widehat{\mathrm{cay}}_t^{\mathrm{NDS}} = c_t^{\mathrm{NDS}} - \widehat{\beta}_{a,t}^{\mathrm{NDS}} a_t - \widehat{\beta}_{y,t}^{\mathrm{NDS}} y_t.$$

We compare forecasting performance based on these measures with those constructed from the fixed coefficient specification estimated by DOLS. More specifically, for each measure $\widehat{\mathrm{cay}}_t^{(j)}$, where $j \in \{\mathrm{PCE}, \mathrm{NDS}\}$, we estimate and construct the forecasts based on the following predictive regression model:

$$r_{t+1} = \alpha + \beta \widehat{\mathrm{cay}}_t^{(j)} + \varepsilon_{t+1}, \tag{36}$$

where r_{t+1} denotes the risk premium. The OOS evaluation begins in 1995Q1 and ends in 2019Q4. All estimators use only observations available up to the relevant forecast origin. Forecasts are generated recursively using an expansion window. Forecasting performance is evaluated using the OOS R^2 statistic (in percent), computed relative to the historical mean benchmark:

$$R^2_{\mathrm{OOS}} = 1 - \frac{\sum_{t=T_0}^{T-1} (r_{t+1} - \widehat{r}_{t+1|t})^2}{\sum_{t=T_0}^{T-1} (r_{t+1} - \overline{r}_{t+1|t})^2}. \tag{37}$$

Here, T_0 and $T-1$ denote the first and last periods of the OOS forecast evaluation, respectively; r_{t+1} denotes the realized equity premium; $\widehat{r}_{t+1|t}$ is the one-step-ahead forecast from the predictive regression using a cay measure constructed from either the fixed-coefficient or time-varying cointegration regression; and $\overline{r}_{t+1|t}$ denotes the historical mean forecast. To assess the statistical significance of differences in mean squared forecast errors (MSFE) between the cay-based forecasts and the historical mean benchmark, we report p-values of Clark and West (2007)'s test, with HAC

long-run variance estimated using a Bartlett kernel with four lags.[4]

Table 1 reports the OOS forecasting results obtained using alternative measures of cay. For PCE consumption, the forecast based on cay constructed from the SDOLS estimator in which k is selected by GCV achieves an overall OOS R^2 of 0.537%, compared with -3.639% from the DOLS estimator and -0.983% from the SDOLS estimator with fixed $k = 2$. The gains from the SDOLS with k selected by GCV criterion are concentrated in recessions, during which its OOS R^2 reaches 11.373%. In contrast, SDOLS with $k = 2$ modestly outperforms the historical mean benchmark during expansions but not over the full evaluation period. For NDS consumption, SDOLS with $k = 2$ performs the best and gains are also statistically significant. It produces an overall OOS R^2 of 2.579%, with a Clark-West p-value of 0.036, and an OOS R^2 of 4.114% during expansions, with a p-value of 0.018. The performance of the SDOLS with k selected by GCV criterion also improves substantially upon DOLS for NDS consumption, although it does not outperform the historical mean benchmark over the full evaluation period. Overall, allowing the cointegrating coefficients to vary over time improves forecasting performance relative to fixed coefficient counterpart for both consumption measures, although the specification that performs best differs across the consumption measures, empirical specifications, and business-cycle states.

To provide further insight into when forecasts based on the time-varying measures of cay outperform the competing alternatives, in Figures 1 and 2, we plot cumulative differences in mean squared forecast errors (MSFE) relative to the historical mean benchmark. Positive values indicate that the corresponding forecasts have a lower cumulative MSFE compared to the benchmark. For PCE consumption, measures constructed from SDOLS with k selected by GCV record a sharp gain around the 2007-2009 financial crisis and remains better than the benchmark afterwards, whereas the performance of fixed coefficient counterpart deteriorates markedly toward the end of the evaluation period. SDOLS with $k = 2$ performs well early in the evaluation period, but its cumulative gains subsequently diminish and eventually become negative. For NDS consumption, SDOLS with $k = 2$ maintains a positive gain for most of the evaluation period and performs particularly well after the financial crisis. Although SDOLS with k selected by GCV falls below the historical mean benchmark during the crisis, it still substantially outperforms the fixed coefficient

4. The recession-specific p-values should be interpreted with caution since the OOS evaluation period contains relatively few recession observations.

Table 1 — Quarterly U.S. equity premium out-of-sample forecasting results

	Overall		Recession		Expansion	
Predictor	R^2%	p-value	R^2%	p-value	R^2%	p-value
Panel A: PCE consumption						
$\widehat{\text{cay}}_{\text{DOLS}}$	-3.639	0.329	-0.511	0.507	-4.685	0.323
$\widehat{\text{cay}}_{\text{SDOLS,GCV}}$	0.537	0.195	11.373	0.073	-3.086	0.608
$\widehat{\text{cay}}_{\text{SDOLS},k=2}$	-0.983	0.577	-5.435	1.000	0.506	0.290
Panel B: NDS consumption						
$\widehat{\text{cay}}_{\text{DOLS}}$	-9.773	0.378	6.306	0.017	-15.148	0.502
$\widehat{\text{cay}}_{\text{SDOLS,GCV}}$	-1.826	0.704	-9.162	0.979	0.626	0.133
$\widehat{\text{cay}}_{\text{SDOLS},k=2}$	2.579	0.036	-2.015	0.977	4.114	0.018

Notes. The table reports out-of-sample forecasting performance for the quarterly U.S. equity premium using cay measures constructed from DOLS and two SDOLS estimators. For $\widehat{\text{cay}}_{\text{SDOLS,GCV}}$, the series truncation order k is selected by GCV criteria at each forecast origin. For $\widehat{\text{cay}}_{\text{SDOLS},k=2}$, k is fixed at $k = 2$. All specifications use four leads and lags. Panel A uses PCE consumption, while Panel B uses NDS consumption. Forecast performance is evaluated using the OOS R^2 statistic (in percent) defined in (37), relative to the historical mean benchmark. The p-values are from Clark and West (2007)'s tests in which HAC long-run variances are computed using a Bartlett kernel with four lags. The evaluation period is 1995Q1-2019Q4. The first forecast uses data through 1994Q4, and all forecasts are generated recursively using an expanding estimation window. Results are reported for the full evaluation period and separately for NBER recession and expansion periods.

counterpart throughout most of the evaluation period.

To assess the sensitivity of the forecasting performance on the choice of series truncation order k and lead-lag augmentation order K, Table 2 examines it across $k \in \{2, \ldots, 12\}$ and $K \in \{0, 1, 2, 4\}$. Forecasts constructed from each SDOLS estimator are compared with ones from a recursively estimated DOLS estimator using the same K. For NDS consumption, SDOLS improves compared to its corresponding DOLS counterpart in all 44 cases. The improvement is especially pronounced for the parsimonious $k = 2$ specification, for which the OOS R^2(%) gain ranges from 9.810 to 12.351 percentage points across the four different choices of K. For PCE consumption, SDOLS improves in 27 of the 44 cases. The PCE gains are strongest at low or high series orders and are less uniform for intermediate values of k. Overall, it seems that the forecasting gains from the time-varying cay measure relative to its fixed-coefficient counterpart are more robust to specification choices for NDS consumption.

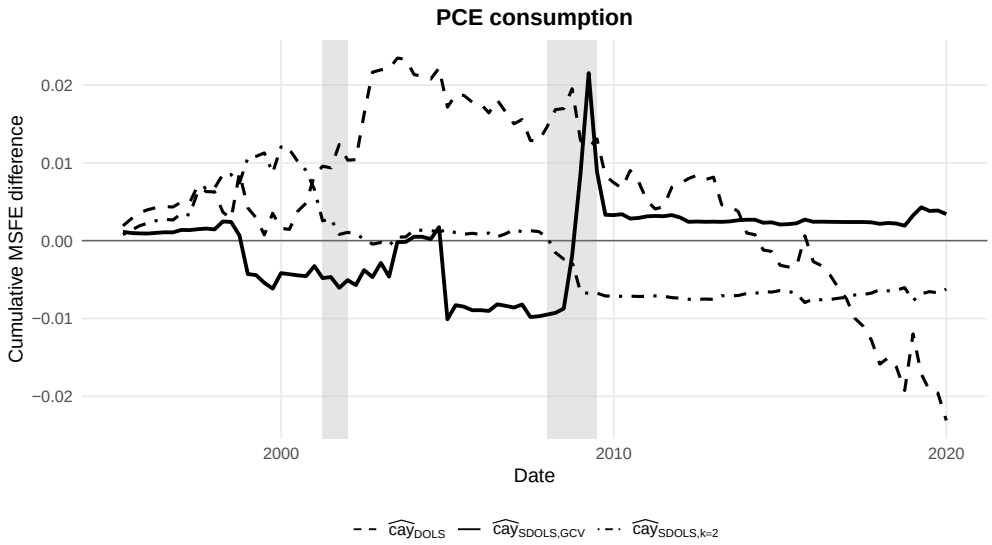


Figure 1 — Cumulative difference in MSFE relative to the historical mean benchmark

Notes. The figure reports cumulative differences in MSFE relative to the historical mean for PCE consumption. Positive values indicate that the corresponding predictor outperforms the historical mean benchmark. The legend uses the same predictor notation as in Table 1. All specifications use $K = 4$, and shaded areas indicate NBER recessions.

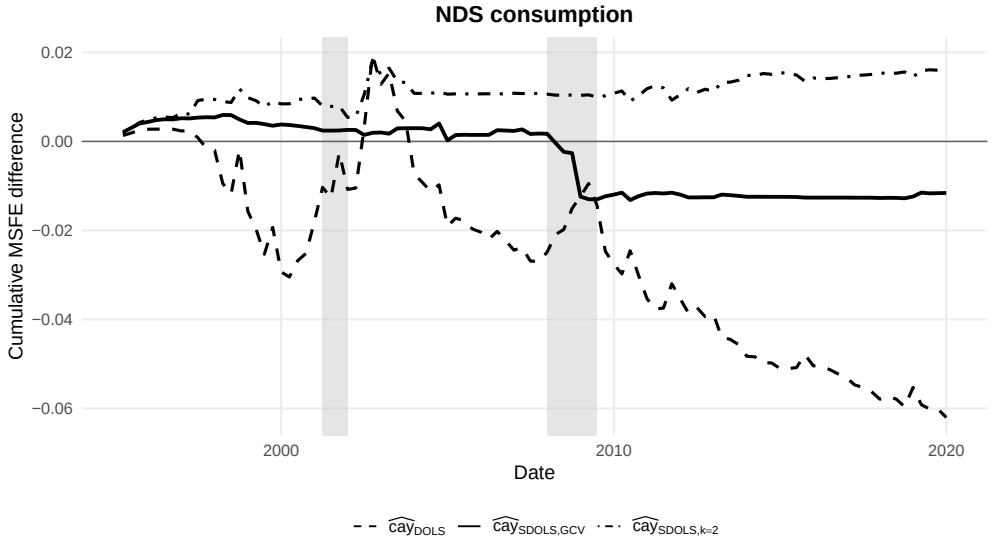


Figure 2 — Cumulative difference in MSFE relative to the historical mean benchmark

Notes. The figure reports cumulative differences in MSFE relative to the historical mean for NDS consumption. Positive values indicate that the corresponding predictor outperforms the historical mean benchmark. The legend uses the same predictor notation as in Table 1. All specifications use $K = 4$, and shaded areas indicate NBER recessions.

Table 2 — Sensitivity of forecasting performance to the series and lead-lag orders

(a) No leads or lags ($K = 0$)

	PCE consumption			NDS consumption		
k	SDOLS $R^2(\%)$	DOLS $R^2(\%)$	ΔR^2 (%)	SDOLS $R^2(\%)$	DOLS $R^2(\%)$	ΔR^2 (%)
2	-0.562	-2.427	1.865	1.987	-7.823	9.810
3	-0.779	-2.427	1.648	-1.683	-7.823	6.140
4	-0.798	-2.427	1.629	-3.217	-7.823	4.606
5	-2.582	-2.427	-0.155	-6.183	-7.823	1.640
6	-1.328	-2.427	1.099	-3.766	-7.823	4.057
7	-1.253	-2.427	1.173	-3.462	-7.823	4.361
8	-2.089	-2.427	0.338	-2.675	-7.823	5.147
9	-2.332	-2.427	0.095	-1.548	-7.823	6.274
10	-2.008	-2.427	0.419	-1.322	-7.823	6.501
11	-0.922	-2.427	1.505	-0.541	-7.823	7.282
12	0.049	-2.427	2.476	1.184	-7.823	9.007

(b) One lead and lag ($K = 1$)

	PCE consumption			NDS consumption		
k	SDOLS $R^2(\%)$	DOLS $R^2(\%)$	ΔR^2 (%)	SDOLS $R^2(\%)$	DOLS $R^2(\%)$	ΔR^2 (%)
2	-0.740	-2.842	2.103	1.899	-8.460	10.359
3	-0.919	-2.842	1.923	-1.517	-8.460	6.943
4	-1.314	-2.842	1.528	-3.565	-8.460	4.895
5	-3.588	-2.842	-0.746	-6.777	-8.460	1.683
6	-2.561	-2.842	0.282	-4.966	-8.460	3.494
7	-3.351	-2.842	-0.509	-4.743	-8.460	3.717
8	-4.240	-2.842	-1.398	-4.423	-8.460	4.037
9	-4.553	-2.842	-1.710	-2.619	-8.460	5.841
10	-3.999	-2.842	-1.157	-2.547	-8.460	5.913
11	-1.428	-2.842	1.414	-0.880	-8.460	7.580
12	-0.469	-2.842	2.373	0.773	-8.460	9.233

Table 2 — Sensitivity of forecasting performance to the series and lead-lag orders (continued)

(c) Two leads and lags ($K = 2$)

	PCE consumption			NDS consumption		
k	SDOLS $R^2(\%)$	DOLS $R^2(\%)$	ΔR^2 (%)	SDOLS $R^2(\%)$	DOLS $R^2(\%)$	ΔR^2 (%)
2	-0.873	-3.161	2.288	1.957	-9.079	11.035
3	-1.010	-3.161	2.151	-1.226	-9.079	7.853
4	-2.150	-3.161	1.012	-4.527	-9.079	4.552
5	-4.750	-3.161	-1.589	-7.946	-9.079	1.132
6	-4.263	-3.161	-1.101	-6.770	-9.079	2.309
7	-5.387	-3.161	-2.225	-6.444	-9.079	2.635
8	-6.261	-3.161	-3.100	-6.857	-9.079	2.222
9	-6.740	-3.161	-3.579	-4.205	-9.079	4.874
10	-6.019	-3.161	-2.857	-3.617	-9.079	5.461
11	-1.793	-3.161	1.368	-1.597	-9.079	7.482
12	-0.270	-3.161	2.891	-0.833	-9.079	8.245

(d) Four leads and lags ($K = 4$)

	PCE consumption			NDS consumption		
k	SDOLS $R^2(\%)$	DOLS $R^2(\%)$	ΔR^2 (%)	SDOLS $R^2(\%)$	DOLS $R^2(\%)$	ΔR^2 (%)
2	-0.983	-3.639	2.656	2.579	-9.773	12.351
3	-1.015	-3.639	2.624	-0.963	-9.773	8.810
4	-3.548	-3.639	0.091	-6.453	-9.773	3.319
5	-6.381	-3.639	-2.742	-9.513	-9.773	0.259
6	-5.761	-3.639	-2.122	-8.331	-9.773	1.441
7	-6.600	-3.639	-2.962	-8.627	-9.773	1.145
8	-6.084	-3.639	-2.445	-9.168	-9.773	0.605
9	-6.985	-3.639	-3.346	-6.934	-9.773	2.838
10	-2.019	-3.639	1.620	-2.605	-9.773	7.167
11	-1.383	-3.639	2.256	-0.928	-9.773	8.845
12	0.766	-3.639	4.405	-1.918	-9.773	7.854

Notes. The four panels report overall OOS R^2 statistics, in percent, for every SDOLS specification with $k \in \{2, \ldots, 12\}$ and $K \in \{0, 1, 2, 4\}$. The associated DOLS column uses a recursively estimated fixed coefficient DOLS with the same lead-lag order K. The difference is $\Delta R^2 = R^2_{\text{OOS,SDOLS}} - R^2_{\text{OOS,DOLS}}$ in percentage points; positive entries therefore indicate an improvement from the fixed coefficient specifcation. The evaluation period is 1995Q1-2019Q4 and the historical mean is the common benchmark used to construct both OOS R^2 statistics.

5 — Conclusion

Although cointegrating relationships characterize long-run equilibria and may appear permanent, the parameters governing these relationships can evolve gradually over time in response to changes in policy regimes, institutions, technologies, or preferences. Allowing for such parameter evolution is important not only for characterizing changes in long-run equilibria but also for empirical applications, such as forecasting, that rely on the estimated cointegrating relationship.

This paper develops a nonparametric series estimation framework for cointegrated systems with coefficients that evolve smoothly over time. We first establish the limit theory for the series-based OLS (SOLS) estimator under nonstationary (unconditional) volatility of unknown form. Then, we develop an asymptotic efficiency theory for a broad class of series-based estimators and propose a series-based dynamic OLS (SDOLS) estimator that attains the efficiency bound and is straightforward to implement. The framework provides a flexible and computationally convenient approach to estimating the time-varying cointegrating relationship, while remaining as agnostic as possible on the type of parameter evolution.

In an empirical application to the consumption-wealth ratio, we find that time-varying cay measures improve equity-premium forecasting performance relative to their fixed-coefficient counterparts and the historical mean benchmark, although the magnitude of the gains depends on implementation choices and varies across business-cycle states. The Monte Carlo simulations presented in Section F of the Online Supplement show that the SDOLS estimator generally improves finite-sample performance relative to existing alternatives.

References

Amihud, Y. and Hurvich, C. M. (2004). Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis*, 39(4):813–841.

Amihud, Y., Hurvich, C. M., and Wang, Y. (2008). Multiple-predictor regressions: Hypothesis testing. *The Review of Financial Studies*, 22(1):413–434.

Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proceedings of the American Mathematical Society*, 6(2):170–176.

Andrews, D. W., Lee, I., and Ploberger, W. (1996). Optimal changepoint tests for normal linear regression. *Journal of Econometrics*, 70(1):9–38.

Bai, J., Lumsdaine, R. L., and Stock, J. H. (1998). Testing for and dating common breaks in multivariate time series. *The Review of Economic Studies*, 65(3):395–432.

Balke, N. S. and Fomby, T. B. (1997). Threshold cointegration. *International Economic Review*, pages 627–645.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925.

Basawa, I. V. and Scott, D. J. (1983). *Asymptotic optimal inference for non-ergodic models*. New York: Springer Verlag.

Bierens, H. J. (1997). Testing the unit root with drift hypothesis against nonlinear trend stationarity, with an application to the US price level and interest rate. *Journal of Econometrics*, 81(1):29–64.

Bierens, H. J. and Martins, L. F. (2010). Time-varying cointegration. *Econometric Theory*, 26(5):1453–1490.

Boswijk, H. P., Cavaliere, G., Rahbek, A., and Taylor, A. R. (2016). Inference on co-integration parameters in heteroskedastic vector autoregressions. *Journal of Econometrics*, 192(1):64–85.

Boyd, J. P. (2001). *Chebyshev and Fourier spectral methods*. Courier Corporation.

Brillinger, D. R. (1975). *Time series: data analysis and theory*. New York: Holt, Rinehart and Winston.

Cai, Z. and Wang, Y. (2014). Testing predictive regression models with nonstationary regressors. *Journal of Econometrics*, 178:4–14.

Cavaliere, G., Rahbek, A., and Taylor, A. R. (2010). Testing for co-integration in vector autoregressions with non-stationary volatility. *Journal of Econometrics*, 158(1):7–24.

Cavaliere, G. and Taylor, A. M. R. (2006). Testing the null of co-integration in the presence of variance breaks. *Journal of Time Series Analysis*, 27(4):613–636.

Cavaliere, G. and Taylor, A. R. (2007). Testing for unit roots in time series models with non-stationary volatility. *Journal of Econometrics*, 140(2):919–947.

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1):291–311.

Dauber, M. and Lawrenz, J. (2025). The decay of cay. *Journal of Empirical Finance*, page 101668.

Davidson, J. (2021). *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press, Oxford, 2 edition.

Dong, C. and Gao, J. (2025). *Modern Series Methods in Econometrics and Statistics*. Springer, New York.

Einmahl, U. (1989). Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *Journal of Multivariate Analysis*, 28(1):20–68.

Friz, P. K. and Kifer, Y. (2024). Almost sure diffusion approximation in averaging via rough paths theory. *Electronic Journal of Probability*, 29:1–56.

Gao, J., Peng, B., and Yan, Y. (2025). Time-varying vector error-correction models: Estimation and inference. *Journal of Econometrics*, 251:106035.

Gao, J., Tong, H., and Wolff, R. (2002). Model specification tests in nonparametric stochastic regression models. *Journal of Multivariate Analysis*, 83(2):324–359.

Hall, S. G., Psaradakis, Z., and Sola, M. (1997). Cointegration and changes in regime: the Japanese consumption function. *Journal of Applied Econometrics*, 12(2):151–168.

Hansen, B. E. (1992). Tests for parameter instability in regressions with I(1) processes. *Journal of Business & Economic Statistics*, pages 321–335.

Hansen, H. and Johansen, S. (1999). Some tests for parameter constancy in cointegrated VAR models. *The Econometrics Journal*, 2(2):306–333.

Kifer, Y. (2026). Limit theorems for signatures. *Electronic Journal of Probability*, 31:1–53.

Koop, G., Leon-Gonzalez, R., and Strachan, R. W. (2011). Bayesian inference in a time varying cointegration model. *Journal of Econometrics*, 165(2):210–220.

Kuo, B.-S. (1998). Test for partial parameter instability in regressions with I(1) processes. *Journal of Econometrics*, 86(2):337–368.

Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance*, 56(3):815–849.

Li, D., Phillips, P. C., and Gao, J. (2020). Kernel-based inference in time-varying coefficient cointegrating regression. *Journal of Econometrics*, 215(2):607–632.

Park, J. Y. (1992). Canonical cointegrating regressions. *Econometrica: Journal of the Econometric Society*, pages 119–143.

Park, J. Y. and Hahn, S. B. (1999). Cointegrating regressions with time varying coefficients. *Econometric Theory*, 15(5):664–703.

Park, J. Y. and Phillips, P. C. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, 4(3):468–497.

Phillips, P. C. (1991). Optimal inference in cointegrated systems. *Econometrica: Journal of the Econometric Society*, pages 283–306.

Phillips, P. C. and Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *The Review of Economic Studies*, 53(4):473–495.

Phillips, P. C., Li, D., and Gao, J. (2017). Estimating smooth structural change in cointegration models. *Journal of Econometrics*, 196(1):180–195.

Phillips, P. C. and Loretan, M. (1991). Estimating long-run economic equilibria. *The Review of Economic Studies*, 58(3):407–436.

Phillips, P. C. and Solo, V. (1992). Asymptotics for linear processes. *The Annals of Statistics*, pages 971–1001.

Phillips, P. C. and Wang, Y. (2023). When bias contributes to variance: True limit theory in functional coefficient cointegrating regression. *Journal of Econometrics*, 232(2):469–489.

Quintos, C. E. (1998). Stability tests in error correction models. *Journal of Econometrics*, 82(2):289–315.

Rudin, W. (1987). *Real and Complex Analysis*. McGraw-Hill International Editions, New York, 3 edition.

Saikkonen, P. (1991). Asymptotically efficient estimation of cointegration regressions. *Econometric Theory*, 7(1):1–21.

Saikkonen, P. and Choi, I. (2004). Cointegrating smooth transition regressions. *Econometric Theory*, 20(2):301–340.

Seo, B. (1998). Tests for structural change in cointegrated systems. *Econometric Theory*, 14(2):222–259.

Shreve, S. E. et al. (2004). *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer.

Stock, J. H. and Watson, M. W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica: Journal of the Econometric Society*, pages 783–820.

Xiao, Z. (2009). Functional-coefficient cointegration models. *Journal of Econometrics*, 152(2):81–92.

Yan, Y., Gao, J., and Peng, B. (2025). Asymptotics for time-varying vector MA(∞) processes. *Econometric Theory*, 41(3):584–616.

A — Data Appendix

The data construction broadly follows the empirical implementation in Dauber and Lawrenz (2025). We provide a detailed description of the data sources and variable definitions here for completeness.

Consumption and labor income data are obtained from the National Income and Product Accounts (NIPA) published by the Bureau of Economic Analysis (BEA). Two alternative measures of consumption are considered. The first is the conventional nondurables and services consumption measure excluding clothing and footwear (NDS). The second is total personal consumption expenditures (PCE). Each series is separately deflated to 2017-dollars with its respective deflator. Labor income is defined as the sum of wages and salaries, personal current transfer receipts, and employer contributions for employee pension and insurance funds minus employee contributions for government social insurance, and taxes.

Aggregate wealth is defined as the net worth of households and nonprofit organizations. The source is the Board of Governors of the Federal Reserve System, Financial Accounts. The aggregate wealth and labor income series are all deflated to 2017-dollars using the PCE chain-type price deflator.

Consumption, aggregate wealth and labor income series are all transformed into per capita terms and expressed in logarithmics. The population data are retrieved from FRED, Federal Reserve Economic Data.

Stock returns are measured as the CRSP Value-Weighted index return. The equity premium is defined as the index return in excess of the lagged three-month Treasury-bill rate. They are obtained from Amit Goyal's website (https://docs.google.com/spreadsheets/d/1qwpl2R_DNujpU5YUkk8lacP1tTeMb9iJ/edit?gid=134094480#gid=134094480).

This online supplement is organized as follows. Sections B and C provide proofs of the main results in Sections 2 and 3, respectively. Sections D and E collect the supporting lemmas and auxiliary results, together with their proofs, for Sections 2 and 3, respectively. Section F presents a Monte Carlo study to assess the finite-sample performance of the proposed estimators.

Notation. Unless otherwise stated, Sections B-E follow the notation and conventions introduced in the main text. In Sections B and D, we retain the general volatility profile $\mathbf{H}(\cdot)$ and the continuous martingale $\underline{\mathbf{B}}(\cdot)$. In Sections C and E, as the innovations are assumed to be stationary, we use $\mathbf{B}(\cdot)$ to denote the vector Brownian motion. The relevant results from Section 2 continue to hold in the setting of Section 3 with $\underline{\mathbf{B}}(\cdot)$ replaced by $\mathbf{B}(\cdot)$.

B — Proofs of main results in Section 2

B.1. Proof of Lemma 1

By Parseval's identity (Theorem 4.18 (iv) of Rudin (1987)),

$$\int_0^1 |a(r) - a_k(r)|^2 \, dr = \sum_{j=k}^{\infty} b_{1,j}^2.$$

Integrations by parts q times gives, for $j \geq 1$,

$$b_{1,j} = \int_0^1 a(r)\psi_j(r)\, dr = \sqrt{2} \sum_{\ell=1}^{\lfloor q/2 \rfloor} \frac{(-1)^{\ell+1}}{(j\pi)^{2\ell}} \, a^{(2\ell-1)}(r) \cos(j\pi r) \Big|_0^1 + \mathcal{R}_{q,j}.$$

where

$$\mathcal{R}_{q,j} = \begin{cases} \dfrac{(-1)^{\frac{q+1}{2}}}{(j\pi)^q} \displaystyle\int_0^1 a^{(q)}(r)\sqrt{2}\sin(j\pi r)\, dr, & \text{if } q \text{ is odd,} \\ \dfrac{(-1)^{\frac{q}{2}}}{(j\pi)^q} \displaystyle\int_0^1 a^{(q)}(r)\sqrt{2}\cos(j\pi r)\, dr, & \text{if } q \text{ is even.} \end{cases} \tag{B.1}$$

The endpoint restrictions in Assumption 2 make every displayed boundary term equal to zero. Consequently, $b_{1,j} = \mathcal{R}_{q,j}$. When q is odd,

$$\begin{aligned} \sum_{j=k}^{\infty} b_{1,j}^2 &= \sum_{j=k}^{\infty} \frac{1}{(j\pi)^{2q}} \left(\int_0^1 a^{(q)}(r)\sqrt{2}\sin(j\pi r)\, dr \right)^2 \\ &\leq \frac{1}{\pi^{2q}k^{2q}} \int_0^1 \left| a^{(q)}(r) \right|^2 dr = O\left(k^{-2q}\right), \end{aligned}$$

where the inequality follows by Bessel's inequality as in Theorem 4.17 of Rudin (1987). Similarly, when q is even, we can show that $\sum_{j=k}^{\infty} b_{1,j}^2 = O\left(k^{-2q}\right)$. This completes the proof. □

B.2. Proof of Lemma 2

Let $\mathbf{V}_{n,t}^x = \mathbf{\Phi}_x(1)\frac{1}{\sqrt{n}}\sum_{s=1}^{t}\boldsymbol{\varepsilon}_s$, so that $\mathbf{V}_n^x(r) = \mathbf{V}_{n,\lfloor nr \rfloor}^x$ and $\mathbf{V}_{n,t}^x = \mathbf{V}_n^x(\tau_t)$. Using the BN decomposition as in Phillips and Solo (1992), we have

$$\mathbf{X}_t - \mathbf{X}_{t-1} = \mathbf{u}_{xt} = \mathbf{\Phi}_x(1)\boldsymbol{\varepsilon}_t - \Delta\widetilde{\mathbf{u}}_{x,t},$$

where $\widetilde{\mathbf{u}}_{x,t} = \sum_{j=0}^{\infty}\widetilde{\mathbf{\Phi}}_{xj}\boldsymbol{\varepsilon}_{t-j}$ with $\widetilde{\mathbf{\Phi}}_{xj} = \sum_{k=j+1}^{\infty}\mathbf{\Phi}_{xk}$. Then,

$$\mathbf{X}_t = \mathbf{\Phi}_x(1)\sum_{s=1}^{t}\boldsymbol{\varepsilon}_s + \widetilde{\mathbf{u}}_{x,0} - \widetilde{\mathbf{u}}_{x,t} + \mathbf{X}_0 = \sqrt{n}\mathbf{V}_{n,t}^x + \widetilde{\mathbf{u}}_{x,0} - \widetilde{\mathbf{u}}_{x,t} + \mathbf{X}_0.$$

We have

$$\begin{aligned}
&\frac{1}{n^2}\sum_{t=1}^{n}\mathbf{X}_t^{(k)}\mathbf{X}_t^{(k)\top} \\
&= \frac{1}{n^2}\sum_{t=1}^{n}\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{X}_t\right)\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{X}_t\right)^{\top} \\
&= \frac{1}{n^2}\sum_{t=1}^{n}\left(\boldsymbol{\psi}_{(k)}(\tau_t)\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\otimes\mathbf{X}_t\mathbf{X}_t^{\top}\right) \\
&= \frac{1}{n^2}\sum_{t=1}^{n}\left(\boldsymbol{\psi}_{(k)}(\tau_t)\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\otimes\left(\sqrt{n}\mathbf{V}_{n,t}^x\right)\left(\sqrt{n}\mathbf{V}_{n,t}^x\right)^{\top}\right) \\
&\quad + \frac{1}{n^2}\sum_{t=1}^{n}\left(\boldsymbol{\psi}_{(k)}(\tau_t)\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\otimes\left(\sqrt{n}\mathbf{V}_{n,t}^x\right)\left(\widetilde{\mathbf{u}}_{x,0} - \widetilde{\mathbf{u}}_{x,t} + \mathbf{X}_0\right)^{\top}\right) \\
&\quad + \frac{1}{n^2}\sum_{t=1}^{n}\left(\boldsymbol{\psi}_{(k)}(\tau_t)\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\otimes\left(\widetilde{\mathbf{u}}_{x,0} - \widetilde{\mathbf{u}}_{x,t} + \mathbf{X}_0\right)\left(\sqrt{n}\mathbf{V}_{n,t}^x\right)^{\top}\right) \\
&\quad + \frac{1}{n^2}\sum_{t=1}^{n}\left(\boldsymbol{\psi}_{(k)}(\tau_t)\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\otimes\left(\widetilde{\mathbf{u}}_{x,0} - \widetilde{\mathbf{u}}_{x,t} + \mathbf{X}_0\right)\left(\widetilde{\mathbf{u}}_{x,0} - \widetilde{\mathbf{u}}_{x,t} + \mathbf{X}_0\right)^{\top}\right) \\
&:= \mathbf{\Lambda}_n(1) + \mathbf{\Lambda}_n(2) + \mathbf{\Lambda}_n(3) + \mathbf{\Lambda}_n(4),
\end{aligned}$$

so that

$$\left\|\frac{1}{n^2}\sum_{t=1}^{n}\mathbf{X}_t^{(k)}\mathbf{X}_t^{(k)\top} - \mathbf{\Lambda}_n(1)\right\| \leq \|\mathbf{\Lambda}_n(2)\| + \|\mathbf{\Lambda}_n(3)\| + \|\mathbf{\Lambda}_n(4)\|.$$

We first show that $\|\mathbf{\Lambda}_n(2)\| = O_p\left(\frac{k}{n^a}\right)$, $\|\mathbf{\Lambda}_n(3)\| = O_p\left(\frac{k}{n^a}\right)$, and $\|\mathbf{\Lambda}_n(4)\| = O_p\left(\frac{k}{n^{2a}}\right)$, where $a = \frac{p-2}{2p}$. This would imply that

$$\frac{1}{n^2}\sum_{t=1}^{n}\mathbf{X}_t^{(k)}\mathbf{X}_t^{(k)\top} = \mathbf{\Lambda}_n(1) + O_p\left(\frac{k}{n^a}\right). \tag{B.2}$$

Let us start with $\|\mathbf{\Lambda}_n(2)\|$. Notice that

$$\begin{aligned}
\|\mathbf{\Lambda}_n(2)\| &\leq \frac{1}{n^2}\sum_{t=1}^{n}\left\|\boldsymbol{\psi}_{(k)}(\tau_t)\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\right\|\left\|(\sqrt{n}\,\mathbf{V}_{n,t}^{x})(\widetilde{\mathbf{u}}_{x,0}-\widetilde{\mathbf{u}}_{x,t}+\mathbf{X}_0)^{\top}\right\| \\
&\leq \frac{1}{n^2}\sum_{t=1}^{n}\left\|\boldsymbol{\psi}_{(k)}(\tau_t)\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\right\|\sqrt{n}\,\left\|\mathbf{V}_{n,t}^{x}\right\|\|\widetilde{\mathbf{u}}_{x,0}-\widetilde{\mathbf{u}}_{x,t}+\mathbf{X}_0\| \\
&\leq \max_{1\leq t\leq n}\left\|\frac{1}{\sqrt{n}}(\widetilde{\mathbf{u}}_{x,0}-\widetilde{\mathbf{u}}_{x,t}+\mathbf{X}_0)\right\|\cdot\frac{1}{n^2}\sum_{t=1}^{n}\left\|\boldsymbol{\psi}_{(k)}(\tau_t)\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\right\|\cdot n\,\left\|\mathbf{V}_{n,t}^{x}\right\|.
\end{aligned}$$

By standard Riemann–sum approximation for the integral, we have

$$\frac{1}{n}\sum_{t=1}^{n}\left\|\boldsymbol{\psi}_{(k)}(\tau_t)\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\right\| = \int_0^1\left\|\boldsymbol{\psi}_{(k)}(r)\boldsymbol{\psi}_{(k)}^{\top}(r)\right\|dr + O(1/n) = O(1)k.$$

In addition, by Lemma D.3, we have

$$\max_{1\leq t\leq n}\left\|\frac{1}{\sqrt{n}}(\widetilde{\mathbf{u}}_{x,0}-\widetilde{\mathbf{u}}_{x,t}+\mathbf{X}_0)\right\| = O_p\left(\frac{1}{n^a}\right),$$

for $a = \frac{p-2}{2p}$. Substituting all these results above yields

$$\|\mathbf{\Lambda}_n(2)\| = O_p\left(\frac{k}{n^a}\right).$$

Similar arguments lead to $\|\mathbf{\Lambda}_n(3)\| = O_p\left(\frac{k}{n^a}\right)$ and $\|\mathbf{\Lambda}_n(4)\| = O_p\left(\frac{k}{n^{2a}}\right)$.

We now move on to $\mathbf{\Lambda}_n(1)$. By construction, $\mathbf{V}_{n,0} = \mathbf{0}$. We have

$$\begin{aligned}
&\frac{1}{n^2}\sum_{t=1}^{n}\left(\boldsymbol{\psi}_{(k)}(\tau_t)\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\otimes\left(\sqrt{n}\mathbf{V}_{n,t}^{x}\right)\left(\sqrt{n}\mathbf{V}_{n,t}^{x}\right)^{\top}\right) \\
&= \int_0^1\left(\boldsymbol{\psi}_{(k)}(\lfloor nr\rfloor/n)\boldsymbol{\psi}_{(k)}^{\top}(\lfloor nr\rfloor/n)\otimes\mathbf{V}_{n,\lfloor nr\rfloor}^{x}\mathbf{V}_{n,\lfloor nr\rfloor}^{x\top}\right)dr + \frac{1}{n}\left(\boldsymbol{\psi}_{(k)}(1)\boldsymbol{\psi}_{(k)}^{\top}(1)\otimes\mathbf{V}_{n,n}^{x}\mathbf{V}_{n,n}^{x\top}\right) \\
&= \int_0^1\left(\boldsymbol{\psi}_{(k)}(\lfloor nr\rfloor/n)\boldsymbol{\psi}_{(k)}^{\top}(\lfloor nr\rfloor/n)\otimes\mathbf{V}_{n,\lfloor nr\rfloor}^{x}\mathbf{V}_{n,\lfloor nr\rfloor}^{x\top}\right)dr + O_p(k/n),
\end{aligned}$$

where $\frac{1}{n}\left(\boldsymbol{\psi}_{(k)}(1)\boldsymbol{\psi}_{(k)}^{\top}(1)\otimes\mathbf{V}_{n,n}^{x}\mathbf{V}_{n,n}^{x\top}\right)=O_p(k/n)=o_p(1)$, due to the fact that $\left\|\boldsymbol{\psi}_{(k)}(1)\boldsymbol{\psi}_{(k)}(1)^{\top}\right\|=O(1)\,k$ and (5). Since

$$\begin{aligned}\left\|\boldsymbol{\psi}_{(k)}(\lfloor nr\rfloor/n)-\boldsymbol{\psi}_{(k)}(r)\right\|^2 &= 2\sum_{j=1}^{k-1}\left(\cos\left(j\pi\frac{\lfloor nr\rfloor}{n}\right)-\cos(j\pi r)\right)^2\\ &\leq O(1)\left|\frac{\lfloor nr\rfloor}{n}-r\right|^2\left(\sum_{j=1}^{k-1}j^2\right)\leq O(1)k^3/n^2,\end{aligned}$$

we have

$$\left\|\boldsymbol{\psi}_{(k)}(\lfloor nr\rfloor/n)-\boldsymbol{\psi}_{(k)}(r)\right\|=O\left(\frac{k^{3/2}}{n}\right).$$

This implies that

$$\begin{aligned}&\left\|\int_0^1\left(\boldsymbol{\psi}_{(k)}(\lfloor nr\rfloor/n)\boldsymbol{\psi}_{(k)}^{\top}(\lfloor nr\rfloor/n)-\boldsymbol{\psi}_{(k)}(r)\boldsymbol{\psi}_{(k)}^{\top}(r)\right)dr\right\|\\ &\leq O(1)\int_0^1\left(\left\|\boldsymbol{\psi}_{(k)}(\lfloor nr\rfloor/n)\right\|+\left\|\boldsymbol{\psi}_{(k)}(r)\right\|\right)\left\|\boldsymbol{\psi}_{(k)}(\lfloor nr\rfloor/n)-\boldsymbol{\psi}_{(k)}(r)\right\|dr\\ &=O\left(\frac{k^2}{n}\right).\end{aligned}$$

Thus,

$$\begin{aligned}\boldsymbol{\Lambda}_n(1) &= \int_0^1\left(\boldsymbol{\psi}_{(k)}(\lfloor nr\rfloor/n)\boldsymbol{\psi}_{(k)}^{\top}(\lfloor nr\rfloor/n)-\boldsymbol{\psi}_{(k)}(r)\boldsymbol{\psi}_{(k)}^{\top}(r)\right)\otimes\left(\mathbf{V}_{n,\lfloor nr\rfloor}^{x}\mathbf{V}_{n,\lfloor nr\rfloor}^{x\top}\right)dr\\ &\quad+\int_0^1\left(\boldsymbol{\psi}_{(k)}(r)\boldsymbol{\psi}_{(k)}^{\top}(r)\otimes\mathbf{V}_{n,\lfloor nr\rfloor}^{x}\mathbf{V}_{n,\lfloor nr\rfloor}^{x\top}\right)dr \qquad \text{(B.3)}\\ &=\int_0^1\left(\boldsymbol{\psi}_{(k)}(r)\boldsymbol{\psi}_{(k)}^{\top}(r)\otimes\mathbf{V}_{n,\lfloor nr\rfloor}^{x}\mathbf{V}_{n,\lfloor nr\rfloor}^{x\top}\right)dr+O_p\left(\frac{k^2}{n}\right).\end{aligned}$$

By combining (B.2) and (B.3), we conclude that

$$\frac{1}{n^2}\sum_{t=1}^{n}\mathbf{X}_t^{(k)}\mathbf{X}_t^{(k)\top}=\int_0^1\left(\boldsymbol{\psi}_{(k)}(r)\boldsymbol{\psi}_{(k)}^{\top}(r)\otimes\mathbf{V}_{n,\lfloor nr\rfloor}^{x}\mathbf{V}_{n,\lfloor nr\rfloor}^{x\top}\right)dr+O_p\left(\frac{k}{n^a}\right)+O_p\left(\frac{k^2}{n}\right).$$

Finally, we show that

$$\left\|\int_0^1\left(\boldsymbol{\psi}_{(k)}(r)\boldsymbol{\psi}_{(k)}^{\top}(r)\otimes\mathbf{V}_{n,\lfloor nr\rfloor}^{x}\mathbf{V}_{n,\lfloor nr\rfloor}^{x\top}\right)dr-\int_0^1\left[\boldsymbol{\psi}_{(k)}(r)\boldsymbol{\psi}_{(k)}^{\top}(r)\otimes\underline{\mathbf{B}}_X(r)\underline{\mathbf{B}}_X^{\top}(r)\right]dr\right\|_{sp}=O_p(n^{-a}), \qquad \text{(B.4)}$$

where $\underline{\mathbf{B}}_X(r) = \mathbf{\Phi}_x(1)\mathbf{B}_H(r)$. For any unit vector $\mathbf{z} \in \mathbb{R}^{kd_x}$, where $\mathbf{z} = \text{vec}(\mathbf{Z})$, we note that

$$\begin{aligned}
&\left|\mathbf{z}^\top \left(\int_0^1 \left\{\boldsymbol{\psi}_{(k)}(r)\boldsymbol{\psi}_{(k)}^\top(r) \otimes \left(\mathbf{V}^x_{n,\lfloor nr\rfloor}\mathbf{V}^{x\top}_{n,\lfloor nr\rfloor} - \underline{\mathbf{B}}_X(r)\underline{\mathbf{B}}_X^\top(r)\right)\right\} dr\right)\mathbf{z}\right| \\
&= \left\|\int_0^1 \left(\mathbf{Z}\boldsymbol{\psi}_{(k)}(r)\right)^\top \left(\mathbf{V}^x_{n,\lfloor nr\rfloor}\mathbf{V}^{x\top}_{n,\lfloor nr\rfloor} - \underline{\mathbf{B}}_X(r)\underline{\mathbf{B}}_X^\top(r)\right)\left(\mathbf{Z}\boldsymbol{\psi}_{(k)}(r)\right) dr\right\| \\
&\leq \sup_{r\in[0,1]} \left\|\mathbf{V}^x_{n,\lfloor nr\rfloor}\mathbf{V}^{x\top}_{n,\lfloor nr\rfloor} - \underline{\mathbf{B}}_X(r)\underline{\mathbf{B}}_X^\top(r)\right\| \int_0^1 \left\|\mathbf{Z}\boldsymbol{\psi}_{(k)}(r)\right\|^2 dr \\
&\leq \sup_{r\in[0,1]} \left\|\mathbf{V}^x_{n,\lfloor nr\rfloor}\mathbf{V}^{x\top}_{n,\lfloor nr\rfloor} - \underline{\mathbf{B}}_X(r)\underline{\mathbf{B}}_X^\top(r)\right\| \text{tr}\left(\mathbf{Z}^\top\mathbf{Z}\int_0^1 \boldsymbol{\psi}_{(k)}(r)\boldsymbol{\psi}_{(k)}^\top(r)dr\right) \\
&\leq \sup_{r\in[0,1]} \left\|\mathbf{V}^x_{n,\lfloor nr\rfloor}\mathbf{V}^{x\top}_{n,\lfloor nr\rfloor} - \underline{\mathbf{B}}_X(r)\underline{\mathbf{B}}_X^\top(r)\right\|,
\end{aligned}$$

where the last inequality follows from the fact that

$$\text{tr}\left(\mathbf{Z}^\top\mathbf{Z}\int_0^1 \boldsymbol{\psi}_{(k)}(r)\boldsymbol{\psi}_{(k)}^\top(r)dr\right) = \text{tr}\left(\mathbf{Z}^\top\mathbf{Z}\right) = \|\mathbf{Z}\|^2 = 1.$$

In addition, Lemma D.3 gives

$$\sup_{r\in[0,1]} \left\|\mathbf{V}^x_{n,\lfloor nr\rfloor} - \underline{\mathbf{B}}_X(r)\right\| = O_P(n^{-a}).$$

Moreover,

$$\begin{aligned}
\mathbb{E}\left\|\underline{\mathbf{B}}_X(r)\right\| &\leq \left(\mathbb{E}\left\|\underline{\mathbf{B}}_X(r)\right\|^2\right)^{1/2} \\
&= \left(\int_0^r \left\|\mathbf{\Phi}_x(1)\mathbf{H}(s)\right\|^2 ds\right)^{1/2} \leq \left(\int_0^1 \left\|\mathbf{\Phi}_x(1)\mathbf{H}(s)\right\|^2 ds\right)^{1/2} < \infty,
\end{aligned}$$

where the bound holds uniformly over $r \in [0,1]$. Since $\underline{\mathbf{B}}_X(\cdot)$ has continuous sample paths $a.s.$, the triangle inequality and the uniform convergence above give $\sup_{r\in[0,1]} \left\|\mathbf{V}^x_{n,\lfloor nr\rfloor}\right\| = O_P(1)$. Thus,

$$\begin{aligned}
&\sup_{r\in[0,1]} \left\|\mathbf{V}^x_{n,\lfloor nr\rfloor}\mathbf{V}^{x\top}_{n,\lfloor nr\rfloor} - \underline{\mathbf{B}}_X(r)\underline{\mathbf{B}}_X^\top(r)\right\| \\
&\leq \left(\sup_{r\in[0,1]} \left\|\mathbf{V}^x_{n,\lfloor nr\rfloor}\right\| + \sup_{r\in[0,1]} \left\|\underline{\mathbf{B}}_X(r)\right\|\right)\left(\sup_{r\in[0,1]} \left\|\mathbf{V}^x_{n,\lfloor nr\rfloor} - \underline{\mathbf{B}}_X(r)\right\|\right) = O_P(n^{-a}).
\end{aligned}$$

We now show that

$$\sup_{k\geq 1} \left\|\mathbf{\Lambda}^{(k)}_{xx}\right\|_{sp} = O_p(1).$$

Let $\mathbf{z}$ be a unit vector and $\mathrm{vec}(\mathbf{Z}) = \mathbf{z}$. Then,

$$\begin{aligned}\mathbf{z}^\top \mathbf{\Lambda}_{xx}^{(k)} \mathbf{z} &= \int_0^1 \left(\underline{\mathbf{B}}_X^\top(r) \mathbf{Z} \boldsymbol{\psi}_{(k)}(r) \right)^2 dr \\ &\leq \sup_{r \in [0,1]} \|\underline{\mathbf{B}}_X(r)\|^2 \int_0^1 \left\| \mathbf{Z} \boldsymbol{\psi}_{(k)}(r) \right\|^2 dr.\end{aligned}$$

By orthonormality, $\int_0^1 \boldsymbol{\psi}_{(k)}(r) \boldsymbol{\psi}_{(k)}^\top(r) dr = \mathbf{I}_k$. This implies that $\int_0^1 \left\| \mathbf{Z} \boldsymbol{\psi}_{(k)}(r) \right\|^2 dr = \mathrm{tr}\left(\mathbf{Z}^\top \mathbf{Z} \right) = \|\mathbf{z}\|^2 = 1$. Therefore,

$$\mathbf{z}^\top \mathbf{\Lambda}_{xx}^{(k)} \mathbf{z} \leq \sup_{r \in [0,1]} \|\underline{\mathbf{B}}_X(r)\|^2 = O_p(1),$$

uniformly over k. This completes the proof of (i).

We now move on to part (ii). For any $\mathbf{z} \neq \mathbf{0}_{kd_x \times 1}$, let $\mathrm{vec}(\mathbf{Z}) = \mathbf{z}$. Using the fact that

$$\left(\boldsymbol{\psi}_{(k)}(r) \boldsymbol{\psi}_{(k)}^\top(r) \right) \otimes \left(\underline{\mathbf{B}}_X(r) \underline{\mathbf{B}}_X^\top(r) \right) = \left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r) \right) \left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r) \right)^\top$$

and the identity

$$\left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r) \right)^\top \mathrm{vec}(\mathbf{Z}) = \mathrm{vec}\left(\underline{\mathbf{B}}_X^\top(r) \mathbf{Z} \boldsymbol{\psi}_{(k)}(r) \right),$$

we obtain

$$\begin{aligned}&\mathbf{z}^\top \left(\int_0^1 \left(\boldsymbol{\psi}_{(k)}(r) \boldsymbol{\psi}_{(k)}^\top(r) \right) \otimes \left(\underline{\mathbf{B}}_X(r) \underline{\mathbf{B}}_X^\top(r) \right) dr \right) \mathbf{z} \\ &= \int_0^1 \mathbf{z}^\top \Big(\left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r) \right) \left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r) \right)^\top \Big) \mathbf{z} \, dr \\ &= \int_0^1 \left(\left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r) \right)^\top \mathrm{vec}(\mathbf{Z}) \right)^\top \left(\left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r) \right)^\top \mathrm{vec}(\mathbf{Z}) \right) dr \\ &= \int_0^1 \left(\mathrm{vec}\left(\underline{\mathbf{B}}_X^\top(r) \mathbf{Z} \boldsymbol{\psi}_{(k)}(r) \right) \right)^\top \left(\mathrm{vec}\left(\underline{\mathbf{B}}_X^\top(r) \mathbf{Z} \boldsymbol{\psi}_{(k)}(r) \right) \right) dr \\ &= \int_0^1 \left(\underline{\mathbf{B}}_X^\top(r) \mathbf{Z} \boldsymbol{\psi}_{(k)}(r) \right)^2 dr \geq 0 \quad a.s.\end{aligned}$$

To establish the strict positive definitiveness, consider the event $\mathcal{A}_k := \left\{ \omega : \mathbf{\Lambda}_{xx}^{(k)}(\omega) \text{ is singular} \right\}$. Suppose that $\mathbb{P}(\mathcal{A}_k) > 0$. For almost every $\omega \in \mathcal{A}_k$, choose a unit vector $\mathbf{z}(\omega)$ in the null space of $\mathbf{\Lambda}_{xx}^{(k)}(\omega)$ and write $\mathrm{vec}(\mathbf{Z}(\omega)) = \mathbf{z}(\omega)$. Then, by the above calculation, and the fact that $\underline{\mathbf{B}}_X(\cdot)$ is $a.s.$ continuous,

$$\underline{\mathbf{B}}_X^\top(r, \omega) \mathbf{Z}(\omega) \boldsymbol{\psi}_{(k)}(r) = 0, \qquad \forall r \in [0,1].$$

For this sample path, let $\widetilde{Y}_k(r,\omega) = \underline{\mathbf{B}}_X^\top(r,\omega)\mathbf{Z}(\omega)\boldsymbol{\psi}_{(k)}(r)$. Since $\mathbf{Z}(\omega)$ does not depend on r, by standard quadratic variation and covariation formulas for Brownian motion (see Shreve et al. (2004)),

$$\begin{aligned}\left[\widetilde{Y}_k, \widetilde{Y}_k\right](1,\omega) &= \int_0^1 \boldsymbol{\psi}_{(k)}^\top(r)\mathbf{Z}(\omega)^\top d\left[\underline{\mathbf{B}}_X, \underline{\mathbf{B}}_X\right](r,\omega)\mathbf{Z}(\omega)\boldsymbol{\psi}_{(k)}(r) \\ &= \int_0^1 \boldsymbol{\psi}_{(k)}^\top(r)\mathbf{Z}(\omega)^\top \boldsymbol{\Phi}_x(1)\boldsymbol{\Sigma}_\varepsilon(r)\boldsymbol{\Phi}_x^\top(1)\mathbf{Z}(\omega)\boldsymbol{\psi}_{(k)}(r)\, dr \\ &\geq \int_0^1 \lambda_{\min}\left(\boldsymbol{\Phi}_x(1)\boldsymbol{\Sigma}_\varepsilon(r)\boldsymbol{\Phi}_x^\top(1)\right)\left\|\mathbf{Z}(\omega)\boldsymbol{\psi}_{(k)}(r)\right\|^2 > 0,\end{aligned}$$

where the last equality follows from the orthonormality of $\{\psi_j(r)\}$ and Assumption 1(iii) also guarantees that $\boldsymbol{\Sigma}_\varepsilon(r)$ is uniformly bounded over $r \in [0,1]$. This contradicts the fact that $\widetilde{Y}_k(r,\omega)$ is zero for every $r \in [0,1]$, because an identically zero path has zero quadratic variation. Hence, $\mathbb{P}(\mathcal{A}_k) = 0$, and therefore $\boldsymbol{\Lambda}_{xx}^{(k)}(\omega)$ is positive definite $a.s.$ for every fixed $k \geq 1$.

However, the above argument does not provide a lower bound for $\lambda_{\min}\left(\boldsymbol{\Lambda}_{xx}^{(k)}\right)$ which is uniformly over k. We next show that, in fact, we can construct k-dependent unit vector $\mathbf{z}_k$, such that

$$\mathbf{z}_k^\top \boldsymbol{\Lambda}_{xx}^{(k)}\mathbf{z}_k \to 0 \qquad \text{a.s.}$$

as $k \to \infty$. Define

$$\mathbf{c}_k := \frac{\boldsymbol{\psi}_{(k)}(0)}{\left\|\boldsymbol{\psi}_{(k)}(0)\right\|} = \frac{\left(1, \sqrt{2}, \ldots, \sqrt{2}\right)^\top}{\sqrt{2k-1}},$$

and let

$$h_k(r) := \mathbf{c}_k^\top \boldsymbol{\psi}_{(k)}(r).$$

By the orthonormality of $\{\psi_j(\cdot)\}$, we have

$$\int_0^1 h_k^2(r)dr = \|\mathbf{c}_k\|^2 = 1.$$

Moreover, using the trigonometric identity

$$1 + 2\sum_{j=1}^{k-1}\cos(jx) = \frac{\sin\left\{\left(k-\frac{1}{2}\right)x\right\}}{\sin(x/2)},$$

we have

$$h_k(r) = \frac{\sin\left\{\left(k-\frac{1}{2}\right)\pi r\right\}}{\sqrt{2k-1}\sin(\pi r/2)}.$$

Consequently, for every fixed $\delta > 0$,

$$\int_{\delta}^{1} h_k^2(r)dr \leq \frac{1}{(2k-1)\sin^2(\pi\delta/2)} \to 0,$$

as $k \to \infty$.

Now, fix any unit vector $\mathbf{e} \in \mathbb{R}^{d_x}$ and define $\mathbf{z}_k := \mathbf{c}_k \otimes \mathbf{e}$. Then, $\|\mathbf{z}_k\| = 1$. We have

$$\begin{aligned} 0 \leq \lambda_{\min}\left(\mathbf{\Lambda}_{xx}^{(k)}\right) &\leq \mathbf{z}_k^\top \mathbf{\Lambda}_{xx}^{(k)} \mathbf{z}_k \\ &= \int_0^1 h_k^2(r) \left\{\mathbf{e}^\top \underline{\mathbf{B}}_X(r)\right\}^2 dr \leq \int_0^1 h_k^2(r) \left\|\underline{\mathbf{B}}_X(r)\right\|^2 dr. \end{aligned}$$

Given that $\int_0^1 h_k^2(r)dr = 1$ and $\int_\delta^1 h_k^2(r)dr \to 0$ for every fixed $\delta > 0$, the unit mass of $h_k^2(r)$ would be concentrated in arbitrarily small neighborhoods of $r = 0$. Since $\underline{\mathbf{B}}_X(\cdot)$ is continuous $a.s.$ and $\underline{\mathbf{B}}_X(0) = \mathbf{0}_{d_x \times 1}$, it follows that

$$\int_0^1 h_k^2(r) \left\|\underline{\mathbf{B}}_X(r)\right\|^2 dr \to 0, \quad a.s.$$

Since $\mathbf{z}_k^\top \mathbf{\Lambda}_{xx}^{(k)} \mathbf{z}_k \geq 0$, we have

$$\mathbf{z}_k^\top \mathbf{\Lambda}_{xx}^{(k)} \mathbf{z}_k \to 0 \quad a.s.$$

Therefore,

$$\lambda_{\min}\left(\mathbf{\Lambda}_{xx}^{(k)}\right) \to 0 \quad a.s.$$

Since $\mathbf{\Lambda}_{xx}^{(k)}$ is positive definite $a.s.$ for every fixed $k \geq 1$, it follows that

$$\left\|\left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1}\right\|_{sp} = \frac{1}{\lambda_{\min}\left(\mathbf{\Lambda}_{xx}^{(k)}\right)} \to \infty \qquad \text{a.s.}$$

□

B.3. Proof of Proposition 1

Recall that the estimator for $\mathbf{B}^{(k)}$ is given by

$$\widehat{\mathbf{B}}_n^{(k)} = \left(\sum_{t=1}^{n} \mathbf{y}_t \mathbf{X}_t^{(k)\top}\right)\left(\sum_{t=1}^{n} \mathbf{X}_t^{(k)} \mathbf{X}_t^{(k)\top}\right)^{-1}.$$

Simple algebra gives

$$\begin{aligned} n \cdot \left(\widehat{\mathbf{B}}_n^{(k)} - \mathbf{B}^{(k)}\right) &= \left(\frac{1}{n}\sum_{t=1}^{n} \mathbf{u}_{0t}^{(k)} \mathbf{X}_t^{(k)\top}\right) \left(\frac{1}{n^2}\sum_{t=1}^{n} \mathbf{X}_t^{(k)} \mathbf{X}_t^{(k)\top}\right)^{-1} \\ &= \left(\frac{1}{n}\sum_{t=1}^{n} \mathbf{u}_{0t} \mathbf{X}_t^{(k)\top}\right) \left(\frac{1}{n^2}\sum_{t=1}^{n} \mathbf{X}_t^{(k)} \mathbf{X}_t^{(k)\top}\right)^{-1} \\ &\quad + \left(\frac{1}{n}\sum_{t=1}^{n} \left(\mathbf{A}_t - \mathbf{A}_k(\tau_t)\right) \mathbf{X}_t \mathbf{X}_t^{(k)\top}\right) \left(\frac{1}{n^2}\sum_{t=1}^{n} \mathbf{X}_t^{(k)} \mathbf{X}_t^{(k)\top}\right)^{-1} \\ &:= \mathbf{\Gamma}_{n,ux}^{(k)} \left(\mathbf{\Lambda}_{n,xx}^{(k)}\right)^{-1} + \mathbf{\Delta}_{n,xx}^{(k)} \left(\mathbf{\Lambda}_{n,xx}^{(k)}\right)^{-1}. \end{aligned}$$

Using the fact that

$$\begin{aligned} &\left\| \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1/2} \left(\mathbf{\Lambda}_{n,xx}^{(k)} - \mathbf{\Lambda}_{xx}^{(k)}\right) \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1/2} \right\|_{sp} \leq \left\| \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1} \right\|_{sp} \left\| \mathbf{\Lambda}_{n,xx}^{(k)} - \mathbf{\Lambda}_{xx}^{(k)} \right\|_{sp} \\ &= O_p\left(\varrho_k \left(\frac{k}{n^a} + \frac{k^2}{n}\right)\right) = o_p(1), \end{aligned}$$

together with the identity

$$\mathbf{\Lambda}_{n,xx}^{(k)} = \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{1/2} \left\{ \mathbf{I}_{kd_x} + \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1/2} \left(\mathbf{\Lambda}_{n,xx}^{(k)} - \mathbf{\Lambda}_{xx}^{(k)}\right) \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1/2} \right\} \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{1/2},$$

we obtain

$$\begin{aligned} &\left\| \left(\mathbf{\Lambda}_{n,xx}^{(k)}\right)^{-1} \right\|_{sp} \leq \left\| \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1/2} \right\|_{sp} \left\| \left(\mathbf{I}_{kd_x} + \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1/2} \left(\mathbf{\Lambda}_{n,xx}^{(k)} - \mathbf{\Lambda}_{xx}^{(k)}\right) \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1/2} \right)^{-1} \right\|_{sp} \left\| \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1/2} \right\|_{sp} \\ &\leq \left\| \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1} \right\|_{sp} \frac{1}{1 - \left\| \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1/2} \left(\mathbf{\Lambda}_{n,xx}^{(k)} - \mathbf{\Lambda}_{xx}^{(k)}\right) \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1/2} \right\|_{sp}} = O_p(\varrho_k). \end{aligned}$$

This, together with Lemma D.2, implies that

$$\left\| \mathbf{\Delta}_{n,xx}^{(k)} \left(\mathbf{\Lambda}_{n,xx}^{(k)}\right)^{-1} \right\|_{sp} \leq \left\| \mathbf{\Delta}_{n,xx}^{(k)} \right\|_{sp} \left\| \left(\mathbf{\Lambda}_{n,xx}^{(k)}\right)^{-1} \right\|_{sp} = O_p\left(\varrho_k \frac{n}{k^q}\right) = o_p(1).$$

Observe that

$$\begin{aligned} \left\| \left(\mathbf{\Lambda}_{n,xx}^{(k)}\right)^{-1} - \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1} \right\|_{sp} &\leq \left\| \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1} \right\|_{sp} \left\| \mathbf{\Lambda}_{n,xx}^{(k)} - \mathbf{\Lambda}_{xx}^{(k)} \right\|_{sp} \left\| \left(\mathbf{\Lambda}_{n,xx}^{(k)}\right)^{-1} \right\|_{sp} \\ &= O_p\left(\varrho_k^2 \left(\frac{k}{n^a} + \frac{k^2}{n}\right)\right) = o_p(1), \end{aligned}$$

together with Lemma D.1:

$$\left\| \frac{1}{n} \sum_{t=1}^{n} \mathbf{u}_{0t} \mathbf{X}_t^{(k)\top} - \mathbf{\Gamma}_{ux}^{(k)} \right\|_{sp} = O_p\left(\frac{k^{3/2}}{n^{\min(\delta,1/2)}} \right) = o_p(1),$$

we obtain

$$\begin{aligned}
&\left\| \mathbf{\Gamma}_{n,ux}^{(k)} \left(\mathbf{\Lambda}_{n,xx}^{(k)} \right)^{-1} - \mathbf{\Gamma}_{ux}^{(k)} \left(\mathbf{\Lambda}_{xx}^{(k)} \right)^{-1} \right\|_{sp} \\
\leq\ & \left\| \mathbf{\Gamma}_{n,ux}^{(k)} - \mathbf{\Gamma}_{ux}^{(k)} \right\|_{sp} \left\| \left(\mathbf{\Lambda}_{n,xx}^{(k)} \right)^{-1} \right\|_{sp} + \left\| \mathbf{\Gamma}_{ux}^{(k)} \right\|_{sp} \left\| \left(\mathbf{\Lambda}_{n,xx}^{(k)} \right)^{-1} - \left(\mathbf{\Lambda}_{xx}^{(k)} \right)^{-1} \right\|_{sp} \\
=\ & O_p\left(\varrho_k \frac{k^{3/2}}{n^{\min(\delta,1/2)}} \right) + O_p\left(\sqrt{k} \varrho_k^2 \left(\frac{k}{n^a} + \frac{k^2}{n} \right) \right) = o_p(1),
\end{aligned}$$

as claimed. □

B.4. Proof of Theorem 1

First, let $a(z_0)$, $a_k(z_0)$ be a generic element in $\mathbf{A}(z_0)$, $\mathbf{A}_k(z_0)$, respectively. Then, when q is even, we have

$$\begin{aligned}
|a_k(z_0) - a(z_0)| &\leq \sqrt{2} \sum_{j=k}^{\infty} |b_{1,j}| = \sqrt{2} \sum_{j=k}^{\infty} \frac{1}{(j\pi)^q} \left| \int_0^1 a^{(q)}(r) \sqrt{2} \cos(j\pi r)\, dr \right| \\
&\leq \sqrt{2} \left(\sum_{j=k}^{\infty} \frac{1}{(j\pi)^{2q}} \right)^{1/2} \left(\sum_{j=k}^{\infty} \left(\int_0^1 a^{(q)}(r) \sqrt{2} \cos(j\pi r)\, dr \right)^2 \right)^{1/2} \\
&\leq O(1) \left(\sum_{j=k}^{\infty} \frac{1}{(j\pi)^{2q}} \right)^{1/2} = O\left(k^{-q+1/2} \right),
\end{aligned}$$

where $b_{1,j}$ is the Fourier coefficient defined in (10), and the equality uses $b_{1,j} = \mathcal{R}_{q,j}$ together with the even-q definition of $\mathcal{R}_{q,j}$ in (B.1), which is established in the proof of Lemma 1. The second inequality follows by the Cauchy–Schwarz inequality, and the final inequality follows from Bessel's inequality as in Theorem 4.17 of Rudin (1987) and the fact that

$$\int_k^{\infty} \frac{1}{x^{2q}} dx \leq \sum_{j=k}^{\infty} \frac{1}{j^{2q}} \leq \int_{k-1}^{\infty} \frac{1}{x^{2q}} dx,$$

so that final rate is obtained by calculating the associated integrals $\int_k^\infty \frac{1}{x^{2q}}dx$ and $\int_{k-1}^\infty \frac{1}{x^{2q}}dx$. Similarly, when q is odd, we can also show that

$$|a_k(z_0) - a(z_0)| = O\left(k^{-q+1/2}\right).$$

This implies that

$$n \cdot \|\mathbf{A}_k(z_0) - \mathbf{A}(z_0)\| = O\left(\sqrt{k}\frac{n}{k^q}\right) = o\left(\sqrt{k}\right).$$

Together with Proposition 1 and the triangular inequality, we obtain

$$\left\| n \cdot \mathrm{vec}\left(\widehat{\mathbf{A}}_n(z_0) - \mathbf{A}(z_0)\right) - \left(\mathbf{T}_{(k)}^\top(z_0)\left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1} \otimes \mathbf{I}_{d_y}\right) \mathrm{vec}\left\{\mathbf{\Gamma}_{ux}^{(k)}\right\} \right\| = o_p(\sqrt{k}).$$

If for almost every $r \in [0,1]$, $\mathbf{\Omega}_{0x}(r) = \mathbf{0}_{d_y \times d_x}$ and $\mathbf{\Lambda}_{0x}(r) + \mathbf{\Sigma}_{0x}(r) = \mathbf{0}_{d_y \times d_x}$,

$$\mathbf{\Gamma}_{ux}^{(k)} = \int_0^1 d\underline{\mathbf{B}}_0(r)\left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r)\right)^\top.$$

As the continuous time martingale $\underline{\mathbf{B}}_0(r)$ and $\underline{\mathbf{B}}_X(r)$ are uncorrelated and hence independently distributed, we can proceed in the same way as Lemma 5.1 of Park and Phillips (1988) to show that

$$\mathrm{vec}\left(\mathbf{\Gamma}_{ux}^{(k)}\right)|\mathcal{F}_X(1) \equiv \mathcal{N}\left(\mathbf{0}_{d_y k d_x \times 1}, \int_0^1 \left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r)\right)\left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r)\right)^\top \otimes \mathbf{\Omega}_{00}(r)dr\right), \qquad a.s.$$

where $\mathcal{F}_X(r) := \sigma\{\underline{\mathbf{B}}_X(s) : 0 \le s \le r\}$, for $r \in [0,1]$.

It is useful to note that, although the dimension of this conditional covariance matrix increases with k, its spectral norm remains uniformly bounded. Indeed, by Lemma 2,

$$\begin{aligned}
&\sup_{k\ge 1}\left\|\int_0^1 \left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r)\right)\left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r)\right)^\top \otimes \mathbf{\Omega}_{00}(r)dr\right\|_{sp} \\
\le &\left(\sup_{r\in[0,1]}\|\mathbf{\Omega}_{00}(r)\|_{sp}\right)\left(\sup_{k\ge 1}\left\|\mathbf{\Lambda}_{xx}^{(k)}\right\|_{sp}\right) = O_p(1).
\end{aligned}$$

Let

$$\begin{aligned}
&\mathbb{V}_{(k)}(z_0) \\
&= \int_0^1 \left\{\left(\mathbf{T}_{(k)}^\top(z_0)\left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1}\left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r)\right)\right)\left(\mathbf{T}_{(k)}^\top(z_0)\left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1}\left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r)\right)\right)^\top\right\} \otimes \mathbf{\Omega}_{00}(r)dr.
\end{aligned}$$

Therefore

$$\left(\mathbb{V}_{(k)}(z_0)\right)^{-1/2}\left(\mathbf{T}_{(k)}^{\top}(z_0)\left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1}\otimes\mathbf{I}_{d_y}\right)\operatorname{vec}\left\{\mathbf{\Gamma}_{ux}^{(k)}\right\}|\mathcal{F}_X(1)\equiv\mathcal{N}\left(\mathbf{0}_{d_yd_x\times1},\mathbf{I}_{d_yd_x}\right),\qquad a.s.$$

Since the conditional distribution on the left-hand side is $a.s.$ $\mathcal{N}\left(\mathbf{0}_{d_yd_x\times1},\mathbf{I}_{d_yd_x}\right)$, which does not depend on $\mathcal{F}_X(1)$, the unconditional distribution is also $\mathcal{N}\left(\mathbf{0}_{d_yd_x\times1},\mathbf{I}_{d_yd_x}\right)$.

What remains is to verify that $\left\|\left(\mathbb{V}_{(k)}(z_0)\right)^{-1/2}\right\|_{sp}=O_p\left(k^{-1/2}\right)$. By Lemma 2, $\left\|\mathbf{\Lambda}_{xx}^{(k)}\right\|_{sp}=O_p(1)$ uniformly over k. Then, we have

$$\left\|\left(\mathbb{V}_{(k)}(z_0)\right)^{-1/2}\right\|_{sp}\leq\left(\sup_{r\in[0,1]}\left\|\mathbf{\Omega}_{00}^{-1/2}(r)\right\|_{sp}\right)\left\|\left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{1/2}\right\|_{sp}\left\|\boldsymbol{\psi}_{(k)}(z_0)\right\|^{-1}=O_p(k^{-1/2}).$$

Consequently,

$$n\left(\mathbb{V}_{(k)}(z_0)\right)^{-1/2}\operatorname{vec}\left(\widehat{\mathbf{A}}_n(z_0)-\mathbf{A}(z_0)\right)=\left(\mathbb{V}_{(k)}(z_0)\right)^{-1/2}\left\{\mathbf{T}_{(k)}^{\top}(z_0)\left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1}\otimes\mathbf{I}_{d_y}\right\}\operatorname{vec}\left(\mathbf{\Gamma}_{ux}^{(k)}\right)+o_p(1).$$

Since the first term on the right-hand side has an $\mathcal{N}\left(\mathbf{0}_{d_yd_x\times1},\mathbf{I}_{d_yd_x}\right)$ distribution for every n, the result follows from Slutsky's theorem:

$$n\left(\mathbb{V}_{(k)}(z_0)\right)^{-1/2}\operatorname{vec}\left(\widehat{\mathbf{A}}_n(z_0)-\mathbf{A}(z_0)\right)\Rightarrow\mathcal{N}\left(\mathbf{0}_{d_yd_x\times1},\mathbf{I}_{d_yd_x}\right).\qquad\square$$

C — Proofs of the main results in Section 3

C.1. Proof of Theorem 2

Let $\widetilde{\mathbf{A}}_n(z_0)\in\mathcal{S}(z_0)$ be an arbitrary estimator of $\mathbf{A}(z_0)$. Then, by Definition 1, we have the following

$$\left\|n\cdot\left(\widetilde{\mathbf{A}}_n(z_0)-\mathbf{A}(z_0)\right)-\left[\left(\int_0^1 d\mathbf{B}_0(r)\mathbf{L}_{(k)}^{\top}(r)\right)\mathbf{V}_{(k)}+\mathbf{U}_{(k)}\right]\mathbf{T}_{(k)}(z_0)\right\|_{sp}=o_p(\sqrt{k}),$$

where $\mathbf{L}_{(k)}(r)$, $\mathbf{V}_{(k)}$, $\mathbf{U}_{(k)}$, and $\mathbf{T}_{(k)}(z_0)$ are defined as in Definition 1. Furthermore, let $\mathbf{B}_{0\cdot x}(r)=\mathbf{B}_0(r)-\mathbf{\Omega}_{0x}\mathbf{\Omega}_{xx}^{-1}\mathbf{B}_X(r)$. We have

$$\int_0^1 d\mathbf{B}_0(r)\mathbf{L}_{(k)}^{\top}(r)=\int_0^1 d\mathbf{B}_{0\cdot x}(r)\mathbf{L}_{(k)}^{\top}(r)+\mathbf{\Omega}_{0x}\mathbf{\Omega}_{xx}^{-1}\int_0^1 d\mathbf{B}_X(r)\mathbf{L}_{(k)}^{\top}(r).$$

Write the leading term as

$$\mathbf{R}_{(k)}(z_0) = \left[\left(\int_0^1 d\mathbf{B}_{0\cdot x}(r)\mathbf{L}_{(k)}^\top(r)\right)\mathbf{V}_{(k)}\right]\mathbf{T}_{(k)}(z_0) + \boldsymbol{\Omega}_{0x}\boldsymbol{\Omega}_{xx}^{-1}\left(\int_0^1 d\mathbf{B}_X(r)\mathbf{L}_{(k)}^\top(r)\right)\mathbf{V}_{(k)}\mathbf{T}_{(k)}(z_0) + \mathbf{U}_{(k)}\mathbf{T}_{(k)}(z_0).$$

In addition, recall that

$$\mathbb{V}_{(k)}^e(z_0) = \left\{\mathbf{T}_{(k)}^\top(z_0)\left(\boldsymbol{\Lambda}_{xx}^{(k)}\right)^{-1}\mathbf{T}_{(k)}(z_0)\right\}\otimes\boldsymbol{\Omega}_{0\cdot x},$$

where $\boldsymbol{\Omega}_{0\cdot x} = \boldsymbol{\Omega}_{00} - \boldsymbol{\Omega}_{0x}\boldsymbol{\Omega}_{xx}^{-1}\boldsymbol{\Omega}_{x0}$. Since

$$\left\|\left(\mathbb{V}_{(k)}^e(z_0)\right)^{-1/2}\right\|_{sp} \leq \left\|(\boldsymbol{\Omega}_{0\cdot x})^{-1/2}\right\|_{sp}\left\|\left(\boldsymbol{\Lambda}_{xx}^{(k)}\right)^{1/2}\right\|_{sp}\left\|\boldsymbol{\psi}_{(k)}(z_0)\right\|^{-1} = O_p(k^{-1/2}),$$

we have

$$\left\|n\left(\mathbb{V}_{(k)}^e(z_0)\right)^{-1/2}\mathrm{vec}\left(\widetilde{\mathbf{A}}_n^{(k)}(z_0) - \mathbf{A}(z_0)\right) - \boldsymbol{\theta}_{(k)}(z_0)\right\| = o_p(1),$$

where $\boldsymbol{\theta}_{(k)}(z_0) = \left(\mathbb{V}_{(k)}^e(z_0)\right)^{-1/2}\mathrm{vec}\left(\mathbf{R}_{(k)}(z_0)\right)$. Let $\mathcal{F}_X(r) := \sigma\{\mathbf{B}_X(s) : 0 \leq s \leq r\}$, for $r \in [0,1]$. Then, by proceeding as in the proof of Lemma 5.1 in Park and Phillips (1988), we have

$$\boldsymbol{\theta}_{(k)}(z_0)|\mathcal{F}_X(1) \equiv \mathcal{N}\left(\boldsymbol{\mu}_{(k)}(z_0), \boldsymbol{\Sigma}_{(k)}(z_0)\right), \qquad a.s.,$$

where

$$\boldsymbol{\mu}_{(k)}(z_0) = \left(\mathbb{V}_{(k)}^e(z_0)\right)^{-1/2}\mathrm{vec}\left(\left[\boldsymbol{\Omega}_{0x}\boldsymbol{\Omega}_{xx}^{-1}\left(\int_0^1 d\mathbf{B}_X(r)\mathbf{L}_{(k)}^\top(r)\right)\mathbf{V}_{(k)} + \mathbf{U}_{(k)}\right]\mathbf{T}_{(k)}(z_0)\right),$$

$$\boldsymbol{\Sigma}_{(k)}(z_0) = \left(\mathbb{V}_{(k)}^e(z_0)\right)^{-1/2}\left\{\mathbf{T}_{(k)}^\top(z_0)\mathbf{V}_{(k)}^\top\left(\int_0^1 \mathbf{L}_{(k)}(r)\mathbf{L}_{(k)}^\top(r)dr\right)\mathbf{V}_{(k)}\mathbf{T}_{(k)}(z_0)\otimes\boldsymbol{\Omega}_{0\cdot x}\right\}\left(\mathbb{V}_{(k)}^e(z_0)\right)^{-1/2}.$$

Recall that, the efficient leading term $\boldsymbol{\phi}_{(k)}(z_0)$ is defined according to

$$\boldsymbol{\phi}_{(k)}(z_0) := \left(\mathbb{V}_{(k)}^e(z_0)\right)^{-1/2}\mathrm{vec}\left(\left[\int_0^1 d\mathbf{B}_{0\cdot x}(r)\left(\boldsymbol{\psi}_{(k)}(r)\otimes\mathbf{B}_X(r)\right)^\top\right]\left(\boldsymbol{\Lambda}_{xx}^{(k)}\right)^{-1}\mathbf{T}_{(k)}(z_0)\right),$$

where

$$\boldsymbol{\phi}_{(k)}(z_0)|\mathcal{F}_X(1) \equiv \mathcal{N}\left(\mathbf{0}_{d_y d_x\times 1}, \mathbf{I}_{d_y d_x}\right), \qquad a.s.$$

Given that conditioning on $\mathcal{F}_X(1)$, both $\boldsymbol{\theta}_{(k)}(z_0)$ and $\boldsymbol{\phi}_{(k)}(z_0)$ are affine transformations of $\mathbf{B}_{0\cdot x}(r)$. Hence,

they are jointly normally distributed conditional on $\mathcal{F}_X(1)$, $a.s.$ Let

$$\boldsymbol{\eta}_{(k)}(z_0) := \boldsymbol{\theta}_{(k)}(z_0) - \boldsymbol{\phi}_{(k)}(z_0).$$

Since $\boldsymbol{\eta}_{(k)}(z_0)$ is a linear transformation of the joint conditional Normal vector $\left(\boldsymbol{\theta}_{(k)}^\top(z_0), \boldsymbol{\phi}_{(k)}^\top(z_0)\right)^\top$, it follows that $\left(\boldsymbol{\eta}_{(k)}^\top(z_0), \boldsymbol{\phi}_{(k)}^\top(z_0)\right)^\top$ is also jointly conditional Normal, $a.s.$

Note that

$$\begin{aligned}
&\mathbb{E}\left[\boldsymbol{\theta}_{(k)}(z_0)\boldsymbol{\phi}_{(k)}^\top(z_0)|\mathcal{F}_X(1)\right] \\
&= \left(\mathbb{V}_{(k)}^e(z_0)\right)^{-1/2}\left\{\mathbf{T}_{(k)}^\top(z_0)\mathbf{V}_{(k)}^\top\left(\int_0^1 \mathbf{L}_{(k)}(r)\mathbf{q}_{(k)}^\top(r)\,dr\right)\left(\boldsymbol{\Lambda}_{xx}^{(k)}\right)^{-1}\mathbf{T}_{(k)}(z_0)\otimes\boldsymbol{\Omega}_{0\cdot x}\right\}\left(\mathbb{V}_{(k)}^e(z_0)\right)^{-1/2}, a.s.
\end{aligned}$$

where $\mathbf{q}_{(k)}(r) := \boldsymbol{\psi}_{(k)}(r)\otimes\mathbf{B}_X(r)$. By (23), we have

$$\mathbf{T}_{(k)}^\top(z_0)\mathbf{V}_{(k)}^\top\left(\int_0^1 \mathbf{L}_{(k)}(r)\mathbf{q}_{(k)}^\top(r)\,dr\right) = \mathbf{T}_{(k)}^\top(z_0), \qquad a.s.$$

This, together with the definition of $\mathbb{V}_{(k)}^e(z_0)$, implies that

$$\mathbb{E}\left[\boldsymbol{\theta}_{(k)}(z_0)\boldsymbol{\phi}_{(k)}^\top(z_0)|\mathcal{F}_X(1)\right] = \mathbf{I}_{d_y d_x}, \qquad a.s.$$

Then, recall that $\boldsymbol{\eta}_{(k)}(z_0) = \boldsymbol{\theta}_{(k)}(z_0) - \boldsymbol{\phi}_{(k)}(z_0)$, we have

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{\eta}_{(k)}(z_0)\boldsymbol{\phi}_{(k)}^\top(z_0)|\mathcal{F}_X(1)\right] &= \mathbb{E}\left[\boldsymbol{\theta}_{(k)}(z_0)\boldsymbol{\phi}_{(k)}^\top(z_0)|\mathcal{F}_X(1)\right] - \mathbb{E}\left[\boldsymbol{\phi}_{(k)}(z_0)\boldsymbol{\phi}_{(k)}^\top(z_0)|\mathcal{F}_X(1)\right] \\
&= \mathbf{0}_{d_y d_x}, \qquad a.s.
\end{aligned} \tag{C.1}$$

As $\left(\boldsymbol{\eta}_{(k)}^\top(z_0), \boldsymbol{\phi}_{(k)}^\top(z_0)\right)^\top$ is also jointly conditional Normal, $a.s.$, (C.1) also implies that $\boldsymbol{\eta}_{(k)}(z_0)$ and $\boldsymbol{\phi}_{(k)}(z_0)$ are conditionally independent given $\mathcal{F}_X(1)$, $a.s.$. Given that the conditional distribution of $\boldsymbol{\phi}_{(k)}(z_0)$ given $\mathcal{F}_X(1)$ is standard Normal and does not depend on $\mathcal{F}_X(1)$, they are also unconditionally independent.

Then, for each $k \geq 1$, by Theorem 2 of Anderson (1955) or Lemma 1 in Basawa and Scott (1983), we have

$$\mathbb{P}\left(\boldsymbol{\theta}_{(k)}(z_0) \in \mathcal{C}|\mathcal{F}_X(1)\right) \leq \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{C}\right), \qquad a.s.$$

where $\boldsymbol{\xi} \sim \mathcal{N}\left(\mathbf{0}_{d_y d_x \times 1}, \mathbf{I}_{d_y d_x}\right)$. Taking expectations on both sides yields

$$\mathbb{E}\left[\mathbb{P}\left(\boldsymbol{\theta}_{(k)}(z_0) \in \mathcal{C} | \mathcal{F}_X(1)\right)\right] = \mathbb{P}\left(\boldsymbol{\theta}_{(k)}(z_0) \in \mathcal{C}\right) \leq \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{C}\right), \tag{C.2}$$

so the unconditional inequality also holds by the law of iterated expectations.

Fix $\varepsilon > 0$, define $\mathcal{C}_\varepsilon := \mathcal{C} + \left\{\mathbf{x} \in \mathbb{R}^{d_y d_x} : \|\mathbf{x}\| \leq \varepsilon\right\}$. Recall that

$$\left\| n \left(\mathbb{V}^e_{(k)}(z_0)\right)^{-1/2} \widetilde{\mathbf{R}}_n(z_0) - \boldsymbol{\theta}_{(k)}(z_0) \right\| = o_p(1),$$

where $\widetilde{\mathbf{R}}_n(z_0) = \operatorname{vec}\left(\widetilde{\mathbf{A}}^{(k)}_n(z_0) - \mathbf{A}(z_0)\right)$. Then,

$$\begin{aligned}
&\mathbb{P}\left(n \left(\mathbb{V}^e_{(k)}(z_0)\right)^{-1/2} \widetilde{\mathbf{R}}_n(z_0) \in \mathcal{C}\right) \\
&\leq \mathbb{P}\left(\boldsymbol{\theta}_{(k)}(z_0) \in \mathcal{C}_\varepsilon\right) + \mathbb{P}\left(\left\| n \left(\mathbb{V}^e_{(k)}(z_0)\right)^{-1/2} \widetilde{\mathbf{R}}_n(z_0) - \boldsymbol{\theta}_{(k)}(z_0) \right\| > \varepsilon\right).
\end{aligned}$$

By the preceding expansion, the second term above is $o(1)$ as $n \to \infty$. Thus, by (C.2), we have

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{P}\left(n \left(\mathbb{V}^e_{(k)}(z_0)\right)^{-1/2} \widetilde{\mathbf{R}}_n(z_0) \in \mathcal{C}\right) &\leq \limsup_{n\to\infty} \mathbb{P}\left(\boldsymbol{\theta}_{(k)}(z_0) \in \mathcal{C}_\varepsilon\right) \\
&\leq \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{C}_\varepsilon\right).
\end{aligned}$$

Since the above inequality holds for any fixed $\varepsilon > 0$, we have

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{P}\left(n \left(\mathbb{V}^e_{(k)}(z_0)\right)^{-1/2} \widetilde{\mathbf{R}}_n(z_0) \in \mathcal{C}\right) &\leq \inf_{\varepsilon>0} \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{C}_\varepsilon\right) \\
&= \lim_{\varepsilon \downarrow 0} \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{C}_\varepsilon\right) = \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{C}\right),
\end{aligned}$$

the proof of the first part of the theorem is now completed.

To prove the second part, let

$$n \left(\mathbb{V}^e_{(k)}(z_0)\right)^{-1/2} \widetilde{\mathbf{R}}_n(z_0) = \boldsymbol{\phi}_{(k)}(z_0) + \boldsymbol{\eta}_{(k)}(z_0) + \mathbf{r}_{(k)}(z_0),$$

where $\left\|\mathbf{r}_{(k)}(z_0)\right\| = o_p(1)$. Let $\mathcal{C}_\varepsilon$ be defined as before. Let $\varepsilon_0 > 0$ be as in (26) and define the event $\mathcal{A}_n := \left\{\left\|\boldsymbol{\eta}_{(k)}(z_0)\right\| > \varepsilon_0/2\right\}$. Since $\mathbf{r}_{(k)}(z_0) = o_p(1)$, the triangle inequality and (26) imply that $\liminf_{n\to\infty} \mathbb{P}(\mathcal{A}_n) > 0$. By Theorem 2 in Anderson (1955) or Lemma 1 in Basawa and Scott (1983) and the

fact that $\boldsymbol{\eta}_{(k)}(z_0)$ and $\boldsymbol{\phi}_{(k)}(z_0)$ are independent, we have

$$\mathbb{P}\left(\boldsymbol{\phi}_{(k)}(z_0)+\boldsymbol{\eta}_{(k)}(z_0)\in\mathcal{C}_\varepsilon|\boldsymbol{\eta}_{(k)}(z_0)\right)\leq\mathbb{P}\left(\boldsymbol{\xi}\in\mathcal{C}_\varepsilon\right),\quad a.s.$$

On $\mathcal{A}_n$, we have

$$\mathbb{P}\left(\boldsymbol{\phi}_{(k)}(z_0)+\boldsymbol{\eta}_{(k)}(z_0)\in\mathcal{C}_\varepsilon|\boldsymbol{\eta}_{(k)}(z_0)\right)\leq\max_{\|\mathbf{a}\|\geq\varepsilon_0/2}\mathbb{P}\left(\boldsymbol{\xi}+\mathbf{a}\in\mathcal{C}_\varepsilon\right),\quad a.s.$$

Let $\Delta_\varepsilon(\varepsilon_0):=\mathbb{P}\left(\boldsymbol{\xi}\in\mathcal{C}_\varepsilon\right)-\max_{\|\mathbf{a}\|\geq\varepsilon_0/2}\mathbb{P}\left(\boldsymbol{\xi}+\mathbf{a}\in\mathcal{C}_\varepsilon\right)$. Since $\mathcal{C}_\varepsilon$ is bounded, the strict form of Anderson's inequality gives $\mathbb{P}(\boldsymbol{\xi}+\mathbf{a}\in\mathcal{C}_\varepsilon)<\mathbb{P}(\boldsymbol{\xi}\in\mathcal{C}_\varepsilon)$ for every $\mathbf{a}\neq\mathbf{0}$. Moreover, the function $\mathbf{a}\mapsto\mathbb{P}(\boldsymbol{\xi}+\mathbf{a}\in\mathcal{C}_\varepsilon)$ is continuous and converges to zero as $\|\mathbf{a}\|\to\infty$. Hence, its supremum over $\{\mathbf{a}:\|\mathbf{a}\|\geq\varepsilon_0/2\}$ can be restricted to a compact set of the form $\{\mathbf{a}:\varepsilon_0/2\leq\|\mathbf{a}\|\leq R\}$ for some sufficiently large R, and is therefore attained. Consequently, $\Delta_\varepsilon(\varepsilon_0)>0$.

Obviously, By the law of iterated expectations, for sufficiently large n, we have

$$\begin{aligned}
\mathbb{P}\left(\boldsymbol{\phi}_{(k)}(z_0)+\boldsymbol{\eta}_{(k)}(z_0)\in\mathcal{C}_\varepsilon\right)&=\mathbb{E}\left[\mathbb{P}\left(\boldsymbol{\phi}_{(k)}(z_0)+\boldsymbol{\eta}_{(k)}(z_0)\in\mathcal{C}_\varepsilon|\boldsymbol{\eta}_{(k)}(z_0)\right)\right]\\
&\leq\mathbb{P}\left(\boldsymbol{\xi}\in\mathcal{C}_\varepsilon\right)\mathbb{P}\left(\mathcal{A}_n^c\right)+\max_{\|\mathbf{a}\|\geq\varepsilon_0/2}\mathbb{P}\left(\boldsymbol{\xi}+\mathbf{a}\in\mathcal{C}_\varepsilon\right)\mathbb{P}\left(\mathcal{A}_n\right)\\
&=\mathbb{P}\left(\boldsymbol{\xi}\in\mathcal{C}_\varepsilon\right)-\Delta_\varepsilon(\varepsilon_0)\mathbb{P}\left(\left\|\boldsymbol{\eta}_{(k)}(z_0)\right\|>\varepsilon_0/2\right)\\
&\leq\mathbb{P}\left(\boldsymbol{\xi}\in\mathcal{C}_\varepsilon\right)-\Delta_\varepsilon(\varepsilon_0)c_0,
\end{aligned}$$

where $c_0>0$ can be set to be $\frac{1}{2}\liminf_{n\to\infty}\mathbb{P}\left(\mathcal{A}_n\right)$.

Note that, on $\mathcal{C}$, we have

$$\mathbb{P}\left(n\left(\mathbb{V}_{(k)}^e(z_0)\right)^{-1/2}\widetilde{\mathbf{R}}_n(z_0)\in\mathcal{C}\right)\leq\mathbb{P}\left(\boldsymbol{\phi}_{(k)}(z_0)+\boldsymbol{\eta}_{(k)}(z_0)\in\mathcal{C}_\varepsilon\right)+\mathbb{P}\left(\left\|\mathbf{r}_{(k)}(z_0)\right\|>\varepsilon\right).$$

As $n\to\infty$, $\mathbb{P}\left(\left\|\mathbf{r}_{(k)}(z_0)\right\|>\varepsilon\right)\to0$. Then,

$$\limsup_{n\to\infty}\mathbb{P}\left(n\left(\mathbb{V}_{(k)}^e(z_0)\right)^{-1/2}\widetilde{\mathbf{R}}_n(z_0)\in\mathcal{C}\right)\leq\mathbb{P}\left(\boldsymbol{\xi}\in\mathcal{C}_\varepsilon\right)-\Delta_\varepsilon(\varepsilon_0)c_0.$$

Since $\lim_{\varepsilon\downarrow0}\mathbb{P}\left(\boldsymbol{\xi}\in\mathcal{C}_\varepsilon\right)=\mathbb{P}\left(\boldsymbol{\xi}\in\mathcal{C}\right)$, there exists $\rho>0$, such that for sufficiently small ε, we have $\mathbb{P}\left(\boldsymbol{\xi}\in\mathcal{C}_\varepsilon\right)-\mathbb{P}\left(\boldsymbol{\xi}\in\mathcal{C}\right)<\rho$. Given that $\Delta_\varepsilon(\varepsilon_0)\to\Delta_0(\varepsilon_0)>0$ as $\varepsilon\downarrow0$, for sufficiently small ε, we have

$|\Delta_\varepsilon(\varepsilon_0) - \Delta_0(\varepsilon_0)| < \frac{1}{2}\Delta_0(\varepsilon_0)$ so that $\Delta_\varepsilon(\varepsilon_0) > \frac{1}{2}\Delta_0(\varepsilon_0) > 0$. By choosing $\rho = \frac{c_0}{4}\Delta_0(\varepsilon_0)$, we obtain

$$\begin{aligned}\mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{C}_\varepsilon\right) - \Delta_\varepsilon(\varepsilon_0)c_0 &< \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{C}\right) + \rho - \Delta_\varepsilon(\varepsilon_0)c_0 \\ &< \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{C}\right) + \frac{c_0}{4}\Delta_0(\varepsilon_0) - \frac{c_0}{2}\Delta_0(\varepsilon_0) < \mathbb{P}\left(\boldsymbol{\xi} \in \mathcal{C}\right),\end{aligned}$$

as claimed. This completes the proof of the strict inequality part. □

C.2. Proof of Theorem 3

We begin by justifying (28) and deriving properties of the associated projection error. First, we show that the autocovariance matrix of $\mathbf{u}_t$: $\boldsymbol{\Gamma}_u(k) := \sum_{j=0}^{\infty} \boldsymbol{\Phi}_j \boldsymbol{\Sigma}_\varepsilon \boldsymbol{\Phi}_{j+k}^\top$, is absolute summable, where $k \geq 0$. Under Assumption 1, $\boldsymbol{\Sigma}_\varepsilon$ is positive definite and $\sum_{m=0}^{\infty} \|\boldsymbol{\Phi}_m\| < \infty$, we have

$$\begin{aligned}\sum_{k=0}^{\infty} \|\boldsymbol{\Gamma}_u(k)\| &\leq \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \|\boldsymbol{\Phi}_j\| \, \|\boldsymbol{\Sigma}_\varepsilon\| \left\|\boldsymbol{\Phi}_{j+k}^\top\right\| \\ &= \|\boldsymbol{\Sigma}_\varepsilon\| \sum_{j=0}^{\infty} \|\boldsymbol{\Phi}_j\| \left(\sum_{k=0}^{\infty} \left\|\boldsymbol{\Phi}_{j+k}^\top\right\|\right) \\ &\leq \|\boldsymbol{\Sigma}_\varepsilon\| \left(\sum_{m=0}^{\infty} \|\boldsymbol{\Phi}_m\|\right)^2 < \infty.\end{aligned}$$

As $\boldsymbol{\Gamma}_u(-k) = \boldsymbol{\Gamma}_u^\top(k)$ for all $k \geq 0$, we have $\sum_{k=-\infty}^{\infty} \|\boldsymbol{\Gamma}_u(k)\| < \infty$.

Given that $\sum_{k=0}^{\infty} \|\boldsymbol{\Gamma}_u(k)\| < \infty$ and $\inf_{\lambda\in[-\pi,\pi]} \lambda_{\min}\left(\mathbf{f}_{uu}(\lambda)\right) > 0$, the conditions in Theorem 8.3.1 in Brillinger (1975) are satisfied. We may write

$$\mathbf{u}_{0t} = \sum_{j=-\infty}^{\infty} \boldsymbol{\Pi}_j \mathbf{u}_{x,t-j} + \mathbf{v}_t,$$

where $\boldsymbol{\Pi}_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} \boldsymbol{\Pi}(\lambda) e^{ij\lambda} d\lambda$ satisfies $\sum_{j=-\infty}^{\infty} \|\boldsymbol{\Pi}_j\| < \infty$. $\boldsymbol{\Pi}(\lambda) = \mathbf{f}_{u_0u_x}(\lambda)\mathbf{f}_{u_xu_x}^{-1}(\lambda)$ is the complex regression coefficient of $\mathbf{u}_{0t}$ on $\mathbf{u}_{xt}$ at frequency λ. The error series $\mathbf{v}_t$ has spectral density matrix

$$\mathbf{f}_{vv}(\lambda) = \mathbf{f}_{u_0u_0}(\lambda) - \mathbf{f}_{u_0u_x}(\lambda)\mathbf{f}_{u_xu_x}^{-1}(\lambda)\mathbf{f}_{u_xu_0}(\lambda).$$

Using the fact that $\boldsymbol{\Omega}_{ab} = 2\pi\mathbf{f}_{ab}(0)$, we obtain

$$\boldsymbol{\Omega}_{vv} = \boldsymbol{\Omega}_{00} - \boldsymbol{\Omega}_{0x}\boldsymbol{\Omega}_{xx}^{-1}\boldsymbol{\Omega}_{x0}, \tag{C.3}$$

which is exactly $\mathbf{\Omega}_{0\cdot x}$.

Finally, since the cross spectral density matrix of $\mathbf{u}_{xt}$ and $\mathbf{v}_t$ is given by

$$\mathbf{f}_{vu_x}(\lambda) = \mathbf{f}_{u_0u_x}(\lambda) - \mathbf{\Pi}(\lambda)\mathbf{f}_{u_xu_x}(\lambda) = \mathbf{0}_{d_y\times d_x}, \ \ \forall\lambda,$$

we have

$$\mathbb{E}\left[\mathbf{v}_t\mathbf{u}_{x,t-k}^\top\right] = \int_{-\pi}^{\pi}\mathbf{f}_{vu_x}(\lambda)e^{ik\lambda}\,d\lambda = \mathbf{0}_{d_y\times d_x}, \ \ \forall k.$$

In addition, since $\sum_{j=-\infty}^{\infty}\|\mathbf{\Pi}_j\| < \infty$ and $\mathbf{u}_{xt}$ is weakly stationary, the process $\sum_{j=-\infty}^{\infty}\mathbf{\Pi}_j\mathbf{u}_{x,t-j}$ is also weakly stationary, so does $\mathbf{v}_t$.

Recall that our regression model for SDOLS is given by

$$\mathbf{y}_t = \mathbf{B}^{(k)}\mathbf{X}_t^{(k)} + \sum_{j=-K}^{K}\mathbf{\Pi}_j\Delta\mathbf{X}_{t-j} + \widetilde{\mathbf{v}}_t^{(k)}, \tag{C.4}$$

where $\widetilde{\mathbf{v}}_t^{(k)} = \mathbf{v}_t + \sum_{|j|>K}\mathbf{\Pi}_j\mathbf{u}_{x,t-j} + \left(\mathbf{A}_t - \mathbf{A}_k(\tau_t)\right)\mathbf{X}_t$. Let $\mathbf{Z}_t^{(K)} = \left(\Delta\mathbf{X}_{t+K}^\top, \cdots, \Delta\mathbf{X}_{t-K}^\top\right)^\top$ and $\mathbf{\Pi}^{(K)} = [\mathbf{\Pi}_{-K}, \cdots, \mathbf{\Pi}_K]$. In view of the Frisch–Waugh–Lovell theorem, $\mathbf{B}^{(k)}$ can be estimated from the regression

$$\check{\mathbf{y}}_t = \mathbf{B}^{(k)}\check{\mathbf{X}}_t^{(k)} + \check{\mathbf{u}}_{0t}^{(k)},$$

where

$$\check{\mathbf{y}}_t = \mathbf{y}_t - \widehat{\mathbf{B}}_{YZ}\mathbf{Z}_t^{(K)}, \quad \check{\mathbf{X}}_t^{(k)} = \mathbf{X}_t^{(k)} - \widehat{\mathbf{\Pi}}_{XZ}\mathbf{Z}_t^{(K)},$$

with

$$\widehat{\mathbf{B}}_{YZ} = \left(\sum_{t=K+1}^{n-K}\mathbf{y}_t\mathbf{Z}_t^{(K)\top}\right)\left(\sum_{t=K+1}^{n-K}\mathbf{Z}_t^{(K)}\mathbf{Z}_t^{(K)\top}\right)^{-1},$$
$$\widehat{\mathbf{\Pi}}_{XZ} = \left(\sum_{t=K+1}^{n-K}\mathbf{X}_t^{(k)}\mathbf{Z}_t^{(K)\top}\right)\left(\sum_{t=K+1}^{n-K}\mathbf{Z}_t^{(K)}\mathbf{Z}_t^{(K)\top}\right)^{-1},$$

and $\check{\mathbf{u}}_{0t}^{(k)} = \mathbf{v}_t + \sum_{|j|>K}\mathbf{\Pi}_j\mathbf{u}_{x,t-j} + \left(\mathbf{A}_t - \mathbf{A}_k(\tau_t)\right)\mathbf{X}_t + \left(\mathbf{B}^{(k)}\widehat{\mathbf{\Pi}}_{XZ} - \widehat{\mathbf{B}}_{YZ} + \mathbf{\Pi}^{(K)}\right)\mathbf{Z}_t^{(K)}$. By con-

struction of $\check{\mathbf{X}}_t^{(k)}$, we have

$$\sum_{t=K+1}^{n-K} \check{\mathbf{X}}_t^{(k)} \mathbf{Z}_t^{(K)\top} = \mathbf{0}_{kd_x \times (2K+1)d_x},$$
$$\left(\mathbf{B}^{(k)} \widehat{\mathbf{\Pi}}_{XZ} - \widehat{\mathbf{B}}_{YZ} + \mathbf{\Pi}^{(K)}\right) \sum_{t=K+1}^{n-K} \mathbf{Z}_t^{(K)} \check{\mathbf{X}}_t^{(k)\top} = \mathbf{0}_{d_y \times kd_x}.$$

Therefore,

$$\begin{aligned}
\widetilde{\mathbf{B}}_n^{(k)} - \mathbf{B}^{(k)} &= \left(\sum_{t=K+1}^{n-K} \check{\mathbf{u}}_{0t}^{(k)} \check{\mathbf{X}}_t^{(k)\top}\right) \left(\sum_{t=K+1}^{n-K} \check{\mathbf{X}}_t^{(k)} \check{\mathbf{X}}_t^{(k)\top}\right)^{-1} \\
&= \left(\sum_{t=K+1}^{n-K} \widetilde{\mathbf{v}}_t \check{\mathbf{X}}_t^{(k)\top}\right) \left(\sum_{t=K+1}^{n-K} \check{\mathbf{X}}_t^{(k)} \check{\mathbf{X}}_t^{(k)\top}\right)^{-1} \\
&\quad + \left(\sum_{t=K+1}^{n-K} \left(\mathbf{A}_t - \mathbf{A}_k(\tau_t)\right) \mathbf{X}_t \check{\mathbf{X}}_t^{(k)\top}\right) \left(\sum_{t=K+1}^{n-K} \check{\mathbf{X}}_t^{(k)} \check{\mathbf{X}}_t^{(k)\top}\right)^{-1},
\end{aligned} \tag{C.5}$$

where $\widetilde{\mathbf{v}}_t = \mathbf{v}_t + \sum_{|j|>K} \mathbf{\Pi}_j \mathbf{u}_{x,t-j}$. For later use, define

$$\widetilde{\mathbf{\Gamma}}_{n,vx}^{(k)} := \frac{1}{n-2K} \sum_{t=K+1}^{n-K} \widetilde{\mathbf{v}}_t \check{\mathbf{X}}_t^{(k)\top}, \quad \check{\mathbf{\Lambda}}_{n,xx}^{(k)} := \frac{1}{(n-2K)^2} \sum_{t=K+1}^{n-K} \check{\mathbf{X}}_t^{(k)} \check{\mathbf{X}}_t^{(k)\top}.$$

Also, define the sample and population covariance matrices of $\mathbf{Z}_t^{(K)}$ by

$$\widehat{\mathbf{\Sigma}}_{zz}^{(K)} := \frac{1}{n-2K} \sum_{t=K+1}^{n-K} \mathbf{Z}_t^{(K)} \mathbf{Z}_t^{(K)\top}, \qquad \mathbf{\Sigma}_{zz}^{(K)} := \mathbb{E}\left(\mathbf{Z}_t^{(K)} \mathbf{Z}_t^{(K)\top}\right).$$

First, note that

$$
\begin{aligned}
&\frac{1}{(n-2K)^2}\sum_{t=K+1}^{n-K}\check{\mathbf{X}}_t^{(k)}\check{\mathbf{X}}_t^{(k)\top}\\
=&\frac{1}{(n-2K)^2}\sum_{t=K+1}^{n-K}\mathbf{X}_t^{(k)}\mathbf{X}_t^{(k)\top}\\
&-\left(\frac{1}{(n-2K)^2}\sum_{t=K+1}^{n-K}\mathbf{X}_t^{(k)}\mathbf{Z}_t^{(K)\top}\right)\left(\widehat{\mathbf{\Sigma}}_{zz}^{(K)}\right)^{-1}\left(\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{Z}_t^{(K)}\mathbf{X}_t^{(k)\top}\right)\\
=&\frac{1}{(n-2K)^2}\sum_{t=K+1}^{n-K}\mathbf{X}_t^{(k)}\mathbf{X}_t^{(k)\top}\\
&-\left(\frac{1}{(n-2K)^2}\sum_{t=K+1}^{n-K}\mathbf{X}_t^{(k)}\mathbf{Z}_t^{(K)\top}\right)\left[\left(\widehat{\mathbf{\Sigma}}_{zz}^{(K)}\right)^{-1}-\left(\mathbf{\Sigma}_{zz}^{(K)}\right)^{-1}\right]\left(\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{Z}_t^{(K)}\mathbf{X}_t^{(k)\top}\right)\\
&-\left(\frac{1}{(n-2K)^2}\sum_{t=K+1}^{n-K}\mathbf{X}_t^{(k)}\mathbf{Z}_t^{(K)\top}\right)\left(\mathbf{\Sigma}_{zz}^{(K)}\right)^{-1}\left(\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{Z}_t^{(K)}\mathbf{X}_t^{(k)\top}\right)\\
:=&\frac{1}{(n-2K)^2}\sum_{t=K+1}^{n-K}\mathbf{X}_t^{(k)}\mathbf{X}_t^{(k)\top}+\mathbf{R}_{xx,1}+\mathbf{R}_{xx,2}.
\end{aligned}
$$

By Lemma E.6, we have

$$
\begin{aligned}
&\|\mathbf{R}_{xx,1}\|\\
\le&\left\|\frac{1}{(n-2K)^2}\sum_{t=K+1}^{n-K}\mathbf{X}_t^{(k)}\mathbf{Z}_t^{(K)\top}\right\|\left\|\left(\widehat{\mathbf{\Sigma}}_{zz}^{(K)}\right)^{-1}-\left(\mathbf{\Sigma}_{zz}^{(K)}\right)^{-1}\right\|_{sp}\left\|\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{Z}_t^{(K)}\mathbf{X}_t^{(k)\top}\right\|\\
=&O_p\left(\frac{\sqrt{Kk}}{n}\right)O_p\left(\frac{K}{\sqrt{n}}\right)O_p\left(\sqrt{Kk}\right)=O_p\left(\frac{K^2k}{n^{3/2}}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
\|\mathbf{R}_{xx,2}\|\le&\left\|\frac{1}{(n-2K)^2}\sum_{t=K+1}^{n-K}\mathbf{X}_t^{(k)}\mathbf{Z}_t^{(K)\top}\right\|\left\|\left(\mathbf{\Sigma}_{zz}^{(K)}\right)^{-1}\right\|_{sp}\left\|\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{Z}_t^{(K)}\mathbf{X}_t^{(k)\top}\right\|\\
=&O_p\left(\frac{\sqrt{Kk}}{n}\right)O(1)O_p\left(\sqrt{Kk}\right)=O_p\left(\frac{Kk}{n}\right).
\end{aligned}
$$

Since $\varrho_k K\sqrt{\frac{k}{n}}\to 0$ implies $K^2/n^{1-\kappa}\to 0$, we have $K/\sqrt{n}\to 0$. This implies that

$$
\left\|\frac{1}{(n-2K)^2}\sum_{t=K+1}^{n-K}\check{\mathbf{X}}_t^{(k)}\check{\mathbf{X}}_t^{(k)\top}-\frac{1}{n^2}\sum_{t=1}^{n}\mathbf{X}_t^{(k)}\mathbf{X}_t^{(k)\top}\right\|_{sp}=O_p\left(\frac{Kk}{n}\right).
$$

Next, by Lemma 2,

$$\left\| \frac{1}{n^2} \sum_{t=1}^{n} \mathbf{X}_t^{(k)} \mathbf{X}_t^{(k)\top} - \mathbf{\Lambda}_{xx}^{(k)} \right\|_{sp} = O_p\left(\frac{k}{n^a}\right) + O_p\left(\frac{k^2}{n}\right),$$

where $\mathbf{\Lambda}_{xx}^{(k)}$ is defined as in Lemma 2 with $\underline{\mathbf{B}}_X(\cdot)$ replaced by $\mathbf{B}_X(\cdot)$. This further implies that

$$\left\| \frac{1}{(n-2K)^2} \sum_{t=K+1}^{n-K} \check{\mathbf{X}}_t^{(k)} \check{\mathbf{X}}_t^{(k)\top} - \mathbf{\Lambda}_{xx}^{(k)} \right\|_{sp} = O_p\left(\frac{Kk}{n} + \frac{k}{n^a} + \frac{k^2}{n}\right).$$

Under the conditions $\frac{\varrho_k^2 k^{3/2}}{n^a} \to 0$ and $\varrho_k K\sqrt{\frac{k}{n}} \to 0$, we have $\varrho_k \frac{k}{n^a} = o\left(n^{-a/2+\kappa/4}\right) = o(1)$, where we have used $k \asymp n^\kappa$ and $\kappa < a$. Moreover, $\varrho_k \frac{Kk}{n} = \left(\varrho_k K\sqrt{\frac{k}{n}}\right)\sqrt{\frac{k}{n}} = o(1)$, since $k/n \to 0$. Consequently, $\varrho_k\left(\frac{Kk}{n} + \frac{k}{n^a}\right) \to 0$. Therefore,

$$\begin{aligned}
&\left\| \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1/2} \left(\frac{1}{(n-2K)^2} \sum_{t=K+1}^{n-K} \check{\mathbf{X}}_t^{(k)} \check{\mathbf{X}}_t^{(k)\top} - \mathbf{\Lambda}_{xx}^{(k)} \right) \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1/2} \right\| \\
\le\ &\left\| \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1} \right\|_{sp} \left\| \frac{1}{(n-2K)^2} \sum_{t=K+1}^{n-K} \check{\mathbf{X}}_t^{(k)} \check{\mathbf{X}}_t^{(k)\top} - \mathbf{\Lambda}_{xx}^{(k)} \right\|_{sp} = O_p\left(\varrho_k\left(\frac{Kk}{n} + \frac{k}{n^a} + \frac{k^2}{n}\right)\right) = o_p(1),
\end{aligned}$$

since $k^2/n = o(k/n^a)$. As

$$\begin{aligned}
&\frac{1}{(n-2K)^2} \sum_{t=K+1}^{n-K} \check{\mathbf{X}}_t^{(k)} \check{\mathbf{X}}_t^{(k)\top} \\
=\ &\left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{1/2} \left(\mathbf{I}_{kd_x} + \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1/2} \left(\frac{1}{(n-2K)^2} \sum_{t=K+1}^{n-K} \check{\mathbf{X}}_t^{(k)} \check{\mathbf{X}}_t^{(k)\top} - \mathbf{\Lambda}_{xx}^{(k)} \right) \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1/2} \right) \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{1/2},
\end{aligned}$$

we conclude that

$$\left\| \left(\frac{1}{(n-2K)^2} \sum_{t=K+1}^{n-K} \check{\mathbf{X}}_t^{(k)} \check{\mathbf{X}}_t^{(k)\top} \right)^{-1} \right\|_{sp} = O_p(\varrho_k).$$

Finally,

$$\begin{aligned}
&\left\|\left(\frac{1}{(n-2K)^2}\sum_{t=K+1}^{n-K}\check{\mathbf{X}}_t^{(k)}\check{\mathbf{X}}_t^{(k)\top}\right)^{-1}-\left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1}\right\|_{sp}\\
\leq&\left\|\left(\frac{1}{(n-2K)^2}\sum_{t=K+1}^{n-K}\check{\mathbf{X}}_t^{(k)}\check{\mathbf{X}}_t^{(k)\top}\right)^{-1}\right\|_{sp}\left\|\frac{1}{(n-2K)^2}\sum_{t=K+1}^{n-K}\check{\mathbf{X}}_t^{(k)}\check{\mathbf{X}}_t^{(k)\top}-\mathbf{\Lambda}_{xx}^{(k)}\right\|_{sp}\left\|\left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1}\right\|_{sp}\\
=&O_p\left(\varrho_k^2\frac{k}{n^a}\right),
\end{aligned}$$

as under the maintained assumptions $\frac{k^2}{n}=o\left(\frac{k}{n^a}\right)$ and $\frac{Kk}{n}=o\left(\frac{k}{n^a}\right)$.

Second, note that

$$\begin{aligned}
&\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\widetilde{\mathbf{v}}_t\check{\mathbf{X}}_t^{(k)\top}\\
=&\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\widetilde{\mathbf{v}}_t\mathbf{X}_t^{(k)\top}\\
&-\left(\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\widetilde{\mathbf{v}}_t\mathbf{Z}_t^{(K)\top}\right)\left(\widehat{\mathbf{\Sigma}}_{zz}^{(K)}\right)^{-1}\left(\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{Z}_t^{(K)}\mathbf{X}_t^{(k)\top}\right)\\
=&\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\widetilde{\mathbf{v}}_t\mathbf{X}_t^{(k)\top}\\
&-\left(\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\widetilde{\mathbf{v}}_t\mathbf{Z}_t^{(K)\top}\right)\left[\left(\widehat{\mathbf{\Sigma}}_{zz}^{(K)}\right)^{-1}-\left(\mathbf{\Sigma}_{zz}^{(K)}\right)^{-1}\right]\left(\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{Z}_t^{(K)}\mathbf{X}_t^{(k)\top}\right)\\
&-\left(\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\widetilde{\mathbf{v}}_t\mathbf{Z}_t^{(K)\top}\right)\left(\mathbf{\Sigma}_{zz}^{(K)}\right)^{-1}\left(\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{Z}_t^{(K)}\mathbf{X}_t^{(k)\top}\right)\\
:=&\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\widetilde{\mathbf{v}}_t\mathbf{X}_t^{(k)\top}+\mathbf{R}_{vx,1}+\mathbf{R}_{vx,2}.
\end{aligned}$$

It follows again from Lemma E.6 that

$$\begin{aligned}
\|\mathbf{R}_{vx,1}\|\leq&\left\|\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\widetilde{\mathbf{v}}_t\mathbf{Z}_t^{(K)\top}\right\|\left\|\left(\widehat{\mathbf{\Sigma}}_{zz}^{(K)}\right)^{-1}-\left(\mathbf{\Sigma}_{zz}^{(K)}\right)^{-1}\right\|_{sp}\left\|\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{Z}_t^{(K)}\mathbf{X}_t^{(k)\top}\right\|\\
=&O_p\left(\sqrt{\frac{K}{n}}\right)O_p\left(\frac{K}{\sqrt{n}}\right)O_p\left(\sqrt{Kk}\right)=O_p\left(\frac{K^2\sqrt{k}}{n}\right),
\end{aligned}$$

and

$$\|\mathbf{R}_{vx,2}\| \leq \left\| \frac{1}{n-2K} \sum_{t=K+1}^{n-K} \widetilde{\mathbf{v}}_t \mathbf{Z}_t^{(K)\top} \right\| \left\| \left(\mathbf{\Sigma}_{zz}^{(K)} \right)^{-1} \right\|_{sp} \left\| \frac{1}{n-2K} \sum_{t=K+1}^{n-K} \mathbf{Z}_t^{(K)} \mathbf{X}_t^{(k)\top} \right\|$$
$$= O_p\left(\sqrt{\frac{K}{n}} \right) O(1) O_p\left(\sqrt{Kk} \right) = O_p\left(\frac{K\sqrt{k}}{\sqrt{n}} \right).$$

Therefore,

$$\left\| \frac{1}{n-2K} \sum_{t=K+1}^{n-K} \widetilde{\mathbf{v}}_t \check{\mathbf{X}}_t^{(k)\top} - \frac{1}{n-2K} \sum_{t=K+1}^{n-K} \widetilde{\mathbf{v}}_t \mathbf{X}_t^{(k)\top} \right\|_{sp} = O_p\left(\frac{K\sqrt{k}}{\sqrt{n}} \right).$$

In addition,

$$\frac{1}{n-2K} \sum_{t=K+1}^{n-K} \widetilde{\mathbf{v}}_t \mathbf{X}_t^{(k)\top} = \frac{1}{n-2K} \sum_{t=K+1}^{n-K} \mathbf{v}_t \mathbf{X}_t^{(k)\top} + \frac{1}{n-2K} \sum_{t=K+1}^{n-K} \left(\sum_{|j|>K} \mathbf{\Pi}_j \mathbf{u}_{x,t-j} \right) \mathbf{X}_t^{(k)\top}$$
$$= \frac{1}{n-2K} \sum_{t=K+1}^{n-K} \mathbf{v}_t \mathbf{X}_t^{(k)\top} + O_p\left(\sqrt{kn} \sum_{|j|>K} \|\mathbf{\Pi}_j\| \right),$$

which follows by Cauchy–Schwarz inequality

$$\left\| \frac{1}{n-2K} \sum_{t=K+1}^{n-K} \left(\sum_{|j|>K} \mathbf{\Pi}_j \mathbf{u}_{x,t-j} \right) \mathbf{X}_t^{(k)\top} \right\|_{sp}$$
$$\leq \left(\frac{1}{n-2K} \sum_{t=K+1}^{n-K} \left\| \sum_{|j|>K} \mathbf{\Pi}_j \mathbf{u}_{x,t-j} \right\|^2 \right)^{1/2} \left(\frac{1}{n-2K} \sum_{t=K+1}^{n-K} \left\| \mathbf{X}_t^{(k)} \right\|^2 \right)^{1/2}$$
$$= O_p\left(\sqrt{kn} \sum_{|j|>K} \|\mathbf{\Pi}_j\| \right).$$

Let $\widetilde{\mathbf{\Gamma}}_{vx}^{(k)} = \int_0^1 d\mathbf{B}_v(r) \left(\boldsymbol{\psi}_{(k)}(r) \otimes \mathbf{B}_X(r) \right)^\top$. By combining all results above, together with Lemma E.6(iv), we obtain

$$\left\| \frac{1}{n-2K} \sum_{t=K+1}^{n-K} \widetilde{\mathbf{v}}_t \check{\mathbf{X}}_t^{(k)\top} - \widetilde{\mathbf{\Gamma}}_{vx}^{(k)} \right\|_{sp} = O_p\left(\frac{k^{3/2}}{n^{\delta_*}} + K\sqrt{\frac{k}{n}} + \sqrt{kn} \sum_{|j|>K} \|\mathbf{\Pi}_j\| \right).$$

Finally, let $\mathbf{u} \in \mathbb{R}^{d_y}$ and $\mathbf{v} \in \mathbb{R}^{d_x k}$ be arbitrary unit vectors. By definition,

$$\left\| \frac{1}{n-2K} \sum_{t=K+1}^{n-K} \left(\mathbf{A}_t - \mathbf{A}_k(\tau_t) \right) \mathbf{X}_t \check{\mathbf{X}}_t^{(k)\top} \right\|_{sp} = \max_{\|\mathbf{u}\|=1, \|\mathbf{v}\|=1} \left| \mathbf{u}^\top \left(\frac{1}{n-2K} \sum_{t=K+1}^{n-K} \left(\mathbf{A}_t - \mathbf{A}_k(\tau_t) \right) \mathbf{X}_t \check{\mathbf{X}}_t^{(k)\top} \right) \mathbf{v} \right|.$$

Note that, by the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\left|\mathbf{u}^\top \left\{ \frac{1}{n-2K} \sum_{t=K+1}^{n-K} \left(\mathbf{A}_t - \mathbf{A}_k(\tau_t)\right) \mathbf{X}_t \check{\mathbf{X}}_t^{(k)\top} \right\} \mathbf{v} \right| \\
&\leq \left\{ \frac{1}{n-2K} \sum_{t=K+1}^{n-K} \left[\mathbf{u}^\top \left(\mathbf{A}_t - \mathbf{A}_k(\tau_t)\right) \mathbf{X}_t \right]^2 \right\}^{1/2} \left\{ \frac{1}{n-2K} \sum_{t=K+1}^{n-K} \left(\check{\mathbf{X}}_t^{(k)\top} \mathbf{v} \right)^2 \right\}^{1/2} \\
&\leq \max_{K+1 \leq t \leq n-K} \|\mathbf{X}_t\| \left\{ \frac{1}{n-2K} \sum_{t=K+1}^{n-K} \left\| \mathbf{u}^\top \left(\mathbf{A}_t - \mathbf{A}_k(\tau_t)\right) \right\|^2 \right\}^{1/2} \left\{ (n-2K) \mathbf{v}^\top \check{\mathbf{\Lambda}}_{n,xx}^{(k)} \mathbf{v} \right\}^{1/2}.
\end{aligned}$$

Standard results from the unit root literature (see, e.g. Phillips and Durlauf, 1986) gives

$$\max_{K+1 \leq t \leq n-K} \|\mathbf{X}_t\| = O_p(\sqrt{n}).$$

In addition, since $K = o(n)$,

$$\frac{1}{n-2K} \sum_{t=K+1}^{n-K} \left\| \mathbf{u}^\top \left(\mathbf{A}_t - \mathbf{A}_k(\tau_t)\right) \right\|^2 \leq \frac{n}{n-2K} \cdot \frac{1}{n} \sum_{t=1}^{n} \|\mathbf{A}_t - \mathbf{A}_k(\tau_t)\|^2 = O\left(k^{-2q}\right),$$

where the details for

$$\frac{1}{n} \sum_{t=1}^{n} \|\mathbf{A}_t - \mathbf{A}_k(\tau_t)\|^2 = O\left(k^{-2q}\right)$$

is postponed in Lemma D.2. In addition, by the preceding result as in Lemma 2,

$$\mathbf{v}^\top \check{\mathbf{\Lambda}}_{n,xx}^{(k)} \mathbf{v} \leq \left\| \check{\mathbf{\Lambda}}_{n,xx}^{(k)} \right\|_{sp} \leq \left\| \check{\mathbf{\Lambda}}_{n,xx}^{(k)} - \mathbf{\Lambda}_{n,xx}^{(k)} \right\|_{sp} + \left\| \mathbf{\Lambda}_{n,xx}^{(k)} \right\|_{sp} = O_p\left(\frac{Kk}{n}\right) + O_p(1) = O_p(1),$$

uniformly over k. It then follows by taking the maximum over all such $\mathbf{u}$ and $\mathbf{v}$ that

$$\left\| \frac{1}{n-2K} \sum_{t=K+1}^{n-K} \left(\mathbf{A}_t - \mathbf{A}_k(\tau_t)\right) \mathbf{X}_t \check{\mathbf{X}}_t^{(k)\top} \right\|_{sp} = O_p\left(\frac{n}{k^q}\right).$$

Now, substituting all results back to (C.5) yields

$$
\begin{aligned}
& n \cdot \left(\widetilde{\mathbf{B}}_n^{(k)} - \mathbf{B}^{(k)}\right) \\
&= \left(\frac{n}{n-2K}\right)\left(\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\widetilde{\mathbf{v}}_t\check{\mathbf{X}}_t^{(k)\top}\right)\left(\frac{1}{(n-2K)^2}\sum_{t=K+1}^{n-K}\check{\mathbf{X}}_t^{(k)}\check{\mathbf{X}}_t^{(k)\top}\right)^{-1} \\
&\quad + \left(\frac{n}{n-2K}\right)\left(\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\left(\mathbf{A}_t - \mathbf{A}_k(\tau_t)\right)\mathbf{X}_t\check{\mathbf{X}}_t^{(k)\top}\right)\left(\frac{1}{(n-2K)^2}\sum_{t=K+1}^{n-K}\check{\mathbf{X}}_t^{(k)}\check{\mathbf{X}}_t^{(k)\top}\right)^{-1} \\
&= \left(\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\widetilde{\mathbf{v}}_t\check{\mathbf{X}}_t^{(k)\top}\right)\left(\frac{1}{(n-2K)^2}\sum_{t=K+1}^{n-K}\check{\mathbf{X}}_t^{(k)}\check{\mathbf{X}}_t^{(k)\top}\right)^{-1} + o_p(1) \\
&= \widetilde{\mathbf{\Gamma}}_{n,vx}^{(k)}\left(\check{\mathbf{\Lambda}}_{n,xx}^{(k)}\right)^{-1} + o_p(1).
\end{aligned}
$$

The second equality follows from the fact that

$$
\begin{aligned}
& \left\| \frac{n}{n-2K}\left\{\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\left(\mathbf{A}_t - \mathbf{A}_k(\tau_t)\right)\mathbf{X}_t\check{\mathbf{X}}_t^{(k)\top}\right\}\left(\frac{1}{(n-2K)^2}\sum_{t=K+1}^{n-K}\check{\mathbf{X}}_t^{(k)}\check{\mathbf{X}}_t^{(k)\top}\right)^{-1}\right\|_{sp} \\
&= O_p\left(\frac{\varrho_k n}{k^q}\right) = o_p(1),
\end{aligned}
$$

where we have used $n/(n-2K) = 1 + o(1)$ and $\left\|\left(\check{\mathbf{\Lambda}}_{n,xx}^{(k)}\right)^{-1}\right\|_{sp} = O_p(\varrho_k)$.

Then, proceeding in the similar way as used for Proposition 1, we obtain

$$
\begin{aligned}
& \left\|\widetilde{\mathbf{\Gamma}}_{n,vx}^{(k)}\left(\check{\mathbf{\Lambda}}_{n,xx}^{(k)}\right)^{-1} - \widetilde{\mathbf{\Gamma}}_{vx}^{(k)}\left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1}\right\|_{sp} \\
&\leq \left\|\widetilde{\mathbf{\Gamma}}_{n,vx}^{(k)} - \widetilde{\mathbf{\Gamma}}_{vx}^{(k)}\right\|_{sp}\left\|\left(\check{\mathbf{\Lambda}}_{n,xx}^{(k)}\right)^{-1}\right\|_{sp} + \left\|\widetilde{\mathbf{\Gamma}}_{vx}^{(k)}\right\|_{sp}\left\|\left(\check{\mathbf{\Lambda}}_{n,xx}^{(k)}\right)^{-1} - \left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1}\right\|_{sp} \\
&= O_p\left(\frac{\varrho_k k^{3/2}}{n^{\delta_*}} + \varrho_k K\sqrt{\frac{k}{n}} + \varrho_k\sqrt{kn}\sum_{|j|>K}\|\mathbf{\Pi}_j\| + \frac{\varrho_k^2 k^{3/2}}{n^a}\right) = o_p(1).
\end{aligned}
$$

This, together with the fact

$$
n \cdot \|\mathbf{A}_k(z_0) - \mathbf{A}(z_0)\| = o\left(\sqrt{k}\right),
$$

as we derived in Theorem 1, further implies that

$$
\left\| n \cdot \mathrm{vec}\left(\widetilde{\mathbf{A}}_n(z_0) - \mathbf{A}(z_0)\right) - \left(\mathbf{T}_{(k)}^{\top}(z_0)\left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1} \otimes \mathbf{I}_{d_y}\right)\mathrm{vec}\left\{\widetilde{\mathbf{\Gamma}}_{vx}^{(k)}\right\}\right\| = o_p(\sqrt{k}).
$$

The Brownian motions $\mathbf{B}_v(r)$ and $\mathbf{B}_X(r)$ are uncorrelated and hence independently distributed. Then, by Lemma 5.1 of Park and Phillips (1988), we can show that

$$\text{vec}\left(\widetilde{\mathbf{\Gamma}}_{vx}^{(k)}\right)|\mathcal{F}_X(1) \equiv \mathcal{N}\left(\mathbf{0}_{d_y k d_x \times 1}, \mathbf{\Lambda}_{xx}^{(k)} \otimes \mathbf{\Omega}_{0\cdot x}\right), \qquad a.s.,$$

where the spectral norm of the conditional covariance matrix $\mathbf{\Lambda}_{xx}^{(k)} \otimes \mathbf{\Omega}_{0\cdot x}$ is uniformly bounded over k. Let

$$\mathbb{V}_{(k)}^{e}(z_0) = \left\{\mathbf{T}_{(k)}^{\top}(z_0)\left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1}\mathbf{T}_{(k)}(z_0)\right\} \otimes \mathbf{\Omega}_{0\cdot x}.$$

Then, obviously

$$\left(\mathbb{V}_{(k)}^{e}(z_0)\right)^{-1/2}\left(\mathbf{T}_{(k)}^{\top}(z_0)\left(\mathbf{\Lambda}_{xx}^{(k)}\right)^{-1} \otimes \mathbf{I}_{d_y}\right)\text{vec}\left\{\widetilde{\mathbf{\Gamma}}_{vx}^{(k)}\right\}|\mathcal{F}_X(1) \equiv \mathcal{N}\left(\mathbf{0}_{d_y d_x \times 1}, \mathbf{I}_{d_y d_x}\right), \qquad a.s.$$

Since the conditional distribution on the left-hand side is $a.s.$ $\mathcal{N}\left(\mathbf{0}_{d_y d_x \times 1}, \mathbf{I}_{d_y d_x}\right)$, which does not depend on $\mathcal{F}_X(1)$, the unconditional distribution is also $\mathcal{N}\left(\mathbf{0}_{d_y d_x \times 1}, \mathbf{I}_{d_y d_x}\right)$. Following the proof of Theorem 1, we can verify that $\left\|\left(\mathbb{V}_{(k)}^{e}(z_0)\right)^{-1/2}\right\|_{sp} = O_p\left(k^{-1/2}\right)$. Therefore, we conclude that

$$n\left(\mathbb{V}_{(k)}^{e}(z_0)\right)^{-1/2}\text{vec}\left(\widetilde{\mathbf{A}}_n(z_0) - \mathbf{A}(z_0)\right) \Rightarrow \mathcal{N}\left(\mathbf{0}_{d_y d_x \times 1}, \mathbf{I}_{d_y d_x}\right). \qquad \square$$

D — Supporting lemmas and auxiliary results for Section 2

Lemma D.1. *Let $k = \lfloor c \cdot n^{\kappa} \rfloor$ for some $c > 0$, where $0 < \kappa < 1$. Under Assumption 1 and Assumption 4, we have, as $n \to \infty$*

$$\left\|\frac{1}{n}\sum_{t=1}^{n}\mathbf{u}_{0t}\mathbf{X}_t^{(k)\top} - \mathbf{\Gamma}_{ux}^{(k)}\right\|_{sp} = O_p\left(\frac{k^{3/2}}{n^{\min(\delta,1/2)}}\right),$$

where $\delta > 0$ is given in Assumption 4 and

$$\mathbf{\Gamma}_{ux}^{(k)} := \int_0^1 d\underline{\mathbf{B}}_0(r)\left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r)\right)^{\top} + \int_0^1 \boldsymbol{\psi}_{(k)}^{\top}(r) \otimes \left(\mathbf{\Lambda}_{0x}(r) + \mathbf{\Sigma}_{0x}(r)\right) dr$$

satisfies $\left\|\mathbf{\Gamma}_{ux}^{(k)}\right\| = O_p(\sqrt{k})$.

Proof. Let $\mathbf{U}_n(r) = \mathbf{\Phi}_0(1)\frac{1}{\sqrt{n}}\sum_{s=1}^{\lfloor nr \rfloor}\boldsymbol{\varepsilon}_s$. We note that

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\mathbf{u}_{0t}\mathbf{X}_t^{(k)\top} &= \frac{1}{n}\sum_{t=1}^{n}\mathbf{u}_{0t}\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{X}_{t-1}\right)^\top + \frac{1}{n}\sum_{t=1}^{n}\mathbf{u}_{0t}\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{u}_{xt}\right)^\top \\ &:= \mathbf{\Gamma}_n(1) + \mathbf{\Gamma}_n(2).\end{aligned}$$

Let us start with $\mathbf{\Gamma}_n(1)$. BN decomposition gives

$$\begin{aligned}\mathbf{\Gamma}_n(1) =& \mathbf{\Phi}_0(1)\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\varepsilon}_t\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{X}_{t-1}\right)^\top - \frac{1}{n}\sum_{t=1}^{n}\Delta\widetilde{\mathbf{u}}_{0t}\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{X}_{t-1}\right)^\top \\ =& \mathbf{\Phi}_0(1)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{\varepsilon}_t\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{V}_{n,t-1}^{x}\right)^\top + \mathbf{\Phi}_0(1)\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\varepsilon}_t\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\left(\widetilde{\mathbf{u}}_{x,0}-\widetilde{\mathbf{u}}_{x,t-1}+\mathbf{X}_0\right)\right)^\top \\ &- \frac{1}{n}\sum_{t=1}^{n}\Delta\widetilde{\mathbf{u}}_{0t}\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{X}_{t-1}\right)^\top \\ :=& \mathbf{\Gamma}_n(1,1)+\mathbf{\Gamma}_n(1,2)+\mathbf{\Gamma}_n(1,3).\end{aligned}$$

Let $\mathbf{V}_n^x(r) = \mathbf{V}_{n,\lfloor nr\rfloor}^x$ and $\mathbf{V}_n^x(r-) = \lim_{s\uparrow r}\mathbf{V}_n^x(s)$. In particular, $\mathbf{V}_n^x(\tau_t-) = \mathbf{V}_{n,t-1}^x$. Let $\mathcal{F}_t := \sigma\left(\boldsymbol{\varepsilon}_s : s\le t\right)$ denote the natural filtration generated by the innovations. Then,

$$\begin{aligned}\mathbf{\Gamma}_n(1,1) &= \sum_{t=1}^{n}\left(\mathbf{\Phi}_0(1)\frac{1}{\sqrt{n}}\boldsymbol{\varepsilon}_t\right)\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{V}_{n,t-1}^{x}\right)^\top \\ &= \int_0^1 d\mathbf{U}_n(r)\,\left(\boldsymbol{\psi}_{(k)}(r)\otimes\mathbf{V}_n^x(r-)\right)^\top,\end{aligned}$$

where the second equality follows from the definition of stochastic integration with respect to $\mathbf{U}_n(\cdot)$, whose jumps occur at $r=\tau_t$. Moving on to $\mathbf{\Gamma}_n(1,2)$. We note that

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\varepsilon}_t\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\left(\widetilde{\mathbf{u}}_{x,0}-\widetilde{\mathbf{u}}_{x,t-1}\right)\right)^\top|\mathcal{F}_{t-1}\right] &= \mathbb{E}\left[\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}\right]\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\left(\widetilde{\mathbf{u}}_{x,0}-\widetilde{\mathbf{u}}_{x,t-1}\right)\right)^\top \\ &= \mathbf{0}_{(d_y+d_x)\times kd_x},\end{aligned}$$

so that $\left\{\boldsymbol{\varepsilon}_t\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\left(\widetilde{\mathbf{u}}_{x,0}-\widetilde{\mathbf{u}}_{x,t-1}+\mathbf{X}_0\right)\right)\right\}_t$ forms a martingale difference array. By the moment

bound calculation in Lemma D.3, $\max_{0\le t\le n}\mathbb{E}\left\|\widetilde{\mathbf{u}}_{x,t}\right\|^2<\infty$. Then,

$$\begin{aligned}
&\mathbb{E}\left\|\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\varepsilon}_t\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\left(\widetilde{\mathbf{u}}_{x,0}-\widetilde{\mathbf{u}}_{x,t-1}\right)\right)^\top\right\|^2\\
&\le\frac{1}{n^2}\sum_{t=1}^{n}\mathbb{E}\left\|\boldsymbol{\varepsilon}_t\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\left(\widetilde{\mathbf{u}}_{x,0}-\widetilde{\mathbf{u}}_{x,t-1}\right)\right)^\top\right\|^2\\
&\le\frac{1}{n^2}\sum_{t=1}^{n}\left\|\boldsymbol{\psi}_{(k)}(\tau_t)\right\|^2\mathbb{E}\left\|\boldsymbol{\varepsilon}_t\right\|^2\mathbb{E}\left\|\widetilde{\mathbf{u}}_{x,0}-\widetilde{\mathbf{u}}_{x,t-1}\right\|^2=O\left(\frac{k}{n}\right),
\end{aligned}$$

which follows from the fact that $\left\|\boldsymbol{\psi}_{(k)}(\tau_t)\right\|=O(1)\sqrt{k}$. In addition,

$$\left\|\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\varepsilon}_t\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{X}_0\right)^\top\right\|\le\left\|\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\varepsilon}_t\boldsymbol{\psi}_{(k)}(\tau_t)\right\|\|\mathbf{X}_0\|=O_P\left(\sqrt{\frac{k}{n}}\right).$$

This together with triangular ineqaulity implies that $\|\boldsymbol{\Gamma}_n(1,2)\|=O_p\left(\sqrt{\frac{k}{n}}\right)=o_p(1)$.

Finally, we consider $\boldsymbol{\Gamma}_n(1,3)$. Summation by parts leads to

$$\begin{aligned}
\boldsymbol{\Gamma}_n(1,3)&=-\frac{1}{n}\sum_{t=1}^{n}\Delta\widetilde{\mathbf{u}}_{0t}\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{X}_{t-1}\right)^\top\\
&=\frac{1}{n}\sum_{t=1}^{n}\widetilde{\mathbf{u}}_{0t}\left\{\left(\boldsymbol{\psi}_{(k)}(\tau_{t+1})\otimes\mathbf{X}_t\right)-\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{X}_{t-1}\right)\right\}^\top+o_p(1)\\
&=\frac{1}{n}\sum_{t=1}^{n}\widetilde{\mathbf{u}}_{0t}\left\{\left(\boldsymbol{\psi}_{(k)}(\tau_{t+1})-\boldsymbol{\psi}_{(k)}(\tau_t)\right)\otimes\mathbf{X}_t+\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\left(\mathbf{X}_t-\mathbf{X}_{t-1}\right)\right\}^\top+o_p(1)\\
&=\frac{1}{n}\sum_{t=1}^{n}\widetilde{\mathbf{u}}_{0t}\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{u}_{xt}\right)^\top+O_p\left(\frac{k^{3/2}}{\sqrt{n}}\right),
\end{aligned}$$

where the final equality follows from the Lipschitz-continuity of the base function, which implies that $\left\|\boldsymbol{\psi}_{(k)}(\tau_{t+1})-\boldsymbol{\psi}_{(k)}(\tau_t)\right\|=O(k^{3/2}/n)$, together with Cauchy–Schwarz inequality and the fact that $\frac{1}{n}\sum_{t=1}^{n}\left\|\mathbf{X}_t\right\|^2=O_p(n)$,[5] so that

$$\left\|\frac{1}{n}\sum_{t=1}^{n}\widetilde{\mathbf{u}}_{0t}\left\{\left(\boldsymbol{\psi}_{(k)}(\tau_{t+1})-\boldsymbol{\psi}_{(k)}(\tau_t)\right)\otimes\mathbf{X}_t\right\}^\top\right\|=O_p\left(\frac{k^{3/2}}{\sqrt{n}}\right).$$

5. By Lemma D.3 and (5), we have $\frac{1}{\sqrt{n}}\mathbf{X}_{\lfloor n\cdot\rfloor}\Rightarrow\underline{\mathbf{B}}_X(\cdot)$, where $\underline{\mathbf{B}}_X(\cdot)=\boldsymbol{\Phi}_x(1)\mathbf{B}_H(\cdot)$. Then, the desired rate is obtained by continuous mapping theorem (CMT).

In addition, by the W.L.L.N. established in Lemma D.4, we have

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\widetilde{\mathbf{u}}_{0t}\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{u}_{xt}\right)^{\top} &= \frac{1}{n}\sum_{t=1}^{n}\left(\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\otimes\widetilde{\mathbf{u}}_{0t}\mathbf{u}_{x,t}^{\top}\right)\\ &= \int_0^1 \boldsymbol{\psi}_{(k)}^{\top}(r)\otimes\left(\sum_{j=0}^{\infty}\widetilde{\boldsymbol{\Phi}}_{0j}\boldsymbol{\Sigma}_{\varepsilon}(r)\boldsymbol{\Phi}_{xj}^{\top}\right)dr + o_P(1).\end{aligned}$$

By combining all results obtained above, we get

$$\left\|\boldsymbol{\Gamma}_n(1)-\int_0^1 d\mathbf{U}_n(r)\,\left(\boldsymbol{\psi}_{(k)}(r)\otimes\mathbf{V}_n^x(r-)\right)^{\top}-\int_0^1\boldsymbol{\psi}_{(k)}^{\top}(r)\otimes\left(\sum_{j=0}^{\infty}\widetilde{\boldsymbol{\Phi}}_{0j}\boldsymbol{\Sigma}_{\varepsilon}(r)\boldsymbol{\Phi}_{xj}^{\top}\right)dr\right\| = O_p\left(\frac{k^{3/2}}{\sqrt{n}}\right).$$

We now show that

$$\left\|\int_0^1 d\mathbf{U}_n(r)\,\left(\boldsymbol{\psi}_{(k)}(r)\otimes\mathbf{V}_n^x(r-)\right)^{\top}-\int_0^1 d\underline{\mathbf{B}}_0(r)\,\left(\boldsymbol{\psi}_{(k)}(r)\otimes\underline{\mathbf{B}}_X(r)\right)^{\top}\right\|_{sp} = O_p\left(\frac{k^{3/2}}{n^{\delta}}\right), \tag{D.1}$$

where δ is defined as in Assumption 4. We shall prove this based on the inequality

$$\begin{aligned}&\left\|\int_0^1 d\mathbf{U}_n(r)\,\left(\boldsymbol{\psi}_{(k)}(r)\otimes\mathbf{V}_n^x(r-)\right)^{\top}-\int_0^1 d\underline{\mathbf{B}}_0(r)\,\left(\boldsymbol{\psi}_{(k)}(r)\otimes\underline{\mathbf{B}}_X(r)\right)^{\top}\right\|_{sp}\\ &\le\sqrt{\sum_{j=0}^{k-1}\left\|\int_0^1\psi_j(r)d\mathbf{U}_n(r)\mathbf{V}_n^{x\top}(r-)-\int_0^1\psi_j(r)d\underline{\mathbf{B}}_0(r)\underline{\mathbf{B}}_X^{\top}(r)\right\|_{sp}^2},\end{aligned}$$

where $j=0,1,\cdots,k-1$. Let

$$\mathbf{R}_n(r)=\int_0^r d\mathbf{U}_n(s)\mathbf{V}_n^{x\top}(s-)-\int_0^r d\underline{\mathbf{B}}_0(s)\underline{\mathbf{B}}_X^{\top}(s).$$

Then,

$$\int_0^1\psi_j(r)d\mathbf{U}_n(r)\mathbf{V}_n^{x\top}(r-)-\int_0^1\psi_j(r)d\underline{\mathbf{B}}_0(r)\underline{\mathbf{B}}_X^{\top}(r)=\int_0^1\psi_j(r)d\mathbf{R}_n(r).$$

Summation by parts gives

$$\int_0^1\psi_j(r)d\mathbf{R}_n(r)=\psi_j(1)\mathbf{R}_n(1)-\int_0^1\mathbf{R}_n(r)d\psi_j(r),$$

since $\mathbf{R}_n(0) = \mathbf{0}_{d_y \times d_x}$ and there is no quadratic variation between $\psi_j(\cdot)$ and $\mathbf{R}_n(\cdot)$. Assumption 4 gives

$$\sup_{r\in[0,1]} \|\mathbf{R}_n(r)\| = O_p\left(n^{-\delta}\right),$$

for some $\delta > 0$. This implies that

$$\begin{aligned}\left\|\int_0^1 \psi_j(r) d\mathbf{R}_n(r)\right\|_{sp} &\leq \|\psi_j(1)\mathbf{R}_n(1)\|_{sp} + \left\|\int_0^1 \mathbf{R}_n(r) d\psi_j(r)\right\|_{sp} \\ &= O_p\left(n^{-\delta}\right)\left(1 + \int_0^1 |d\psi_j(r)|\right) = \left(\frac{1+j}{n^\delta}\right) O_p(1),\end{aligned}$$

which yields

$$\begin{aligned}&\left\|\int_0^1 d\mathbf{U}_n(r)\ \left(\boldsymbol{\psi}_{(k)}(r) \otimes \mathbf{V}_n^x(r-)\right)^\top - \int_0^1 d\underline{\mathbf{B}}_0(r)\left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r)\right)^\top\right\|_{sp} \\ &\leq \frac{O_p(1)}{n^\delta}\left(\sum_{j=0}^{k-1}(1+j)^2\right)^{1/2} = O_p\left(\frac{k^{3/2}}{n^\delta}\right),\end{aligned}$$

as claimed.

Turning to $\boldsymbol{\Gamma}_n(2)$. This is straightforward, as by invoking again Lemma D.4, we obtain

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^n \mathbf{u}_{0t}\left(\boldsymbol{\psi}_{(k)}(\tau_t) \otimes \mathbf{u}_{xt}\right)^\top &= \frac{1}{n}\sum_{t=1}^n \left(\boldsymbol{\psi}_{(k)}^\top(\tau_t) \otimes \mathbf{u}_{0t}\mathbf{u}_{x,t}^\top\right) \\ &= \int_0^1 \boldsymbol{\psi}_{(k)}^\top(r) \otimes \left(\sum_{j=0}^\infty \boldsymbol{\Phi}_{0j}\boldsymbol{\Sigma}_\varepsilon(r)\boldsymbol{\Phi}_{xj}^\top\right) dr + o_P(1).\end{aligned}$$

Thus, we have

$$\frac{1}{n}\sum_{t=1}^n \mathbf{u}_{0t}\mathbf{X}_t^{(k)\top} = \int_0^1 d\underline{\mathbf{B}}_0(r)\left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r)\right)^\top + \int_0^1 \boldsymbol{\psi}_{(k)}^\top(r) \otimes \left(\boldsymbol{\Lambda}_{0x}(r) + \boldsymbol{\Sigma}_{0x}(r)\right) dr + o_p(1),$$

where

$$\boldsymbol{\Sigma}_{0x}(\cdot) := \sum_{j=0}^\infty \boldsymbol{\Phi}_{0j}\boldsymbol{\Sigma}_\varepsilon(\cdot)\boldsymbol{\Phi}_{xj}^\top, \quad \boldsymbol{\Lambda}_{0x}(\cdot) := \sum_{j=0}^\infty \widetilde{\boldsymbol{\Phi}}_{0j}\boldsymbol{\Sigma}_\varepsilon(\cdot)\boldsymbol{\Phi}_{xj}^\top.$$

To show $\left\|\boldsymbol{\Gamma}_{ux}^{(k)}\right\| = O_p(\sqrt{k})$, we first show that $\left\|\int_0^1 d\underline{\mathbf{B}}_0(r)\left(\boldsymbol{\psi}_{(k)}(r) \otimes \underline{\mathbf{B}}_X(r)\right)^\top\right\| = O_p(\sqrt{k})$. To see

this, we first let

$$\underline{\mathbf{B}}_{0\cdot x}(r) := \underline{\mathbf{B}}_0(r) - \int_0^r \mathbf{\Omega}_{0x}(s)\mathbf{\Omega}_{xx}^{-1}(s)d\underline{\mathbf{B}}_X(s),$$
$$\mathbf{\Omega}_{0\cdot x}(r) := \mathbf{\Omega}_{00}(r) - \mathbf{\Omega}_{0x}(r)\mathbf{\Omega}_{xx}^{-1}(r)\mathbf{\Omega}_{x0}(r).$$

Note that

$$\begin{aligned}
&\int_0^1 d\underline{\mathbf{B}}_0(r)\left(\boldsymbol{\psi}_{(k)}(r)\otimes\underline{\mathbf{B}}_X(r)\right)^\top \\
=&\int_0^1 d\underline{\mathbf{B}}_{0\cdot x}(r)\left(\boldsymbol{\psi}_{(k)}(r)\otimes\underline{\mathbf{B}}_X(r)\right)^\top + \int_0^1 \mathbf{\Omega}_{0x}(r)\mathbf{\Omega}_{xx}^{-1}(r)d\underline{\mathbf{B}}_X(r)\left(\boldsymbol{\psi}_{(k)}(r)\otimes\underline{\mathbf{B}}_X(r)\right)^\top.
\end{aligned}$$

Recall that, in the heteroskedastic setting, $\mathcal{F}_X(r) = \sigma\{\underline{\mathbf{B}}_X(s) : 0 \le s \le r\}$. Since

$$\begin{aligned}
&\mathbb{E}\left[\left\|\int_0^1 d\underline{\mathbf{B}}_{0\cdot x}(r)\left(\boldsymbol{\psi}_{(k)}(r)\otimes\underline{\mathbf{B}}_X(r)\right)^\top\right\|^2 |\mathcal{F}_X(1)\right] \\
=&\int_0^1 \operatorname{tr}\left(\mathbf{\Omega}_{0\cdot x}(r)\right)\left\|\boldsymbol{\psi}_{(k)}(r)\otimes\underline{\mathbf{B}}_X(r)\right\|^2 dr \\
\le&\sup_{r\in[0,1]}\operatorname{tr}\left(\mathbf{\Omega}_{0\cdot x}(r)\right)\sup_{r\in[0,1]}\|\underline{\mathbf{B}}_X(r)\|^2\int_0^1\left\|\boldsymbol{\psi}_{(k)}(r)\right\|^2 dr = O_p(k),
\end{aligned}$$

we have

$$\left\|\int_0^1 d\underline{\mathbf{B}}_{0\cdot x}(r)\left(\boldsymbol{\psi}_{(k)}(r)\otimes\underline{\mathbf{B}}_X(r)\right)^\top\right\| = O_p(\sqrt{k}).$$

By similar arguments, we can show that

$$\begin{aligned}
&\mathbb{E}\left\|\int_0^1 \mathbf{\Omega}_{0x}(r)\mathbf{\Omega}_{xx}(r)^{-1}d\underline{\mathbf{B}}_X(r)\left(\boldsymbol{\psi}_{(k)}(r)\otimes\underline{\mathbf{B}}_X(r)\right)^\top\right\|^2 \\
=&\int_0^1 \operatorname{tr}\left(\mathbf{\Omega}_{0x}(r)\mathbf{\Omega}_{xx}^{-1}(r)\mathbf{\Omega}_{0x}^\top(r)\right)\left\|\boldsymbol{\psi}_{(k)}(r)\right\|^2\mathbb{E}\left\|\underline{\mathbf{B}}_X(r)\right\|^2 dr \\
\le& O(1)k.
\end{aligned}$$

Then, $\left\|\int_0^1 d\underline{\mathbf{B}}_0(r)\left(\boldsymbol{\psi}_{(k)}(r)\otimes\underline{\mathbf{B}}_X(r)\right)^\top\right\| = O_p(\sqrt{k})$ follows by triangular inequality.

□

Lemma D.2. *Let $k = \lfloor c\cdot n^\kappa\rfloor$ for some $c > 0$, where $0 < \kappa < a$, $a = \frac{p-2}{2p}$, and $p > 4$ is as in Assumption 1.*

Under Assumptions 1-2, we have, as $n \to \infty$

$$\left\| \frac{1}{n} \sum_{t=1}^{n} \left(\mathbf{A}_t - \mathbf{A}_k(\tau_t) \right) \mathbf{X}_t \mathbf{X}_t^{(k)\top} \right\|_{sp} = O_p\left(\frac{n}{k^q} \right)$$

Proof. Let $\mathbf{u} \in \mathbb{R}^{d_y}$ and $\mathbf{v} \in \mathbb{R}^{d_x k}$ be arbitrary unit vectors. By definition,

$$\left\| \frac{1}{n^2} \sum_{t=1}^{n} \left(\mathbf{A}_t - \mathbf{A}_k(\tau_t) \right) \mathbf{X}_t \mathbf{X}_t^{(k)\top} \right\|_{sp} = \max_{\|\mathbf{u}\|=1, \|\mathbf{v}\|=1} \left| \mathbf{u}^\top \left(\frac{1}{n^2} \sum_{t=1}^{n} \left(\mathbf{A}_t - \mathbf{A}_k(\tau_t) \right) \mathbf{X}_t \mathbf{X}_t^{(k)\top} \right) \mathbf{v} \right|.$$

Note that, by Cauchy-Schwarz inequality,

$$\begin{aligned}
&\left| \mathbf{u}^\top \left(\frac{1}{n^2} \sum_{t=1}^{n} \left(\mathbf{A}_t - \mathbf{A}_k(\tau_t) \right) \mathbf{X}_t \mathbf{X}_t^{(k)\top} \right) \mathbf{v} \right| \\
&\leq \left\{ \frac{1}{n^2} \sum_{t=1}^{n} \left[\mathbf{u}^\top \left(\mathbf{A}_t - \mathbf{A}_k(\tau_t) \right) \mathbf{X}_t \right]^2 \right\}^{1/2} \left\{ \frac{1}{n^2} \sum_{t=1}^{n} \left(\mathbf{X}_t^{(k)\top} \mathbf{v} \right)^2 \right\}^{1/2} \\
&\leq \left(\max_{1 \leq t \leq n} \frac{\|\mathbf{X}_t\|}{\sqrt{n}} \right) \left(\frac{1}{n} \sum_{t=1}^{n} \left\| \mathbf{u}^\top \left(\mathbf{A}_t - \mathbf{A}_k(\tau_t) \right) \right\|^2 \right)^{1/2} \left(\mathbf{v}^\top \mathbf{\Lambda}_{n,xx}^{(k)} \mathbf{v} \right)^{1/2},
\end{aligned}$$

where $\mathbf{\Lambda}_{n,xx}^{(k)} = \frac{1}{n^2} \sum_{t=1}^{n} \mathbf{X}_t^{(k)} \mathbf{X}_t^{(k)\top}$.

Again, by Lemma D.3 and (5), together with CMT, gives $\max_{1 \leq t \leq n} \frac{\|\mathbf{X}_t\|}{\sqrt{n}} = O_p(1)$. In addition,

$$\begin{aligned}
\mathbf{v}^\top \mathbf{\Lambda}_{n,xx}^{(k)} \mathbf{v} &\leq \left\| \mathbf{\Lambda}_{n,xx}^{(k)} \right\|_{sp} \\
&\leq \left\| \mathbf{\Lambda}_{n,xx}^{(k)} - \mathbf{\Lambda}_{xx}^{(k)} \right\|_{sp} + \left\| \mathbf{\Lambda}_{xx}^{(k)} \right\|_{sp} = O_p(1),
\end{aligned}$$

uniformly over k. This follows from Lemma 2, together with the fact that $k = \lfloor c \cdot n^\kappa \rfloor$ for some $c > 0$, where $0 < \kappa < a$.

Finally, for $r \in [0,1]$, define $\mathbf{e}_k(r) := \mathbf{A}(r) - \mathbf{A}_k(r)$, and $f_k(r) := \|\mathbf{e}_k(r)\|^2$. Thus, $f_k(\tau_t) =$

$\|\mathbf{A}_t - \mathbf{A}_k(\tau_t)\|^2$. Observe that

$$\begin{aligned}
\left|\frac{1}{n}\sum_{t=1}^{n} f_k(\tau_t) - \int_0^1 f_k(r)\,dr\right| &\leq \sum_{t=1}^{n}\int_{\frac{t-1}{n}}^{\frac{t}{n}} |f_k(\tau_t) - f_k(r)|\,dr \\
&\leq \sum_{t=1}^{n}\int_{\frac{t-1}{n}}^{\frac{t}{n}}\int_{r}^{\frac{t}{n}} \left|f_k^{(1)}(s)\right| ds dr \\
&= \sum_{t=1}^{n}\int_{\frac{t-1}{n}}^{\frac{t}{n}}\int_{\frac{t-1}{n}}^{s} \left|f_k^{(1)}(s)\right| dr ds \\
&= \sum_{t=1}^{n}\int_{\frac{t-1}{n}}^{\frac{t}{n}} \left(s - \frac{t-1}{n}\right)\left|f_k^{(1)}(s)\right| ds \\
&\leq \frac{1}{n}\sum_{t=1}^{n}\int_{\frac{t-1}{n}}^{\frac{t}{n}} \left|f_k^{(1)}(s)\right| ds \\
&= \frac{1}{n}\int_0^1 \left|f_k^{(1)}(s)\right| ds \\
&\leq \frac{2}{n}\left(\int_0^1 \|\mathbf{e}_k(r)\|^2\,dr\right)^{1/2}\left(\int_0^1 \left\|\mathbf{e}_k^{(1)}(r)\right\|^2 dr\right)^{1/2} \\
&= O\left(\frac{k}{n}\cdot\frac{1}{k^{2q}}\right) = o\left(k^{-2q}\right),
\end{aligned}$$

which follows from Lemma 1 and Lemma D.5. It then follows from the triangle inequality that

$$\frac{1}{n}\sum_{t=1}^{n}\left\|\mathbf{u}^\top\left(\mathbf{A}_t - \mathbf{A}_k(\tau_t)\right)\right\|^2 \leq \frac{1}{n}\sum_{t=1}^{n}\|\mathbf{A}_t - \mathbf{A}_k(\tau_t)\|^2 = O\left(1/k^{2q}\right).$$

The final result follows by taking maximum over all such $\mathbf{u}$ and $\mathbf{v}$. □

The following Lemma is a generalization of Lemma 1 in Park and Hahn (1999) under nonstationary volatility.

Lemma D.3. *Let $\mathbf{u}_t$ be defined as in (3)-(4) and suppose that Assumption 1 is satisfied. Then,*

$$\sup_{0\leq r\leq 1}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor}\mathbf{u}_t - \mathbf{\Phi}(1)\mathbf{B}_H(r)\right\| = O_P(n^{-a}), \tag{D.2}$$

$$\sup_{0\leq r\leq 1}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor}\mathbf{u}_t - \mathbf{\Phi}(1)\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor}\mathbf{H}_t\mathbf{e}_t\right\| = O_P(n^{-a}), \tag{D.3}$$

where $\mathbf{B}_H(\cdot) = \int_0^{\cdot}\mathbf{H}(s)d\mathbf{B}(s)$, $\mathbf{B}(s)$ is a standard Brownian motion vector, $a = \frac{p-2}{2p}$ and p is defined as in Assumption 1(ii).

Proof. We start with (D.3). By the BN decomposition

$$\mathbf{u}_t = \mathbf{\Phi}(1)\boldsymbol{\varepsilon}_t - \Delta\widetilde{\boldsymbol{u}}_t,$$

we note that, for each $r \in [0,1]$,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor}\mathbf{u}_t - \mathbf{\Phi}(1)\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor}\mathbf{H}_t\mathbf{e}_t = \frac{1}{\sqrt{n}}\left(\widetilde{\boldsymbol{u}}_0 - \widetilde{\boldsymbol{u}}_{\lfloor rn\rfloor}\right). \tag{D.4}$$

For any $\delta > 0$, we have

$$\begin{aligned}\mathbb{P}\left(\max_{1\le t\le n}\frac{1}{n^{1/p}}\left\|\widetilde{\boldsymbol{u}}_t\right\| > \delta\right) &\le \sum_{t=1}^{n}\mathbb{P}\left(\left\|\widetilde{\boldsymbol{u}}_t\right\| > \delta n^{1/p}\right)\\ &\le \frac{1}{\delta^p n}\sum_{t=1}^{n}\mathbb{E}\left\|\widetilde{\boldsymbol{u}}_t\right\|^p \le \frac{1}{\delta^p}\max_{1\le t\le n}\mathbb{E}\left\|\widetilde{\boldsymbol{u}}_t\right\|^p < \infty\end{aligned}$$

where the second inequality follows by the Markov's inequality and the final inequality follows from the fact that

$$\begin{aligned}\left(\mathbb{E}\left\|\widetilde{\boldsymbol{u}}_t\right\|^p\right)^{1/p} &\le \sum_{j=0}^{\infty}\left\|\widetilde{\mathbf{\Phi}}_j\right\|\left(\mathbb{E}\left\|\mathbf{H}_{t-j}\mathbf{e}_{t-j}\right\|^p\right)^{1/p}\\ &\le \sum_{j=0}^{\infty}\left\|\widetilde{\mathbf{\Phi}}_j\right\|\left\|\mathbf{H}_{t-j}\right\|_{sp}\left(\mathbb{E}\left\|\mathbf{e}_{t-j}\right\|^p\right)^{1/p} < \infty,\end{aligned}$$

which holds uniformly over t and the first inequality is obtained by Minkowski's inequality. Consequently, $\max_{1\le t\le n}\left\|\widetilde{\boldsymbol{u}}_t\right\| = O_P(n^{1/p})$ and $\left\|\widetilde{\boldsymbol{u}}_0\right\| = O_P(1)$. Applying the triangle inequality to (D.4), we obtain

$$\sup_{0\le r\le 1}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor}\mathbf{u}_t - \mathbf{\Phi}(1)\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor}\mathbf{H}_t\mathbf{e}_t\right\| \le \frac{1}{\sqrt{n}}\left(\left\|\widetilde{\boldsymbol{u}}_0\right\| + \max_{1\le t\le n}\left\|\widetilde{\boldsymbol{u}}_t\right\|\right) = O_P(n^{1/p-1/2}) = O_P(n^{-a}),$$

which establishes (D.3).

Turing to (D.2), it suffices to show that

$$\sup_{0\le r\le 1}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor}\boldsymbol{\varepsilon}_t - \mathbf{B}_H(r)\right\| = O_P(n^{-a}).$$

Let $\mathbf{E}_n(r) := \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \mathbf{e}_t$. Note that

$$\begin{aligned}\mathbf{H}(r)\mathbf{E}_n(r) - \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \mathbf{H}(\tau_t)\mathbf{e}_t &= \sum_{t=1}^{\lfloor nr \rfloor} \left(\mathbf{H}(r) - \mathbf{H}(\tau_t)\right) \frac{\mathbf{e}_t}{\sqrt{n}} \\ &= \sum_{t=1}^{\lfloor nr \rfloor} \left(\int_{\tau_t}^{r} d\mathbf{H}(s) \right) \frac{\mathbf{e}_t}{\sqrt{n}} \\ &= \int_0^r d\mathbf{H}(s) \left(\sum_{t=1}^{\lfloor nr \rfloor} \mathbb{1}_{\{\tau_t < s\}} \frac{\mathbf{e}_t}{\sqrt{n}} \right) = \int_0^r d\mathbf{H}(s)\mathbf{E}_n(s-).\end{aligned}$$

In addition, we also have

$$\mathbf{B}_H(r) = \int_0^r \mathbf{H}(s) d\mathbf{B}(s) = \mathbf{H}(r)\mathbf{B}(r) - \int_0^r d\mathbf{H}(s)\mathbf{B}(s).$$

By Theorem 1 of Einmahl (1989), together with the Hölder continuity of Brownian paths, on a richer probability space, there exists a suitable coupling such that, retaining the notation $\mathbf{e}_t$ for the distributional equivalent copies,

$$\sup_{0 \le r \le 1} \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \mathbf{e}_t - \mathbf{B}(r) \right\| = O_P(n^{-a}).$$

Observe that

$$\begin{aligned}\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \boldsymbol{\varepsilon}_t - \mathbf{B}_H(r) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \mathbf{H}(\tau_t)\mathbf{e}_t - \mathbf{B}_H(r) \\ &= \mathbf{H}(r) \left(\mathbf{E}_n(r) - \mathbf{B}(r) \right) - \int_0^r d\mathbf{H}(s) \left(\mathbf{E}_n(s-) - \mathbf{B}(s) \right).\end{aligned}$$

We then have

$$\begin{aligned}\sup_{r \in [0,1]} \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \boldsymbol{\varepsilon}_t - \mathbf{B}_H(r) \right\| &\le \left(\sup_{r \in [0,1]} \|\mathbf{H}(r)\| \right) \left(\sup_{r \in [0,1]} \|\mathbf{E}_n(r) - \mathbf{B}(r)\| \right) \\ &\quad + V_H[0,1] \left(\sup_{r \in [0,1]} \|\mathbf{E}_n(r) - \mathbf{B}(r)\| \right) = O_P(n^{-a}).\end{aligned}$$

□

Lemma D.4. *Suppose that the sequence $\boldsymbol{\varepsilon}_t$ is generated according to $\boldsymbol{\varepsilon}_t = \mathbf{H}_t \mathbf{e}_t$, where $\mathbf{e}_t \overset{i.i.d.}{\sim} \left(\mathbf{0}_{(d_y+d_x)\times 1}, \mathbf{I}_{d_y+d_x} \right)$ satisfies Assumption 1(ii) and $\mathbf{H}_t$ satisfies Assumption 1(iii). Let $\mathbf{u}_{1t}$ and $\mathbf{u}_{2t}$ be two generic linear processes*

defined according to

$$\mathbf{u}_{1t} = \sum_{i=0}^{\infty} \mathbf{A}_i \boldsymbol{\varepsilon}_{t-i}, \;\; \mathbf{u}_{2t} = \sum_{i=0}^{\infty} \mathbf{B}_i \boldsymbol{\varepsilon}_{t-i},$$

where the coefficient matrices satisfy $\sum_{j=0}^{\infty} j\|\mathbf{A}_j\| < \infty$ *and* $\sum_{j=0}^{\infty} j\|\mathbf{B}_j\| < \infty$*. Then,*

$$\left\| \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{\psi}_{(k)}^{\top}(\tau_t) \otimes \mathbf{u}_{1t}\mathbf{u}_{2t}^{\top} - \left(\int_0^1 \boldsymbol{\psi}_{(k)}^{\top}(r) \otimes \sum_{i=0}^{\infty} \mathbf{A}_i \boldsymbol{\Sigma}_{\varepsilon}(r) \mathbf{B}_i^{\top} \, dr \right) \right\| = O_P\left(\sqrt{\frac{k}{n}} + \frac{k^{3/2}}{n} \right),$$

where $\boldsymbol{\Sigma}_{\varepsilon}(\cdot) = \mathbf{H}(\cdot)\mathbf{H}^{\top}(\cdot)$*, and* $\boldsymbol{\psi}_{(k)}(\cdot) = (\psi_0(\cdot), \psi_1(\cdot), \cdots, \psi_{k-1}(\cdot))^{\top}$ *is the collection of Chebyshev time polynomials up to order* k*.*

Proof. First, we note that the sequence $\boldsymbol{\varepsilon}_t$ also has pth order uniformed bounded moment. To see this, note that

$$\max_{1\le t\le n} \mathbb{E}\,\|\mathbf{H}_t \mathbf{e}_t\|^p \le \left(\max_{1\le t\le n} \|\mathbf{H}_t\|^p \right) \left(\max_{1\le t\le n} \mathbb{E}\,\|\boldsymbol{e}_t\|^p \right) < \infty,$$

which follows by Assumption 1(ii)-(iii). The proof then follows the steps as in Phillips and Solo (1992) and also Cavaliere and Taylor (2007). Let $\mathbf{Z}$ be a generic matrix of dimension $(d_y + d_x) \times (d_y + d_x)$. Define

$$\mathbf{f}_{0i}(\mathbf{Z}) = \mathbf{A}_i \mathbf{Z} \mathbf{B}_i^{\top}, \;\; \mathbf{f}_{ri}(\mathbf{Z}) = \mathbf{A}_i \mathbf{Z} \mathbf{B}_{i+r}^{\top} + \mathbf{A}_{i+r} \mathbf{Z}^{\top} \mathbf{B}_i^{\top},$$

and $\mathbf{f}_r(z) = \sum_{i=0}^{\infty} \mathbf{f}_{ri} z^i$. Then, we can write

$$\mathbf{u}_{1t}\mathbf{u}_{2t}^{\top} = \mathbf{f}_0(\mathcal{L}) \left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^{\top} \right) + \sum_{r=1}^{\infty} \mathbf{f}_r(\mathcal{L}) \left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-r}^{\top} \right). \tag{D.5}$$

Using the usual BN factorization,

$$\mathbf{f}_r(z) = \mathbf{f}_r(1) - (1-z)\widetilde{\mathbf{f}}_r(z),$$

where

$$\widetilde{\mathbf{f}}_r(z) = \sum_{i=0}^{\infty} \widetilde{\mathbf{f}}_{ri} z^i, \;\; \text{with} \;\; \widetilde{\mathbf{f}}_{ri} = \sum_{s=i+1}^{\infty} \mathbf{f}_{rs},$$

we can rewrite (D.5) according to

$$\mathbf{u}_{1t}\mathbf{u}_{2t}^{\top} = \mathbf{X}_t^a + \mathbf{X}_t^b,$$

where

$$\mathbf{X}_t^a = \mathbf{f}_0(1) \left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^{\top} \right) + \widetilde{\mathbf{X}}_{t-1}^a - \widetilde{\mathbf{X}}_t^a, \;\; \mathbf{X}_t^b = \mathbf{M}_t + \widetilde{\mathbf{X}}_{t-1}^b - \widetilde{\mathbf{X}}_t^b,$$

with

$$\mathbf{M}_t = \sum_{r=1}^{\infty} \mathbf{f}_r(1) \left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-r}^\top \right), \quad \widetilde{\mathbf{X}}_t^a = \widetilde{\mathbf{f}}_0(\mathcal{L}) \left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top \right), \ \widetilde{\mathbf{X}}_t^b = \sum_{r=1}^{\infty} \widetilde{\mathbf{f}}_r(\mathcal{L}) \left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-r}^\top \right).$$

Thus,

$$\begin{aligned}
\frac{1}{n} \sum_{t=1}^{n} \boldsymbol{\psi}_{(k)}^\top(\tau_t) \otimes \mathbf{u}_{1t} \mathbf{u}_{2t}^\top &= \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{\psi}_{(k)}^\top(\tau_t) \otimes \mathbf{X}_t^a + \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{\psi}_{(k)}^\top(\tau_t) \otimes \mathbf{X}_t^b \\
&= \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{\psi}_{(k)}^\top(\tau_t) \otimes \left(\mathbf{f}_0(1) \left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top \right) \right) - \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{\psi}_{(k)}^\top(\tau_t) \otimes \left(\widetilde{\mathbf{X}}_t^a - \widetilde{\mathbf{X}}_{t-1}^a \right) \quad \text{(D.6)} \\
&\quad + \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{\psi}_{(k)}^\top(\tau_t) \otimes \mathbf{M}_t - \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{\psi}_{(k)}^\top(\tau_t) \otimes \left(\widetilde{\mathbf{X}}_t^b - \widetilde{\mathbf{X}}_{t-1}^b \right).
\end{aligned}$$

In the first step, we show that

$$\left\| \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{\psi}_{(k)}^\top(\tau_t) \otimes \left(\widetilde{\mathbf{X}}_t^a - \widetilde{\mathbf{X}}_{t-1}^a \right) \right\| = o_P(1), \ \left\| \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{\psi}_{(k)}^\top(\tau_t) \otimes \left(\widetilde{\mathbf{X}}_t^b - \widetilde{\mathbf{X}}_{t-1}^b \right) \right\| = o_P(1). \quad \text{(D.7)}$$

Note that

$$\begin{aligned}
\mathbb{E} \left\| \widetilde{\mathbf{X}}_t^b \right\|^2 &= \sum_r \sum_i \mathbb{E} \left\| \widetilde{\mathbf{f}}_{ri} \left(\boldsymbol{\varepsilon}_{t-i} \boldsymbol{\varepsilon}_{t-i-r}^\top \right) \right\|^2 \\
&\leq \sum_r \sum_i \left(\sum_{s>i} \left(\|\mathbf{A}_s\| \, \|\mathbf{B}_{s+r}\| + \|\mathbf{A}_{s+r}\| \, \|\mathbf{B}_s\| \right) \right)^2 \mathbb{E} \left\| \boldsymbol{\varepsilon}_{t-i} \right\|^2 \left\| \boldsymbol{\varepsilon}_{t-i-r} \right\|^2 \\
&\leq O(1) \left(\sum_r \sum_i \left\{ \left(\sum_{s>i} \|\mathbf{A}_s\| \, \|\mathbf{B}_{s+r}\| \right)^2 + \left(\sum_{s>i} \|\mathbf{A}_{s+r}\| \, \|\mathbf{B}_s\| \right)^2 \right\} \right).
\end{aligned}$$

Observe that

$$\begin{aligned}
\sum_r \sum_i \left(\sum_{s>i} \|\mathbf{A}_s\| \, \|\mathbf{B}_{s+r}\| \right)^2 &\leq \sum_i \left(\sum_{s>i} \|\mathbf{A}_s\|^2 \right) \left(\sum_r \sum_{s>i} \|\mathbf{B}_{s+r}\|^2 \right) \\
&\leq \left(\sum_s s \, \|\mathbf{B}_s\|^2 \right) \left(\sum_i \sum_{s>i} \|\mathbf{A}_s\|^2 \right) \\
&\leq \left(\left(\sum_s s \, \|\mathbf{B}_s\| \right) \left(\sum_s \|\mathbf{B}_s\| \right) \right) \left(\left(\sum_s s \, \|\mathbf{A}_s\| \right) \left(\sum_s \|\mathbf{A}_s\| \right) \right) < \infty.
\end{aligned}$$

Similarly, we can show that $\sum_r \sum_i \left(\sum_{s>i} \|\mathbf{A}_{s+r}\| \, \|\mathbf{B}_s\| \right)^2 < \infty$. This implies that

$$\mathbb{E} \left\| \widetilde{\mathbf{X}}_t^b \right\|^2 < \infty,$$

which holds uniformly over t. Now, summation by parts gives

$$\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\otimes\left(\widetilde{\mathbf{X}}_t^b-\widetilde{\mathbf{X}}_{t-1}^b\right)=\frac{1}{n}\left(\boldsymbol{\psi}_{(k)}^{\top}(1)\otimes\widetilde{\mathbf{X}}_n^b-\boldsymbol{\psi}_{(k)}^{\top}(1/n)\otimes\widetilde{\mathbf{X}}_0^b\right)$$
$$-\frac{1}{n}\sum_{t=1}^{n-1}\left\{\left(\boldsymbol{\psi}_{(k)}\left(\frac{t+1}{n}\right)-\boldsymbol{\psi}_{(k)}\left(\frac{t}{n}\right)\right)^{\top}\otimes\widetilde{\mathbf{X}}_t^b\right\}.$$

As

$$\begin{aligned}
&\mathbb{E}\left\|\frac{1}{n}\sum_{t=1}^{n-1}\left\{\left(\boldsymbol{\psi}_{(k)}\left(\frac{t+1}{n}\right)-\boldsymbol{\psi}_{(k)}\left(\frac{t}{n}\right)\right)^{\top}\otimes\widetilde{\mathbf{X}}_t^b\right\}\right\|^2\\
\leq&\frac{1}{n^2}\sum_t\sum_s\left\|\boldsymbol{\psi}_{(k)}\left(\frac{t+1}{n}\right)-\boldsymbol{\psi}_{(k)}\left(\frac{t}{n}\right)\right\|\left\|\boldsymbol{\psi}_{(k)}\left(\frac{s+1}{n}\right)-\boldsymbol{\psi}_{(k)}\left(\frac{s}{n}\right)\right\|\mathbb{E}\left(\left\|\widetilde{\mathbf{X}}_t^b\right\|\left\|\widetilde{\mathbf{X}}_s^b\right\|\right)\\
\leq&O(1)\frac{1}{n^2}\left(\sum_{t=1}^{n-1}\left\|\boldsymbol{\psi}_{(k)}\left(\frac{t+1}{n}\right)-\boldsymbol{\psi}_{(k)}\left(\frac{t}{n}\right)\right\|\right)^2=O\left(\frac{k^3}{n^2}\right),
\end{aligned}$$

we conclude that

$$\left\|\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\otimes\left(\widetilde{\mathbf{X}}_t^b-\widetilde{\mathbf{X}}_{t-1}^b\right)\right\|=O_P\left(\frac{k^{3/2}}{n}\right).$$

Using similar arguments, we can show that

$$\begin{aligned}
\mathbb{E}\left\|\widetilde{\mathbf{X}}_t^a\right\|^2&\leq\mathbb{E}\left(\sum_{i=0}^{\infty}\left(\sum_{s>i}\|\mathbf{A}_s\|\,\|\mathbf{B}_s\|\right)\|\boldsymbol{\varepsilon}_{t-i}\|^2\right)^2\\
&\leq O(1)\left(\sum_{i=0}^{\infty}\left(\sum_{s>i}\|\mathbf{A}_s\|\,\|\mathbf{B}_s\|\right)\right)^2\\
&=O(1)\left(\sum_{s=1}^{\infty}s\,\|\mathbf{A}_s\|\,\|\mathbf{B}_s\|\right)^2\\
&\leq O(1)\left(\left(\sum_s s\,\|\mathbf{A}_s\|\right)\left(\sum_s\|\mathbf{B}_s\|\right)\right)^2<\infty.
\end{aligned}$$

Then, using summation by parts, we obtain

$$\left\|\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\otimes\left(\widetilde{\mathbf{X}}_t^a-\widetilde{\mathbf{X}}_{t-1}^a\right)\right\|=o_P(1).$$

This completes the proof of (D.7).

In the next step, we show that

$$\left\|\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\otimes\mathbf{f}_0(1)\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{\top}\right)-\left(\int_0^1\boldsymbol{\psi}_{(k)}^{\top}(r)\otimes\sum_{i=0}^{\infty}\mathbf{A}_i\boldsymbol{\Sigma}_{\varepsilon}(r)\mathbf{B}_i^{\top}dr\right)\right\|=O_P\left(\sqrt{\frac{k}{n}}+\frac{k^{3/2}}{n}\right). \tag{D.8}$$

First, note that

$$\begin{aligned}\mathbf{f}_0(1)\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{\top}\right)&=\sum_{i=0}^{\infty}\mathbf{A}_i\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{\top}\mathbf{B}_i^{\top}\\&=\sum_{i=0}^{\infty}\mathbf{A}_i\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{\top}-\boldsymbol{\Sigma}_{\varepsilon}(\tau_t)\right)\mathbf{B}_i^{\top}+\sum_{i=0}^{\infty}\mathbf{A}_i\boldsymbol{\Sigma}_{\varepsilon}(\tau_t)\mathbf{B}_i^{\top}.\end{aligned}$$

Given that

$$\mathbb{E}\left\|\sum_{i=0}^{\infty}\mathbf{A}_i\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{\top}-\boldsymbol{\Sigma}_{\varepsilon}(\tau_t)\right)\mathbf{B}_i^{\top}\right\|^2\leq\left(\sum_{i=0}^{\infty}\|\mathbf{A}_i\|\,\|\mathbf{B}_i\|\right)^2\left(\mathbb{E}\,\|\boldsymbol{\varepsilon}_t\|^4\right)<\infty,$$

we obtain

$$\begin{aligned}&\mathbb{E}\left\|\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\otimes\left(\sum_{i=0}^{\infty}\mathbf{A}_i\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{\top}-\boldsymbol{\Sigma}_{\varepsilon}(\tau_t)\right)\mathbf{B}_i^{\top}\right)\right\|^2\\&\leq\frac{1}{n^2}\sum_{t=1}^{n}\left\|\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\right\|^2\mathbb{E}\left\|\sum_{i=0}^{\infty}\mathbf{A}_i\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{\top}-\boldsymbol{\Sigma}_{\varepsilon}(\tau_t)\right)\mathbf{B}_i^{\top}\right\|^2=O\left(\frac{k}{n}\right).\end{aligned}$$

In addition, as for any $r,s\in[0,1]$, $\|\boldsymbol{\Sigma}_{\varepsilon}(r)-\boldsymbol{\Sigma}_{\varepsilon}(s)\|\leq O(1)\,\|\mathbf{H}(r)-\mathbf{H}(s)\|$, we have $V_{\boldsymbol{\Sigma}_{\varepsilon}}\leq O(1)V_H[0,1]<\infty$. Then,

$$\begin{aligned}\int_0^1\|\boldsymbol{\Sigma}_{\varepsilon}(\lfloor nr\rfloor/n)-\boldsymbol{\Sigma}_{\varepsilon}(r)\|^2\,dr&=\sum_{j=0}^{n-1}\int_{\frac{j}{n}}^{\frac{j+1}{n}}\|\boldsymbol{\Sigma}_{\varepsilon}(\lfloor nr\rfloor/n)-\boldsymbol{\Sigma}_{\varepsilon}(r)\|^2\,dr\\&\leq\sum_{j=0}^{n-1}\int_{\frac{j}{n}}^{\frac{j+1}{n}}V_{\boldsymbol{\Sigma}_{\varepsilon}}^2\left[\frac{j}{n},\frac{j+1}{n}\right]dr\\&\leq\frac{1}{n}\sum_{j=0}^{n-1}V_{\boldsymbol{\Sigma}_{\varepsilon}}^2\left[\frac{j}{n},\frac{j+1}{n}\right]\\&\leq\frac{1}{n}\left(\sum_{j=0}^{n-1}V_{\boldsymbol{\Sigma}_{\varepsilon}}\left[\frac{j}{n},\frac{j+1}{n}\right]\right)^2=\frac{1}{n}V_{\boldsymbol{\Sigma}_{\varepsilon}}^2[0,1].\end{aligned}$$

This further implies that

$$
\begin{aligned}
&\left\|\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\otimes\left(\sum_{i=0}^{\infty}\mathbf{A}_i\boldsymbol{\Sigma}_{\varepsilon}(\tau_t)\mathbf{B}_i^{\top}\right)-\left(\int_0^1\boldsymbol{\psi}_{(k)}^{\top}(r)\otimes\sum_{i=0}^{\infty}\mathbf{A}_i\boldsymbol{\Sigma}_{\varepsilon}(r)\mathbf{B}_i^{\top}dr\right)\right\| \\
\leq&\left\|\int_0^1\left(\boldsymbol{\psi}_{(k)}(\lfloor nr\rfloor/n)-\boldsymbol{\psi}_{(k)}(r)\right)^{\top}\otimes\left(\sum_{i=0}^{\infty}\mathbf{A}_i\boldsymbol{\Sigma}_{\varepsilon}(\lfloor nr\rfloor/n)\mathbf{B}_i^{\top}\right)dr\right\| \\
&+\left\|\int_0^1\boldsymbol{\psi}_{(k)}^{\top}(r)\otimes\left(\sum_{i=0}^{\infty}\mathbf{A}_i\left(\boldsymbol{\Sigma}_{\varepsilon}(\lfloor nr\rfloor/n)-\boldsymbol{\Sigma}_{\varepsilon}(r)\right)\mathbf{B}_i^{\top}\right)\right\|+O\left(\frac{\sqrt{k}}{n}\right) \\
\leq&O(1)\left(\max_{r\in[0,1]}\left\|\boldsymbol{\psi}_{(k)}(\lfloor nr\rfloor/n)-\boldsymbol{\psi}_{(k)}(r)\right\|\right)+\left(\int_0^1\left\|\sum_{i=0}^{\infty}\mathbf{A}_i\left(\boldsymbol{\Sigma}_{\varepsilon}(\lfloor nr\rfloor/n)-\boldsymbol{\Sigma}_{\varepsilon}(r)\right)\mathbf{B}_i^{\top}\right\|^2dr\right)^{1/2}+O\left(\frac{\sqrt{k}}{n}\right) \\
=&O\left(\frac{k^{3/2}}{n}\right)+O\left(\frac{1}{\sqrt{n}}\right)+O\left(\frac{\sqrt{k}}{n}\right)=O\left(\frac{k^{3/2}}{n}+\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

which completes the proof of (D.8).

In the final step, we show that

$$
\left\|\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\otimes\mathbf{M}_t\right\|=O_P\left(\sqrt{\frac{k}{n}}\right). \tag{D.9}
$$

Recall that, $\mathbf{M}_t=\sum_{r=1}^{\infty}\mathbf{f}_r(1)\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t-r}^{\top}\right)$ is a M.D.S. with respect to the filtration $\mathcal{F}_t$:

$$
\mathbb{E}\left(\mathbf{M}_t|\mathcal{F}_{t-1}\right)=\mathbf{0}_{(d_y+d_x)\times(d_y+d_x)}.
$$

Then,

$$
\begin{aligned}
\mathbb{E}\left\|\mathbf{M}_t\right\|^2&\leq\mathbb{E}\left\|\boldsymbol{\varepsilon}_t\right\|^2\mathbb{E}\left(\sum_{r=1}^{\infty}\left\|\mathbf{f}_r(1)\right\|\left\|\boldsymbol{\varepsilon}_{t-r}\right\|\right)^2 \\
&\leq O(1)\left(\sum_r\sum_s\left\|\mathbf{f}_r(1)\right\|\left\|\mathbf{f}_s(1)\right\|\mathbb{E}\left(\left\|\boldsymbol{\varepsilon}_{t-r}\right\|\left\|\boldsymbol{\varepsilon}_{t-s}\right\|\right)\right) \\
&\leq O(1)\sum_r\sum_s\left\|\mathbf{f}_r(1)\right\|\left\|\mathbf{f}_s(1)\right\|\leq O(1)\left(\sum_r\left\|\mathbf{f}_r(1)\right\|\right)^2<\infty<\infty,
\end{aligned}
$$

which follows from the fact that

$$
\sum_r\left\|\mathbf{f}_r(1)\right\|\leq\sum_{i,j,i\neq j}\left\|\mathbf{A}_i\right\|\left\|\mathbf{B}_j\right\|\leq\left(\sum_{i=0}^{\infty}\left\|\mathbf{A}_i\right\|\right)\left(\sum_{i=0}^{\infty}\left\|\mathbf{B}_i\right\|\right)<\infty.
$$

Thus,

$$\mathbb{E}\left\|\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\psi}_{(k)}^{\top}(\tau_t)\otimes\mathbf{M}_t\right\|^2 \leq \frac{1}{n^2}\sum_{t=1}^{n}\left\|\boldsymbol{\psi}_{(k)}(\tau_t)\right\|^2\mathbb{E}\left\|\mathbf{M}_t\right\|^2 \leq O(1)\frac{k}{n}.$$

This completes the proof of (D.9). Then, the final results follow by applying triangular inequality on (D.6).

□

Lemma D.5. *Let $a(\cdot)\in L^2[0,1]$ and $\widetilde{\psi}_j(r)=\sqrt{2}\sin(j\pi r)$, for $j=1,2,\cdots$. Suppose further that $a(\cdot)$ satisfies Assumption 2. Then,*

$$\int_0^1\left|a^{(1)}(r)-a_k^{(1)}(r)\right|^2dr=O\left(k^{-2(q-1)}\right),$$

where $a_k^{(1)}(r)=\sum_{j=1}^{k-1}d_{1,j}\widetilde{\psi}_j(r)$ and $q>1$ is an integer.

Proof. First, $\{\widetilde{\psi}_j(r)\}$ is a complete orthonormal basis in the Hilbert space $L^2[0,1]$, where completeness follows from that of the trigonometric system by odd extension; see Section 4.23 of Rudin (1987). Second, notice that as $a(r)=\sum_{j=0}^{\infty}b_{1,j}\psi_j(r)$, $b_{1,j}=\int_0^1a(r)\psi_j(r)dr$, whereas $d_{1,j}=\int_0^1a^{(1)}(r)\widetilde{\psi}_j(r)dr$. Since

$$d_{1,j}=\int_0^1a^{(1)}(r)\widetilde{\psi}_j(r)dr=-\pi j\int_0^1a(r)\psi_j(r)dr=-\pi jb_{1,j},$$

we have

$$a^{(1)}(r)=\sum_{j=1}^{\infty}d_{1,j}\widetilde{\psi}_j(r)=-\sum_{j=1}^{\infty}(j\pi)b_{1,j}\widetilde{\psi}_j(r),\quad a_k^{(1)}(r)=\sum_{j=1}^{k-1}d_{1,j}\widetilde{\psi}_j(r)=-\sum_{j=1}^{k-1}(j\pi)b_{1,j}\widetilde{\psi}_j(r).$$

Now, Parseval's identity (Theorem 4.18 (iv) of Rudin (1987)) gives

$$\int_0^1\left|a^{(1)}(r)-a_k^{(1)}(r)\right|^2dr=\sum_{j=k}^{\infty}(j\pi)^2b_{1,j}^2.$$

What remains is to follow the proof of Lemma 1. For instance, when q is even, we get

$$\begin{aligned}\int_0^1\left|a^{(1)}(r)-a_k^{(1)}(r)\right|^2dr&=\sum_{j=k}^{\infty}(j\pi)^2b_{1,j}^2\\&=\sum_{j=k}^{\infty}\frac{(j\pi)^2}{(j\pi)^{2q}}\left(\int_0^1a^{(q)}(r)\psi_j(r)dr\right)^2\\&\leq\frac{1}{\pi^{2(q-1)}k^{2(q-1)}}\int_0^1\left|a^{(q)}(r)\right|^2dr=O\left(k^{-2(q-1)}\right),\end{aligned}$$

where the inequality follows by Bessel's inequality as in Theorem 4.17 of Rudin (1987). The case when q is odd is similar so details are omitted. □

E — Supporting lemma for Section 3

Lemma E.6. *Let* $\mathbf{Z}_t^{(K)} = \left(\Delta\mathbf{X}_{t+K}^\top, \cdots, \Delta\mathbf{X}_{t-K}^\top\right)^\top$, $K = o(\sqrt{n})$, $\sqrt{n}\sum_{|j|>K}\|\mathbf{\Pi}_j\| \to 0$, *and* $k = \lfloor c \cdot n^\kappa \rfloor$ *for some* $c > 0$, *where* $0 < \kappa < 1$. *Under Assumptions 1, 2, 5, and the condition* $\inf_{\lambda\in[-\pi,\pi]} \lambda_{\min}\left(\mathbf{f}_{uu}(\lambda)\right) > 0$, *we have, as* $n \to \infty$

(i) $\left\|\frac{1}{n-2K}\sum_{t=K+1}^{n-K} \widetilde{\mathbf{v}}_t \mathbf{Z}_t^{(K)\top}\right\| = O_p\left(\sqrt{K/n}\right)$;

(ii) $\left\|\left(\frac{1}{n-2K}\sum_{t=K+1}^{n-K} \mathbf{Z}_t^{(K)}\mathbf{Z}_t^{(K)\top}\right)^{-1} - \left(\mathbf{\Sigma}_{zz}^{(K)}\right)^{-1}\right\|_{sp} = O_p\left(K/\sqrt{n}\right), \left\|\left(\mathbf{\Sigma}_{zz}^{(K)}\right)^{-1}\right\|_{sp} = O(1)$;

(iii) $\left\|\frac{1}{n-2K}\sum_{t=K+1}^{n-K} \mathbf{Z}_t^{(K)}\mathbf{X}_t^{(k)\top}\right\| = O_p\left(\sqrt{Kk}\right)$;

(iv)
$$\left\|\frac{1}{n-2K}\sum_{t=K+1}^{n-K} \mathbf{v}_t\mathbf{X}_t^{(k)\top} - \int_0^1 d\mathbf{B}_v(r)\left(\boldsymbol{\psi}_{(k)}(r)\otimes\mathbf{B}_X(r)\right)^\top\right\|_{sp} = O_p\left(\frac{k^{3/2}}{n^{\delta_*}} + K\sqrt{\frac{k}{n}}\right),$$

where $\left\|\int_0^1 d\mathbf{B}_v(r)\left(\boldsymbol{\psi}_{(k)}(r)\otimes\mathbf{B}_X(r)\right)^\top\right\| = O_p(\sqrt{k})$ *and* δ_* *is defined as in Assumption 5.*

Proof. *Proof of* (i). Recall that $\widetilde{\mathbf{v}}_t = \mathbf{v}_t + \sum_{|j|>K}\mathbf{\Pi}_j\Delta\mathbf{X}_{t-j}$. Then,

$$\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\widetilde{\mathbf{v}}_t\mathbf{Z}_t^{(K)\top} = \frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{v}_t\mathbf{Z}_t^{(K)\top} + \frac{1}{n-2K}\sum_{t=K+1}^{n-K}\left(\sum_{|j|>K}\mathbf{\Pi}_j\Delta\mathbf{X}_{t-j}\right)\mathbf{Z}_t^{(K)\top}.$$

By Lemma A5 in Saikkonen (1991), we have $\left\|\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\left(\sum_{|j|>K}\mathbf{\Pi}_j\Delta\mathbf{X}_{t-j}\right)\mathbf{Z}_t^{(K)\top}\right\| = o_p\left(\sqrt{K/n}\right)$. In addition, Lemma A6 in Saikkonen (1991) gives $\left\|\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{v}_t\mathbf{Z}_t^{(K)\top}\right\| = O_p\left(\sqrt{K/n}\right)$. (i) follows accordingly.

Proof of (ii). This is just Lemma A3 and Lemma A4 in Saikkonen (1991) so details are omitted.

Proof of (iii). Since d_x is fixed and $\mathbf{X}_0 = O_p(1)$, for notation simplicity we shall prove for the scalar case $d_x = 1$ and assume that $X_0 = 0$. Note that

$$\begin{aligned}\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{Z}_t^{(K)}\mathbf{X}_t^{(k)\top} &= \frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{Z}_t^{(K)}\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes X_t\right)^\top \\ &= \frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{Z}_t^{(K)}X_t\boldsymbol{\psi}_{(k)}^\top(\tau_t),\end{aligned}$$

which is of dimension $(2K+1)\times k$. Consider a generic element $\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\psi_r(\tau_t)X_t u_{x,t+j}$, where $j=K,\cdots,-K$, $r=0,1,\cdots,k-1$. Observe that

$$\begin{aligned}\mathbb{E}\left(X_t u_{x,t+j}X_s u_{x,s+j}\right) &= \sum_{\ell_1=1}^{t}\sum_{\ell_2=1}^{s}\mathbb{E}\left(u_{x,\ell_1}u_{x,t+j}u_{x,\ell_2}u_{x,s+j}\right)\\ &= \sum_{\ell_1=1}^{t}\sum_{\ell_2=1}^{s}\kappa_x(t+j-\ell_1,\ell_2-\ell_1,s+j-\ell_1)+\sum_{\ell_1=1}^{t}\sum_{\ell_2=1}^{s}\gamma_x(t+j-\ell_1)\gamma_x(s+j-\ell_2)\\ &\quad+\sum_{\ell_1=1}^{t}\sum_{\ell_2=1}^{s}\gamma_x(\ell_2-\ell_1)\gamma_x(s-t)+\sum_{\ell_1=1}^{t}\sum_{\ell_2=1}^{s}\gamma_x(s+j-\ell_1)\gamma_x(t+j-\ell_2),\end{aligned}$$

where $\kappa_x(m_1,m_2,m_3):=\text{cum}\left(u_{x,t},u_{x,t+m_1},u_{x,t+m_2},u_{x,t+m_3}\right)$ is the fourth order cumulant of $u_{x,t}$ and $\gamma_x(j)$ is the jth autocovariance of $u_{x,t}$. Then,

$$\begin{aligned}&\mathbb{E}\left(\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\psi_r(\tau_t)X_t u_{x,t+j}\right)^2\\ &=\frac{1}{(n-2K)^2}\sum_t\sum_s\psi_r(\tau_t)\psi_r(\tau_s)\mathbb{E}\left(X_t u_{x,t+j}X_s u_{x,s+j}\right)\\ &=\frac{1}{(n-2K)^2}\sum_t\sum_s\psi_r(\tau_t)\psi_r(\tau_s)\sum_{\ell_1}\sum_{\ell_2}\kappa_x(t+j-\ell_1,\ell_2-\ell_1,s+j-\ell_1)\\ &\quad+\frac{1}{(n-2K)^2}\sum_t\sum_s\psi_r(\tau_t)\psi_r(\tau_s)\sum_{\ell_1}\sum_{\ell_2}\gamma_x(t+j-\ell_1)\gamma_x(s+j-\ell_2)\\ &\quad+\frac{1}{(n-2K)^2}\sum_t\sum_s\psi_r(\tau_t)\psi_r(\tau_s)\sum_{\ell_1}\sum_{\ell_2}\gamma_x(\ell_2-\ell_1)\gamma_x(s-t)\\ &\quad+\frac{1}{(n-2K)^2}\sum_t\sum_s\psi_r(\tau_t)\psi_r(\tau_s)\sum_{\ell_1}\sum_{\ell_2}\gamma_x(s+j-\ell_1)\gamma_x(t+j-\ell_2).\end{aligned}$$

Now consider the first term above. Note that $|\psi_r(\tau_t)|\le\sqrt{2}$ uniformly over any r and t. We have

$$\begin{aligned}&\left|\frac{1}{(n-2K)^2}\sum_t\sum_s\psi_r(\tau_t)\psi_r(\tau_s)\sum_{\ell_1}\sum_{\ell_2}\kappa_x(t+j-\ell_1,\ell_2-\ell_1,s+j-\ell_1)\right|\\ &\le\frac{O(1)}{(n-2K)^2}\sum_t\sum_s\sum_{\ell_1}\sum_{m_2\in\mathbb{Z}}\left|\kappa_x(t+j-\ell_1,m_2,s+j-\ell_1)\right|\\ &\le\frac{O(1)}{(n-2K)^2}\sum_t\sum_s\sum_{m_1\in\mathbb{Z}}\sum_{m_2\in\mathbb{Z}}\left|\kappa_x(m_1,m_2,m_1+s-t)\right|\\ &\le\frac{O(1)}{(n-2K)}\sum_{m_1\in\mathbb{Z}}\sum_{m_2\in\mathbb{Z}}\sum_{m_3\in\mathbb{Z}}\left|\kappa_x(m_1,m_2,m_3)\right|=O\left(\frac{1}{n-2K}\right),\end{aligned}$$

since Assumption 1 guarantees that $\sum_{m_1\in\mathbb{Z}}\sum_{m_2\in\mathbb{Z}}\sum_{m_3\in\mathbb{Z}}\ \left|\kappa_x(m_1,m_2,m_3)\right|<\infty$. Moving on to the second term:

$$\begin{aligned}
&\left|\frac{1}{(n-2K)^2}\sum_t\sum_s\psi_r(\tau_t)\psi_r(\tau_s)\sum_{\ell_1}\sum_{\ell_2}\gamma_x(t+j-\ell_1)\gamma_x(s+j-\ell_2)\right| \\
&\leq \frac{O(1)}{(n-2K)^2}\left(\sum_t\sum_{\ell_1}|\gamma_x(t+j-\ell_1)|\right)\left(\sum_s\sum_{\ell_2}|\gamma_x(s+j-\ell_2)|\right) \\
&\leq \frac{O(1)}{(n-2K)^2}\left\{\sum_t\left(\sum_{h_1\in\mathbb{Z}}|\gamma_x(h_1)|\right)\right\}\left\{\sum_s\left(\sum_{h_2\in\mathbb{Z}}|\gamma_x(h_2)|\right)\right\}=O(1),
\end{aligned}$$

since Assumption 1(i) guarantees that $\sum_{h\in\mathbb{Z}}|\gamma_x(h)|<\infty$. Using similar arguments, we could show that the third and fourth terms above are also $O(1)$. This implies that $\mathbb{E}\left|\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\psi_r(\tau_t)X_tu_{x,t+j}\right|^2=O(1)$, uniformly over j,r. Consequently,

$$\mathbb{E}\left\|\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{Z}_t^{(K)}\mathbf{X}_t^{(k)\top}\right\|^2=\sum_{j=-K}^{K}\sum_{r=1}^{k}\mathbb{E}\left|\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\psi_r(\tau_t)X_tu_{x,t+j}\right|^2=O(Kk)$$

It then follows from Markov's inequality and $\|\mathbf{M}\|_{sp}\leq\|\mathbf{M}\|$ that

$$\left\|\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{Z}_t^{(K)}\mathbf{X}_t^{(k)\top}\right\|_{sp}=O_p\left(\sqrt{Kk}\right).$$

This completes the proof of part (iii).

Proof of (iv). Let us first consider the full sum:

$$\frac{1}{n}\sum_{t=1}^{n}\mathbf{v}_t\mathbf{X}_t^{(k)\top},$$

and proceed similarly as in the proof of Lemma D.1. First, we note that

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\mathbf{v}_t\mathbf{X}_t^{(k)\top}&=\frac{1}{n}\sum_{t=1}^{n}\mathbf{v}_t\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{X}_{t-1}\right)^\top+\frac{1}{n}\sum_{t=1}^{n}\mathbf{v}_t\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{u}_{xt}\right)^\top \\
&:=\widetilde{\boldsymbol{\Gamma}}_n(1)+\widetilde{\boldsymbol{\Gamma}}_n(2).
\end{aligned}$$

Observe that

$$\mathbb{E}\left\|\widetilde{\mathbf{\Gamma}}_n(2)\right\|^2 = \frac{1}{n^2}\sum_{t=1}^{n}\left\|\boldsymbol{\psi}_{(k)}(\tau_t)\right\|^2 \mathbb{E}\left\|\mathbf{v}_t\mathbf{u}_{xt}^\top\right\|^2 + \frac{2}{n^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\boldsymbol{\psi}_{(k)}^\top(\tau_t)\boldsymbol{\psi}_{(k)}(\tau_s)\,\mathrm{tr}\left[\mathbf{\Gamma}_{vx}(t-s)\right],$$

where $\mathbf{\Gamma}_{vx}(h) = \mathbb{E}\left(\mathbf{v}_t\mathbf{u}_{xt}^\top\mathbf{u}_{x,t-h}\mathbf{v}_{t-h}^\top\right)$. By Cauchy-Schwarz inequality,

$$\begin{aligned}\frac{1}{n^2}\sum_{t=1}^{n}\left\|\boldsymbol{\psi}_{(k)}(\tau_t)\right\|^2 \mathbb{E}\left\|\mathbf{v}_t\mathbf{u}_{xt}^\top\right\|^2 &\leq \frac{1}{n^2}\sum_{t=1}^{n}\left\|\boldsymbol{\psi}_{(k)}(\tau_t)\right\|^2 \left(\mathbb{E}\left\|\mathbf{v}_t\right\|^4\right)^{1/2}\left(\mathbb{E}\left\|\mathbf{u}_{xt}\right\|^4\right)^{1/2} \\ &\leq \frac{O(1)}{n^2}\sum_{t=1}^{n}\left\|\boldsymbol{\psi}_{(k)}(\tau_t)\right\|^2 = O\left(\frac{k}{n}\right).\end{aligned}$$

For the second term, we have

$$\begin{aligned}\left|\frac{2}{n^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\boldsymbol{\psi}_{(k)}^\top(\tau_t)\boldsymbol{\psi}_{(k)}(\tau_s)\,\mathrm{tr}\left[\mathbf{\Gamma}_{vx}(t-s)\right]\right| &= \left|\frac{2}{n^2}\sum_{h=1}^{n-1}\mathrm{tr}\left[\mathbf{\Gamma}_{vx}(h)\right]\sum_{s=1}^{n-h}\boldsymbol{\psi}_{(k)}^\top(\tau_{s+h})\boldsymbol{\psi}_{(k)}(\tau_s)\right| \\ &\leq \frac{2}{n^2}\sum_{h=1}^{n-1}\left|\mathrm{tr}\left[\mathbf{\Gamma}_{vx}(h)\right]\right|\left(\sum_{s=1}^{n}\left\|\boldsymbol{\psi}_{(k)}(\tau_s)\right\|^2\right) = O\left(\frac{k}{n}\right),\end{aligned}$$

which follows by the absolute summability of $\|\mathbf{\Gamma}_{vx}(h)\|$. Then, $\left\|\widetilde{\mathbf{\Gamma}}_n(2)\right\| = O_P\left(\sqrt{\frac{k}{n}}\right)$ by the Chebyshev's inequality.

Moving on to $\widetilde{\mathbf{\Gamma}}_n(1)$. Again, as in the proof of Lemma D.1, by BN decomposition, we have

$$\widetilde{\mathbf{\Gamma}}_n(1) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{v}_t\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{V}_{n,t-1}^x\right)^\top + \frac{1}{n}\sum_{t=1}^{n}\mathbf{v}_t\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\left(\widetilde{\mathbf{u}}_{x,0}-\widetilde{\mathbf{u}}_{x,t-1}+\mathbf{X}_0\right)\right)^\top.$$

Using similar arguments for (D.1), we could show that

$$\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{v}_t\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\mathbf{V}_{n,t-1}^x\right)^\top - \int_0^1 d\mathbf{B}_v(r)\left(\boldsymbol{\psi}_{(k)}(r)\otimes\mathbf{B}_X(r)\right)^\top\right\|_{sp} = O_P\left(\frac{k^{3/2}}{n^{\delta_*}}\right).$$

In addition, following the analysis of $\mathbf{\Gamma}_n(1,2)$ in Lemma D.1, we obtain

$$\left\|\frac{1}{n}\sum_{t=1}^{n}\mathbf{v}_t\left(\boldsymbol{\psi}_{(k)}(\tau_t)\otimes\left(\widetilde{\mathbf{u}}_{x,0}-\widetilde{\mathbf{u}}_{x,t-1}+\mathbf{X}_0\right)\right)^\top\right\|_{sp} = O_P\left(\sqrt{\frac{k}{n}}\right).$$

What remains is to analyze the difference:

$$\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{v}_t\mathbf{X}_t^{(k)\top}-\frac{1}{n}\sum_{t=1}^{n}\mathbf{v}_t\mathbf{X}_t^{(k)\top}=\frac{2K}{n(n-2K)}\sum_{t=K+1}^{n-K}\mathbf{v}_t\mathbf{X}_t^{(k)\top}-\frac{1}{n}\sum_{t=1}^{K}\mathbf{v}_t\mathbf{X}_t^{(k)\top}-\frac{1}{n}\sum_{t=n-K+1}^{n}\mathbf{v}_t\mathbf{X}_t^{(k)\top}.$$

Note that

$$\left\|\frac{1}{n}\sum_{t=1}^{K}\mathbf{v}_t\mathbf{X}_t^{(k)\top}\right\|\leq\frac{O(1)\sqrt{k}}{n}\left(\max_{1\leq t\leq n}\|\mathbf{X}_t\|\right)\left(\sum_{t=1}^{K}\|\mathbf{v}_t\|\right)=O_P\left(K\sqrt{\frac{k}{n}}\right).$$

Similarly, we can show that

$$\left\|\frac{1}{n}\sum_{t=n-K+1}^{n}\mathbf{v}_t\mathbf{X}_t^{(k)\top}\right\|=O_P\left(K\sqrt{\frac{k}{n}}\right),\quad\left\|\sum_{t=K+1}^{n-K}\mathbf{v}_t\mathbf{X}_t^{(k)\top}\right\|=O_P\left(\sqrt{kn}(n-2K)\right).$$

This implies that

$$\left\|\frac{1}{n-2K}\sum_{t=K+1}^{n-K}\mathbf{v}_t\mathbf{X}_t^{(k)\top}-\frac{1}{n}\sum_{t=1}^{n}\mathbf{v}_t\mathbf{X}_t^{(k)\top}\right\|=O_P\left(K\sqrt{\frac{k}{n}}\right),$$

as desired. Finally, the bound $\left\|\int_0^1 d\mathbf{B}_v(r)\left(\boldsymbol{\psi}_{(k)}(r)\otimes\mathbf{B}_X(r)\right)^\top\right\|=O_p(\sqrt{k})$ follows by an argument analogous to that one $\left\|\mathbf{\Gamma}_{ux}^{(k)}\right\|$ in Lemma D.1 so the details are omitted. □

F — Monte Carlo simulations

The Monte Carlo experiments specialize the general model to $d_y=d_x=1$. To make the scalar nature of the simulation design explicit, we write y_t, $\beta_t:=\beta(\tau_t)$ with $\tau_t=t/n$, and x_t in place of $\mathbf{y}_t$, $\mathbf{A}_t$, and $\mathbf{X}_t$, respectively, while retaining the notation for innovations u_{0t} and u_{xt} used in the main text.

We use the data generating process (D.G.P.):

$$y_t=\beta_t x_t+u_{0t},$$

where β_t takes one the following functional forms

- M1: $\beta_t=1+\tau_t$;
- M2: $\beta_t=e^{-\tau_t}$;
- M3: $\beta_t=e^{-\frac{1-\cos(\pi\tau_t)}{2}}$.

It is worth mentioning that among the three coefficient designs, only M3 satisfies the endpoint restrictions in Assumption 2. In particular, M1 and M2 violate these restrictions already at the first derivative, since $\beta^{(1)}(0) = \beta^{(1)}(1) = 1$ for M1 and $\beta^{(1)}(r) = -e^{-r} \neq 0$ at $r = 0, 1$. By contrast, M3 is infinitely differentiable and satisfies $\beta^{(2\ell-1)}(0) = \beta^{(2\ell-1)}(1) = 0$ for every positive integer ℓ. Thus, M3 is the design directly covered by the approximation theory, while M1 and M2 are included to examine robustness to violations of the endpoint restrictions.

The regressor x_t is generated according to $x_t = x_{t-1} + u_{xt}$. We generate $\mathbf{u}_t = (u_{0t}, u_{xt})^\top$ from a bivariate MA(1) process

$$\mathbf{u}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\Theta}\boldsymbol{\varepsilon}_{t-1}, \quad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathcal{N}(0, \boldsymbol{\Sigma}_\varepsilon)$$

with

$$\boldsymbol{\Theta} = \begin{pmatrix} 0.3 & -0.4 \\ \theta_{21} & 0.6 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_\varepsilon = \begin{pmatrix} 1 & \sigma_{21} \\ \sigma_{21} & 1 \end{pmatrix},$$

and θ_{21}, σ_{21} are allowed to vary. We consider values of $\{0.8, 0.4, 0, -0.8\}$ for θ_{21} and $\{-0.85, -0.5, 0.5\}$ for σ_{21}, as in Phillips and Loretan (1991).

Let

$$\boldsymbol{b}^{(k)} = (b_0, \ldots, b_{k-1})^\top, \qquad \beta_k(r) = \boldsymbol{b}^{(k)\top}\boldsymbol{\psi}_{(k)}(r),$$

and define the $k \times 1$ series-augmented regressor

$$\boldsymbol{x}_t^{(k)} = x_t\boldsymbol{\psi}_{(k)}(\tau_t).$$

Then, we can write

$$y_t = \boldsymbol{b}^{(k)\top}\boldsymbol{x}_t^{(k)} + u_{0t}^{(k)}, \tag{F.1}$$

where $u_{0t}^{(k)} = u_{0t} + \{\beta_t - \beta_k(\tau_t)\}\, x_t$. Then, the series-based OLS (SOLS) estimator is computed by $\widehat{\beta}_k(r) = \widehat{\boldsymbol{b}}^{(k)\top}\boldsymbol{\psi}_{(k)}(r)$, where $\widehat{\boldsymbol{b}}^{(k)}$ denotes the OLS estimator from (F.1).

For SDOLS estimator introduced in Section 3.3, we consider a slightly more flexible regression model compared to (29) in which the number of lags (K_a) and leads (K_e) may differ[6]:

$$y_t = \boldsymbol{b}^{(k)\top}\boldsymbol{x}_t^{(k)} + \sum_{j=-K_e}^{K_a} \gamma_j \Delta x_{t-j} + \widetilde{v}_t. \tag{F.2}$$

6. As pointed out in Saikkonen (1991), using the same truncation value K for both leads and lags of $\Delta\mathbf{X}_t$ is mainly for ease of theoretical development. In practice, different values of K_a and K_e may be used.

We consider several combinations of the numbers of lags (K_a) and leads (K_e): (i) $K_a = K_e = 0$; (ii) $K_a = 2, K_e = 0$; (iii) $K_a = 4, K_e = 0$; (iv) $K_a = 2, K_e = 1$; (v) $K_a = 4, K_e = 2$; (vi) $K_a = K_e = 1$; (vii) $K_a = K_e = 4$; and (viii) $K_a = K_e = 8$.[7] We also consider the series CCR (SCCR) estimator proposed by Park and Hahn (1999).

Regarding the truncation parameter k, we use the generalized cross validation (GCV) method as in Gao et al. (2002) to select an optimal value $\widehat{k}$:

$$\widehat{k} = \underset{k \in \mathcal{K}_n}{\arg\min} \left(1 - \frac{df(k, K_a, K_e)}{n}\right)^{-2} \frac{1}{n} \sum_{t=1}^{n} \left(\widehat{e}_t^{(k;K_a,K_e)}\right)^2,$$

where $df(k, K_a, K_e)$ is the total number of parameters to be estimated, $\mathcal{K}_n = \{1, 2, \cdots, 10, 11, 12\}$, and $\widehat{e}_t^{(k;K_a,K_e)}$ is the fitted residual from (F.2). For the SOLS regression (F.1), the arguments K_a and K_e are omitted, so the criterion uses $df(k)$ and the fitted residual $\widehat{e}_t^{(k)}$. The GCV criterion is minimized separately for each estimator and, for SDOLS, each augmentation specification (K_a, K_e), accordingly. Then, we write $\widehat{\beta}_t := \widehat{\beta}_{\widehat{k}}(\tau_t)$.

The performance of the estimators is evaluated by the mean squared error (MSE),

$$\frac{1}{n} \sum_{t=1}^{n} \left(\widehat{\beta}_t - \beta_t\right)^2,$$

and the empirical coverage probabilities of the associated 95% confidence intervals,

$$\left(\widehat{\beta}_t - 1.96 \,\text{std}\left(\widehat{\beta}_t\right), \widehat{\beta}_t + 1.96 \,\text{std}\left(\widehat{\beta}_t\right)\right),$$

at the point $t = \lfloor \frac{n}{2} \rfloor$. std $\left(\widehat{\beta}_t\right)$ is the estimated standard error constructed based on the asymptotic results in either Theorem 1 or Theorem 3, with long-run variance estimated using Bartlett kernel with 1 lag. All simulation results are based on 2,000 Monte Carlo replications.

Tables F1-F3 report the average MSE under designs M1-M3, respectively. For each combination of $(\theta_{21}, \sigma_{21})$ and sample size n, we compare among SOLS estimator, SCCR estimator, and SDOLS estimator across a range of choices for the numbers of lags (K_a) and leads (K_e). The MSEs are in levels for SOLS, whereas all other entries are ratios relative to the baseline performance. Values below 1 indicate an improvement from augmentation, and boxed entries mark the best performing estimator.

Several patterns emerge. First, gains from the efficient estimators are substantially larger when $\sigma_{21} < 0$.

7. Note that, in all these specifications based on (F.2), Δx_t is always included in the model.

In these cases, SDOLS delivers the lowest MSE in most designs. Second, among all estimators, SDOLS provides the best performance in 84 out of 108 cases, compared with 12 cases each for SOLS and SCCR. The best performing SDOLS typically involves small to modest augmentation, and including leads of Δx_t generally provides additional improvements. SDOLS with leads delivers the lowest MSE in 28 out of 36 cases under M1, 25 out of 36 cases under M2, and 17 out of 36 cases under M3. Overall, the simulation evidence suggests that SDOLS with small to modest augmentation orders improves estimation accuracy, especially in situations with stronger negative contemporaneous correlation.

Tables F4-F6 report the empirical coverage probabilities evaluated at the point $t = \lfloor n/2 \rfloor$ under designs M1-M3, respectively. All entries are reported in levels, with each row corresponding to the same implementation of various estimators as in Tables F1-F3. Boxed entries indicate the estimator whose coverage is closest to the nominal 95% level for each combination of $(\theta_{21}, \sigma_{21})$ and sample size n. Across the 108 designs, SDOLS delivers coverage closest to the nominal level in 93 cases, compared with 3 cases for SOLS and 12 cases for SCCR. SDOLS with leads delivers the best performance in 25 out of 36 cases under M1, 28 out of 36 cases under M2, and 24 out of 36 cases under M3. Moderate augmentation schemes such as $(K_a, K_e) = (2, 1)$, $(4, 2)$, and $(4, 4)$ frequently perform well, while $(K_a, K_e) = (8, 8)$ is selected most often under M3. Overall, the results indicate that SDOLS not only improves estimation accuracy, but also provides more reliable pointwise inference than both SOLS and SCCR.

Finally, it is worth mentioning again that since M3 satisfies Assumption 2, its results provide direct simulation evidence consistent with the asymptotic theory. The broadly similar performance under M1 and M2 suggests that the estimators can remain effective even when the endpoint compatibility conditions are violated. This should be interpreted as finite sample robustness beyond the sufficient conditions of the theory.

Table F1 — Small sample properties of the series estimator of β_t under M1: $\beta_t = 1 + \tau_t$

	$n = 200$				$n = 500$				$n = 1000$			
θ_{21}	0.8	0.4	0.0	−0.8	0.8	0.4	0.0	−0.8	0.8	0.4	0.0	−0.8
						$\sigma_{21} = -0.85$						
SOLS	0.028	0.015	0.007	0.002	0.007	0.003	0.001	0.000	0.002	0.001	0.000	0.000
SCCR	0.510	0.317	0.205	0.182	0.370	0.247	0.177	0.195	0.315	0.233	0.208	0.284
(0,0)	0.442	0.239	0.149	0.167	0.429	0.260	0.162	0.210	0.425	0.258	0.184	0.280
(2,0)	0.279	0.205	0.154	0.148	0.233	0.200	0.160	0.183	0.216	0.191	0.182	0.250
(4,0)	0.297	0.223	0.170	0.159	0.237	0.202	0.163	0.180	0.217	0.193	0.184	0.246
(2,1)	0.268	0.202	0.155	0.127	0.219	0.196	0.160	0.148	0.203	0.191	0.182	0.212
(4,2)	0.287	0.222	0.171	0.126	0.225	0.200	0.164	0.136	0.204	0.193	0.184	0.199
(1,1)	0.260	0.200	0.156	0.123	0.217	0.201	0.161	0.147	0.202	0.198	0.185	0.212
(4,4)	0.298	0.228	0.170	0.124	0.225	0.201	0.165	0.133	0.205	0.194	0.185	0.192
(8,8)	0.432	0.316	0.223	0.157	0.244	0.216	0.178	0.145	0.212	0.200	0.191	0.197
						$\sigma_{21} = -0.5$						
SOLS	0.010	0.009	0.006	0.002	0.002	0.002	0.001	0.000	0.001	0.000	0.000	0.000
SCCR	0.777	0.672	0.535	0.416	0.648	0.566	0.489	0.404	0.603	0.543	0.476	0.479
(0,0)	0.725	0.601	0.449	0.352	0.721	0.569	0.453	0.430	0.725	0.581	0.469	0.504
(2,0)	0.645	0.578	0.463	0.382	0.591	0.506	0.450	0.446	0.573	0.506	0.457	0.524
(4,0)	0.713	0.628	0.503	0.416	0.609	0.519	0.461	0.457	0.578	0.504	0.457	0.528
(2,1)	0.621	0.575	0.469	0.333	0.549	0.493	0.448	0.387	0.506	0.485	0.458	0.439
(4,2)	0.688	0.637	0.510	0.357	0.565	0.506	0.461	0.370	0.510	0.490	0.459	0.425
(1,1)	0.600	0.563	0.456	0.328	0.544	0.496	0.447	0.379	0.508	0.493	0.459	0.433
(4,4)	0.702	0.650	0.525	0.356	0.570	0.510	0.460	0.368	0.511	0.490	0.463	0.419
(8,8)	1.029	0.949	0.742	0.456	0.628	0.537	0.495	0.394	0.527	0.509	0.477	0.428
						$\sigma_{21} = 0.5$						
SOLS	0.001	0.002	0.002	0.003	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
SCCR	0.976	1.072	1.142	0.939	0.919	1.002	1.064	0.911	0.906	0.982	1.032	0.866
(0,0)	1.211	1.251	1.261	1.009	1.223	1.183	1.194	0.952	1.172	1.164	1.116	0.900
(2,0)	1.044	1.116	1.258	1.116	0.972	0.980	1.076	1.046	0.958	0.941	0.979	1.006
(4,0)	1.081	1.196	1.351	1.196	0.957	0.967	1.113	1.073	0.937	0.914	0.981	1.022
(2,1)	1.055	1.155	1.246	0.946	0.773	0.911	1.090	0.902	0.771	0.846	0.974	0.834
(4,2)	1.105	1.237	1.347	1.027	0.776	0.942	1.113	0.921	0.749	0.864	0.984	0.840
(1,1)	1.056	1.153	1.240	0.916	0.802	0.968	1.116	0.887	0.815	0.907	1.026	0.834
(4,4)	1.152	1.281	1.351	1.045	0.765	0.946	1.128	0.927	0.734	0.854	0.993	0.839
(8,8)	1.597	1.720	1.854	1.392	0.829	1.000	1.185	0.992	0.753	0.881	1.013	0.860

Notes. SOLS: series OLS estimator. SCCR: series CCR estimator. All other rows are based on the SDOLS estimator, with different combinations of (K_a, K_e). Entries in the row "SOLS" report MSE in levels (rounded to the 3rd decimal), while the remaining entries are ratios relative to SOLS performance. For each combination of $(\theta_{21}, \sigma_{21})$ and sample size n, boxed entries indicate the best performance.

Table F2 — Small sample properties of the series estimator of β_t under M2: $\beta_t = e^{-\tau_t}$

	$n = 200$				$n = 500$				$n = 1000$			
θ_{21}	0.8	0.4	0.0	−0.8	0.8	0.4	0.0	−0.8	0.8	0.4	0.0	−0.8
					$\sigma_{21} = -0.85$							
SOLS	0.027	0.013	0.006	0.002	0.006	0.003	0.001	0.000	0.002	0.001	0.000	0.000
SCCR	0.506	0.338	0.246	0.292	0.378	0.276	0.242	0.256	0.338	0.265	0.258	0.327
(0,0)	0.430	0.241	0.161	0.137	0.420	0.244	0.160	0.178	0.421	0.259	0.177	0.192
(2,0)	0.281	0.212	0.165	0.126	0.232	0.178	0.158	0.163	0.214	0.187	0.170	0.179
(4,0)	0.309	0.227	0.180	0.130	0.237	0.180	0.161	0.165	0.216	0.189	0.170	0.178
(2,1)	0.265	0.216	0.167	0.111	0.221	0.176	0.159	0.141	0.201	0.184	0.170	0.155
(4,2)	0.299	0.236	0.185	0.111	0.228	0.178	0.162	0.127	0.204	0.187	0.172	0.143
(1,1)	0.259	0.212	0.166	0.108	0.220	0.176	0.157	0.139	0.202	0.186	0.166	0.155
(4,4)	0.309	0.246	0.189	0.112	0.233	0.182	0.163	0.124	0.206	0.188	0.172	0.139
(8,8)	0.446	0.350	0.256	0.145	0.248	0.195	0.175	0.134	0.213	0.194	0.176	0.144
					$\sigma_{21} = -0.5$							
SOLS	0.010	0.008	0.005	0.002	0.002	0.002	0.001	0.000	0.001	0.000	0.000	0.000
SCCR	0.797	0.665	0.557	0.473	0.659	0.571	0.498	0.444	0.608	0.543	0.511	0.478
(0,0)	0.742	0.578	0.437	0.320	0.722	0.557	0.445	0.403	0.715	0.569	0.479	0.417
(2,0)	0.660	0.561	0.449	0.332	0.594	0.500	0.445	0.414	0.556	0.490	0.462	0.424
(4,0)	0.712	0.608	0.471	0.362	0.604	0.510	0.452	0.422	0.558	0.492	0.463	0.427
(2,1)	0.609	0.569	0.455	0.316	0.546	0.490	0.449	0.374	0.483	0.476	0.463	0.375
(4,2)	0.673	0.614	0.482	0.338	0.561	0.506	0.454	0.366	0.489	0.478	0.465	0.352
(1,1)	0.597	0.543	0.441	0.300	0.547	0.484	0.438	0.357	0.479	0.474	0.459	0.355
(4,4)	0.700	0.642	0.493	0.340	0.569	0.512	0.465	0.365	0.491	0.480	0.469	0.349
(8,8)	1.004	0.901	0.677	0.454	0.616	0.550	0.499	0.391	0.509	0.496	0.489	0.363
					$\sigma_{21} = 0.5$							
SOLS	0.001	0.001	0.002	0.003	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
SCCR	0.979	1.061	1.079	0.896	0.958	1.012	1.074	0.871	0.953	1.008	1.048	0.898
(0,0)	1.170	1.297	1.258	1.055	1.240	1.247	1.259	1.012	1.265	1.249	1.204	0.985
(2,0)	1.047	1.122	1.262	1.102	1.018	1.064	1.133	1.064	0.999	0.985	1.032	1.044
(4,0)	1.122	1.194	1.349	1.129	1.002	1.059	1.152	1.087	0.974	0.949	1.024	1.051
(2,1)	1.033	1.192	1.290	1.012	0.849	1.032	1.139	0.897	0.783	0.871	1.029	0.881
(4,2)	1.122	1.297	1.367	1.056	0.870	1.066	1.156	0.931	0.772	0.893	1.023	0.886
(1,1)	1.003	1.148	1.265	0.985	0.860	1.087	1.098	0.884	0.793	0.880	1.025	0.870
(4,4)	1.170	1.357	1.392	1.080	0.862	1.076	1.175	0.940	0.759	0.885	1.030	0.887
(8,8)	1.725	1.899	2.007	1.458	0.920	1.151	1.261	0.984	0.787	0.908	1.062	0.921

Notes. See Notes in Table F1.

Table F3 — Small sample properties of the series estimator of β_t under M3: $\beta_t = e^{-\frac{1-\cos(\pi\tau_t)}{2}}$

	$n = 200$				$n = 500$				$n = 1000$			
θ_{21}	0.8	0.4	0.0	−0.8	0.8	0.4	0.0	−0.8	0.8	0.4	0.0	−0.8
					$\sigma_{21} = -0.85$							
SOLS	0.027	0.013	0.006	0.002	0.006	0.002	0.001	0.000	0.002	0.001	0.000	0.000
SCCR	0.515	0.319	0.206	0.153	0.371	0.229	0.141	0.090	0.308	0.188	0.118	0.110
(0,0)	0.426	0.215	0.119	0.052	0.398	0.214	0.120	0.064	0.390	0.205	0.114	0.070
(2,0)	0.280	0.198	0.121	0.044	0.230	0.175	0.121	0.054	0.211	0.169	0.118	0.057
(4,0)	0.298	0.216	0.137	0.045	0.236	0.180	0.123	0.054	0.212	0.170	0.118	0.055
(2,1)	0.272	0.201	0.124	0.043	0.222	0.174	0.123	0.051	0.199	0.168	0.117	0.049
(4,2)	0.291	0.224	0.138	0.045	0.228	0.181	0.127	0.048	0.201	0.171	0.119	0.051
(1,1)	0.263	0.195	0.126	0.043	0.220	0.174	0.121	0.052	0.199	0.172	0.119	0.049
(4,4)	0.300	0.232	0.140	0.046	0.234	0.182	0.129	0.051	0.203	0.172	0.118	0.051
(8,8)	0.443	0.322	0.185	0.068	0.255	0.199	0.140	0.056	0.210	0.178	0.124	0.052
					$\sigma_{21} = -0.5$							
SOLS	0.010	0.008	0.005	0.002	0.002	0.001	0.001	0.000	0.000	0.000	0.000	0.000
SCCR	0.774	0.662	0.526	0.368	0.639	0.543	0.440	0.287	0.584	0.496	0.395	0.278
(0,0)	0.720	0.581	0.382	0.191	0.711	0.519	0.394	0.224	0.707	0.531	0.381	0.220
(2,0)	0.654	0.573	0.413	0.210	0.592	0.478	0.395	0.234	0.575	0.477	0.377	0.229
(4,0)	0.718	0.631	0.445	0.220	0.597	0.491	0.403	0.242	0.580	0.481	0.381	0.233
(2,1)	0.631	0.585	0.416	0.186	0.559	0.488	0.400	0.219	0.524	0.475	0.380	0.217
(4,2)	0.705	0.652	0.452	0.214	0.572	0.504	0.414	0.233	0.529	0.481	0.386	0.219
(1,1)	0.612	0.565	0.397	0.180	0.559	0.489	0.399	0.211	0.523	0.478	0.379	0.208
(4,4)	0.741	0.672	0.473	0.219	0.580	0.512	0.419	0.233	0.536	0.486	0.393	0.216
(8,8)	1.104	1.004	0.715	0.321	0.625	0.550	0.442	0.250	0.555	0.505	0.402	0.213
					$\sigma_{21} = 0.5$							
SOLS	0.001	0.001	0.001	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SCCR	0.924	1.047	1.103	0.914	0.883	0.973	1.028	0.839	0.842	0.930	1.017	0.854
(0,0)	1.273	1.296	1.406	1.093	1.257	1.332	1.362	1.049	1.226	1.287	1.329	1.085
(2,0)	1.050	1.239	1.466	1.226	1.042	1.146	1.356	1.140	0.990	1.158	1.321	1.163
(4,0)	1.113	1.379	1.590	1.306	1.034	1.177	1.397	1.165	0.968	1.160	1.377	1.172
(2,1)	1.352	1.428	1.497	1.073	1.249	1.169	1.371	0.970	1.021	1.215	1.330	1.006
(4,2)	1.523	1.621	1.606	1.195	1.317	1.333	1.409	0.994	1.085	1.279	1.393	1.023
(1,1)	1.333	1.398	1.420	1.033	1.294	1.181	1.340	0.958	1.076	1.257	1.372	0.991
(4,4)	1.640	1.751	1.630	1.245	1.344	1.367	1.473	0.994	1.114	1.320	1.398	1.037
(8,8)	2.282	2.363	2.418	1.722	1.454	1.425	1.631	1.084	1.158	1.353	1.415	1.062

Notes. See Notes in Table F1.

Table F4 — Small sample properties of the series estimator of β_t under M1: $\beta_t = 1 + \tau_t$

	$n = 200$				$n = 500$				$n = 1000$			
θ_{21}	0.8	0.4	0.0	−0.8	0.8	0.4	0.0	−0.8	0.8	0.4	0.0	−0.8
						$\sigma_{21} = -0.85$						
SOLS	0.74	0.79	0.82	0.89	0.78	0.83	0.88	0.92	0.81	0.87	0.91	0.95
SCCR	0.82	0.82	0.84	0.92	0.85	0.85	0.88	0.94	0.88	0.89	0.88	0.94
(0,0)	0.79	0.83	0.85	0.93	0.81	0.85	0.90	0.95	0.83	0.88	0.90	0.96
(2,0)	0.87	0.87	0.85	0.94	0.88	0.88	0.90	0.96	0.91	0.91	0.91	0.97
(4,0)	0.87	0.86	0.85	0.94	0.88	0.88	0.90	0.96	0.91	0.91	0.91	0.97
(2,1)	0.89	0.86	0.84	0.96	0.90	0.88	0.90	0.95	0.92	0.91	0.91	0.94
(4,2)	0.89	0.86	0.84	0.97	0.90	0.88	0.90	0.97	0.92	0.91	0.91	0.98
(1,1)	0.88	0.86	0.85	0.96	0.89	0.87	0.90	0.95	0.92	0.91	0.90	0.94
(4,4)	0.88	0.86	0.84	0.97	0.89	0.88	0.90	0.98	0.92	0.91	0.91	0.97
(8,8)	0.89	0.87	0.84	0.97	0.90	0.87	0.89	0.98	0.92	0.90	0.91	0.97
						$\sigma_{21} = -0.5$						
SOLS	0.78	0.79	0.83	0.88	0.82	0.84	0.87	0.92	0.85	0.88	0.89	0.93
SCCR	0.81	0.81	0.85	0.91	0.86	0.87	0.89	0.95	0.88	0.89	0.89	0.94
(0,0)	0.80	0.81	0.84	0.91	0.85	0.86	0.88	0.94	0.86	0.88	0.90	0.94
(2,0)	0.82	0.83	0.83	0.91	0.85	0.87	0.89	0.94	0.86	0.89	0.90	0.94
(4,0)	0.81	0.82	0.84	0.90	0.85	0.87	0.88	0.94	0.86	0.89	0.90	0.94
(2,1)	0.85	0.83	0.84	0.93	0.88	0.88	0.89	0.96	0.88	0.89	0.90	0.94
(4,2)	0.84	0.83	0.84	0.93	0.88	0.88	0.88	0.97	0.89	0.88	0.90	0.95
(1,1)	0.84	0.83	0.84	0.93	0.88	0.87	0.89	0.95	0.88	0.88	0.90	0.94
(4,4)	0.84	0.82	0.84	0.94	0.88	0.88	0.89	0.96	0.88	0.89	0.90	0.95
(8,8)	0.85	0.83	0.84	0.94	0.87	0.88	0.89	0.96	0.89	0.88	0.90	0.95
						$\sigma_{21} = 0.5$						
SOLS	0.87	0.88	0.86	0.87	0.90	0.91	0.90	0.89	0.92	0.91	0.92	0.91
SCCR	0.87	0.87	0.85	0.87	0.91	0.91	0.90	0.90	0.92	0.91	0.91	0.92
(0,0)	0.84	0.86	0.84	0.85	0.89	0.89	0.89	0.88	0.91	0.90	0.91	0.90
(2,0)	0.85	0.88	0.86	0.86	0.90	0.91	0.91	0.89	0.92	0.91	0.92	0.90
(4,0)	0.86	0.87	0.86	0.86	0.89	0.91	0.91	0.89	0.91	0.90	0.92	0.91
(2,1)	0.91	0.88	0.86	0.89	0.93	0.92	0.91	0.92	0.94	0.92	0.92	0.93
(4,2)	0.91	0.88	0.85	0.89	0.94	0.92	0.91	0.91	0.94	0.92	0.92	0.93
(1,1)	0.91	0.88	0.85	0.89	0.93	0.92	0.90	0.92	0.94	0.91	0.92	0.93
(4,4)	0.92	0.88	0.86	0.89	0.94	0.92	0.90	0.92	0.95	0.92	0.92	0.93
(8,8)	0.91	0.88	0.87	0.88	0.94	0.92	0.90	0.91	0.94	0.91	0.92	0.93

Notes. SOLS: series OLS estimator. SCCR: series CCR estimator. All other rows are based on the SDOLS estimator, with different combinations of (K_a, K_e). Entries report coverage probabilities of 95% pointwise confidence intervals evaluated at $t = \lfloor n/2 \rfloor$, so that $\tau_{\lfloor n/2 \rfloor} = 0.5$ for the reported sample sizes. Boxed entries indicate the specification whose coverage is closest to the nominal 95% level for each combination of $(\theta_{21}, \sigma_{21})$ and sample size n.

Table F5 — Small sample properties of the series estimator of β_t under M2: $\beta_t = e^{-\tau_t}$

	$n = 200$				$n = 500$				$n = 1000$			
θ_{21}	0.8	0.4	0.0	−0.8	0.8	0.4	0.0	−0.8	0.8	0.4	0.0	−0.8
						$\sigma_{21} = -0.85$						
SOLS	0.71	0.78	0.82	0.85	0.74	0.79	0.85	0.86	0.77	0.81	0.88	0.90
SCCR	0.78	0.77	0.73	0.76	0.81	0.82	0.83	0.89	0.85	0.84	0.86	0.91
(0,0)	0.75	0.78	0.83	0.92	0.78	0.82	0.89	0.93	0.81	0.85	0.91	0.95
(2,0)	0.85	0.83	0.84	0.93	0.88	0.88	0.90	0.94	0.89	0.89	0.91	0.96
(4,0)	0.85	0.84	0.84	0.93	0.89	0.88	0.89	0.94	0.90	0.89	0.90	0.96
(2,1)	0.86	0.84	0.84	0.95	0.88	0.89	0.90	0.96	0.90	0.89	0.91	0.97
(4,2)	0.87	0.85	0.84	0.96	0.88	0.89	0.90	0.97	0.90	0.89	0.90	0.98
(1,1)	0.86	0.84	0.84	0.95	0.88	0.88	0.90	0.96	0.90	0.88	0.91	0.97
(4,4)	0.87	0.84	0.84	0.96	0.89	0.88	0.90	0.97	0.90	0.89	0.91	0.98
(8,8)	0.87	0.84	0.84	0.96	0.88	0.88	0.89	0.97	0.90	0.89	0.91	0.98
						$\sigma_{21} = -0.5$						
SOLS	0.75	0.79	0.82	0.85	0.79	0.82	0.84	0.89	0.82	0.84	0.86	0.91
SCCR	0.78	0.79	0.80	0.81	0.82	0.83	0.86	0.90	0.85	0.86	0.88	0.93
(0,0)	0.78	0.79	0.82	0.88	0.81	0.84	0.88	0.91	0.84	0.86	0.89	0.93
(2,0)	0.81	0.81	0.82	0.88	0.82	0.86	0.88	0.91	0.85	0.87	0.89	0.93
(4,0)	0.81	0.81	0.83	0.87	0.82	0.86	0.89	0.90	0.86	0.87	0.89	0.93
(2,1)	0.83	0.81	0.82	0.91	0.85	0.87	0.88	0.92	0.89	0.88	0.89	0.96
(4,2)	0.83	0.82	0.83	0.92	0.86	0.87	0.88	0.93	0.89	0.88	0.89	0.96
(1,1)	0.83	0.81	0.81	0.92	0.85	0.87	0.89	0.92	0.89	0.88	0.89	0.96
(4,4)	0.83	0.82	0.83	0.93	0.86	0.86	0.88	0.94	0.88	0.88	0.89	0.96
(8,8)	0.83	0.82	0.82	0.92	0.85	0.86	0.88	0.94	0.88	0.88	0.89	0.96
						$\sigma_{21} = 0.5$						
SOLS	0.87	0.85	0.85	0.79	0.88	0.89	0.88	0.88	0.90	0.90	0.91	0.91
SCCR	0.88	0.86	0.84	0.82	0.89	0.89	0.87	0.89	0.91	0.92	0.90	0.91
(0,0)	0.84	0.84	0.83	0.79	0.86	0.87	0.85	0.87	0.88	0.89	0.90	0.89
(2,0)	0.87	0.87	0.85	0.80	0.88	0.89	0.88	0.88	0.90	0.91	0.90	0.90
(4,0)	0.86	0.87	0.85	0.80	0.87	0.90	0.88	0.88	0.90	0.91	0.91	0.90
(2,1)	0.88	0.86	0.84	0.86	0.92	0.90	0.88	0.90	0.92	0.92	0.90	0.92
(4,2)	0.90	0.87	0.85	0.86	0.92	0.91	0.88	0.90	0.93	0.92	0.91	0.92
(1,1)	0.88	0.86	0.84	0.85	0.92	0.90	0.88	0.90	0.93	0.93	0.90	0.92
(4,4)	0.90	0.88	0.85	0.86	0.93	0.91	0.88	0.90	0.94	0.93	0.90	0.92
(8,8)	0.89	0.88	0.85	0.87	0.93	0.91	0.88	0.90	0.94	0.93	0.90	0.92

Notes. See Notes in Table F4.

Table F6 — Small sample properties of the series estimator of β_t under M3: $\beta_t = e^{-\frac{1-\cos(\pi\tau_t)}{2}}$

	$n = 200$				$n = 500$				$n = 1000$			
θ_{21}	0.8	0.4	0.0	−0.8	0.8	0.4	0.0	−0.8	0.8	0.4	0.0	−0.8
						$\sigma_{21} = -0.85$						
SOLS	0.72	0.78	0.80	0.85	0.76	0.81	0.83	0.88	0.76	0.83	0.86	0.91
SCCR	0.79	0.78	0.75	0.88	0.84	0.87	0.87	0.92	0.86	0.88	0.87	0.89
(0,0)	0.75	0.79	0.83	0.94	0.78	0.84	0.86	0.92	0.79	0.85	0.86	0.94
(2,0)	0.85	0.84	0.84	0.95	0.88	0.87	0.86	0.94	0.88	0.87	0.86	0.95
(4,0)	0.85	0.84	0.84	0.95	0.88	0.87	0.86	0.93	0.88	0.87	0.86	0.95
(2,1)	0.87	0.85	0.84	0.96	0.88	0.87	0.86	0.96	0.88	0.87	0.87	0.97
(4,2)	0.87	0.84	0.83	0.97	0.88	0.87	0.86	0.96	0.88	0.88	0.87	0.96
(1,1)	0.87	0.84	0.84	0.96	0.89	0.87	0.86	0.96	0.88	0.87	0.86	0.97
(4,4)	0.87	0.85	0.84	0.97	0.89	0.87	0.86	0.97	0.88	0.88	0.87	0.97
(8,8)	0.88	0.85	0.84	0.97	0.89	0.88	0.85	0.97	0.88	0.88	0.87	0.97
						$\sigma_{21} = -0.5$						
SOLS	0.76	0.77	0.79	0.85	0.81	0.84	0.83	0.90	0.81	0.84	0.85	0.90
SCCR	0.78	0.78	0.79	0.88	0.84	0.86	0.87	0.94	0.85	0.87	0.88	0.92
(0,0)	0.78	0.79	0.81	0.91	0.83	0.85	0.86	0.92	0.83	0.86	0.87	0.92
(2,0)	0.81	0.81	0.81	0.91	0.83	0.84	0.86	0.92	0.84	0.87	0.86	0.92
(4,0)	0.79	0.80	0.81	0.91	0.84	0.84	0.86	0.91	0.84	0.87	0.87	0.92
(2,1)	0.82	0.82	0.81	0.92	0.86	0.86	0.86	0.93	0.87	0.87	0.87	0.93
(4,2)	0.82	0.81	0.81	0.93	0.87	0.85	0.86	0.93	0.87	0.88	0.86	0.94
(1,1)	0.82	0.81	0.81	0.92	0.87	0.86	0.85	0.92	0.87	0.87	0.86	0.93
(4,4)	0.83	0.81	0.82	0.93	0.86	0.86	0.86	0.93	0.87	0.88	0.86	0.94
(8,8)	0.83	0.82	0.81	0.93	0.87	0.86	0.86	0.94	0.87	0.87	0.87	0.94
						$\sigma_{21} = 0.5$						
SOLS	0.86	0.87	0.84	0.81	0.89	0.88	0.89	0.89	0.86	0.88	0.88	0.89
SCCR	0.88	0.86	0.84	0.83	0.89	0.89	0.88	0.90	0.88	0.89	0.89	0.92
(0,0)	0.85	0.85	0.83	0.80	0.86	0.86	0.88	0.87	0.85	0.88	0.87	0.88
(2,0)	0.86	0.87	0.84	0.81	0.87	0.88	0.88	0.88	0.86	0.89	0.88	0.88
(4,0)	0.86	0.87	0.85	0.81	0.87	0.88	0.88	0.88	0.86	0.89	0.88	0.88
(2,1)	0.89	0.85	0.85	0.87	0.90	0.89	0.88	0.90	0.89	0.90	0.89	0.91
(4,2)	0.90	0.86	0.85	0.86	0.91	0.90	0.88	0.90	0.89	0.90	0.89	0.91
(1,1)	0.88	0.86	0.83	0.86	0.90	0.89	0.87	0.90	0.89	0.90	0.88	0.91
(4,4)	0.90	0.87	0.84	0.86	0.91	0.90	0.88	0.90	0.90	0.90	0.88	0.91
(8,8)	0.90	0.87	0.85	0.85	0.91	0.90	0.88	0.90	0.91	0.90	0.88	0.91

Notes. See Notes in Table F4.