Supplementary Materials for “Principal Component Analysis for a Mix of Stationary and Nonstationary Variables”

James D. Hamilton*

Xinwei Ma†

Jin Xi‡

April 8, 2026

Abstract

This supplementary material provides proofs of the results in the main paper and their extensions, discusses additional methodological and technical results, and reports additional simulation evidence.

*Email: jhamilton@ucsd.edu
†Email: x1ma@ucsd.edu
‡Email: xijin@amss.ac.cn

A Proofs

A.1 Proof of Theorem 1

Recall that the operator norm of a symmetric matrix S is defined by

$$\|S\|_{op} = \sup_{\gamma \in \Gamma} \frac{\gamma' S \gamma}{\gamma' \gamma}.$$

When S is positive semidefinite, this is the largest eigenvalue: $\|S\|_{op} = \varrho_{\max}(S)$. Since the trace of S (the sum of the diagonal elements) is equal to the sum of the eigenvalues, it follows immediately that for any symmetric positive semidefinite matrix S,

$$\sup_{\gamma \in \Gamma} \frac{\gamma' S \gamma}{\gamma' \gamma} \leq \sum_{i=1}^{N} s_{ii}. \tag{A-1}$$

Proof of Theorem 1(i)-(ii)

Note first that

$$\begin{aligned}
& \left\| (NT)^{-1} \sum_{t=1}^{T} \hat{C}_t \hat{C}_t' - (NT)^{-1} \sum_{t=1}^{T} C_t C_t' \right\|_{\text{op}} \\
= \;& \sup_{\gamma \in \Gamma} \frac{\left| \gamma' \left((NT)^{-1} \sum_{t=1}^{T} \hat{C}_t \hat{C}_t' - (NT)^{-1} \sum_{t=1}^{T} C_t C_t' \right) \gamma \right|}{\gamma' \gamma} \\
= \;& \sup_{\gamma \in \Gamma} \left| (N^2 T)^{-1} \gamma' \sum_{t=1}^{T} \hat{C}_t \hat{C}_t' \gamma - (N^2 T)^{-1} \gamma' \sum_{t=1}^{T} C_t C_t' \gamma \right| \\
\leq \;& \sup_{\gamma \in \Gamma} (N^2 T)^{-1} \gamma' \sum_{t=1}^{T} \hat{V}_t \hat{V}_t' \gamma + 2 \sup_{\gamma \in \Gamma} (N^2 T)^{-1} \left| \gamma' \sum_{t=1}^{T} C_t \hat{V}_t' \gamma \right|,
\end{aligned} \tag{A-2}$$

where the last line uses the relation $\hat{C}_t = C_t + \hat{V}_t$. We first show that the two terms in the last line of (A-2) converge in probability to 0. To show this for the first term, notice from (A-1) that

$$\begin{aligned}
\sup_{\gamma \in \Gamma} (N^2 T)^{-1} \gamma' \sum_{t=1}^{T} \hat{V}_t \hat{V}_t' \gamma &= \sup_{\gamma \in \Gamma} (NT)^{-1} \frac{\gamma' \sum_{t=1}^{T} \hat{V}_t \hat{V}_t' \gamma}{\gamma' \gamma} \\
&\leq (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{v}_{it}^2 \; \leq \; \max_{1 \leq i \leq N} T^{-1} \sum_{t=1}^{T} \hat{v}_{it}^2,
\end{aligned} \tag{A-3}$$

which tends to zero in probability by Assumption 3(ii).

For the last term in (A-2),

$$\sup_{\gamma\in\Gamma}\left|(N^2T)^{-1}\gamma'\textstyle\sum_{t=1}^T C_t\hat{V}_t'\gamma\right| \leq \left[\sup_{\gamma\in\Gamma}(N^2T)^{-1}\gamma'\textstyle\sum_{t=1}^T C_tC_t'\gamma\right]^{1/2}\left[\sup_{\gamma\in\Gamma}(N^2T)^{-1}\gamma'\textstyle\sum_{t=1}^T \hat{V}_t\hat{V}_t'\gamma\right]^{1/2}.$$

The first term converges in probability to a number no larger than $\omega_{11}^{1/2}$ from result (R8) in Stock and Watson (2002) and the second converges in probability to 0 from (A-3). Thus

$$\sup_{\gamma\in\Gamma}\left|(N^2T)^{-1}\gamma'\textstyle\sum_{t=1}^T C_t\hat{V}_t'\gamma\right| \xrightarrow{p} 0.$$

We thus conclude from (A-2) that

$$\sup_{\gamma\in\Gamma}\left|(N^2T)^{-1}\gamma'\textstyle\sum_{t=1}^T \hat{C}_t\hat{C}_t'\gamma - (N^2T)^{-1}\gamma'\textstyle\sum_{t=1}^T C_tC_t'\gamma\right| \xrightarrow{p} 0, \tag{A-4}$$

or equivalently

$$\left\|(NT)^{-1}\textstyle\sum_{t=1}^T \hat{C}_t\hat{C}_t' - (NT)^{-1}\textstyle\sum_{t=1}^T C_tC_t'\right\|_{\text{op}} \xrightarrow{p} 0.$$

Recall that $T^{-1}\sum_{t=1}^T \hat{f}_{jt}^2$ and $T^{-1}\sum_{t=1}^T \tilde{f}_{jt}^2$ denote the jth largest eigenvalues of $(NT)^{-1}\sum_{t=1}^T \hat{C}_t\hat{C}_t'$ and $(NT)^{-1}\sum_{t=1}^T C_tC_t'$, respectively, and that $T^{-1}\sum_{t=1}^T \tilde{f}_{jt}^2$ converges in probability to ω_{ij}. Therefore, by Weyl's inequality on perturbation, it is true for all j that

$$|T^{-1}\textstyle\sum_{t=1}^T \hat{f}_{jt}^2 - T^{-1}\sum_{t=1}^T \tilde{f}_{jt}^2| \leq \left\|(NT)^{-1}\textstyle\sum_{t=1}^T \hat{C}_t\hat{C}_t' - (NT)^{-1}\sum_{t=1}^T C_tC_t'\right\|_{\text{op}},$$

which implies that $T^{-1}\sum_{t=1}^T \hat{f}_{jt}^2$ converges in probability to ω_{jj} for $j = 1, 2, \ldots, r$, and to 0 for $j = r+1, \ldots, k$.

Proof of Theorem 1(iii)

Notice that $(N^2T)^{-1}\hat{\Lambda}'\sum_{t=1}^T \hat{C}_t\hat{C}_t'\hat{\Lambda}$ is a diagonal matrix for all N and T by the definition of $\hat{\Lambda}$ with diagonal elements converging in probability to ω_{jj} by result (i):

$$(N^2T)^{-1}\hat{\Lambda}'\textstyle\sum_{t=1}^T \hat{C}_t\hat{C}_t'\hat{\Lambda} \xrightarrow{p} \Omega_{FF}. \tag{A-5}$$

Equation (A-4) then establishes that $(N^2T)^{-1}\hat{\Lambda}'\sum_{t=1}^T C_tC_t'\hat{\Lambda} \xrightarrow{p} \Omega_{FF}$. We also know from results (R2)-(R6) in Stock and Watson (2002) that

$$(N^2T)^{-1}\gamma'\sum_{t=1}^T C_tC_t'\gamma - (N^2T)^{-1}\gamma'\sum_{t=1}^T \Lambda'F_tF_t'\Lambda\gamma \xrightarrow{p} 0$$

where the convergence is uniform for $\gamma \in \Gamma$, meaning

$$\text{plim}\left\{(N^2T)^{-1}\hat{\Lambda}'\sum_{t=1}^T C_tC_t'\right\} = \text{plim}\left\{(N^2T)^{-1}\hat{\Lambda}'\sum_{t=1}^T \Lambda F_tF_t'\Lambda'\hat{\Lambda}\right\} = H\Omega_{FF}H' \tag{A-6}$$

for $H = \text{plim}\ (N^{-1}\hat{\Lambda}'\Lambda)$. Combining results (A-4)-(A-6),

$$\Omega_{FF} = H\Omega_{FF}H'. \tag{A-7}$$

Let $\hat{h}_j'$ denote the jth row of $\hat{\Lambda}'\Lambda/N$,

$$\underset{(1\times r)}{\hat{h}_j'} = \underset{(1\times N)}{\hat{\lambda}_j'}\underset{(N\times r)}{\Lambda}/N,$$

for $\hat{\lambda}_j'$ the jth row of $\hat{\Lambda}'$. Then

$$\hat{h}_j'\hat{h}_j = \frac{\hat{\lambda}_j'}{\sqrt{N}}\frac{\Lambda\Lambda'}{N}\frac{\hat{\lambda}_j}{\sqrt{N}}.$$

This is less than or equal to the largest eigenvalue of $\Lambda\Lambda'/N$, which converges to 1. Letting $h_j' = (h_{j1}, h_{j2}, ..., h_{jr})'$ denote the jth row of H, we thus have

$$\hat{h}_j'\hat{h}_j \xrightarrow{p} h_{j1}^2 + h_{j2}^2 + \cdots h_{jr}^2 \leq 1.$$

The (1,1) element of (A-7) states

$$h_1'\Omega_{FF}h_1 = h_{11}^2\omega_{11} + h_{12}^2\omega_{22} + \cdots + h_{1r}^2\omega_{rr} = \omega_{11}.$$

Since $\omega_{11} > \omega_{22} > \cdots > \omega_{rr} > 0$, this requires $h_{11}^2 = 1$ and $h_{12} = \cdots = h_{1r} = 0$. Thus the (1,1) element of $\hat{\Lambda}'\Lambda/N$ converges in probability to ± 1 and other elements of the first row converge to zero.

The (2,2) element of (A-7) states

$$h_{21}^2\omega_{11} + h_{22}^2\omega_{22} + \cdots + h_{2r}^2\omega_{rr} = \omega_{22} \tag{A-8}$$

where $h_{21} = \text{plim}\ \hat{\lambda}_2\lambda_1/N$. Regress λ_1 on $\hat{\lambda}_1$ with residual q_1:

$$\lambda_1 = \hat{k}_1\hat{\lambda}_1 + q_1 \tag{A-9}$$

$$\hat{k}_1 = (\hat{\lambda}_1'\hat{\lambda}_1/N)^{-1}(\hat{\lambda}_1'\lambda_1/N)$$

$$q_1'\hat{\lambda}_1 = 0$$

$$\lambda_1'\lambda_1/N = \hat{k}_1^2(\hat{\lambda}_1'\hat{\lambda}_1/N) + q_1'q_1/N.$$

We saw above that $\hat{k}_1^2 \xrightarrow{p} 1$, which along with $\lambda_1'\lambda_1/N \to 1$ and $\hat{\lambda}_1'\hat{\lambda}_1/N = 1$ establishes $q_1'q_1/N \xrightarrow{p} 0$. Premultiply (A-9) by $\hat{\lambda}_2'/N$:

$$\hat{\lambda}_2'\lambda_1/N = \hat{k}_1\hat{\lambda}_2'\hat{\lambda}_1/N + \hat{\lambda}_2'q_1/N = \hat{\lambda}_2'q_1/N.$$

But from Cauchy-Schwarz

$$(\hat{\lambda}_2'q_1/N)^2 \leq (\hat{\lambda}_2'\hat{\lambda}_2/N)(q_1'q_1/N) \xrightarrow{p} 0.$$

Thus $\hat{\lambda}_2'\lambda_1/N \xrightarrow{p} h_{21} = 0$ and (A-8) becomes

$$h_{22}^2\omega_{22} + h_{23}^2\omega_{33} + \cdots + h_{2r}^2\omega_{rr} = \omega_{22}.$$

Since $\omega_{22} > \omega_{33} > \cdots > \omega_{rr}$ and $h_{22}^2 + h_{23}^2 + \cdots + h_{2r}^2 \leq 1$, this requires $h_{22}^2 = 1$ and all other elements of the second row of H to be zero, establishing the second row of the claim in Theorem 1(*iii*). Proceeding iteratively through rows 3,4,..,r establishes the rest of the result in (*iii*).

Proof of Theorem 1(iv)

Write

$$\begin{aligned}\hat{\Xi}\hat{F}_t - F_t &= N^{-1}\hat{\Xi}\hat{\Lambda}'\hat{C}_t - F_t \\ &= N^{-1}\hat{\Xi}\hat{\Lambda}'(\Lambda F_t + e_t + \hat{V}_t) - F_t \\ &= (N^{-1}\hat{\Xi}\hat{\Lambda}'\Lambda - I_r)F_t + N^{-1}\hat{\Xi}\hat{\Lambda}'e_t + N^{-1}\hat{\Xi}\hat{\Lambda}'\hat{V}_t. \end{aligned} \tag{A-10}$$

The task is to show that all three terms in (A-10) have plim 0. That $(N^{-1}\hat{\Xi}\hat{\Lambda}'\Lambda - I_r)F_t \xrightarrow{p} 0$ follows immediately from result (*iii*). For the second term,

$$N^{-1}\hat{\Xi}\hat{\Lambda}'e_t = N^{-1}(\hat{\Xi}\hat{\Lambda}' - \Lambda')e_t + N^{-1}\hat{\Xi}\Lambda'e_t. \tag{A-11}$$

Consider the square of the jth element of the first term in (A-11):

$$\left[\frac{(\hat{\Xi}_j\hat{\lambda}_j' - \lambda_j')e_t}{N}\right]^2 \leq \left[\frac{(\hat{\Xi}_j\hat{\lambda}_j' - \lambda_j')(\hat{\Xi}_j\hat{\lambda}_j - \lambda_j)}{N}\right]\left[\frac{e_t'e_t}{N}\right]. \tag{A-12}$$

The first term in (A-12) is

$$\frac{(\hat{\Xi}_j\hat{\lambda}_j' - \lambda_j')(\hat{\Xi}_j\hat{\lambda}_j - \lambda_j)}{N} = \frac{\hat{\Xi}_j^2\hat{\lambda}_j'\hat{\lambda}_j}{N} - \frac{\lambda_j'\hat{\Xi}_j\hat{\lambda}_j}{N} - \frac{\hat{\Xi}_j\hat{\lambda}_j'\lambda_j}{N} + \frac{\lambda_j'\lambda_j}{N},$$

which converges in probability to zero by Theorem 1(*iii*). The second term in (A-12) is $O_p(1)$, by result (R1) in Stock and Watson (2002), meaning the plim of (A-12) is zero. The second term in (A-11) also converges in probability to zero as in Stock and Watson (2002) Result (R15). Hence $N^{-1}\hat{\Xi}\hat{\Lambda}'e_t \xrightarrow{p} 0$.

For the third term in (A-10), $N^{-1}\hat{\Xi}\hat{\Lambda}'\hat{V}_t$, note that the jth element is $N^{-1}\hat{\Xi}_j\hat{\lambda}_j'\hat{V}_t$ whose square is

$$N^{-2}\hat{\lambda}_j'\hat{V}_t\hat{V}_t'\hat{\lambda}_j \leq N^{-1}\textstyle\sum_{i=1}^N \hat{v}_{it}^2 \xrightarrow{p} 0 \tag{A-13}$$

due to our Assumption 3.

A.2 Proof of Theorem 2

We first verify Assumption 3(ii) that $\max_{1\le i\le N} T^{-1}\sum_{t=1}^{T}\hat{v}_{it}^2 \xrightarrow{p} 0$. Write (16) as $\sum_{t=1}^{T}\hat{v}_{it}^2 = q_{iT}'Q_{iT}^{-1}q_{iT}$ where

$$q_{iT} = \sum_{t=1}^{T} \Upsilon_{iT}^{-1} z_{i,t-h} e_{it} \text{ and } Q_{iT} = \sum_{t=1}^{T} \Upsilon_{iT}^{-1} z_{i,t-h} z_{i,t-h}' \Upsilon_{iT}^{-1}.$$

Then

$$\max_{1\le i\le N} \frac{1}{T}\sum_{t=1}^{T}\hat{v}_{it}^2 \le \left(\frac{1}{T}\max_{1\le i\le N}\|q_{iT}\|^2\right)\cdot\left(\min_{1\le i\le N}\varrho_{\min}(Q_{iT})\right)^{-1}. \tag{A-14}$$

By Condition 4(i) and a standard maximal inequality argument (see, for example, Lemma 2.2.2 of van der Vaart and Wellner 1996),

$$\frac{1}{T}\max_{1\le i\le N}\|q_{iT}\|^2 = O_p(N/T). \tag{A-15}$$

To control the minimum eigenvalue in (A-14), we employ Weyl's inequality on perturbation, which bounds the difference in minimum eigenvalues by the norm of the matrix difference:

$$\left|\varrho_{\min}(Q_{iT}) - \varrho_{\min}(Q_i)\right| \le \|Q_{iT} - Q_i\|.$$

An implication of this is that if $\min_{1\le i\le N}\varrho_{\min}(Q_{iT}) < \delta_N$ for some δ_N and $\max_{1\le i\le N}\|Q_{iT}-Q_i\| \le \mathfrak{c}_6(N)$, then $\min_{1\le i\le N}\varrho_{\min}(Q_i) < \delta_N + \mathfrak{c}_6(N)$. We can then use Condition 4(iii) to conclude

$$\begin{aligned}
Prob\Big(\min_{1\le i\le N}\varrho_{\min}(Q_{iT}) < \delta_N\Big) &= Prob\Big(\min_{1\le i\le N}\varrho_{\min}(Q_{iT}) < \delta_N \text{ and } \max_{1\le i\le N}\|Q_{iT}-Q_i\| > \mathfrak{c}_6(N)\Big)\\
&\quad + Prob\Big(\min_{1\le i\le N}\varrho_{\min}(Q_{iT}) < \delta_N \text{ and } \max_{1\le i\le N}\|Q_{iT}-Q_i\| \le \mathfrak{c}_6(N)\Big)\\
&\le Prob\Big(\max_{1\le i\le N}\|Q_{iT}-Q_i\| > \mathfrak{c}_6(N)\Big)\\
&\quad + Prob\Big(\min_{1\le i\le N}\varrho_{\min}(Q_i) < \delta_N + \mathfrak{c}_6(N)\Big)\\
&= o(1) + Prob\Big(\min_{1\le i\le N}\varrho_{\min}(Q_i) < \delta_N + \mathfrak{c}_6(N)\Big)\\
&\le o(1) + N\cdot\max_{1\le i\le N} Prob\Big(\varrho_{\min}(Q_i) < \delta_N + \mathfrak{c}_6(N)\Big).
\end{aligned}$$

Now set $\delta_N = C\log^{-\mathfrak{c}_4}(N)$ for some small constant C satisfying $\mathfrak{c}_3/(2C)^{1/\mathfrak{c}_4} > 1$, where $\mathfrak{c}_4$ is defined in Condition 4(ii). Since $\mathfrak{c}_6(N) = o(\delta_N)$, the above is further bounded by

$$
\begin{aligned}
Prob\Big(\min_{1\le i\le N}\varrho_{\min}(Q_{iT}) < C\log^{-\mathfrak{c}_4}(N)\Big) \le\ & o(1) + N\cdot \max_{1\le i\le N} Prob\Big(\varrho_{\min}(Q_i) < 2\delta_N\Big)\\
&\le o(1) + N\cdot \mathfrak{c}_2\exp\Big\{-\frac{\mathfrak{c}_3}{(2C\log^{-\mathfrak{c}_4}(N))^{1/\mathfrak{c}_4}}\Big\}\\
&= o(1) + \mathfrak{c}_2\exp\Big\{\Big(-\frac{\mathfrak{c}_3}{(2C)^{1/\mathfrak{c}_4}}+1\Big)\log(N)\Big\} \to 0.
\end{aligned}
$$

As a result, we have that

$$
\Big(\min_{1\le i\le N}\varrho_{\min}(Q_{iT})\Big)^{-1} = O_p\Big(\log^{\mathfrak{c}_4}(N)\Big). \tag{A-16}
$$

Results (A-14), (A-15), and (A-16) establish that Assumption 3(ii) follows from Assumption 4 provided that $N\log^{\mathfrak{c}_4}(N)/T = o(1)$.

We next verify Assumption 3(i) that $N^{-1}\sum_{i=1}^N \hat{v}_{it}^2 \xrightarrow{p} 0$ for each t. To start, we have

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^N \hat{v}_{it}^2 &\le \frac{1}{N}\sum_{i=1}^N\Big\{\Big\|\sqrt{T}\Upsilon_{iT}^{-1}z_{i,t-h}\Big\|^2\cdot\Big\|\Big(\sum_{t=1}^T\Upsilon_{iT}^{-1}z_{i,t-h}z_{i,t-h}'\Upsilon_{iT}^{-1}\Big)^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^T\Upsilon_{iT}^{-1}z_{i,t-h}c_{it}\Big)\Big\|^2\Big\}\\
&\le \Big(\max_{1\le i\le N}\Big\|Q_{iT}^{-1}\Big(\frac{1}{\sqrt{T}}q_{iT}\Big)\Big\|^2\Big)\Big(\frac{1}{N}\sum_{i=1}^N\|\sqrt{T}\Upsilon_{iT}^{-1}z_{i,t-h}\|^2\Big)\\
&\le \Big(\min_{1\le i\le N}\varrho_{\min}(Q_{iT})\Big)^{-2}\Big(\frac{1}{T}\max_{1\le i\le N}\|q_{iT}\|^2\Big)\Big(\frac{1}{N}\sum_{i=1}^N\|\sqrt{T}\Upsilon_{iT}^{-1}z_{i,t-h}\|^2\Big).
\end{aligned}
$$

From (A-15) and (A-16), we have

$$
\Big(\min_{1\le i\le N}\varrho_{\min}(Q_{iT})\Big)^{-2}\Big(\frac{1}{T}\max_{1\le i\le N}\|q_{iT}\|^2\Big) = O_p\Big(\frac{N\log^{2\mathfrak{c}_4}(N)}{T}\Big).
$$

Finally, Markov's inequality and Condition 4(i) together imply that

$$
\frac{1}{N}\sum_{i=1}^N\|\sqrt{T}\Upsilon_{iT}^{-1}z_{i,t-h}\|^2 = O_p\Big(\frac{1}{N}\sum_{i=1}^N E(\|\sqrt{T}\Upsilon_{iT}^{-1}z_{i,t-h}\|^2)\Big) = O_p(1),
$$

which concludes the proof.

A.3 Proof of Theorem 3

Recall that $u_{it} = (1-L)^{d_i} y_{it} - \mu_i = \sum_{\ell=0}^{\infty} \psi_{i\ell} \eta_{i,t-\ell}$. The asymptotic distribution is different depending on whether: (1) y_{it} is stationary ($d_i = 0$); (2) y_{it} has a unit root but no time trend ($d_i = 1$ and $\mu_i = 0$); (3) y_{it} has a unit root with time trend bounded away from zero ($d_i = 1$ and $\exists \mathfrak{c}_{12} > 0$ such that $|\mu_i| \geq \mathfrak{c}_{12}$ for all i). If there is a subset of observations for which μ_i is nonzero but extremely close to zero, as long as the number of such observations does not increase with N and T, the existence of such a $\mathfrak{c}_{12}$ is immediate. For completeness we also discuss the case when the number of observations with μ_i arbitrarily close to zero grows with T: (4) $\mu_i = \tilde{\mu}_i / \sqrt{T}$ for $\tilde{\mu}_i \neq 0$. For ease of exposition, we will partition $\{1, 2, \ldots, N\}$ into $\mathcal{N}_1$, $\mathcal{N}_2$, $\mathcal{N}_3$, and $\mathcal{N}_4$ corresponding to the four settings.

Case 1 (stationary y_{it})

In the stationary case, there is no need to introduce any additional scaling or rotation of the regressors, so $u_{it} = y_{it} - \mu_i$ and $\Upsilon_{iT}^{-1} = T^{-1/2} I_{p+1}$ where I_{p+1} is the identity matrix. For this case Assumption 5 guarantees that

$$Q_{iT} = \textstyle\sum_{t=1}^{T} \Upsilon_{iT}^{-1} z_{i,t-h} z_{i,t-h}' \Upsilon_{iT}^{-1} = T^{-1} \sum_{t=1}^{T} z_{i,t-h} z_{i,t-h}' \xrightarrow{p} E(z_{i,t-h} z_{i,t-h}') := Q_i.$$

Below we will verify Conditions 4(i–iii).

For the first part of 4(i), our Assumption 5 directly implies that $E(y_{it}^2)$ is uniformly bounded. To show the second part of 4(i), it suffices to bound the following terms individually:

$$E\Big(\Big|\frac{1}{\sqrt{T}} \sum_{t=1}^{T} c_{it}\Big|^2\Big) = var\Big(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} c_{it}\Big),$$
$$\text{and } E\Big(\Big|\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{i,t-l} c_{it}\Big|^2\Big) = var\Big(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{i,t-l} c_{it}\Big), \text{ for } l = h, \ldots, h+p-1.$$

As c_{it} is a linear combination of $u_{it}, u_{i,t-h}, \ldots, u_{i,t-h-p+1}$, the above can further be bounded by the following variances:

$$var\Big(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{it}\Big), \text{ and } var\Big(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{i,t-l} u_{i,t-l'}\Big),$$

for $l = h, \dots, h+p-1$ and $l' = 0, h, \dots, h+p-1$. Given our linear process specification, the above is bounded. See, for example, Appendix A of Shumway and Stoffer (2017) for an exact calculation. The same argument can be applied to show that $E(\|Q_{iT} - Q_i\|^2) = O(1/T)$.

Condition 4(ii) holds automatically given our assumption that the minimum eigenvalue of $\tilde{Q}_i$ is bounded away from zero (i.e., it is satisfied for any $\mathfrak{c}_4 > 0$).

Finally we verify 4(iii). We adopt Markov's inequality with a union bound, which implies that

$$\begin{aligned} Prob\left(\max_{i \in \mathcal{N}_0} \left\|Q_{iT} - Q_i\right\| > \sqrt{\frac{N \log^{1/2}(N)}{T}}\right) &\leq \sum_{i \in \mathcal{N}_0} Prob\left(\left\|Q_{iT} - Q_i\right\| > \sqrt{\frac{N \log^{1/2}(N)}{T}}\right) \\ &\leq C \frac{|\mathcal{N}_0|}{N} \frac{1}{\log^{1/2}(N)} \leq C \frac{1}{\log^{1/2}(N)}, \end{aligned}$$

where C is some constant and $|\mathcal{N}_0|$ is the cardinality of $\mathcal{N}_0$.

Case 2 (unit-root y_{it} with no time trend)

For these observations, we use $\Upsilon_{iT}^{-1} = \tilde{\Upsilon}_{iT}^{-1} R_i^{-1}$, where $\tilde{\Upsilon}_{iT}$ is a diagonal scaling matrix

$$\tilde{\Upsilon}_{iT} = \begin{bmatrix} \sqrt{T} I_p & 0 \\ 0 & T \end{bmatrix} \tag{A-17}$$

and R_i is a nonsingular rotation matrix such that $\Upsilon_{iT}^{-1} z_{i,t-h} = \tilde{\Upsilon}_{iT}^{-1} \tilde{z}_{i,t-h}$ with

$$\tilde{z}_{i,t-h} = (u_{i,t-h},\ u_{i,t-h-1},\ \dots,\ u_{i,t-h-p+2},\ 1,\ y_{i,t-h}).$$

Notice that if $p = 1$, we simply set $\tilde{z}_{i,t-h} = z_{i,t-h} = (1,\ y_{i,t-h})$ so no rotation of the regressors is needed. Assumption 5 guarantees that as in Hamilton (1994, eq. (17.7.18)),

$$Q_{iT} = \textstyle\sum_{t=1}^{T} \Upsilon_{iT}^{-1} z_{i,t-h} z_{i,t-h}' \Upsilon_{iT}^{-1} \xrightarrow{d} \begin{bmatrix} \tilde{Q}_i & 0 & 0 \\ 0 & 1 & \psi_i(1) \int_0^1 W_i(r) dr \\ 0 & \psi_i(1) \int_0^1 W_i(r) dr & \psi_i(1)^2 \int_0^1 W_i(r)^2 dr \end{bmatrix} := Q_i \tag{A-18}$$

where $\psi_i(1) = \sum_{l=0}^{\infty} \psi_{il}$, and $W_i(r)$ is a standard Brownian motion for each i.

For Condition 4(i), we first note that for any stationary component of $\tilde{z}_{i,t-h}$, the condition holds by standard variance calculation (see the proof of the stationary case). For the nonstationary component of $\tilde{z}_{i,t-h}$, namely $y_{i,t-h}$, we rewrite it as

$$y_{i,t-h} = \psi_i(1) \sum_{s=1-h}^{t-h} \eta_{is} - \tilde{\varepsilon}_{i,t-h} + \tilde{\varepsilon}_{i,-h}, \text{ where } \tilde{\varepsilon}_{i,t-h} = \sum_{l=0}^{\infty} \tilde{\psi}_{il} \eta_{i,t-h-l} \text{ and } \tilde{\psi}_{il} = \sum_{l'=l+1}^{\infty} \psi_{il'}; \quad \text{(A-19)}$$

see for example Hamilton (1994, eq. (17.5.3)). Thus to bound moments of $T^{-1} \sum_{t=1}^{T} y_{i,t-h} u_{it}$ it is sufficient to note that

$$E\left(\left| \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{s=1-h}^{t-h} \eta_{is} \right) \left(\sum_{l=0}^{\infty} \psi_{il} \eta_{i,t-l} \right) \right|^2 \right) = O(1),$$

where the bound is uniform for $i \in \mathcal{N}_1$.

We next verify Condition 4(ii). Assumption 5(iv) bounds the smallest eigenvalue of the upper-left block of (A-18), while Assumption 5(ii) allows us to ignore the additional scaling $\psi_i(1)$ for the Brownian motion. Thus it is sufficient to establish a bound on the smallest eigenvalue of the (2×2) matrix

$$Q_i = \begin{bmatrix} 1 & \int_0^1 W_i(r)dr \\ \int_0^1 W_i(r)dr & \int_0^1 W_i(r)^2 dr \end{bmatrix}$$

as a slight abuse of notation.

To provide a probabilistic bound on its minimum eigenvalue, notice that the determinant of Q_i is given by

$$|Q_i| = \int_0^1 \left(W_i(r) - \int_0^1 W_i(s)ds \right)^2 dr = \varrho_{\min}(Q_i) \varrho_{\max}(Q_i)$$

while the norm $\|Q_i\| \geq \varrho_{\max}(Q_i)$. This means

$$\varrho_{\min}(Q_i) \geq \|Q_i\|^{-1} \int_0^1 \left(W_i(r) - \int_0^1 W_i(s)ds \right)^2 dr.$$

Beghin et al. (2005) showed in their Section 3 that

$$Prob\left(\int_0^1 \left(W_i(r) - \int_0^1 W_i(s)ds\right)^2 dr < \epsilon\right) = O(\exp\{-C\epsilon^{-1}\}).$$

To bound the matrix norm $\|Q_i\|$, we notice that the supremum of a Brownian motion is sub-Gaussian by the reflection principle (see, for example, Çinlar 2011, Chapter VIII, 1.21), and therefore the entries of Q_i are at most sub-exponential, which leads to the following:

$$Prob\left(\|Q_i\| > \frac{1}{\epsilon}\right) = O(\exp\{-C\epsilon^{-1}\}).$$

As a result, Condition 4(ii) holds with $\mathfrak{c}_4 = 2$.

Finally, we verify Condition 4(iii) by establishing a uniform distributional approximation result. We can write

$$Q_{iT} = \sum_{t=1}^{T} \Upsilon_{iT}^{-1} z_{i,t-h} z'_{i,t-h} \Upsilon_{iT}^{-1} = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} x_{i,t-h}x'_{i,t-h} & \frac{1}{T}\sum_{t=1}^{T} x_{i,t-h} & \frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{i,t-h}x_{i,t-h} \\ \frac{1}{T}\sum_{t=1}^{T} x'_{i,t-h} & 1 & \frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{it} \\ \frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{i,t-h}x'_{i,t-h} & \frac{1}{T^{3/2}}\sum_{t=1}^{T} y_{i,t-h} & \frac{1}{T^2}\sum_{t=1}^{T} y_{i,t-h}^2 \end{bmatrix},$$

where $x_{i,t-h} = (u_{i,t-h}, u_{i,t-h-1}, \ldots, u_{i,t-h-p+2})$. We will only illustrate convergence of the bottom-right 2×2 block of Q_{iT}, as the other entries converge in probability, which can be demonstrated by standard variance calculation.

To start, we recall that the decomposition in (A-19) expresses y_{it} as a sum of independent components and two remainder terms. Our assumption implies that $\tilde{\psi}_{il}$ are absolutely summable, from which we conclude that $\tilde{\epsilon}_{it}$ have bounded fourth moments. As a result,

$$\max_{i\in\mathcal{N}_1}\max_{1\le t\le T}\left|\tilde{\epsilon}_{i,t-h}\right| = O_p((NT)^{1/4}) = o_p\left(\frac{\sqrt{T}}{\log^3(N)}\right).$$

In the above, the last step follows from the condition $N\log^{12}(N)/T \to 0$. We next employ Theorem 4 of Komlós et al. (1976) with a union bound: setting $x = \sqrt{T}/\log^3(N)$ and $H(x) = x^4$ in their theorem, we have

$$Prob\left(\max_{i\in\mathcal{N}_1}\max_{1\le t\le T}\left|y^{\star}_{i,t-h}-\tilde{W}_i(t)\right|>\frac{\sqrt{T}}{\log^3(N)}\right)=O\left(\frac{N\log^{12}(N)}{T}\right)+o(1)=o(1). \tag{A-20}$$

Here, $\tilde{W}_i(t)$ and $y^{\star}_{it}$ are defined on a possibly new probability space, such that $\tilde{W}_i(t)$ is a standard Brownian motion scaled by $\psi_i(1)$, and that $y^{\star}_{it}$ and the original y_{it} have the same distribution. To improve readability, however, we slightly abuse notation and do not distinguish $y^{\star}_{it}$ and y_{it}. Given this theoretical device, we will show below that under the condition $N\log^{12}(N)/T\to 0$, the following hold:

$$Prob\left(\max_{i\in\mathcal{N}_1}\left|\frac{1}{T}\sum_{t=1}^{T}\frac{y_{i,t-h}}{\sqrt{T}}-\int_0^1\tilde{W}_i(t)dt\right|>\frac{1}{\log^3(N)}\right)=o(1) \tag{A-21}$$

$$Prob\left(\max_{i\in\mathcal{N}_1}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_{i,t-h}}{\sqrt{T}}\right)^2-\int_0^1\tilde{W}_i(t)^2dt\right|>\frac{1}{\log^{9/4}(N)}\right)=o(1). \tag{A-22}$$

Since $\log^3(N)>\log^{9/4}(N)$, we conclude that if we set

$$\mathfrak{c}_6(N)\propto\frac{1}{\log^{9/4}(N)}=o\left(\frac{1}{\log^{\mathfrak{c}_4}(N)}\right)\text{ with }\mathfrak{c}_4=2,$$

then Condition 4(iii) would hold.

We first prove (A-21). Write

$$\max_{i\in\mathcal{N}_1}\left|\frac{1}{T}\sum_{t=1}^{T}\frac{y_{i,t-h}}{\sqrt{T}}-\int_0^1\tilde{W}_i(t)dt\right|\le\underbrace{\max_{i\in\mathcal{N}_1}\left|\frac{1}{T}\sum_{t=1}^{T}\frac{y_{i,t-h}}{\sqrt{T}}-\frac{1}{T}\sum_{t=1}^{T}\tilde{W}_i(t/T)\right|}_{\text{(I)}}+\underbrace{\max_{i\in\mathcal{N}_1}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{W}_i(t/T)-\int_0^1\tilde{W}_i(t)dt\right|}_{\text{(II)}}.$$

A bound for (I) directly follows from our earlier discussion, which implies

$$Prob\left(\text{(I)}>\frac{1}{\log^3(N)}\right)=o(1).$$

The term (II) is simply the approximation error of the Riemann sum of the Brownian path, which satisfies

$$\text{(II)} \le \max_{i\in\mathcal{N}_1} \frac{1}{T}\sum_{t=1}^{T} \sup_{\frac{t}{T}\le s<\frac{t+1}{T}} |\tilde{W}_i(t/T) - \tilde{W}_i(s)|.$$

To provide a probabilistic bound, we notice that the supremum $\sqrt{T}\sup_{\frac{t}{T}\le s<\frac{t+1}{T}} |\tilde{W}_i(t/T) - \tilde{W}_i(s)|$ is sub-Gaussian by the reflection principal. Therefore, by a maximal inequality argument, one has

$$Prob\bigg(\text{(II)} > \frac{\log(N)}{\sqrt{T}}\bigg) = o(1).$$

The conclusion (A-21) then follows from the rates we established for (I) and (II) as $\log(N)/\sqrt{T}$ is negligible compared to $\log^{-3}(N)$.

We now show (A-22). We first write

$$\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_{i,t-h}}{\sqrt{T}}\right)^2 - \int_0^1 \tilde{W}_i(t)^2 dt\right| \le \left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_{i,t-h}}{\sqrt{T}}\right)^2 - \frac{1}{T}\sum_{t=1}^{T}\tilde{W}_i(t/T)^2\right| + \left|\frac{1}{T}\sum_{t=1}^{T}\tilde{W}_i(t/T)^2 - \int_0^1 \tilde{W}_i(t)^2 dt\right|.$$

Next, we further bound the first term on the right-hand side by

$$\begin{aligned}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_{i,t-h}}{\sqrt{T}}\right)^2 - \frac{1}{T}\sum_{t=1}^{T}\tilde{W}_i(t/T)^2\right| &\le \frac{2}{T}\sum_{t=1}^{T}\left|\frac{y_{i,t-h}}{\sqrt{T}} - \tilde{W}_i(t/T)\right|\tilde{W}_i(t/T) + \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_{i,t-h}}{\sqrt{T}} - \tilde{W}_i(t/T)\right|^2 \\ &\le 2\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_{i,t-h}}{\sqrt{T}} - \tilde{W}_i(t/T)\right|^2}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\tilde{W}_i(t/T)^2} + \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_{i,t-h}}{\sqrt{T}} - \tilde{W}_i(t/T)\right|^2.\end{aligned}$$

Therefore,

$$\begin{aligned}\max_{i\in\mathcal{N}_1}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_{i,t-h}}{\sqrt{T}}\right)^2 - \int_0^1 \tilde{W}_i(t)^2 dt\right| &\le 2\underbrace{\sqrt{\max_{i\in\mathcal{N}_1}\frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_{i,t-h}}{\sqrt{T}} - \tilde{W}_i(t/T)\right|^2}}_{\text{(III)}}\underbrace{\sqrt{\max_{i\in\mathcal{N}_1}\frac{1}{T}\sum_{t=1}^{T}\tilde{W}_i(t/T)^2}}_{\text{(IV)}} \\ &+ \underbrace{\max_{i\in\mathcal{N}_1}\frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_{i,t-h}}{\sqrt{T}} - \tilde{W}_i(t/T)\right|^2}_{\text{(V)}} + \underbrace{\max_{i\in\mathcal{N}_1}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{W}_i(t/T)^2 - \int_0^1 \tilde{W}_i(t)^2 dt\right|}_{\text{(VI)}}.\end{aligned}$$

Clearly from (A-20),

$$Prob\left(\text{(III)} > \frac{1}{\log^3(N)}\right) = o(1), \quad Prob\left(\text{(V)} > \frac{1}{\log^6(N)}\right) = o(1).$$

By a maximal inequality for sub-Gaussian random variables, one has

$$Prob\left(\text{(IV)} > \log^{3/4}(N)\right) = o(1).$$

Term (VI) is again the approximation error of a Riemann sum, which is bounded by

$$\text{(VI)} \leq \max_{i\in\mathcal{N}_1} \frac{1}{T}\sum_{t=1}^{T} \sup_{\frac{t}{T}\leq s<\frac{t+1}{T}} |\tilde{W}_i(t/T)^2 - \tilde{W}_i(s)^2| \leq \max_{i\in\mathcal{N}_1}\left(\frac{1}{T}\sum_{t=1}^{T} \sup_{\frac{t}{T}\leq s<\frac{t+1}{T}} |\tilde{W}_i(t/T) - \tilde{W}_i(s)|\right)\left(2\sup_{1\leq s\leq T}|\tilde{W}_i(s)|\right),$$

and therefore it satisfies

$$Prob\left(\text{(VI)} > \frac{\log^2(N)}{\sqrt{T}}\right) = o(1).$$

Finally, we note that the rate for the product (III)·(IV) is $\log^{-9/4}(N)$, which dominates the rates for (V) and (VI).

Case 3 (unit-root y_{it} with linear time trend)

With a time trend, $\Upsilon_{iT}^{-1} = \tilde{\Upsilon}_{iT}^{-1} R_i^{-1}$ with

$$\tilde{\Upsilon}_{iT} = \begin{bmatrix} \sqrt{T} I_p & 0 \\ 0 & T^{3/2} \end{bmatrix}$$

and $\Upsilon_{iT}^{-1} z_{i,t-h} = \tilde{\Upsilon}_{iT}^{-1}\tilde{z}_{i,t-h}$ with $\tilde{z}_{i,t-h} = (u_{i,t-h},\, u_{i,t-h-1},\, \ldots,\, u_{i,t-h-p+2},\, 1,\, y_{i,t-h})$. Here

$$\textstyle\sum_{t=1}^{T} \Upsilon_{iT}^{-1} z_{i,t-h} z_{i,t-h}' \Upsilon_{iT}^{-1} \xrightarrow{p} \begin{bmatrix} \tilde{Q}_i & 0 & 0 \\ 0 & 1 & \mu_i/2 \\ 0 & \mu_i/2 & \mu_i^2/3 \end{bmatrix} := Q_i$$

To verify Condition 4(i), it suffices to compute $E((\sqrt{T}T^{-3/2})^2 y_{i,t-h}^2)$, which is uniformly bounded in i and t, and

$$E\left(\left|\frac{1}{T^{3/2}}\sum_{t=1}^{T}t\left(\sum_{l=1}^{\infty}\psi_{il}\eta_{i,t-l}\right)\right|^2\right)=\frac{1}{T^3}\sum_{t=1}^{T}\sum_{t'=1}^{T}tt'cov(u_{it},u_{it'})$$
$$\leq\frac{1}{T^2}\sum_{t=1}^{T}t\sum_{t'=1}^{T}cov(u_{it},u_{it'})\leq C\frac{1}{T^2}\sum_{t=1}^{T}t$$

for some constant C, making the above bounded.

Condition 4(ii) is satisfied by Assumption 5, and we take $\mathfrak{c}_4=1/2$.

Finally we check Condition 4(iii). The upper diagonal block of Q_{iT} converges to $\tilde{Q}_i$, which has been discussed in Case 1 in this proof. Take any stationary component $u_{i,t-l}$ of $x_{i,t-h}$, then we can consider

$$E\left(\left|\frac{1}{T^2}\sum_{t=1}^{T}y_{i,t-h}u_{i,t-l}\right|^2\right)\leq 2E\left(\left|\frac{1}{T^2}\sum_{t=1}^{T}(y_{i,t-h}-\mu_i t)u_{i,t-l}\right|^2\right)+2\mu_i^2E\left(\left|\frac{1}{T^2}\sum_{t=1}^{T}tu_{i,t-l}\right|^2\right).$$

The first term on the right-hand side is of order T^{-2} by our discussion in Case 2, and the second term is of order T^{-1} by Condition 4(i), which we verified earlier. Therefore, $T^{-2}\sum_{t=1}^{T}y_{i,t-h}u_{i,t-l}=O_p(T^{-1/2})$, where the bound is uniform for $i\in\mathcal{N}_2$. This establishes the convergence of too-right and bottom-left entries of Q_{iT}.

Now consider $T^{-2}\sum_{t=1}^{T}y_{i,t-h}$. We write it as the sum $T^{-2}\sum_{t=1}^{T}(y_{i,t-h}-\mu_i t)$ and $T^{-2}\sum_{t=1}^{T}\mu_i t$. Clearly the former is of order $T^{-1/2}$ by standard variance calculation.

Finally, we consider

$$\frac{1}{T^3}\sum_{t=1}^{T}y_{i,t-h}^2=\frac{1}{T^3}\sum_{t=1}^{T}\left((y_{i,t-h}-\mu_i t)+\mu_i t\right)^2=\frac{1}{T^3}\sum_{t=1}^{T}(y_{i,t-h}-\mu_i t)^2+\frac{2\mu_i}{T^3}\sum_{t=1}^{T}(y_{i,t-h}-\mu_i t)t+\frac{\mu_i^2}{T^3}\sum_{t=1}^{T}t^2.$$

The first term on the right-hand side is of order T^{-1} by mean calculation. The second term has order $T^{-1/2}$ by variance calculation.

Case 4 (unit root with time trend arbitrarily close to zero)

Finally we consider observations for which $\mu_i=\tilde{\mu}_i/\sqrt{T}$ for $\tilde{\mu}_i\neq 0$. For this case (A-19) becomes

$$y_{i,t-h}=\xi_{i,t-h}+\tilde{\mu}_i(t-h)/\sqrt{T}$$

for $\xi_{i,t-h} = \psi_i(1)\sum_{s=1-h}^{t-h}\eta_{is} - \tilde{e}_{i,t-h} + \tilde{e}_{i,-h}$. It turns out for this case we can use the same scaling matrix as in (A-17). The (3,2) element in (A-18) would then be

$$T^{-3/2}\sum_{t=1}^{T}\left[\xi_{i,t-h} + \tilde{\mu}_i(t-h)/\sqrt{T}\right] \xrightarrow{d} \psi_i(1)\int_0^1 W_i(r)dr + \tilde{\mu}_i/2.$$

The (3,3) element is

$$\begin{aligned} & T^{-2}\sum_{t=1}^{T}\left[\xi_{i,t-h} + \tilde{\mu}_i(t-h)/\sqrt{T}\right]^2 \\ = \; & T^{-2}\sum_{t=1}^{T}\xi_{i,t-h}^2 + 2T^{-5/2}\sum_{t=1}^{T}\xi_{i,t-h}\tilde{\mu}_i(t-h) + T^{-3}\sum_{t=1}^{T}\tilde{\mu}_i^2(t-h)^2 \\ & \xrightarrow{d} [\psi_i(1)]^2\int_0^1 [W_i(r)dr]^2 + 2\tilde{\mu}_i\psi_i(1)\int_0^1 rW_i(r)dr + \tilde{\mu}_i^2/3 \end{aligned}$$

using results (h), (i), and (k) in Proposition 17.3 in Hamilton (1994). Thus for Case 4, the bottom-right block of the limiting matrix Q_i takes the form

$$\begin{aligned} & \begin{bmatrix} 1 & \psi_i(1)\int_0^1 W_i(r)dr + \tilde{\mu}_i/2 \\ \psi_i(1)\int_0^1 W_i(r)dr + \tilde{\mu}_i/2 & \psi_i(1)^2\int_0^1 W_i(r)^2dr + 2\tilde{\mu}_i\psi_i(1)\int_0^1 rW_i(r)dr + \tilde{\mu}_i^2/3 \end{bmatrix} \\ = & \begin{bmatrix} 1 & \int_0^1 [\psi_i(1)W_i(r) + \tilde{\mu}_i r]\,dr \\ \int_0^1 [\psi_i(1)W_i(r) + \tilde{\mu}_i r]\,dr & \int_0^1 [\psi_i(1)W_i(r) + \tilde{\mu}_i r]^2\,dr \end{bmatrix}. \end{aligned}$$

The determinant of this (2×2) matrix is

$$\int_0^1\left[\psi_i(1)W_i(r) + \tilde{\mu}_i r - \int_0^1(\psi_i(1)W_i(s) + \tilde{\mu}_i s)ds\right]^2 dr.$$

Its left-tail behavior follows from L_2 small-ball probabilities of Brownian motions with drifts. Finally, note that all moment bounds established for earlier cases continue to hold.

A.4 Extension to the local-to-unity case

In this section we extend our uniform distributional approximation result to series that are local to unity. To be specific, let $\mathcal{I}_{\text{LTU}}$ be a collection of series where $i \in \mathcal{I}_{\text{LTU}}$ takes the form

$$y_{it} = \rho_{iT}y_{i,t-1} + u_{it}, \qquad \text{with } \rho_{iT} = 1 - \frac{c_i}{T},$$

and $c_i \in [0, \mathfrak{c}_{13}]$ for some constant $0 < \mathfrak{c}_{13} < \infty$, and u_{it} satisfies Assumption 5. As before, we define

$$Q_{iT} = \sum_{t=1}^{T} \Upsilon_{iT}^{-1} z_{i,t-h} z_{i,t-h}' \Upsilon_{iT}^{-1}$$

Our goal is to show that

$$Prob\left(\max_{i \in \mathcal{I}_{\text{LTU}}} \left\| Q_{iT} - Q_i(c_i) \right\| \le c_6(N) \right) \to 1,$$

and hence Assumption 4(iii) holds for the local-to-unity case. Throughout the proof we impose the same growth condition as in Appendix A.3, namely

$$\frac{N \log^{12}(N)}{T} \to 0.$$

We will set $c_6(N) \asymp \log^{-9/4} N$.

We first discuss the limiting matrix $Q_i(c_i)$. Define the continuous-time Brownian motion

$$\mathbb{W}_i(r) = \frac{\widetilde{W}_i(\lfloor Tr \rfloor)}{\sqrt{T}}, \qquad r \in [0, 1],$$

where $\lfloor \cdot \rfloor$ denotes the floor function, and $\widetilde{W}_i(\cdot)$ is the Gaussian process (scaled Brownian motion) coupled to the partial sums as in (A-20) of Appendix A.3. For each $c \in [0, \mathfrak{c}_{13}]$ consider the Ornstein–Uhlenbeck (OU) functional

$$J_{i,c}(r) = \int_0^r e^{-c(r-s)}\, d\mathbb{W}_i(s), \qquad r \in [0, 1].$$

We then let $Q_i(c)$ be the same limiting matrix as in Appendix A.3, Case 2, with the modification that every appearance of the Brownian path $\widetilde{W}_i(\cdot)$ is replaced by the OU path $J_{i,c}(\cdot)$. Concretely, in the bottom-right 2×2 block, the entries $\int_0^1 \widetilde{W}_i(r)\, dr$ and $\int_0^1 \widetilde{W}_i(r)^2\, dr$ are replaced by $\int_0^1 J_{i,c}(r)\, dr$ and $\int_0^1 J_{i,c}(r)^2\, dr$.

Let Υ_{iT} be the same scaling matrix as in the $d_i = 1$ case (the one used in Appendix A.3, Case 2). Under this normalization, Appendix A.3, Case 2 shows that it suffices to verify convergence of the bottom-right 2×2 block of Q_{iT}, because all other entries converge in probability by standard variance calculations (the argument is unchanged here since it only uses moment bounds implied by Assumption 5 and the same scaling). Hence it remains to establish uniform approximation of

$$\frac{1}{T}\sum_{t=1}^{T}\frac{y_{i,t-h}}{\sqrt{T}} \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_{i,t-h}}{\sqrt{T}}\right)^2$$

by the corresponding OU functionals $\int_0^1 J_{i,c_i}(r)\,dr$ and $\int_0^1 J_{i,c_i}(r)^2\,dr$.

Fix $i \in \mathcal{I}_{\text{LTU}}$ and write $\rho = \rho_{iT} = 1 - c_i/T$ to save notation. Iterating $y_{it} = \rho y_{i,t-1} + u_{it}$ gives

$$y_{it} = \rho^t y_{i0} + \sum_{s=1}^{t}\rho^{t-s}u_{is}.$$

Let $S_{it} = \sum_{s=1}^{t} u_{is}$ and set $S_{i0} = 0$. Since $u_{is} = S_{is} - S_{i,s-1}$, we can rewrite the convolution via summation by parts:

$$\begin{aligned}
\sum_{s=1}^{t}\rho^{t-s}u_{is} &= \sum_{s=1}^{t}\rho^{t-s}(S_{is} - S_{i,s-1}) = \sum_{s=1}^{t}\rho^{t-s}S_{is} - \sum_{s=1}^{t}\rho^{t-s}S_{i,s-1} \\
&= \sum_{s=1}^{t}\rho^{t-s}S_{is} - \sum_{r=0}^{t-1}\rho^{t-1-r}S_{ir} = S_{it} + \sum_{s=1}^{t-1}\left(\rho^{t-s} - \rho^{t-1-s}\right)S_{is} \\
&= S_{it} - (1-\rho)\sum_{s=1}^{t-1}\rho^{t-1-s}S_{is} = S_{it} - \frac{c_i}{T}\sum_{s=1}^{t-1}\rho^{t-1-s}S_{is}.
\end{aligned}$$

Combining results in the two displays above, one has that, for every $t \le T$,

$$\frac{y_{it}}{\sqrt{T}} = \rho^t\frac{y_{i0}}{\sqrt{T}} + \frac{S_{it}}{\sqrt{T}} - \frac{c_i}{T}\sum_{s=1}^{t-1}\rho^{t-1-s}\frac{S_{is}}{\sqrt{T}}.$$

Let $Y_{iT}(r) = y_{i,\lfloor Tr\rfloor}/\sqrt{T}$. Substituting $\mathbb{W}_{iT}$ for the scaled partial sums, we obtain the decomposition

$$Y_{iT}(r) = \rho^t\frac{y_{i0}}{\sqrt{T}} + \mathbb{W}_{iT}(r) - \frac{c_i}{T}\sum_{s=1}^{t-1}\rho^{t-1-s}\mathbb{W}_{iT}(s/T) + R_{iT}(r),$$

where the remainder is

$$R_{iT}(r) = \left(\frac{S_{it}}{\sqrt{T}} - \mathbb{W}_{iT}(r)\right) - \frac{c_i}{T}\sum_{s=1}^{t-1}\rho^{t-1-s}\left(\frac{S_{is}}{\sqrt{T}} - \mathbb{W}_{iT}(s/T)\right).$$

For each $i \in \mathcal{I}_{\text{LTU}}$, define the partial sums

$$S_{it} = \sum_{s=1}^{t} u_{is}, \qquad S_{i0} = 0.$$

Appendix A.3 establishes a uniform strong approximation (their display (A-20)). Applying that result to the partial sums $S_{i,t-h}$, there exist Gaussian processes $\{\widetilde{W}_i(t)\}_{t=0}^{T}$ such that

$$Prob\left(\max_{i \in \mathcal{I}_{\text{LTU}}} \max_{1 \le t \le T} \left| S_{i,t-h} - \widetilde{W}_i(t) \right| > \frac{\sqrt{T}}{\log^3(N)} \right) = o(1).$$

Define the event

$$\mathcal{E}_T = \left\{ \max_{i \in \mathcal{I}_{\text{LTU}}} \max_{1 \le t \le T} \left| S_{i,t-h} - \widetilde{W}_i(t) \right| \le \frac{\sqrt{T}}{\log^3(N)} \right\},$$

then $Prob(\mathcal{E}_T) \to 1$. On $\mathcal{E}_T$,

$$\sup_{i \in \mathcal{I}_{\text{LTU}}} \sup_{0 \le r \le 1} \left| \frac{S_{i,\lfloor Tr \rfloor}}{\sqrt{T}} - \mathbb{W}_{iT}(r) \right| \le \frac{1}{\log^3(N)}.$$

This further implies that for all i and all $r \in [0,1]$,

$$\begin{aligned}
|R_{iT}(r)| &\le \sup_{0 \le r \le 1} \left| \frac{S_{i,\lfloor Tr \rfloor}}{\sqrt{T}} - \mathbb{W}_{iT}(r) \right| + \frac{c_i}{T} \sum_{s=1}^{t-1} \rho^{t-1-s} \sup_{1 \le s \le T} \left| \frac{S_{is}}{\sqrt{T}} - \mathbb{W}_{iT}(s/T) \right| \\
&\le \frac{1}{\log^3(N)} + \left(\frac{c_i}{T} \cdot T \right) \frac{1}{\log^3(N)} \le \frac{1 + \mathfrak{c}_{13}}{\log^3(N)}.
\end{aligned}$$

In other words, on the event $\mathcal{E}_T$,

$$\max_{i \in \mathcal{I}_{\text{LTU}}} \sup_{0 \le r \le 1} \left| Y_{iT}(r) - \left\{ \rho^{\lfloor Tr \rfloor} \frac{y_{i0}}{\sqrt{T}} + \mathbb{W}_{iT}(r) - \frac{c_i}{T} \sum_{s=1}^{\lfloor Tr \rfloor - 1} \rho^{\lfloor Tr \rfloor - 1 - s} \mathbb{W}_{iT}(s/T) \right\} \right| \le \frac{1 + \mathfrak{c}_{13}}{\log^3(N)}.$$

To connect the discrete representation to an OU functional, we now embed the Gaussian increments into a continuous-time Brownian motion. Let $B_i(\cdot)$ be a standard Brownian motion on $[0,1]$ such that

$$B_i(s/T) = \frac{\widetilde{W}_i(s)}{\sqrt{T}}, \qquad s = 0, 1, \ldots, T.$$

(For example, one may take B_i to be the continuous-time version of the Gaussian partial sums

$\widetilde{W}_i(\cdot)/\sqrt{T}$ with independent increments on the grid, and extend it to $[0,1]$ by adding independent Brownian bridges on each interval $[s/T, (s+1)/T]$.)

Define the modulus term

$$\omega_{iT} = \sup_{0\leq r\leq 1} \Big| B_i(r) - B_i(\lfloor Tr \rfloor / T) \Big|.$$

By the proof in Appendix A.3, we have

$$Prob\left(\omega_{iT} > \frac{\log(N)}{\sqrt{T}} \right) = o(1). \tag{A-23}$$

For each $c \in [0, \mathfrak{c}_{13}]$, define the OU process driven by B_i:

$$J_{i,c}(r) = \int_0^r e^{-c(r-s)}\, dB_i(s), \qquad r \in [0,1].$$

By Itô integration by parts,

$$J_{i,c}(r) = B_i(r) - c \int_0^r e^{-c(r-s)} B_i(s)\, ds.$$

Then the discrete OU term can be compared to $J_{i,c_i}(r)$ by adding and subtracting B_i:

$$\begin{aligned}
&\sup_{0\leq r\leq 1} \left| \mathbb{W}_{iT}(r) - \frac{c_i}{T} \sum_{s=1}^{t-1} \rho^{t-1-s} \mathbb{W}_{iT}(s/T) - \left\{ B_i(r) - c_i \int_0^r e^{-c_i(r-u)} B_i(u)\, du \right\} \right| \\
&\leq \sup_{0\leq r\leq 1} |\mathbb{W}_{iT}(r) - B_i(r)| + c_i \sup_{0\leq r\leq 1} \left| \frac{1}{T} \sum_{s=1}^{t-1} \rho^{t-1-s} \mathbb{W}_{iT}(s/T) - \int_0^r e^{-c_i(r-u)} B_i(u)\, du \right|
\end{aligned}$$

The first term is $O_p(\sqrt{\log(N)/T})$ by (A-23). For the second term on the right-hand side of the above display, fix $i \in \mathcal{I}_{\text{LTU}}$ and $r \in [0,1]$, we obtain

$$
\begin{aligned}
&\left| \frac{1}{T} \sum_{s=1}^{t-1} \rho^{t-1-s} \mathbb{W}_{iT}(s/T) - \int_0^r e^{-c_i(r-u)} B_i(u)\, du \right| \\
&\leq \underbrace{\left| \frac{1}{T} \sum_{s=1}^{t-1} \rho^{t-1-s} \Big(\mathbb{W}_{iT}(s/T) - B_i(s/T) \Big) \right|}_{=:A_{iT}(r)} + \underbrace{\left| \frac{1}{T} \sum_{s=1}^{t-1} \rho^{t-1-s} B_i(s/T) - \frac{1}{T} \sum_{s=1}^{t-1} e^{-c_i(r-s/T)} B_i(s/T) \right|}_{=:B_{iT}(r)} \\
&\quad + \underbrace{\left| \frac{1}{T} \sum_{s=1}^{t-1} e^{-c_i(r-u_s)} B_i(s/T) - \int_0^r e^{-c_i(r-u)} B_i(u)\, du \right|}_{=:C_{iT}(r)}.
\end{aligned}
\tag{A-24}
$$

We next bound the three terms uniformly in $r \in [0,1]$.

Bound for $A_{iT}(r)$: Since $0 < \rho < 1$, we have

$$
\sup_{0 \leq r \leq 1} A_{iT}(r) \leq \left(\frac{1}{T} \sum_{s=1}^{t-1} \rho^{t-1-s} \right) \cdot \sup_{0 \leq u \leq 1} \Big| \mathbb{W}_{iT}(u) - B_i(u) \Big| \leq \omega_{iT},
$$

where $\omega_{iT} = \sup_{0 \leq u \leq 1} |\mathbb{W}_{iT}(u) - B_i(u)|$. Hence, by (A-23),

$$
Prob\left(\sup_{0 \leq r \leq 1} A_{iT}(r) > \frac{\log(N)}{\sqrt{T}} \right) = o(1). \tag{A-25}
$$

Bound for $B_{iT}(r)$: For each $r \in [0,1]$ and $s \leq t-1$, write $k = t-1-s \geq 0$. Then

$$
\rho^{t-1-s} = (1 - c_i/T)^k = \exp\Big(k \log(1 - c_i/T) \Big) = \exp\Big(-\frac{c_i k}{T} + R_{kT} \Big),
$$

where $|R_{kT}| \leq C\, k/T^2$ for a constant C depending only on $\mathfrak{c}_{13}$ (by the Taylor expansion of $\log(1-x)$ and $c_i \in [0, \mathfrak{c}_{13}]$). Since $k \leq T$, this implies

$$
\sup_{0 \leq r \leq 1} \max_{1 \leq s \leq t-1} \left| \rho^{t-1-s} - e^{-c_i(r-s/T)} \right| \leq \frac{C'}{T}, \tag{A-26}
$$

with a constant C' depending only on $\mathfrak{c}_{13}$. Therefore,

$$
\sup_{0 \leq r \leq 1} B_{iT}(r) \leq \frac{C'}{T} \cdot \frac{1}{T} \sum_{s=1}^{T} \Big| B_i(s/T) \Big| \leq \frac{C'}{T} \sup_{0 \leq u \leq 1} |B_i(u)|. \tag{A-27}
$$

In particular, $\sup_{0\le u\le 1}|B_i(u)| = O_p(1)$, so $\sup_r B_{iT}(r) = O_p(1/T) = o_P(1)$ uniformly over i.

Bound for $C_{iT}(r)$: The term $C_{iT}(r)$ is the Riemann-sum approximation error for the integral $\int_0^r e^{-c_i(r-u)}B_i(u)\,du$. Since $c_i \in [0, \mathfrak{c}_{13}]$, the weight $u \mapsto e^{-c_i(r-u)}$ is uniformly bounded and uniformly Lipschitz on $[0,1]$. Thus, the proof that bounds the Brownian Riemann-sum error applies here after inserting the bounded weight $e^{-c_i(r-u)}$. Consequently,

$$Prob\left(\max_{i\in\mathcal{I}_{\text{LTU}}}\sup_{0\le r\le 1} C_{iT}(r) > \frac{\log(N)}{\sqrt{T}}\right) = o(1). \tag{A-28}$$

Combining (A-24)–(A-28), we obtain

$$Prob\left(\max_{i\in\mathcal{I}_{\text{LTU}}}\sup_{0\le r\le 1}\left|\frac{1}{T}\sum_{s=1}^{\lfloor Tr\rfloor-1}\rho^{\lfloor Tr\rfloor-1-s}\mathbb{W}_{iT}(s/T) - \int_0^r e^{-c_i(r-u)}B_i(u)\,du\right| > \frac{C'\log(N)}{\sqrt{T}}\right) = o(1)$$

for a constant C' depending only on $\mathfrak{c}_{13}$. Finally, combining the bounds of the three parts,

$$\begin{aligned}\max_{i\in\mathcal{I}_{\text{LTU}}}\sup_{0\le r\le 1}|Y_{iT}(r) - J_{i,c_i}(r)| =& O_p\left(\frac{1}{\log^3(N)} + \frac{\log(N)}{\sqrt{T}} + \sqrt{\frac{\log(NT)}{T}}\right) + \max_{i\in\mathcal{I}_{\text{LTU}}}\left|\frac{y_{i0}}{\sqrt{T}}\right| \\ =& O_p(1/\log^3(N)),\end{aligned}$$

which concludes the proof.

B Data appendix

A balanced panel was created from the 126 variables in the 2024:12 dataset by: using only data over 1960:1-2024:9; dropping the Michigan Survey of Consumer Sentiment (UMCSENT), trade-weighted exchange rate (TWEXAFEGSMTH), and new orders for consumer goods (ACOGNO) and nondefense capital goods (ANDENO), which are the same four series dropped by McCracken and Ng to create a balanced panel from the 2015:4 dataset; dropping the VIX (VIXCLS), which was not included in the 2015:4 dataset and whose first value is July 1962; and dropping the financial commercial paper rate (CP3M) and the commercial paper-fed funds spread (COMPAPFF) which were not reported for April 2020. The particular variables used in our analysis of the 2023 vintage dataset are described in Table B1.

Table B1: R^2 for each variable explained by first and second principal components, 1962:3 to 2023:6

Group 1. Output and income

Index	FRED	Description	PC1	PC1&2
1	RPI	Real Personal Income	0.23	0.40
2	W875RX1	Real personal income ex transfer receipts	0.61	0.75
6	INDPRO	IP Index	0.77	0.85
7	IPFPNSS	IP: Final Products and Nonindustrial Supplies	0.81	0.88
8	IPFINAL	IP: Final Products (Market Group)	0.78	0.82
9	IPCONGD	IP: Consumer Goods	0.52	0.83
10	IPDCONGD	IP: Durable Consumer Goods	0.49	0.82
11	IPNCONGD	IP: Nondurable Consumer Goods	0.41	0.55
12	IPBUSEQ	IP: Business Equipment	0.71	0.71
13	IPMAT	IP: Materials	0.65	0.73
14	IPDMAT	IP: Durable Materials	0.64	0.75
15	IPNMAT	IP: Nondurable Materials	0.62	0.69
16	IPMANSICS	IP: Manufacturing (SIC)	0.77	0.87
17	IPB51222S	IP: Residential Utilities	0.02	0.03
18	IPFUELS	IP: Fuels	0.07	0.07
19	CUMFNS	Capacity Utilization: Manufacturing	0.68	0.73
		Median	0.63	0.74

Table B1 (continued)

Group 2. Labor market

Index	FRED	Description	PC1	PC1&2
20	HWI	Help-Wanted Index for United States	0.55	0.56
21	HWIURATIO	Ratio of Help Wanted/No. Unemployed	0.55	0.55
22	CLF16OV	Civilian Labor Force	0.25	0.34
23	CE16OV	Civilian Employment	0.75	0.75
24	UNRATE	Civilian Unemployment Rate	0.69	0.71
25	UEMPMEAN	Average Duration of Unemployment (Weeks)	0.24	0.24
26	UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	0.38	0.46
27	UEMP5TO14	Civilians Unemployed for 5-14 Weeks	0.65	0.68
28	UEMP15OV	Civilians Unemployed - 15 Weeks and Over	0.66	0.66
29	UEMP15T26	Civilians Unemployed for 15-26 Weeks	0.65	0.66
30	UEMP27OV	Civilians Unemployed for 27 Weeks and Over	0.59	0.59
31	CLAIMSx	Initial Claims	0.42	0.45
32	PAYEMS	All Employees: Total nonfarm	0.81	0.81
33	USGOOD	All Employees: Goods-Producing Industries	0.85	0.85
34	CES1021000001	All Employees: Mining and Logging: Mining	0.03	0.41
35	USCONS	All Employees: Construction	0.67	0.74
36	MANEMP	All Employees: Manufacturing	0.74	0.74
37	DMANEMP	All Employees: Durable goods	0.77	0.77
38	NDMANEMP	All Employees: Nondurable goods	0.52	0.53
39	SRVPRD	All Employees: Service-Providing Industries	0.67	0.68
40	USTPU	All Employees: Trade, Transportation and Utilities	0.80	0.80
41	USWTRADE	All Employees: Wholesale Trade	0.74	0.80
42	USTRADE	All Employees: Retail Trade	0.67	0.68
43	USFIRE	All Employees: Financial Activities	0.42	0.43
44	USGOVT	All Employees: Government	0.08	0.09
45	CES0600000007	Avg Weekly Hours : Goods-Producing	0.33	0.48
46	AWOTMAN	Avg Weekly Overtime Hours : Manufacturing	0.38	0.62
47	AWHMAN	Avg Weekly Hours : Manufacturing	0.31	0.53
115	CES0600000008	Avg Hourly Earnings : Goods-Producing	0.04	0.42
116	CES2000000008	Avg Hourly Earnings : Construction	0.00	0.34
117	CES3000000008	Avg Hourly Earnings : Manufacturing	0.02	0.33
		Median	0.55	0.59

Group 3. Housing

Index	FRED	Description	PC1	PC1&2
48	HOUST	Housing Starts: Total New Privately Owned	0.14	0.37
49	HOUSTNE	Housing Starts, Northeast	0.17	0.39
50	HOUSTMW	Housing Starts, Midwest	0.11	0.42
51	HOUSTS	Housing Starts, South	0.12	0.28
52	HOUSTW	Housing Starts, West	0.11	0.26
53	PERMIT	New Private Housing Permits (SAAR)	0.11	0.35
54	PERMITNE	New Private Housing Permits, Northeast (SAAR)	0.13	0.43
55	PERMITMW	New Private Housing Permits, Midwest (SAAR)	0.10	0.47
56	PERMITS	New Private Housing Permits, South (SAAR)	0.08	0.30
57	PERMITW	New Private Housing Permits, West (SAAR)	0.09	0.24
		Median	0.11	0.36

Table B1 (continued)

Group 4. Consumption, orders, and inventories

Index	FRED	Description	PC1	PC1&2
3	DPCERA3M086SBEA	Real personal consumption expenditures	0.54	0.77
4	CMRMTSPLx	Real Manu. and Trade Industries Sales	0.73	0.89
5	RETAILx	Retail and Food Services Sales	0.46	0.50
58	AMDMNOx	New Orders for Durable Goods	0.69	0.69
59	AMDMUOx	Unfilled Orders for Durable Goods	0.26	0.40
60	BUSINVx	Total Business Inventories	0.27	0.75
61	ISRATIOx	Total Business: Inventories to Sales Ratio	0.22	0.27
		Median	0.46	0.69

Group 5. Money and credit

Index	FRED	Description	PC1	PC1&2
62	M1SL	M1 Money Stock	0.02	0.02
63	M2SL	M2 Money Stock	0.03	0.05
64	M2REAL	Real M2 Money Stock	0.03	0.45
65	BOGMBASE	Monetary Base	0.20	0.20
66	TOTRESNS	Total Reserves of Depository Institutions	0.28	0.28
67	NONBORRES	Reserves Of Depository Institutions	0.00	0.00
68	BUSLOANS	Commercial and Industrial Loans	0.10	0.17
69	REALLN	Real Estate Loans at All Commercial Banks	0.23	0.23
70	NONREVSL	Total Nonrevolving Credit	0.29	0.30
71	CONSPI	Nonrevolving consumer credit to Personal Income	0.09	0.17
118	DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding	0.03	0.03
119	DTCTHFNM	Total Consumer Loans and Leases Outstanding	0.21	0.22
120	INVEST	Securities in Bank Credit at All Commercial Banks	0.02	0.04
		Median	0.09	0.17

Group 6. Interest and exchange rates

Index	FRED	Description	PC1	PC1&2
76	FEDFUNDS	Effective Federal Funds Rate	0.34	0.68
77	TB3MS	3-Month Treasury Bill:	0.37	0.69
78	TB6MS	6-Month Treasury Bill:	0.38	0.71
79	GS1	1-Year Treasury Rate	0.36	0.71
80	GS5	5-Year Treasury Rate	0.16	0.63
81	GS10	10-Year Treasury Rate	0.08	0.59
82	AAA	Moody's Seasoned Aaa Corporate Bond Yield	0.02	0.60
83	BAA	Moody's Seasoned Baa Corporate Bond Yield	0.00	0.61
84	TB3SMFFM	3-Month Treasury C Minus FEDFUNDS	0.06	0.31
85	TB6SMFFM	6-Month Treasury C Minus FEDFUNDS	0.04	0.27
86	T1YFFM	1-Year Treasury C Minus FEDFUNDS	0.00	0.16
87	T5YFFM	5-Year Treasury C Minus FEDFUNDS	0.12	0.34
88	T10YFFM	10-Year Treasury C Minus FEDFUNDS	0.21	0.41
89	AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS	0.33	0.49
90	BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS	0.40	0.49
91	EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	0.00	0.01
92	EXJPUSx	Japan / U.S. Foreign Exchange Rate	0.00	0.03
93	EXUSUKx	U.S. / U.K. Foreign Exchange Rate	0.07	0.08
94	EXCAUSx	Canada / U.S. Foreign Exchange Rate	0.00	0.04
		Median	0.08	0.49

Table B1 (concluded)

Group 7. Prices

Index	FRED	Description	PC1	PC1&2
95	WPSFD49207	PPI: Finished Goods	0.07	0.74
96	WPSFD49502	PPI: Finished Consumer Goods	0.07	0.72
97	WPSID61	PPI: Intermediate Materials	0.06	0.65
98	WPSID62	PPI: Crude Materials	0.11	0.44
99	OILPRICEx	Crude Oil, spliced WTI and Cushing	0.02	0.50
100	PPICMM	PPI: Metals and metal products:	0.17	0.36
101	CPIAUCSL	CPI : All Items	0.09	0.82
102	CPIAPPSL	CPI : Apparel	0.04	0.40
103	CPITRNSL	CPI : Transportation	0.04	0.56
104	CPIMEDSL	CPI : Medical Care	0.11	0.41
105	CUSR0000SAC	CPI : Commodities	0.08	0.73
106	CUSR0000SAD	CPI : Durables	0.00	0.33
107	CUSR0000SAS	CPI : Services	0.01	0.67
108	CPIULFSL	CPI : All Items Less Food	0.04	0.77
109	CUSR0000SA0L2	CPI : All items less shelter	0.07	0.78
110	CUSR0000SA0L5	CPI : All items less medical care	0.10	0.82
111	PCEPI	Personal Cons. Expend.: Chain Index	0.08	0.76
112	DDURRG3M086SBEA	Personal Cons. Exp: Durable goods	0.00	0.37
113	DNDGRG3M086SBEA	Personal Cons. Exp: Nondurable goods	0.06	0.78
114	DSERRG3M086SBEA	Personal Cons. Exp: Services	0.04	0.66
		Median	0.06	0.66

Group 8. Stock market

Index	FRED	Description	PC1	PC1&2
72	S&P 500	S&P's Common Stock Price Index: Composite	0.22	0.33
73	S&P: indust	S&P's Common Stock Price Index: Industrials	0.18	0.28
74	S&P div yield	S&P's Composite Common Stock: Dividend Yield	0.05	0.40
75	S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio	0.11	0.44
		Median	0.15	0.36
		Overall median	0.19	0.50

Notes to Table B1. Index refers to the index number of the variable in our database. FRED refers to variable name in the FRED database. PC1 is the fraction of the variance of the cyclical component of that variable that is explained by the first principal component. PC1&2 is the fraction of the variance of the cyclical component of that variable that is explained by the first and second principal components combined.

C Additional simulations

Here we provide additional details on the simulations.

Cointegration. For this example, the nonstationary variables are cointegrated and there is no factor structure for the stationary variables. The single common factor for the nonstationary variables follows a random walk:

$$F_t = F_{t-1} + v_t \quad t = 1, ..., T;\ F_0 = 0$$

$$y_{it} = F_t + \varepsilon_{it} \quad i = 1, 2, ..., N/2$$

$$y_{it} = \varepsilon_{it} \quad i = (N/2) + 1, ..., N$$

with $v_t \sim N(0, 1)$ and $\varepsilon_{it} \sim N(0, 1)$ independent for all i and t. Notice that the first $(N/2)$ variables are characterized by

$$y_{it} = F_{t-h} + v_{t-h+1} + v_{t-h+2} + \cdots + v_t + \varepsilon_{it}$$

$$\mathbb{P}(y_{it}|1, y_{i,t-h}, y_{i,t-h-1}, ..., y_{i,t-h-p+1}) \simeq F_{t-h}$$

$$c_{it} \simeq v_{t-h+1} + v_{t-h+2} + \cdots + v_t + \varepsilon_{it}.$$

Thus there is a single factor (namely $v_{t-h+1} + v_{t-h+2} + \cdots + v_t$) that is common to the cyclical component of the first $(N/2)$ variables,[1] and the true number of common factors in the sample of N variables is $r = 1$.

Columns (7) and (8) of Table 2 report the results from applying PCA to the raw data. Note that even though half the variables are nonstationary, PCA always correctly concludes that there is a single common factor in these data. Columns (9) and (10) report results from applying PCA to the residuals from 24-period-ahead forecasting regressions. Again the same regression is estimated in the same way for stationary and nonstationary observations. And again PCA on the regression residuals results in the correct answer 100% of the time, even if the sample size is as small as $T = 100$.

Stationary factor with a mix of stationary and nonstationary indicators. In this example the data-

[1] Another way to express this is that the variables are cointegrated with $(N/2) - 1$ linearly independent cointegrating relations given by $y_{it} - y_{1t} \sim I(0)$ for $i = 2, 3, ..., (N/2)$.

generating process features a common stationary factor: $F_t = \rho_F F_{t-1} + v_t$ with $\rho_F = 0.8$. The ith observed variable y_{it} is related to F_t with a weight ω_i, $y_{it} = \omega_i F_t + g_{it}$, where the idiosyncratic components g_{it} are a mix of stationary and nonstationary processes: $g_{it} = \rho_i g_{i,t-1} + e_{it}$ where v_t and e_{it} $(i = 1, ..., N)$ are mutually independent $N(0,1)$. Columns (11)-(14) of Table 2 report results for the following example:

$$\begin{array}{ll} \rho_i = 1, \omega_i = 1 & \text{for } i = 1, ..., N/4 \\ \rho_i = 0.5, \omega_i = 1 & \text{for } i = 1 + N/4, ..., N/2 \\ \rho_i = 1, \omega_i = 0 & \text{for } i = 1 + N/2, ..., 3N/4 \\ \rho_i = 0.5, \omega_i = 0 & \text{for } i = 1 + 3N/4, ..., N \end{array}.$$

Thus for this example half of the N observed variables are nonstationary. The variables are independent of each other apart from their potential common dependence on the $r = 1$-dimensional factor F_t, and this single factor affects some of the observed variables but not others.

Local-to-unit roots. For these simulations we replaced ρ in (21) for half the variables with $\rho_T = 1 - c/T$ with $c = 1$ and T the number of time-series observations in a generated sample. The other half of the generated variables were white noise.

Fractional integration. Here we replaced equation (21) for half the variables with

$$(1 - L)^d y_{it} = \varepsilon_{it}$$

for $i = 1, 2, ..., N/2$ and $d = 0.6$. The other half of the generated variables were again white noise. We generated draws for this process using the MA(∞) representation $y_{it} = \sum_{j=0}^{\infty} \psi_j \varepsilon_{i,t-j}$ and the weights calculated recursively from

$$\psi_j = \frac{d + j - 1}{j} \psi_{j-1}$$

starting from $\psi_0 = 0$. In practice we did this by truncating the sum as $\sum_{j=0}^{M} \psi_j \varepsilon_{i,t-j}$ for $M = 100,000$ using 100,000 pre-sample draws for $\varepsilon_{i,-1}, \varepsilon_{i,-2}, ..., \varepsilon_{i-M}$.

Comparison with the Bai and Ng (2004) PANIC approach. Bai and Ng (2004) recommended taking first-differences of all the observations and then applying PCA to the changes. Columns (4) and

(8) in Table C2 show that this correctly concludes that there are no common factors when applied to the data-generating processes described in Sections 5.1 and 5.2. Columns (3) and (7) show the fraction of the variance of Δy_{it} explained by the first three principal components. These results are very similar to those for our method with $h = 1$, shown in columns (1)-(2) (which reproduce columns (11) and (12) in Table 1) and columns (5)-(6) (which reproduce columns (5) and (6) in Table 2). Note that the R^2 for the $h = 1$ cases refer to the sample variance of the estimated cyclical factors, as it did in the earlier tables.

Table C1: Local to unity and fractionally integrated processes

	Local to unity						Fractionally integrated					
	Raw data		$\hat{c}_t$ (h=24)		$\hat{c}_t$ (h=1)		Raw data		$\hat{c}_t$ (h=24)		$\hat{c}_t$ (h=1)	
j	R^2	r^*	R^2	r^*	R^2	r^*	R^2	r^*	R^2	r^*	R^2	r^*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$T = 100$												
0	—	0	—	0	—	100	—	8	—	99	—	100
1	43.0	98	15.8	3	4.1	0	10.5	88	8.0	1	4.1	0
2	3.7	2	10.6	97	3.8	0	6.7	4	6.5	0	3.8	0
3	2.8	0	5.5	0	3.6	0	4.9	0	5.4	0	3.6	0
$T = 200$												
0	—	0	—	0	—	100	—	0	—	43	—	100
1	44.8	96	11.0	0	2.9	0	9.9	59	7.2	53	2.9	0
2	2.7	4	8.9	0	2.7	0	5.9	41	5.8	4	2.7	0
3	2.2	0	7.2	100	2.6	0	4.4	0	4.5	0	2.6	0
$T = 400$												
0	—	0	—	3	—	100	—	0	—	99	—	100
1	46.1	98	6.4	25	2.2	0	9.3	26	4.8	1	2.2	0
2	2.1	2	5.5	48	2.1	0	5.6	72	4.1	0	2.1	0
3	1.8	0	4.9	24	2.1	0	4.0	2	3.6	0	2.1	0
$T = 600$												
0	—	0	—	89	—	100	—	0	—	100	—	100
1	46.8	98	4.9	11	1.9	0	9.0	18	3.9	0	1.9	0
2	1.9	2	4.3	0	1.9	0	5.4	77	3.4	0	1.9	0
3	1.6	0	3.9	0	1.8	0	3.8	5	3.1	0	1.8	0
$T = 800$												
0	—	0	—	100	—	100	—	0	—	100	—	100
1	47.2	99	4.1	0	1.8	0	8.7	8	3.4	0	1.8	0
2	1.7	1	3.7	0	1.7	0	5.4	78	3.0	0	1.8	0
3	1.5	0	3.4	0	1.7	0	3.8	14	2.7	0	1.7	0
$T = 1000$												
0	—	0	—	100	—	100	—	0	—	100	—	100
1	47.5	99	3.7	0	1.7	0	8.7	8	3.1	0	1.7	0
2	1.6	1	3.3	0	1.7	0	5.3	89	2.7	0	1.7	0
3	1.4	0	3.0	0	1.6	0	3.6	3	2.5	0	1.6	0

Notes to Table C1. R^2 indicates the percentage of total variance accounted for by the jth principal component for $j = 1, 2$ or 3. r^* indicates the percentage of samples for which the criterion (19) selects the number of factors to be $j = 0, 1, 2$, or ≥ 3. In every case, the true number of factors is $r = 0$ and the cross-section dimension is $N = 100$.

Table C2: Comparison of PANIC method with one-period-ahead forecast errors

	Random walk				Stationary			
	$h=1$		PANIC		$h=1$		PANIC	
j	R^2 (1)	r^* (2)	R^2 (3)	r^* (4)	R^2 (5)	r^* (6)	R^2 (7)	r^* (8)
$T=100$								
0	—	100	—	100	—	100	—	100
1	4.1	0	4.2	0	4.1	0	4.2	0
2	3.8	0	3.9	0	3.8	0	3.9	0
3	3.6	0	3.7	0	3.6	0	3.7	0
$T=200$								
0	—	100	—	100	—	100	—	100
1	2.9	0	3.0	0	2.9	0	3.8	0
2	2.7	0	2.8	0	2.7	0	2.9	0
3	2.6	0	2.7	0	2.6	0	2.8	0
$T=400$								
0	—	100	—	100	—	100	—	100
1	2.2	0	2.3	0	2.2	0	3.8	0
2	2.1	0	2.2	0	2.1	0	2.3	0
3	2.1	0	2.1	0	2.1	0	2.2	0
$T=600$								
0	—	100	—	100	—	100	—	100
1	1.9	0	2.0	0	1.9	0	3.4	0
2	1.9	0	1.9	0	1.9	0	2.0	0
3	1.8	0	1.9	0	1.8	0	1.9	0
$T=800$								
0	—	100	—	100	—	100	—	100
1	1.8	0	1.9	0	1.8	0	3.1	0
2	1.7	0	1.8	0	1.7	0	1.8	0
3	1.7	0	1.8	0	1.7	0	1.8	0
$T=1000$								
0	—	100	—	100	—	100	—	100
1	1.7	0	1.8	0	1.7	0	2.8	0
2	1.7	0	1.7	0	1.7	0	1.7	0
3	1.6	0	1.7	0	1.6	0	1.7	0

Notes to Table C2. R^2 indicates the percentage of total variance of the estimated cyclical component (for $h = 1$) or of the observed changes (for PANIC) accounted for by the jth principal component for $j = 1, 2$ or 3. r^* indicates the percentage of samples for which the criterion (19) selects the number of factors to be $j = 0, 1, 2$, or ≥ 3. In every case, the true number of factors is $r = 0$ and the cross-section dimension is $N = 100$.

References

Bai, J. and Ng, S. (2004). A PANIC attack on unit roots and cointegration. *Econometrica*, 72(4):1127–1177.

Beghin, L., Nikitin, Y., and Orsingher, E. (2005). Exact small ball constants for some Gaussian processes under the L^2-norm. *Journal of Mathematical Sciences*, 128(1):2493–2502.

Çinlar, E. (2011). *Probability and Stochastics*. Springer.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Komlós, J., Major, P., and Tusnády, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 34:33–58.

Shumway, R. H. and Stoffer, D. S. (2017). *Time Series Analysis and Its Applications*. Springer.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer.