Alternative asymptotics for cointegration tests in large VARs.

Alexei Onatski*and Chen Wang
Faculty of Economics, University of Cambridge

August 5, 2017

Abstract

Johansen's (1988, 1991) likelihood ratio test for cointegration rank of a vector autoregression (VAR) depends only on the squared sample canonical correlations between current changes and past levels of a simple transformation of the data. We study the asymptotic behavior of the empirical distribution of those squared canonical correlations when the number of observations and the dimensionality of the VAR diverge to infinity simultaneously and proportionally. We find that the distribution weakly converges to the so-called *Wachter distribution*. This finding provides a theoretical explanation for the observed tendency of Johansen's test to find "spurious cointegration".

1 Introduction

Johansen's (1988, 1991) likelihood ratio (LR) test for cointegration rank is a very popular econometric technique. However, it is rarely applied to systems of more than three or four variables. On the other hand, there exist many applications involving much larger systems. For example, Davis (2003) discusses a possibility of applying the test to the data on seven aggregated and individual commodity prices to test Lewbel's (1996) generalization of the Hicks-Leontief composite commodity theorem. In a recent study of exchange rate predictability, Engel et al. (2015) contemplate

*Supported by Keynes Fellowship grant.

a possibility of determining the cointegration rank of a system of seventeen OECD exchange rates. Banerjee et al. (2004) emphasize the importance of testing for no cross-sectional cointegration in panel cointegration analysis (see Breitung and Pesaran (2008) and Choi (2015)), and the cross-sectional dimension of modern macroeconomic panels can easily be as large as forty.

The main reason why the LR test is rarely used in the analysis of relatively large systems is its poor finite sample performance. Even for small systems, the test based on the asymptotic critical values does not perform well (see Johansen (2002)). For large systems, the size distortions become overwhelming, leading to severe over-rejection of the null in favour of too much cointegration as shown in many simulation studies, including Ho and Sorensen (1996) and Gonzalo and Pitarakis (1995, 1999).

In this paper, we study the asymptotic behavior of the sample canonical correlations that the LR statistic is based on, when the number of observations and the system's dimensionality go to infinity simultaneously and proportionally. We show that the empirical distribution of the squared sample canonical correlations converges to the so-called *Wachter distribution*, originally derived by Wachter (1980) as the limit of the empirical spectral distribution of the multivariate beta matrix of growing dimension and degrees of freedom. Our analytical findings explain the observed over-rejection of the null hypothesis by the LR test.

The basic framework for our analysis is standard. Consider a p-dimensional VAR in the error correction form

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t, \tag{1}$$

where D_t and ε_t are vectors of deterministic terms and zero-mean, not necessarily Gaussian, errors with unconstrained covariance matrix, respectively. The (quasi) LR statistic for the test of the null hypothesis of no more than r cointegrating relationships between the p elements of X_t against the alternative of more than r such relationships is given by

$$LR_{r,p,T} = -T \sum_{i=r+1}^{p} \log\left(1 - \lambda_i\right), \tag{2}$$

where T is the sample size, and $\lambda_1 \geq ... \geq \lambda_p$ are the squared sample canonical correlation coefficients between residuals in the regressions of ΔX_t and X_{t-1} on the lagged differences ΔX_{t-i}, $i = 1, ..., k-1$, and the deterministic terms. In the absence of the lagged differences and deterministic terms, the λ's are the eigenvalues of $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$, where S_{00} and S_{11} are the sample covariance matrices of ΔX_t and X_{t-1}, respectively, while S_{01} is the cross sample covariance matrix.

Johansen (1991) shows that the asymptotic distribution of $LR_{r,p,T}$ under the asymptotic regime where $T \to \infty$ while p remains fixed, can be expressed in terms of the eigenvalues of a matrix whose entries are explicit functions of a $(p-r)$-dimensional Brownian motion. Unfortunately, for relatively large p, this asymptotics does not produce good finite sample approximations, as evidenced by the over-rejection phenomenon mentioned above. Therefore, in this paper, we consider a *simultaneous* asymptotic regime $p, T \to_c \infty$ where both p and T diverge to infinity so that

$$p/T \to c \in (0,1]. \tag{3}$$

Our Monte Carlo analysis shows that the corresponding asymptotic approximations are relatively accurate even for such small sample sizes as $p = 10$ and $T = 20$.

The basic specification for the data generating process (1) that we consider has $k = 1$. In the next section, we discuss extensions to more general VARs with low-rank Γ_i matrices and additional common factor terms. We also explain there that our main results hold independently from whether a deterministic vector D_t with fixed or slowly-growing dimension is present or absent from the VAR.

Our study focuses on the behavior of the empirical distribution function (d.f.) of the squared sample canonical correlations,

$$F_p(\lambda) = \frac{1}{p}\sum_{i=1}^{p} \mathbf{1}\{\lambda_i \le \lambda\}, \tag{4}$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. The dependence of $F_p(\lambda)$ on T is suppressed to keep notations simple. We find that, under the null of r cointegrating relationships, as $p, T \to_c \infty$ while $r/p \to 0$,

$$F_p(\lambda) \Rightarrow W_c(\lambda) \equiv W(\lambda; c/(1+c), 2c/(1+c)), \tag{5}$$

where $\Rightarrow$ denotes the weak convergence of d.f.'s, and $W(\lambda; \gamma_1, \gamma_2)$ denotes the *Wachter* d.f. with parameters γ_1 and γ_2, described in detail in the next section.

As explained below, convergence (5) guarantees that the probability of the event

$$LR_{r,p,T}/p^2 \geq -c^{-1} \int \log(1-\lambda)\, \mathrm{d}W_c(\lambda) - \delta \tag{6}$$

converges to one as $p, T \to_c \infty$, for any $\delta > 0$. In contrast, we show that under the standard asymptotic regime, where $T \to \infty$ while p is held fixed, $LR_{r,p,T}/p^2$ concentrates around 2 for relatively large p. A direct calculation reveals that 2 is smaller than the lower bound (6), for all $c > 0$, with the gap growing as c increases. That is, the standard asymptotic distribution of the LR statistic is centered at a too low level, especially for relatively large p. This explains the tendency of the asymptotic LR test to over-reject the null.

The reason for the poor centering delivered by the standard asymptotic approximation is that it classifies terms $(p/T)^j$ in the asymptotic expansion of the LR statistic as $O(T^{-j})$. When p is relatively large, such terms substantially contribute to the finite sample distribution of the statistic, but are ignored as asymptotically negligible. In contrast, the *simultaneous asymptotics* classifies all terms $(p/T)^j$ as $O(1)$. They are not ignored asymptotically, which improves the centering of the simultaneous asymptotic approximation relative to the standard one.

Our study is the first to derive the limit of the empirical d.f. of the squared sample canonical correlations between random walk X_{t-1} and its innovations ΔX_t. Wachter (1980) shows that $W(\lambda; \gamma_1, \gamma_2)$ is the weak limit of the empirical d.f. of the squared sample canonical correlations between q- and m-dimensional independent Gaussian white noises with the size of the sample n, when $q, m, n \to \infty$ so that $q/n \to \gamma_1$ and $m/n \to \gamma_2$. Yang and Pan (2012) show that Wachter's (1980) result holds without the Gaussianity assumption for i.i.d. data with finite second moments. Our proofs do not rely on those previous results. The novelty and difficulty of our setting is that X_t and ΔX_t are not independent processes. This requires original ideas for our proofs.

The rest of this paper is structured as follows. In Section 2, we prove the convergence of $F_p(\lambda)$ to the *Wachter* d.f. under the simultaneous asymptotics. Section 3 derives the sequential limit of $F_p(\lambda)$ as first $T \to \infty$ and then $p \to \infty$. It then uses differences between the sequential and simultaneous limits to explain the over-rejection phenomenon. Section 4 contains a Monte Carlo study. Section 5 concludes. All proofs are given in the Supplementary Material (SM).

2 Convergence to the Wachter distribution

Consider the following basic version of (1)

$$\Delta X_t = \Pi X_{t-1} + \Phi D_t + \varepsilon_t \tag{7}$$

with d_D-dimensional vector of deterministic regressors D_t. We allow innovations ε_t to be i.i.d. vectors with zero mean and a non-singular covariance matrix, not necessarily Gaussian. Let R_{0t} and R_{1t} be the vectors of residuals from the OLS regressions of ΔX_t on D_t, and X_{t-1} on D_t, respectively. Define

$$S_{00} = \frac{1}{T}\sum_{t=1}^{T} R_{0t}R_{0t}', \;\; S_{01} = \frac{1}{T}\sum_{t=1}^{T} R_{0t}R_{1t}', \;\text{ and } S_{11} = \frac{1}{T}\sum_{t=1}^{T} R_{1t}R_{1t}', \tag{8}$$

and let $\lambda_1 \geq ... \geq \lambda_p$ be the eigenvalues of $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$.

The main goal of this section is to establish the weak convergence of the empirical d.f. of the λ's to the Wachter d.f., under the null of r cointegrating relationships, when $p, T \rightarrow_c \infty$ and $r/p \rightarrow 0$. The Wachter distribution with d.f. $W(\lambda; \gamma_1, \gamma_2)$ and parameters $\gamma_1, \gamma_2 \in (0,1)$ has density

$$f_W(\lambda; \gamma_1, \gamma_2) = \frac{1}{2\pi\gamma_1}\frac{\sqrt{(b_+ - \lambda)(\lambda - b_-)}}{\lambda(1-\lambda)} \tag{9}$$

on $[b_-, b_+] \subseteq [0,1]$ with

$$b_\pm = \left(\sqrt{\gamma_1(1-\gamma_2)} \pm \sqrt{\gamma_2(1-\gamma_1)}\right)^2, \tag{10}$$

and atoms of size $\max\{0, 1-\gamma_2/\gamma_1\}$ at zero, and $\max\{0, 1-(1-\gamma_2)/\gamma_1\}$ at unity.

We assume that model (7) may be misspecified in the sense that the data generating process is described by the following generalization of (1)

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Psi F_t + \varepsilon_t, \tag{11}$$

where ε_t, $t = 1, ..., T$, are still i.i.d.$(0, \Sigma)$ with arbitrary $\Sigma > 0$, $\operatorname{rank}\Pi = r$, but k is not necessarily unity, and F_t is a d_F-dimensional vector of deterministic or stochastic variables that does not necessarily coincide with D_t. For example, some of the components of F_t may be common factors not observed and not modelled by the econometrician. Further, we do not put any restrictions on the roots of the characteristic polynomial associated with (11). In particular, explosive behavior and seasonal unit roots are allowed. Finally, no constraints on F_t, and the initial values $X_{1-k}, ..., X_0$, apart from the asymptotic requirements on d_F and k as spelled out in the following theorem, are imposed.

Theorem 1 *Suppose that the data are generated by (11), and let* $\Gamma = [\Gamma_1, ..., \Gamma_{k-1}]$. *If as* $p, T \to_c \infty$,

$$(d_D + d_F + r + k + \operatorname{rank}\Gamma)/p \to 0, \tag{12}$$

then

$$F_p(\lambda) \Rightarrow W_c(\lambda) \equiv W(\lambda; c/(1+c), 2c/(1+c)), \tag{13}$$

in probability. In special cases where innovations ε_t *are Gaussian, convergence (13) holds almost surely.*

The weak convergence in probability of empirical d.f. $F_p(\lambda)$ to $W_c(\lambda)$ can be understood as the usual convergence in probability of the Lévy distance between $F_p(\lambda)$ and $W_c(\lambda)$ to zero (see Billingsley (1995), problem 14.5). Theorem 1 implies that the weak limits of $F_p(\lambda)$ corresponding to the general model (11) and to the basic model $\Delta X_t = \Pi X_{t-1} + \varepsilon_t$ are the same as long as (12) holds.

Condition (12) guarantees that the difference between the general and basic versions of $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$ has rank R that is less than proportional to p (and to T). Then, by the so-called rank inequality (Theorem A.43 in Bai and Silverstein (2010)), the Lévy distance between the general and basic versions of $F_p(\lambda)$ is no larger than R/p, which converges to zero as $p, T \to_c \infty$. For further details, see the proof of Theorem 1 in the SM.

Figure 1 shows quantile plots of $W_c(\lambda)$ for different values of c. For $c = 1/5$, the dimensionality of the data constitutes 20% of the sample size. The upper boundary of support of the corresponding Wachter distribution is above 0.7. In particular, we expect λ_1 be larger than 0.7 for large p and T, even in the absence of any cointegrating relationships. For $c = 1/2$, the upper boundary of support of the Wachter limit is unity. This accords with Gonzalo and Pitarakis' (1995, Lemma 2.3.1) finding that as $T/p \to 2$, $\lambda_1 \to 1$. For $c = 4/5$, the Wachter limit has mass 3/4 at unity.

Wachter (1980) derives $W(\lambda; \gamma_1, \gamma_2)$ as the weak limit of the empirical d.f. of

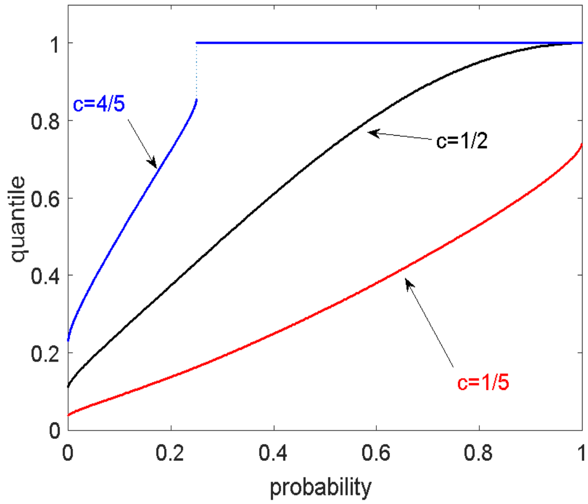

Figure 1: Quantile functions of $W_c(\lambda)$ for $c = 1/5$, $c = 1/2$, and $c = 4/5$.

eigenvalues of the p-dimensional beta[1] matrix $B_p(n_1/2, n_2/2)$ with n_1, n_2 degrees of freedom as $p, n_1, n_2 \to \infty$ so that $p/n_1 \to \gamma_1/\gamma_2$ and $p/n_2 \to \gamma_1/(1-\gamma_2)$. The eigenvalues of multivariate beta matrices are related to many important concepts in multivariate statistics, including canonical correlations, multiple discriminant ratios, and MANOVA. In particular, the squared sample canonical correlations between q- and m-dimensional independent Gaussian samples of size n are jointly distributed as the eigenvalues of $B_q(m/2, (n-m)/2)$, where $q \leq m$ and $n \geq q+m$. Therefore, their empirical d.f. weakly converges to $W(\lambda; \gamma_1, \gamma_2)$ with $\gamma_1 = \lim q/n$ and $\gamma_2 = \lim m/n$.

Note that the latter limit coincides with $W_c(\lambda)$ when $n = T+p$, $q = p$, and $m = 2p$. Hence, Theorem 1 implies that the limiting empirical distribution of the squared sample canonical correlations between T observations of p-dimensional random walk and its own innovations is the same as that between $T+p$ observations of independent p- and $2p$-dimensional white noises. This suggests that there might exist a deep connection between these two settings, which is yet to be discovered.

The weak convergence in probability of $F_p(\lambda)$ established in Theorem 1 implies the convergence in probability of bounded continuous functionals of $F_p(\lambda)$. An example of such a functional is the scaled Pillai-Bartlett statistic for the null of no more than

[1]For the definition of the multivariate beta see Muirhead (1982), p. 110.

r cointegrating relationships (see Gonzalo and Pitarakis (1995))

$$PB_{r,p,T}/p^2 = \left(T/p^2\right) \sum_{j=r+1}^{p} \lambda_j,$$

which is asymptotically equivalent to the scaled LR statistic under the standard asymptotic regime. Since, by definition, $\lambda_j \in [0,1]$, we have

$$PB_{r,p,T}/p^2 = (T/p) \int f(\lambda) \mathrm{d}F_p(\lambda) - \left(T/p^2\right) \sum_{j=1}^{r} \lambda_j, \tag{14}$$

where f is the bounded continuous function

$$f(\lambda) = \begin{cases} 0 & \text{for } \lambda < 0 \\ \lambda & \text{for } \lambda \in [0,1] \\ 1 & \text{for } \lambda > 1. \end{cases}.$$

As long as $r/p \to 0$ as $p, T \to_c \infty$, the second term on the right hand side of (14) converges to zero. Therefore, Theorem 1 implies that $PB_{r,p,T}/p^2$ converges to $c^{-1} \int f(\lambda) \mathrm{d}W_c(\lambda)$ in probability (a.s. in cases of Gaussian ε_t). A direct calculation of the latter integral yields the following corollary.

Corollary 2 *Under the assumptions of Theorem 1, as* $p, T \to_c \infty$*,*

$$PB_{r,p,T}/p^2 \xrightarrow{\text{P}} 2/(1+c) + c^{-1} \max\left\{0, 2 - c^{-1}\right\}.$$

The above convergence in probability becomes the a.s. convergence when ε_t *are Gaussian vectors.*

A similar analysis of the LR statistic (2) is less straightforward because $\log(1-\lambda)$ is unbounded on $\lambda \in [0,1]$. In fact, for $c > 1/2$, the statistic is ill-defined because a non-negligible proportion of the squared sample canonical correlations exactly equal unity. However for $c < 1/2$, we can obtain an asymptotic lower bound on $LR_{r,p,T}/p^2$. Note that for such c, the upper bound of the support of $W_c(\lambda)$ equals

$b_+ = c\left(\sqrt{2} - \sqrt{1-c}\right)^{-2} < 1$. Let

$$\overline{\log}\,(1-\lambda) = \begin{cases} 0 & \text{for } \lambda < 0 \\ \log(1-\lambda) & \text{for } \lambda \in [0, b_+] \\ \log(1-b_+) & \text{for } \lambda > b_+. \end{cases} \tag{15}$$

Clearly, $\overline{\log}\,(1-\lambda)$ is a bounded continuous function and

$$LR_{r,p,T}/p^2 \geq -\left(T/p^2\right) \sum_{j=r+1}^{p} \overline{\log}(1-\lambda_j).$$

As we show in the SM, the latter inequality yields the following asymptotic lower bound on $LR_{r,p,T}/p^2$.

Corollary 3 *Under the assumptions of Theorem 1, for any $c \in (0, 1/2)$ and $\delta > 0$, $\Pr\left\{LR_{r,p,T}/p^2 < \underline{LR}_c - \delta\right\} \to 0$ as $p, T \to_c \infty$, where*

$$\underline{LR}_c = \frac{1+c}{c^2} \ln(1+c) - \frac{1-c}{c^2} \ln(1-c) + \frac{1-2c}{c^2} \ln(1-2c).$$

Furthermore, in cases where ε_t are Gaussian vectors, $\liminf LR_{r,p,T}/p^2 \geq \underline{LR}_c$ a.s.

Corollary 3 implies that an appropriate "centering point" for the scaled LR statistic when p and T are large cannot be lower than $\underline{LR}_c$. As we show in the next section, the standard asymptotic distribution concentrates around a point that is below $\underline{LR}_c$ for large p, which explains the over-rejection phenomenon. To study such a concentration, in the next section, we consider the *sequential* asymptotic regime where first $T \to \infty$, and then $p \to \infty$.

3 Sequential asymptotics and over-rejection

To obtain useful results under the sequential asymptotics, we study eigenvalues of the scaled matrix

$$(T/p)\, S_{01} S_{11}^{-1} S_{01}' S_{00}^{-1}. \tag{16}$$

Under the simultaneous asymptotic regime, the behavior of the scaled and unscaled eigenvalues is the same up to the factor $c^{-1} = \lim T/p$. In contrast, as $T \to \infty$ while p remains fixed, the unscaled eigenvalues converge to zero, while scaled ones do not.

We focus on the basic case where $r = 0$, the data generating process is

$$\Delta X_t = \varepsilon_t,\ t = 1, ..., T,\ \text{with } X_0 = 0, \tag{17}$$

and the only deterministic regressor included by the econometrician in model (7) is constant, that is $d_D = 1$. Then, Johansen's (1988, 1991) results imply that, as $T \to \infty$ while p is held fixed, the eigenvalues of the scaled matrix (16) jointly converge in distribution to the eigenvalues of

$$\frac{1}{p}\int_0^1 (\mathrm{d}B)\, F' \left(\int_0^1 FF'\mathrm{d}u\right)^{-1} \int_0^1 F\, (\mathrm{d}B)', \tag{18}$$

where B is a p-dimensional Brownian motion and F is its demeaned version. We denote the eigenvalues of (18) as $\lambda_{j,0}$, and their empirical d.f. as $F_{p,0}(\lambda)$.

It is reasonable to expect that, as $p \to \infty$, $F_{p,0}(\lambda)$ becomes close to the limit of the empirical d.f. of eigenvalues of (16) under a simultaneous, rather than sequential, asymptotic regime $p, T \to_\gamma \infty$, where γ is close to zero. We denote such a limit as $F_\gamma(\lambda)$. This expectation turns out to be correct in the sense that the following theorem holds.

Theorem 4 *Let $F_0(\lambda)$ be the weak limit of $F_\gamma(\lambda)$ as $\gamma \to 0$. Then, as $p \to \infty$, $F_{p,0}(\lambda) \Rightarrow F_0(\lambda)$, in probability. The d.f. $F_0(\lambda)$ corresponds to a distribution supported on $[a_-, a_+]$ with*

$$a_\pm = \left(1 \pm \sqrt{2}\right)^2, \tag{19}$$

and having density

$$f(\lambda) = \frac{1}{2\pi}\frac{\sqrt{(a_+ - \lambda)(\lambda - a_-)}}{\lambda}. \tag{20}$$

A reader familiar with Large Random Matrix Theory (see Bai and Silverstein (2010)) might recognize $F_0(\lambda)$ as the d.f. of the continuous part of a special case of the *Marchenko-Pastur distribution* (Marchenko and Pastur (1967)). The general Marchenko-Pastur distribution has density

$$f_{MP}\left(\lambda; \kappa, \sigma^2\right) = \frac{1}{2\pi\sigma^2\kappa}\frac{\sqrt{(a_+ - \lambda)(\lambda - a_-)}}{\lambda}$$

over $[a_-, a_+]$ with $a_\pm = \sigma^2 (1 \pm \sqrt{\kappa})^2$ and a point mass $\max\{0, 1 - 1/\kappa\}$ at zero.

Density (20) is two times $f_{MP}(\lambda; \kappa, \sigma^2)$ with $\kappa = 2$ and $\sigma^2 = 1$. The multiplication by two is needed because the mass 1/2 at zero is not a part of the distribution F_0.

Note that, as $T \to \infty$ while p remains fixed,

$$LR_{0,p,T}/p^2 \xrightarrow{d} \frac{1}{p}\sum_{j=1}^{p} \lambda_{j,0} = \int \lambda \mathrm{d}F_{p,0}(\lambda). \tag{21}$$

One may therefore conjecture that under the sequential asymptotics, $LR_{0,p,T}/p^2$ converges in probability to $\int \lambda \mathrm{d}F_0(\lambda)$.

Our next result verifies this conjecture. Since $f(\lambda) \equiv \lambda$ is not a bounded function, the verification cannot rely solely on Theorem 4. In the proof of the next theorem, we show that the tails of $F_{p,0}(\lambda)$ behave sufficiently regularly so that the convergence $\int \lambda \mathrm{d}F_{p,0}(\lambda) \xrightarrow{\mathrm{P}} \int \lambda \mathrm{d}F_0(\lambda)$ does take place.

Theorem 5 *Under the sequential asymptotics, $LR_{0,p,T}/p^2$ converges in probability to $\int \lambda \mathrm{d}F_0(\lambda) = 2$.*

Theorem 5 is consistent with the numerical finding of Johansen et al. (2005, Table 2) that, as T becomes large while p is being fixed, the sample mean of the LR statistic is well approximated by a polynomial $2p^2 + \alpha p$ (see also Johansen (1988) and Gonzalo and Pitarakis (1995)). The value of α depends on how many deterministic regressors are included in the VAR. Our theoretical result justifies the $2p^2$ term in the above approximation. A theoretical analysis of α would require a further study.

The concentration of the LR statistic around $2p^2$ explains why the critical values of the LR test are so large for large values of p. The transformation $LR_{0,p,T} \mapsto LR_{0,p,T}/p - 2p$ makes the LR statistic 'well-behaved' under the sequential asymptotics and leads to more conventional critical values. We report the corresponding transformed 95% critical values alongside the original ones in Table 1.

The transformed critical values resemble 97-99 percentiles of $N(0, 1)$. Since the LR test is one-sided, the resemblance is coincidental. However, we do expect that the sequential asymptotic distribution of the transformed LR statistic is normal (possibly with non-zero mean and non-unit variance). A formal analysis of this conjecture is left for future research.

Corollary 3 and Theorem 5 can be used to explain the over-rejection phenomenon from a theoretical perspective. The reason for finding spurious cointegration when p is relatively large is the discrepancy between simultaneous and sequential asymptotic

p	Original CV	CV/$p-2p$	p	Original CV	CV/$p-2p$
1	4.13	2.13	7	111.79	1.97
2	12.32	2.16	8	143.64	1.96
3	24.28	2.09	9	179.48	1.94
4	40.17	2.04	10	219.38	1.94
5	60.06	2.01	11	263.25	1.93
6	83.94	1.99	12	311.09	1.92

Table 1: The 95% asymptotic critical values (CV) for Johansen's LR test. The original values are taken from the first column of Table II in MacKinnon et al. (1999).

behavior of the LR statistic. As can be seen from Figure 2, the lower bound, $\underline{LR}_c$, for $LR_{0,p,T}/p^2$ under the simultaneous asymptotics is larger than the probability limit, 2, under the sequential one.

The Monte Carlo analysis in the next section shows that 'typical' values of $LR_{0,p,T}/p^2$ in finite samples with comparable p and T are concentrated around $\underline{LR}_c$. In contrast, the standard asymptotic critical values (divided by p^2) are concentrated around two. Hence, the standard asymptotic distribution of the LR statistic is centered at a too low level. As $c \equiv \lim p/T$ increases, the discrepancy $\underline{LR}_c - 2$ grows, and the over-rejection becomes more and more severe.

In addition to $\underline{LR}_c$, Figure 2 shows the probability limit of the scaled Pillai-Bartlett statistic under the simultaneous asymptotics, derived in Corollary 2. In contrast to $\underline{LR}_c$, this limit lies below 2. Therefore, we expect the Pillai-Bartlett test to under-reject, especially in high-dimensional situations. This agrees with the numerical findings of Gonzalo and Pitarakis (1995).

Incidentally, the average of $\underline{LR}_c$ and the probability limit of the Pillai-Bartlett statistic is numerically close to the sequential limit, at least for $c \leq 0.3$. This explains a relatively good performance of the test based on the linear combination $(LR+PB)/2$, proposed by Gonzalo and Pitarakis (1995).

4 Monte Carlo

Throughout this section, the analysis is based on 1000 Monte Carlo (MC) replications. We consider three different distributions for simulated data: Student's $t(3)$, which has only two finite moments; Gaussian; and centered $\chi^2(1)$ distribution, which is skewed to the right. For each of the MC experiments, we report results only for the Student

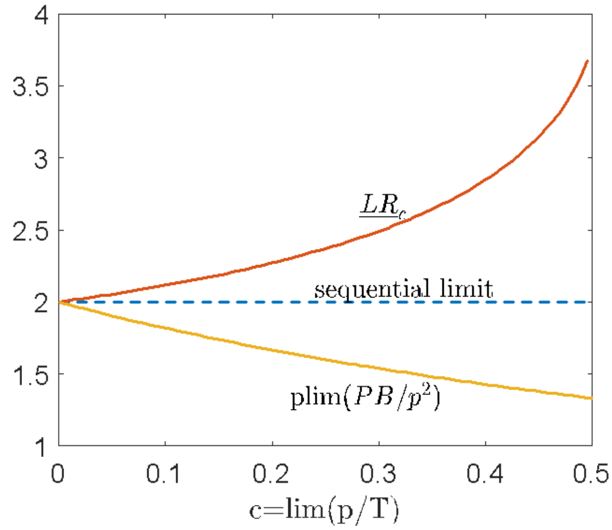

Figure 2: The simultaneous and sequential asymptotic behavior of the scaled (divided by p^2) LR and PB statistics. Dashed line: sequential probability limit of the scaled LR and PB. Upper line: simultaneous asymptotic lower bound on the scaled LR. Lower line: simultaneous probability limit of the scaled PB.

case. The corresponding results for the other two cases turn out to be very similar.

First, we generate pure random walk data with zero starting values for $(p, T) = (10, 100)$ and $(p, T) = (10, 20)$. Figure 3 shows the Tukey boxplots summarizing the MC distribution of each of the λ_{p+1-i}, $i = 1, ..., p$. Indexing λ's by $p + 1 - i$ ensures that $i = 1$ corresponds to the smallest squared sample canonical correlation, λ_p, and $i = p$ corresponds to the largest squared sample canonical correlation, λ_1.

The boxplots are superimposed with the quantile function of the Wachter limit with $c = 1/10$ for the left panel and $c = 1/2$ for the right panel. Precisely, the boxplot for λ_{p+1-i} is compared to the value of the $100\,(i - 1/2)\,/p$ quantile of the Wachter limit. For $i = 1, 2, ..., 10$, these are the 5-th,15-th,...,95-th quantiles of $W_c\,(\lambda)$. We see that, even for such small values of p and T, theoretical quantiles track location of the MC distributions of the empirical quantiles very well.

The dispersion of the MC distributions around the corresponding theoretical quantiles is quite large for the chosen small values of p and T. It is slightly smaller for the Gaussian case, not reported here. To see how such a dispersion changes when p and T increase while p/T remains fixed, we generate pure random walk data with $p = 20, 100$ and $T = 200, 1000$ for $p/T = 1/10$, and $p = 20, 100$ and $T = 40, 200$ for

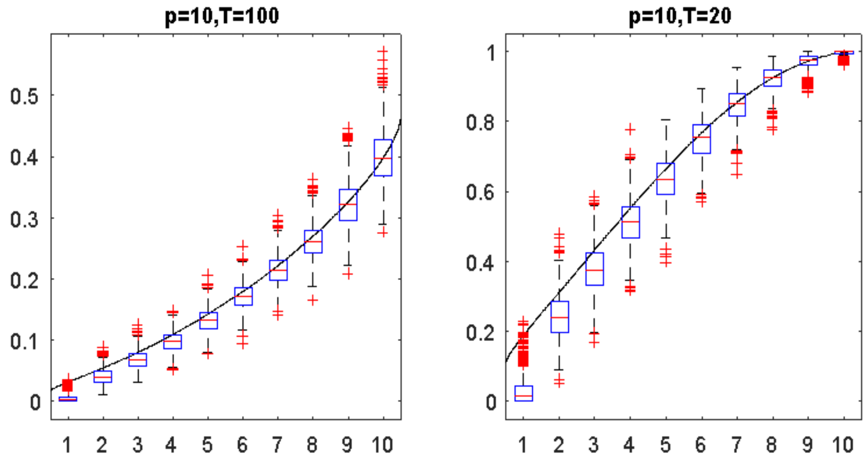

Figure 3: The Tukey boxplots for 1000 MC simulations of ten sample squared canonical correlations correponding to pure random walk data. The boxplots are superimposed with the quantile function of the Wachter limit.

$p/T = 1/2$.

Instead of reporting the Tukey boxplots, we plot only the 5-th and 95-th percentiles of the MC distributions of the λ_{p+1-i}, $i = 1, ..., p$ against $100\,(i - 1/2)\,/p$ quantiles of the corresponding Wachter limit. The plots are shown in Figure 4. We see that the [5%,95%] ranges of the MC distributions of λ_{p+1-i} are still considerably large for $p = 20$. These ranges become much smaller for $p = 100$.

The behavior of the smallest squared canonical correlation λ_p (that is, λ_{p+1-i} with $i = 1$) in Figures 3 and 4 is special in that its MC distribution lies below the corresponding Wachter quantile. This does not contradict our theoretical results because a weak limit of the empirical distribution of λ's is not affected by an arbitrary change in a finite (or slowly growing) number of them.

Our next experiment simulates data with the number of cointegrating relationships, r, equal to $1, 2, 3$, and p. In each case, we set the first r diagonal elements of matrix Π to $\rho = -1$, leaving the other elements equal zero. The sample size is $(p, T) = (20, 200)$. Figure 5 shows the 5-th and 95-th percentiles of the MC distributions of λ_{p+1-i} (solid lines) plotted against the $100(i - 1/2)/p$ quantiles of the corresponding Wachter limit.

Interestingly, exactly r squared canonical correlations deviate from the 45° line. This remains to be the case when we set the first r diagonal elements of Π to $\rho = -0.75$, or when we increase the sample size to $(p, T) = (100, 1000)$. When ρ

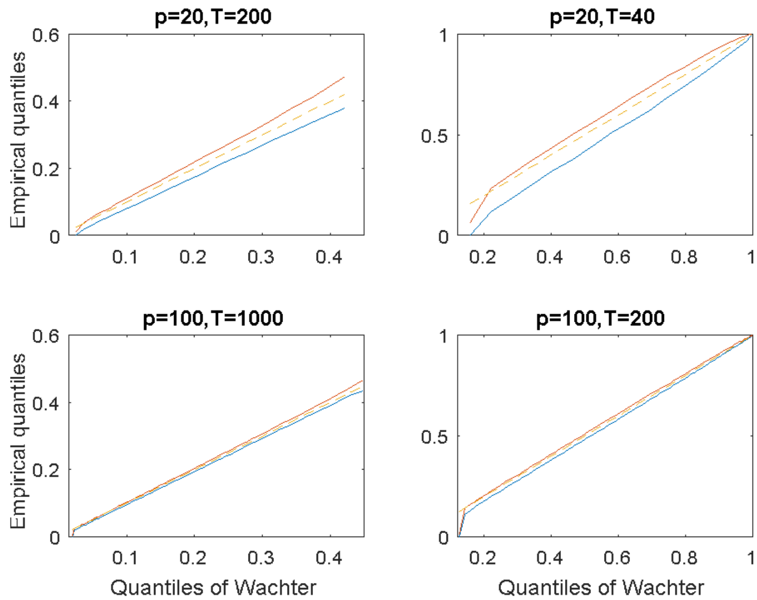

Figure 4: The 5-th and the 95-th percentiles of the MC distributions of λ_{p+1-i}, which are plotted against $100(i-1/2)/p$ quantiles of the Wachter limit. The dashed line is the 45° line. Pure random walk data.

is further increased to -0.5 so that the stationary components of the data become less persistent, the deviations from the 45° line become less pronounced.

The remarkable fact that the number of the squared canonical correlations deviating from the 45° line equals the cointegrating rank cannot be explained by Theorem 1. It is because the limiting empirical distribution of the squared canonical correlations is insensitive to the asymptotic behavior of any finite number of them. We leave asymptotic analysis of individual squared canonical correlations, as opposed to their empirical distribution, for future research.

Plots of squared canonical correlations against the corresponding quantiles of the Wachter distribution are known in the statistical literature as *Wachter plots*. They were proposed by Wachter (1976) in the context of multiple discriminant analysis as a tool to "recognize hopeless from promising analyses at an early stage." Results reported in Figure 5 suggest that counting the number of points where a Wachter plot deviates from the 45° line might be useful for the determination of cointegration rank.

For the interested reader, we now provide details on how to construct a Wachter plot. First, find the squared canonical correlations $\lambda_1 \geq ... \geq \lambda_p$ by computing the

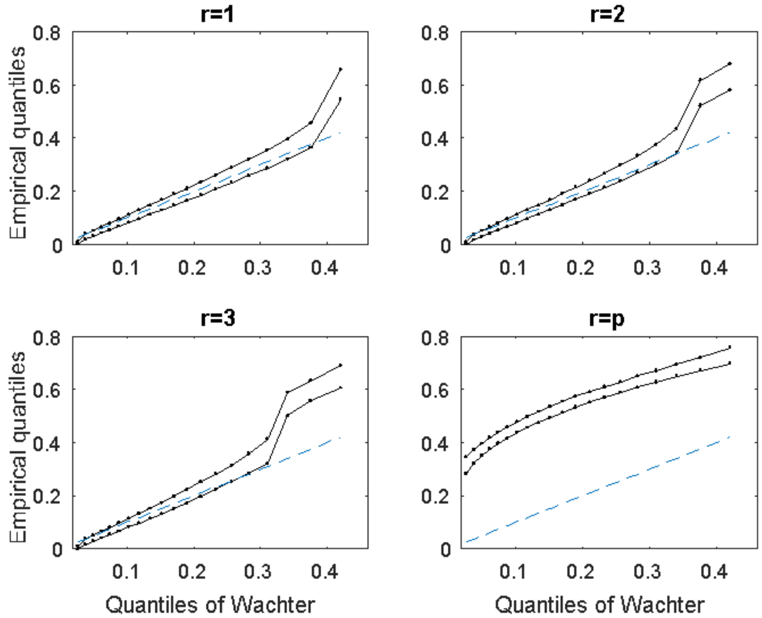

Figure 5: The 5-th and 95-th quantiles of the MC distribution of λ_{p+1-i} plotted against $100(i-1/2)/p$ quantiles of $W_{1/10}(\lambda)$. The number of cointegrating relationships $r \neq 0$. $(p,T)=(20,200)$.

eigenvalues of $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$. Next, set $c=p/T$. Using the Wachter density formula (9) with parameters $\gamma_1=c/(1+c)$ and $\gamma_2=2c/(1+c)$, compute the $100\,(i-1/2)\,/p$ quantiles of $W_c\,(\lambda)$ for $i=1,2,...,p$. Finally, plot points with x-coordinates equal to the computed quantiles and y-coordinates equal to the corresponding λ_{p+1-i}. A MATLAB code for the construction of a Wachter plot is available from the authors upon request.

Our final MC experiment studies the finite sample behavior of $LR_{0,p,T}/p^2$. The simulated data are pure random walk. Figure 6 shows the Tukey boxplots of the MC distributions of $LR_{0,p,T}/p^2$ corresponding to $p/T=1/10,...,5/10$ with $p=10$ (left panel), and $p=100$ (right panel). The boxplots are superimposed with the plot of the simultaneous asymptotic lower bound $\underline{LR}_c$ with c replaced by p/T. For $p=10$, we also show (horizontal dashed line) the standard 95% asymptotic critical value (scaled by $1/p^2$) taken from MacKinnon et al. (1999, Table II). For $p=100$, the standard critical values are not available, and we show the dashed horizontal line at height 2 instead. This is the sequential asymptotic probability limit of $LR_{0,p,T}/p^2$ as established in Theorem 5.

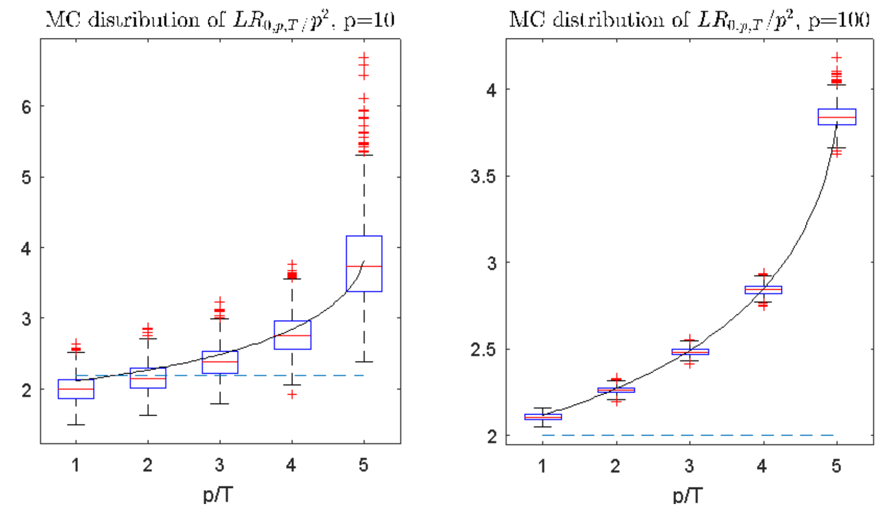

Figure 6: The Tukey boxplots for the MC distributions of $LR_{0,p,T}/p^2$ for various p/T ratios. The boxplots are superimposed with the simultaneous asymptotic lower bound $\underline{LR}_c$. Dashed line in the left panel correspond to 95% critical value for the satandard asymptotic LR test (taken from MacKinnon et al. (1999, Table II)). Dashed line in the right panel has ordinate equal two.

The left panel of Figure 6 illustrates the over-rejection phenomenon. The horizontal dashed line that corresponds to the standard 95% critical value is just above the interquartile range of the MC distribution of $LR_{0,p,T}/p^2$ for $c = 1/10$, is below this range for $c \geq 3/10$, and is below all 1000 MC replications of the scaled LR statistic for $c = 5/10$.

The SM contains two additional MC experiments, where we explore the sensitivity of the empirical distribution of the squared canonical correlations to the nuisance parameters Ψ and Γ. We find that the effect of Ψ and Γ is mostly confined to a few of the largest squared canonical correlations. For example, when Γ_1 is a rank-one matrix with a sufficiently large norm, the largest squared canonical correlation becomes substantially larger than the $100\,(p - 1/2)\,/p$ quantile of the Wachter limit. However, the MC distributions of the other squared canonical correlations do not substantially change, and the entire empirical distribution remains close to the Wachter distribution in terms of the Lévy distance.

5 Conclusion

In this paper, we consider the simultaneous, large-p, large-T, asymptotic behavior of the squared sample canonical correlations between p-dimensional, not necessarily Gaussian, random walk and its innovations. We find that the empirical distribution of these squared sample canonical correlations weakly converges in probability to the so-called *Wachter distribution* with parameters that depend only on the limit of p/T as $p, T \to_c \infty$. In contrast, under the sequential asymptotics, when first $T \to \infty$ and then $p \to \infty$, we establish the convergence in probability to the so-called Marchenko-Pastur distribution. The differences between the limiting distributions under the simultaneous and sequential asymptotics allow us to explain from a theoretical point of view the tendency of the LR test for cointegration to severely over-reject the null when the dimensionality of the data is relatively large.

The Monte Carlo analysis shows that the quantiles of the Wachter distribution constitute very good centering points for the finite sample distributions of the corresponding squared sample canonical correlations. The quality of the centering is excellent even for such small p and T as $p = 10$ and $T = 20$. However, for such small values of p and T, the empirical distribution of the squared sample canonical correlation can considerably fluctuate around the Wachter limit. As p increases to 100, the fluctuations become numerically very small.

This paper opens up many directions for future research. For example, it would be interesting to study the simultaneous asymptotic behavior of a few of the largest sample canonical correlations. As our Monte Carlo analysis suggests, when $r \neq 0$, exactly r of the squared canonical correlations deviate from the corresponding Wachter quantiles. Hence, the Wachter plot may potentially be useful for the determination of the cointegration rank in high dimensional systems.

It would also be interesting to study the first order simultaneous asymptotic behavior of the centered and scaled LR statistic. This paper has established the lower asymptotic bound on LR/p^2. We conjecture that, after centering by this bound and proper scaling, LR/p^2 is distributed normally, at least when ε has sufficiently many moments. We are currently investigating this research direction.

References

[1] Bai, Z. D. and J. W. Silverstein, (2010) *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer Verlag, New York.

[2] Banerjee, A., Marcellino, M., and Osbat, C. (2004) "Some cautions on the use of panel methods for integrated series of macroeconomic data." *The Econometrics Journal* 7 (2), 322–340.

[3] Billingsley, P. (1995) *Probability and Measure,* 3rd Ed., John Wiley & Sons.

[4] Breitung, J., and M.H. Pesaran (2008) "Unit Roots and Cointegration in Panels," in the Econometrics of Panel Data, edited by L. Matyas and P. Sevestre, 279-322. Berlin: Springer.

[5] Choi, I. (2015) Panel Cointegration. In Baltagi B. H. edt. The Oxford Handbook of Panel Data. Oxford University Press.

[6] Davis, G.C. (2003) "The generalized composite commodity theorem: Stronger support in the presence of data limitations," *Review of Economics and Statistics* 85, 476-480.

[7] Engel, C., N.C. Mark, and K.D. West (2015) "Factor Model Forecasts of Exchange Rates," *Econometric Reviews* 34, 32-55.

[8] Gonzalo, J. and J-Y Pitarakis (1995) "Comovements in Large Systems," Working Paper 95-38, Statistics and Econometrics Series 10, Universidad Carlos III de Madrid.

[9] Gonzalo, J. and J-Y Pitarakis (1999) "Dimensionality Effect in Cointegration Analysis," Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive WJ Granger, Oxford University Press, Oxford, 212-229

[10] Ho, M.S., and B.E. Sorensen (1996) "Finding Cointegration Rank in High Dimensional Systems Using the Johansen Test: An Illustration Using Data Based Mote Carlo Simulations," *Review of Economics and Statistics* 78, 726-732.

[11] Johansen, S. (1988) "Statistical Analysis of Cointegrating Vectors," *Journal of Economic Dynamics and Control* 12, 231-254

[12] Johansen, S. (1991) "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica* 59, 1551-1580.

[13] Johansen, S. (2002) "A small sample correction for the test of cointegrating rank in the vector autoregressive model," *Econometrica* 70, 1929-1961.

[14] Johansen, S., H. Hansen, and S. Fachin (2005) "A simulation study of some functionals of random walk," manuscript available at http://www.math.ku.dk/~sjo/.

[15] Lewbel, A. (1996) "Aggregation without separability: a generalized composite commodity theorem," *American Economic Review* 86, 524-543.

[16] MacKinnon, J.G., A. A. Haug, and L. Michelis (1999) "Numerical Distribution Functions of Likelihood ratio Tests for Cointegration," *Journal of Applied Econometrics* 14, 563-577.

[17] Marchenko, V. A. and Pastur, L. A. (1967) "Distribution for some sets of random matrices," *Math. USSR-Sbornik* 1, 457–483.

[18] Muirhead, R.J. (1982) *Aspects of Multivariate Statistical Theory*, John Wiley & Sons.

[19] Wachter, K. W. (1976). "Probability Plotting of Multiple Discriminant Ratios," *Proceedings of the Social Statistics Section of the American Statistical Association, Part II* 830-833. Prindle, Weber and Schmidt, Boston.

[20] Wachter, K. W. (1980). "The limiting empirical measure of multiple discriminant ratios," *Annals of Statistics* 8, 937–957.

[21] Yang, Y. and G. Pan (2012) "The convergence of the empirical distribution of canonical correlation coefficients," *Electronic Journal of Probability* 64, 1-13.

Supplementary Material for "Alternative asymptotics for cointegration tests in large VARs."

Alexei Onatski and Chen Wang
Faculty of Economics, University of Cambridge

March 5, 2018

Abstract

This note contains supplementary material for Onatski and Wang (2017) (OW in what follows). It is lined up with sections in the main text to make it easy to locate the required proofs.

Contents

1 Introduction **2**
1.1 There is no supplementary material for this section of OW. 2

2 Convergence to the Wachter distribution **2**
2.1 Proof of Theorem OW1 2
2.1.1 Reduction to random walk 2
2.1.2 Stieltjes transform 5
2.1.3 Reduction to Gaussianity 7
2.1.4 Identities for Stieltjes transform 15
2.1.5 From identities to a system of approximate equations 23
2.1.6 Solving the system 30
2.2 Proof of Corollary OW2 32
2.3 Proof of Corollary OW3 34

3 Sequential asymptotics and over-rejection **36**
3.1 Proof of Theorem OW4 36
3.1.1 Proof of Lemma 17 (Lévy distance between F_{p,T_γ} and $F_{p,T}$ is small) 37
3.1.2 Proof of Lemma 18 (lower bound on the smallest eigenvalue of pD_γ/T_γ) 39
3.1.3 Proof of Lemma 19 (α_D and α_C are small) 40
3.2 Proof of Theorem OW5 44
3.2.1 Showing that $\int \lambda \mathrm{d}F_0(\lambda) = 2$ 45
3.2.2 Convergence of the second largest eigenvalue 45
3.2.3 Asymptotic behavior of the largest eigenvalue 46
3.2.4 Proof of Lemma 20 (lower bound on $\tilde{\lambda}_{1,pT}$ in terms of $\tilde{\lambda}_{1,pT_\gamma}$) 48
3.2.5 Proof of Lemma 21 (lower bound on $\tilde{\lambda}_{1,pT_\gamma}$) 51

4 Monte Carlo **63**
4.1 Sensitivity to Ψ 63
4.2 Sensitivity to Γ 64

1 Introduction

1.1 There is no supplementary material for this section of OW.

2 Convergence to the Wachter distribution

2.1 Proof of Theorem OW1

Throughout the proof, we will assume without loss of generality that $p, T \to_c \infty$ so that p is strictly increasing, and thus, T is a function of p. This convention allows us to index various quantities that depend on p and T by p only, which simplifies notation. Various constants will be often denoted as K. The value of K may change from one appearance to another.

2.1.1 Reduction to random walk

In this section, we prove that the following three simplifications incur no loss of generality. First, instead of data generating process OW(11), we may consider pure random walk with zero initial values. Second, instead of defining S_{00}, S_{01}, and S_{11} as in OW(8), we may redefine them as

$$S_{00} = \frac{1}{T}\varepsilon M_l \varepsilon', \; S_{01} = \frac{1}{T}\varepsilon M_l U' M_l \varepsilon', \text{ and } S_{11} = \frac{1}{T}\varepsilon M_l U M_l U' M_l \varepsilon', \tag{1}$$

where $\varepsilon = [\varepsilon_1, ..., \varepsilon_T]$, M_l is the projection on the space orthogonal to the constant vector $l = (1, 1, ..., 1)'$, and U is the upper triangular matrix with ones above the diagonal and zeros on the diagonal. As we shall see below, $M_l, M_l U' M_l$, and $M_l U M_l U' M_l$ are circulant matrices (see Golub and Van Loan (1996, ch. 4.7.7)). Therefore, they are simultaneously diagonalizable, which makes the second simplification desirable. Finally, we may assume that the variance of ε_t equals I_p for any $t = 1, ..., T$.

We need the following two auxiliary lemmas. Let $\{G_p(\lambda)\}$ and $\{\tilde{G}_p(\lambda)\}$ be sequences of random distribution functions (d.f.'s). We call these sequences asymptotically equivalent in probability, $G_p \overset{\text{P}}{\sim} \tilde{G}_p$, if the Lévy distance $\mathcal{L}(G_p, \tilde{G}_p)$ converges in probability to zero as $p, T \to_c \infty$. Since Lévy distance metrizes the weak convergence, if $G_p \overset{\text{P}}{\sim} \tilde{G}_p$ and $G_p \overset{\text{P}}{\Rightarrow} F$ (that is, $G_p(\lambda)$ weakly converges to F, in probability), then $\tilde{G}_p \overset{\text{P}}{\Rightarrow} F$ too, and vice versa. We define a.s. asymptotic equivalence similarly, and denote it as $G_p \overset{\text{a.s.}}{\sim} \tilde{G}_p$.

Let S_i and $\tilde{S}_i$ with $i = 0, 1, 2$ be random $p \times p$ matrices, and let S_i^{-1} and $\tilde{S}_i^{-1}$ be their Moore-Penrose generalized inverses (see Horn and Johnson (1985), p. 421).

Lemma 1 *If* $p^{-1} \operatorname{rank}\left(S_i - \tilde{S}_i\right) \overset{\text{a.s.}}{\to} 0$ *as* $p, T \to_c \infty$ *for* $i = 0, 1, 2$, *then* $G_p \overset{\text{a.s.}}{\sim} \tilde{G}_p$, *where* $G_p(\lambda)$ *and* $\tilde{G}_p(\lambda)$ *are the empirical d.f.'s of eigenvalues of* $S_2 S_1^{-1} S_2' S_0^{-1}$ *and* $\tilde{S}_2 \tilde{S}_1^{-1} \tilde{S}_2' \tilde{S}_0^{-1}$, *respectively.*

Proof. Let $R = \operatorname{rank}(S_2 S_1^{-1} S_2' S_0^{-1} - \tilde{S}_2 \tilde{S}_1^{-1} \tilde{S}_2' \tilde{S}_0^{-1})$. The convergence $p^{-1} \operatorname{rank}(S_i - \tilde{S}_i) \overset{\text{a.s.}}{\to} 0$ implies that $R/p \overset{\text{a.s.}}{\to} 0$. On the other hand, by the rank inequality (Theorem A.43 in Bai and Silverstein (2010)), $\mathcal{L}(G_p, \tilde{G}_p) \le R/p$.□

Let $X = [X_{-k+1}, ..., X_T]$, where $X_{-k+1}, ..., X_0$ are arbitrary initial values and X_t with $t \ge 1$ are generated by OW(11), that is

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Psi F_t + \varepsilon_t.$$

Further, let $\tilde{X}_{-k+1}, ..., \tilde{X}_0$ be zero vectors, $\tilde{X}_t = \sum_{s=1}^{t} \varepsilon_t$ for $t \ge 1$, and $\tilde{X} = [\tilde{X}_{-k+1}, ..., \tilde{X}_T]$.

Lemma 2 $\operatorname{rank}\left(X - \tilde{X}\right) \le 2\left(r + \operatorname{rank}\Gamma + k + d_F\right).$

Proof. Write X_t in the VAR(k) form,

$$X_t = \sum_{i=1}^{k} \Pi_i X_{t-i} + \Psi F_t + \varepsilon_t,$$

where Π_i are such that $\Pi = \sum_{i=1}^{k} \Pi_i - I_p$ and $\Gamma_i = -\sum_{j=i+1}^{k} \Pi_j$. Express X_t as a function of the initial values, ε_t and F_t (see Johansen (1995, thm 2.1))

$$X_t = \sum_{s=1}^{k} C_{t-s} \sum_{i=1}^{k-s+1} \Pi_{s+i-1} X_{1-i} + \sum_{j=0}^{t-1} C_j \left(\varepsilon_{t-j} + \Psi F_{t-j}\right), \tag{2}$$

where $C_0 = I$ and C_n is defined recursively by $C_n = \sum_{j=1}^{k\wedge n} C_{n-j}\Pi_j$, $n = 1, 2, ...$ Here $k \wedge n$ denotes the minimum of k and n. Let us denote $\Pi_1 - I_p$ as Π_1^* and let $\Pi_j^* = \Pi_j$ for $j \geq 2$. Then, for $n = 1, 2, ...,$

$$\Delta C_n = C_n - C_{n-1} = \sum_{j=1}^{k\wedge n} C_{n-j}\Pi_j^* = \sum_{j=1}^{n-1} \Delta C_{n-j} \sum_{s=1}^{j\wedge k} \Pi_s^* + \sum_{s=1}^{n\wedge k} \Pi_s^*. \tag{3}$$

Clearly the *column* space of ΔC_1 is spanned by the *column* spaces of Π_j^*, $j = 1, ..., k$. Use this as the basis of induction. Suppose that the column spaces of each of ΔC_j with $j < n$ are spanned by the column spaces of Π_j^*, $j = 1, ..., k$. The identity (3) then implies that the column space of ΔC_n is spanned by the column spaces of Π_j^*, $j = 1, ..., k$, too.

Now rewrite (2) as

$$X_t = \sum_{s=1}^{k} \sum_{h=0}^{t-s} \Delta C_h \sum_{i=1}^{k} \Pi_{s+i-1} X_{1-i} + \sum_{j=0}^{t-1} \sum_{h=0}^{j} \Delta C_h \left(\varepsilon_{t-j} + \Psi F_{t-j}\right),$$

where $\Delta C_0 = C_0 = I_p$, and $\Pi_j = 0$ for $j > k$. Represent X_t as the sum $X_t^{(0)} + X_t^{(1)}$ with

$$\begin{aligned} X_t^{(0)} &= \sum_{s=1}^{k} \sum_{i=1}^{k} \Pi_{s+i-1} X_{1-i} + \sum_{j=0}^{t-1} \left(\varepsilon_{t-j} + \Psi F_{t-j}\right), \text{ and} \\ X_t^{(1)} &= \sum_{s=1}^{k} \sum_{h=1}^{t-s} \Delta C_h \sum_{i=1}^{k} \Pi_{s+i-1} X_{1-i} + \sum_{j=0}^{t-1} \sum_{h=1}^{j} \Delta C_h \left(\varepsilon_{t-j} + \Psi F_{t-j}\right). \end{aligned} \tag{4}$$

Since the column spaces of each of ΔC_h with $h \geq 1$ are spanned by those of Π_j^*, $j = 1, ..., k$, the space spanned by $X_t^{(1)}$, $t = 1, ..., T$ is also spanned by the column spaces of Π_j^*, $j = 1, ..., k$. Since the union of the latter column spaces coincides with the union of the column spaces of Π and Γ, we have

$$\operatorname{rank} X^{(1)} \leq r + \operatorname{rank} \Gamma, \tag{5}$$

where $X^{(1)} = [X_{1-k}^{(1)}, ..., X_T^{(1)}]$ with zero $X_{1-k}^{(1)}, ..., X_0^{(1)}$, and $X_t^{(1)}$ with $t \geq 1$ defined by (4).

Next, represent $X_t^{(0)}$ as the sum $X_t^{(00)} + \tilde{X}_t$, where

$$X_t^{(00)} = \sum_{s=1}^{k} \sum_{i=1}^{k} \Pi_{s+i-1} X_{1-i} + \Psi \sum_{j=0}^{t-1} F_{t-j}, \text{ and } \tilde{X}_t = \sum_{j=0}^{t-1} \varepsilon_{t-j}, \tag{6}$$

and let $X^{(00)} = [X_{1-k}^{(00)}, ..., X_T^{(00)}]$ with $X_t^{(00)} = X_t$ for $t = 1-k, ..., 0$, and $X_t^{(00)}$ with $t \geq 1$ defined by (6). Note that the columns space $X^{(00)}$ is spanned by those of Π_j^*, $j = 1, ..., k$, the column space of the matrix of the initial conditions $[X_{1-k}, ..., X_0]$, and the column space of Ψ. Therefore,

$$\operatorname{rank} X^{(00)} \leq r + \operatorname{rank} \Gamma + k + d_F. \tag{7}$$

Since $X = X^{(1)} + X^{(00)} + \tilde{X}$, inequalities (5) and (7) yield the statement of the lemma.□

Proof of no loss of generality. Now we are ready to prove the absence of a loss of generality in the proposed simplifications. Rewrite definitions OW(8) in the following form

$$S_{00}=\frac{1}{T}\Delta XM_D\Delta X',\ S_{01}=\frac{1}{T}\Delta XM_DX_{-1}',\ \text{and}\ S_{11}=\frac{1}{T}X_{-1}M_DX_{-1}', \tag{8}$$

where $\Delta X=[\Delta X_1,...,\Delta X_T]$, $X_{-1}=[X_0,...,X_{T-1}]$, and M_D is the projection on the space orthogonal to the rows of matrix $[D_1,...,D_T]$. Let $\tilde{S}_{00},\tilde{S}_{01}$, and $\tilde{S}_{11}$ be defined similarly, by replacing ΔX and X_{-1} in (8) by $\Delta\tilde{X}=\left[\Delta\tilde{X}_1,...,\Delta\tilde{X}_T\right]$ and $\tilde{X}_{-1}=\left[\tilde{X}_0,...,\tilde{X}_{T-1}\right]$, respectively.

By the definitions of S_{00} and $\tilde{S}_{00}$,

$$\begin{aligned}\operatorname{rank}\left(S_{00}-\tilde{S}_{00}\right)&=\operatorname{rank}\left\{\frac{1}{T}\left(\Delta X-\Delta\tilde{X}\right)M_D\Delta X'+\frac{1}{T}\Delta\tilde{X}M_D\left(\Delta X-\Delta\tilde{X}\right)'\right\}\\&\leq 2\operatorname{rank}\left(\Delta X-\Delta\tilde{X}\right).\end{aligned}$$

On the other hand,

$$\operatorname{rank}\left(\Delta X-\Delta\tilde{X}\right)\vee\operatorname{rank}\left(X_{-1}-\tilde{X}_{-1}\right)\leq\operatorname{rank}\left(X-\tilde{X}\right),$$

where $a\vee b$ denotes the maximum of a and b. Therefore, by Lemma 2,

$$\operatorname{rank}\left(S_{00}-\tilde{S}_{00}\right)\leq 4\left(r+\operatorname{rank}\Gamma+k+d_F\right).$$

Similarly, we have

$$\operatorname{rank}\left(S_{11}-\tilde{S}_{11}\right)\vee\operatorname{rank}\left(S_{01}-\tilde{S}_{01}\right)\leq 4\left(r+\operatorname{rank}\Gamma+k+d_F\right).$$

Since by assumption, $\left(r+\operatorname{rank}\Gamma+k+d_F\right)/p\to 0$, Lemma 1 implies that the sequences of the empirical d.f.'s of the eigenvalues of $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$ and of $\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{01}'\tilde{S}_{00}^{-1}$ are a.s. asymptotically equivalent. Since $\tilde{S}_{00},\tilde{S}_{01}$, and $\tilde{S}_{11}$ can be thought of as the equivalents of S_{00},S_{01}, and S_{11} after the original data, X, were replaced by pure random walk with zero initial values, $\tilde{X}$, we conclude that such a replacement does not incur any loss of generality.

If the data generating process is pure random walk with zero initial values, then (8) can be rewritten as

$$S_{00}=\frac{1}{T}\varepsilon M_D\varepsilon',\ S_{01}=\frac{1}{T}\varepsilon M_DU'\varepsilon',\ \text{and}\ S_{11}=\frac{1}{T}\varepsilon UM_DU'\varepsilon'.$$

The differences of so defined S_{ij} and their counterparts in (1) are matrices of rank no larger than d_D+3. Indeed, for S_{00}, we have

$$\operatorname{rank}\left(\frac{1}{T}\varepsilon M_D\varepsilon'-\frac{1}{T}\varepsilon M_l\varepsilon'\right)\leq\operatorname{rank}\left(M_D-M_l\right)\leq d_D+1.$$

For S_{01}, we have

$$\begin{aligned}&\operatorname{rank}\left(\frac{1}{T}\varepsilon M_DU'\varepsilon'-\frac{1}{T}\varepsilon M_lU'M_l\varepsilon'\right)\leq\operatorname{rank}\left(M_DU'-M_lU'M_l\right)\\=&\operatorname{rank}\left(\left(M_D-M_l\right)U'+M_lU'\left(I_T-M_l\right)\right)\leq\operatorname{rank}\left(M_D-M_l\right)+\operatorname{rank}\left(I_T-M_l\right)\\\leq&\ d_D+2.\end{aligned}$$

Finally, for S_{11}, we have

$$\begin{aligned}&\operatorname{rank}\left(\frac{1}{T}\varepsilon UM_DU'\varepsilon'-\frac{1}{T}\varepsilon M_lUM_lU'M_l\varepsilon'\right)\leq\operatorname{rank}\left(UM_DU'-M_lUM_lU'M_l\right)\\=&\operatorname{rank}\left(\left(I_T-M_l\right)UM_DU'+M_lUM_DU'\left(I_T-M_l\right)+M_lU\left(M_D-M_l\right)U'M_l\right)\\\leq&\operatorname{rank}\left(M_D-M_l\right)+2\operatorname{rank}\left(I_T-M_l\right)\leq d_D+3.\end{aligned}$$

Therefore, by Lemma 1, there is no loss of generality in redefining S_{00}, S_{01}, and S_{11} as in (1). Hence, in the rest of the proof of Theorem OW1, we use the definitions (1), and assume that the data are p-dimensional pure random walk with zero initial values. Moreover, since the eigenvalues of $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$ are invariant with respect to the transformation $\varepsilon \mapsto \Sigma^{-1/2}\varepsilon$, we will assume that the columns of ε are standardized so that their variance equals I_p. □

Before proceeding to the next section, let us show that matrices involved in the definitions (1) are circulant, as mentioned above.

Lemma 3 *Matrices M_l, $M_lU'M_l$, and $M_lUM_lU'M_l$ are circulant.*

Proof. Matrix M_l is clearly circulant. Further, let e_j denote the j-th column of I_T, and let $Z = [e_2, ..., e_T, e_1]$. By definition, A is circulant if $ZAZ' = A$. Now note that M_l commutes with Z and

$$ZUZ' = U + le_1' - e_1l'.$$

Therefore,

$$ZM_lUM_lZ' = M_lZUZ'M_l = M_lUM_l.$$

Hence, ZUZ' is circulant. It remains to note that the transpose of a circulant matrix is a circulant matrix and the product of two circulant matrices is a circulant matrix.□

As is well known (see e.g. Golub and Van Loan (1996, ch. 4.7.7)), any $T \times T$ circulant matrix V with the first column v admits the diagonalization $V = \frac{1}{T}\mathcal{F}^* \operatorname{diag}(\mathcal{F}v)\mathcal{F}$, where

$$\mathcal{F} = \{\exp(-\mathrm{i}\omega_{s-1}(t-1))\}_{s,t=1}^{T} \tag{9}$$

with $\omega_s = 2\pi s/T$ is the discrete Fourier transform matrix. Here the star superscript denotes transposition and complex conjugation. Note that the first column of M_l equals $e_1 - l/T$, and that of M_lUM_l equals $\tau/T - l(T+1)/(2T)$ with $\tau = (1, 2, ..., T)'$. A direct calculation of the products of $\mathcal{F}$ and these vectors yields

$$\begin{aligned} M_l &= \frac{1}{T}\mathcal{F}^* \operatorname{diag}(0, I_{T-1})\mathcal{F}, \\ M_lUM_l &= \frac{1}{T}\mathcal{F}^* \operatorname{diag}\left(0, \hat{\nabla}^*\right)\mathcal{F}, \\ M_lUM_lU'M_l &= \frac{1}{T}\mathcal{F}^* \operatorname{diag}\left(0, \hat{\nabla}^*\hat{\nabla}\right)\mathcal{F}, \end{aligned} \tag{10}$$

where

$$\hat{\nabla} = \operatorname{diag}\left\{\left(e^{\mathrm{i}\omega_1} - 1\right)^{-1}, ..., \left(e^{\mathrm{i}\omega_{T-1}} - 1\right)^{-1}\right\}. \tag{11}$$

2.1.2 Stieltjes transform

Our proof of the weak convergence $F_p(\lambda) \overset{\mathrm{P}}{\Rightarrow} W_c(\lambda)$ consists of showing that the Stieltjes transforms of $F_p(\lambda)$,

$$m_p(z) = \int \frac{1}{\lambda - z} F_p(\mathrm{d}\lambda), \tag{12}$$

converge in probability pointwise in $z \in \mathbb{C}^+ \equiv \{\zeta : \Im\zeta > 0\}$, where $\Im\zeta$ denotes the imaginary part of a complex number ζ, to the Stieltjes transform $m(z)$ of the Wachter distribution. The fact that a pointwise convergence in probability of Stieltjes transforms implies the weak convergence in probability of the corresponding d.f.'s is mentioned in Chatterjee (2006). However, since we cannot find a proof in the literature, we provide details specific to our problem below.

In fact, we will prove the following more general fact.

Theorem 4 *Let $G_p(\lambda)$ and $\tilde{G}_p(\lambda)$ be two sequences of d.f.'s, supported on a subset of a fixed interval $[-b, b]$ with $b < \infty$, and such that the corresponding Stieltjes transforms $s_p(z)$ and $\tilde{s}_p(z)$ satisfy $|s_p(z) - \tilde{s}_p(z)| \overset{\mathrm{P}}{\to} 0$ as $p \to \infty$, pointwise in $z \in \mathbb{C}^+$. Then, $\mathcal{L}\left(G_p, \tilde{G}_p\right) \overset{\mathrm{P}}{\to} 0$.*

Proof. Suppose, as a matter of contradiction, that $\mathcal{L}\left(G_p, \tilde{G}_p\right)$ does not converge. Then, there exist $\epsilon, \delta > 0$ and a subsequence $\{p_k\}$, along which

$$\Pr\left(\mathcal{L}\left(G_{p_k}, \tilde{G}_{p_k}\right) > \epsilon\right) > \delta. \tag{13}$$

Let $G^c_{p_k}$ and $\tilde{G}^c_{p_k}$ be convolutions of G_{p_k} and $\tilde{G}_{p_k}$ with distribution Φ_σ of a zero mean Gaussian variable with variance σ^2 so small that

$$\Pr\left(\mathcal{L}\left(G^c_{p_k}, \tilde{G}^c_{p_k}\right) > \epsilon/2\right) > \delta. \tag{14}$$

Note that

$$\mathcal{L}\left(G^c_{p_k}, \tilde{G}^c_{p_k}\right) \leq \left\|G^c_{p_k} - \tilde{G}^c_{p_k}\right\| \equiv \sup_\lambda \left|G^c_{p_k}(\lambda) - \tilde{G}^c_{p_k}(\lambda)\right|.$$

By Theorem B.14 of Bai and Silverstein (2010)[1],

$$\begin{aligned}\left\|G^c_{p_k} - \tilde{G}^c_{p_k}\right\| \leq\ & K\int_{-5(1+b)}^{5(1+b)} \left|s^c_{p_k}(z) - \tilde{s}^c_{p_k}(z)\right| \mathrm{d}u + \frac{K}{v}\sup_x \int_{|y|<2\sqrt{3}v} \left|\tilde{G}^c_{p_k}(x+y) - \tilde{G}^c_{p_k}(x)\right| \mathrm{d}y \\ & + \frac{K}{v}\int_{|x|>b+1} \left|G^c_{p_k}(x) - \tilde{G}^c_{p_k}(x)\right| \mathrm{d}x,\end{aligned} \tag{15}$$

where $u \equiv \Re z$ is the real part of z, $v \equiv \Im z$, and $K > 0$ is an absolute constant. Note that the density of $\tilde{G}^c_{p_k}$ is bounded by $\left(\sqrt{2\pi}\sigma\right)^{-1}$, and that[2]

$$\int_{|x|>b+1} \left|G^c_{p_k}(x) - \tilde{G}^c_{p_k}(x)\right| \mathrm{d}x \leq 4\int_{-\infty}^{-1} \Phi_\sigma(y)\,\mathrm{d}y \leq 4\sigma^3 e^{-1/\left(2\sigma^2\right)}/\sqrt{2\pi}.$$

Choosing v, σ, v/σ, and $\sigma^3 e^{-1/\left(2\sigma^2\right)}/v$ all sufficiently small, we can make the sum of the second and the third term on the right hand side of inequality (15) smaller than $\epsilon/4$. Then from (14),

$$\Pr\left(K\int_{-5(1+b)}^{5(1+b)} \left|s^c_{p_k}(z) - \tilde{s}^c_{p_k}(z)\right| \mathrm{d}u > \epsilon/4\right) > \delta. \tag{16}$$

Since for any z_1, z_2 s.t. $\Im z_1 = \Im z_2 = v > 0$,

$$\left|s^c_{p_k}(z_1) - s^c_{p_k}(z_2)\right| \vee \left|\tilde{s}^c_{p_k}(z_1) - \tilde{s}^c_{p_k}(z_2)\right| \leq \left|\Re z_1 - \Re z_2\right|/v^2,$$

the integral in (16) is different from a Riemann sum with a sufficiently large, but finite, number of summands by less than $\epsilon/4$. Therefore, $\left|s^c_{p_k}(z) - \tilde{s}^c_{p_k}(z)\right|$ does not converge to zero in probability pointwise in $z \in \mathbb{C}^+$. But such a convergence does take place.

Indeed, suppose not. Then, there exist $z \in \mathbb{C}^+$ and $\epsilon_1, \delta_1 > 0$ such that

$$\Pr\left(\left|s^c_{p_r}(z) - \tilde{s}^c_{p_r}(z)\right| > \epsilon_1\right) > \delta_1 \tag{17}$$

along a subsequence $\{p_r\}$ of the subsequence $\{p_k\}$. But

$$\begin{aligned} s^c_{p_r}(z) &= \int_{-A}^{A} s_{p_r}(z-y)\,\mathrm{d}\Phi_\sigma(y) + \int_{|y|>A} s_{p_r}(z-y)\,\mathrm{d}\Phi_\sigma(y), \text{ and} \\ \tilde{s}^c_{p_r}(z) &= \int_{-A}^{A} \tilde{s}_{p_r}(z-y)\,\mathrm{d}\Phi_\sigma(y) + \int_{|y|>A} \tilde{s}_{p_r}(z-y)\,\mathrm{d}\Phi_\sigma(y). \end{aligned}$$

[1] In Bai and Silverstein's notation, we choose $a = \sqrt{3}$, $B = b+1$, and $A = 5B$. Such a choice yields $\gamma = 2/3 > 1/2$, and $\kappa = 3/\pi < 1$, so that the conditions of Theorem B.14 are satisifed.

[2] The first of the two displayed inequalities uses the fact that G_{p_k} and $\tilde{G}_{p_k}$ are supported on a subset of $[-b, b]$, whereas the second one uses the bound $\Phi_\sigma(y) \leq \frac{-\sigma}{y\sqrt{2\pi}} e^{-y^2/\left(2\sigma^2\right)}$ for $y < 0$, and an inequality for the incomplete Gamma function (see Olver (1997, p.67)).

The absolute value of the second integrals on the right hand sides of the above equalities can be made less than $\epsilon_1/4$ by choosing a sufficiently large A. The first integrals differ from corresponding Riemann sums with a sufficiently large, but finite, number N of summands by less than $\epsilon_1/8$. Therefore, from (17), we must have

$$\Pr\left(\left|\sum_{i=1}^{N}\left(s_{p_r}\left(z-y_i\right)-\tilde{s}_{p_r}\left(z-y_i\right)\right)\phi_\sigma\left(y_i\right)\left(y_i-y_{i-1}\right)\right|>\epsilon_1/4\right)>\delta_1, \tag{18}$$

where ϕ_σ is the derivative of Φ_σ. However, (18) is impossible because $s_{p_r}(z)-\tilde{s}_{p_r}(z)$ converge to zero in probability, pointwise in $z\in\mathbb{C}^+$.□

2.1.3 Reduction to Gaussianity

This section shows that we may assume Gaussianity of the data without loss of generality. Then, the rotational invariance of the multivariate Gaussian distribution allows us to effectively use the simultaneous diagonalization (10), which is a key element of the proof of Theorem OW1.

Let η be a $p\times T$ matrix with i.i.d. standard normal entries, independent from ε, let

$$S_{G00}=\frac{1}{T}\eta M_l\eta',\ S_{G01}=\frac{1}{T}\eta M_lU'M_l\eta',\ S_{G11}=\frac{1}{T}\eta M_lUM_lU'M_l\eta',$$

and let $F_{Gp}(\lambda)$ be the empirical d.f. of the eigenvalues of $S_{G01}S_{G11}^{-1}S_{G01}'S_{G00}^{-1}$. We would like to establish the asymptotic equivalence $F_p\overset{\mathrm{P}}{\sim}F_{Gp}$. This is achieved in four steps, similarly to Yang and Pan (2012), who study the convergence of the empirical d.f. of the squared canonical correlations between two independent, not necessarily Gaussian, white noise samples.

First, we show that $F_p\overset{\mathrm{P}}{\sim}F_{Gp}$ follows from $\hat{F}_p\overset{\mathrm{P}}{\sim}\hat{F}_{Gp}$, where $\hat{F}_p(\lambda)$ is the empirical d.f. of the eigenvalues of the product of two random projection matrices, and $\hat{F}_{Gp}(\lambda)$ its Gaussian counterpart. Second, we prove that $\hat{F}_p\overset{\mathrm{a.s.}}{\sim}\bar{F}_p$, where $\bar{F}_p$ is a version of $\hat{F}_p$ obtained by truncating, centralizing, and scaling the entries of ε. Third, we perturb the random projections to ensure the boundedness of a few of the related matrices. The size of the perturbation is captured by parameter t, so we denote the perturbed $\bar{F}_p(\lambda)$ as $\bar{F}_{pt}(\lambda)$, and perturbed $\hat{F}_{Gp}(\lambda)$ as $\hat{F}_{Gpt}(\lambda)$. We prove that $\bar{F}_p\overset{\mathrm{P}}{\sim}\hat{F}_{Gp}$ follows from the asymptotic equivalence $\bar{F}_{pt}\overset{\mathrm{P}}{\sim}\hat{F}_{Gpt}$ for all fixed $t>0$. Finally, we establish the asymptotic equivalence $\bar{F}_{pt}\overset{\mathrm{P}}{\sim}\hat{F}_{Gpt}$ for any fixed $t>0$ by using the generalization of the Lindeberg principle due to Chatterjee (2006).

Step 1: sufficiency of $\hat{F}_p\overset{\mathrm{P}}{\sim}\hat{F}_{Gp}$. Note that matrix $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$ and the product of two projections P_1P_2, where

$$\begin{aligned} P_1 &= \frac{1}{T}M_lU'M_l\varepsilon'\left(\frac{1}{T}\varepsilon M_lUM_lU'M_l\varepsilon'\right)^{-1}\varepsilon M_lUM_l \text{ and}\\ P_2 &= \frac{1}{T}M_l\varepsilon'\left(\frac{1}{T}\varepsilon M_l\varepsilon'\right)^{-1}\varepsilon M_l, \end{aligned}$$

have the same non-zero eigenvalues, and P_1P_2 has additional $T-p$ zero eigenvalues. Therefore, the empirical d.f. $\hat{F}_p(\lambda)$ of the eigenvalues of P_1P_2 satisfies

$$\hat{F}_p(\lambda)=\frac{p}{T}F_p(\lambda)+\frac{T-p}{T}\mathbf{1}_{\{\lambda\geq 0\}},$$

where $\mathbf{1}_{\{\lambda\geq 0\}}$ is the indicator function. Hence, it is sufficient to prove that $\hat{F}_p(\lambda)\overset{\mathrm{P}}{\sim}\hat{F}_{G,p}(\lambda)$, where the latter d.f. is the analogue of the former for the Gaussian data.

Step 2: truncation, centralization, and scaling. For each $m=1,2,...$, let T_m be such that $T_m>T_{m-1}$ and for all $T\geq T_m$, we have

$$m^3\mathbb{E}\left(\varepsilon_{11}^2\mathbf{1}_{\{|\varepsilon_{11}|>\sqrt{T}/m\}}\right)<2^{-m}.$$

Let $\delta_T = 1$ for $T < T_1$ and $\delta_T = 1/m$ for all $T \in [T_m, T_{m+1})$. Then, as $T \to \infty$, $\delta_T \to 0$ and $\delta_T \sqrt{T} \to \infty$. Furthermore,

$$T\delta_T^{-1} \Pr(|\varepsilon_{11}| > \delta_T \sqrt{T}) \le \delta_T^{-3} \mathbb{E}\left(\varepsilon_{11}^2 \mathbf{1}_{\{|\varepsilon_{11}| > \delta_T \sqrt{T}\}}\right) \le 1. \tag{19}$$

Let $\tilde{\varepsilon}_{ij} = \varepsilon_{ij} \mathbf{1}_{\{|\varepsilon_{ij}| \le \delta_T \sqrt{T}\}}$ and let P_1^{tr}, P_2^{tr} be the matrices P_1, P_2 with ε_{ij} replaced by $\tilde{\varepsilon}_{ij}$. Denote the empirical d.f. of the eigenvalues of $P_1^{tr} P_2^{tr}$ as $F_p^{tr}(\lambda)$, and $\Pr(|\varepsilon_{11}| > \delta_T \sqrt{T})$ as q_T. Then for any $\delta > 0$, by the rank inequality

$$\begin{aligned}
& \Pr(\mathcal{L}\left(F_p^{tr}, \hat{F}_p\right) > \delta) \le \Pr(\operatorname{rank}(P_1^{tr} P_2^{tr} - P_1 P_2) > \delta T) \\
\le \;& \Pr(\operatorname{rank}(P_1^{tr} - P_1) > \delta T/2) + \Pr(\operatorname{rank}(P_2^{tr} - P_2) > \delta T/2) \\
\le \;& 2 \Pr(\sum_{ij} \mathbf{1}_{\{|\varepsilon_{ij}| > \delta_T \sqrt{T}\}} > \delta T/2) \le 2 \Pr\left(|\sum_{ij} (\mathbf{1}_{\{|\varepsilon_{ij}| > \delta_T \sqrt{T}\}} - q_T)| > pT\left(\frac{\delta}{2p} - q_T\right)\right).
\end{aligned}$$

Applying Bernstein's inequality (see e.g. Bai and Silverstein (2010, p. 21)) to the latter probability, we obtain

$$\Pr(\mathcal{L}\left(F_p^{tr}, \hat{F}_p\right) > \delta) \le 4 \exp\left(-p^2 T (\frac{\delta}{2p} - q_T)^2/\delta\right).$$

By (19), $\delta/(2p) - q_T \ge \delta/(4p)$ for all sufficiently large p and $T(p)$ along the sequence $p, T \to_c \infty$. Therefore, for all sufficiently large p,

$$\Pr(\mathcal{L}\left(F_p^{tr}, \hat{F}_p\right) > \delta) \le 4e^{-bp}$$

for some $b > 0$. It then follows from the Borel-Cantelli lemma that

$$\mathcal{L}\left(F_p^{tr}, \hat{F}_p\right) \overset{\text{a.s.}}{\to} 0 \tag{20}$$

as $p, T \to_c \infty$.

Next, let $\bar{\varepsilon}_{ij} = \tilde{\varepsilon}_{ij} - \mathbb{E}\tilde{\varepsilon}_{ij}$ and $\bar{P}_1, \bar{P}_2$ be the matrices P_1^{tr}, P_2^{tr} with $\tilde{\varepsilon}_{ij}$ replaced by $\bar{\varepsilon}_{ij}$. Denote the empirical d.f. of the eigenvalues of $\bar{P}_1 \bar{P}_2$ as $\bar{F}_p(\lambda)$. Again, by the rank inequality, we have

$$\begin{aligned}
\mathcal{L}\left(\bar{F}_p, F_p^{tr}\right) &\le \frac{1}{T} \operatorname{rank}(\bar{P}_1 \bar{P}_2 - P_1^{tr} P_2^{tr}) \\
&\le \frac{1}{T} \operatorname{rank}(\bar{P}_1 - P_1^{tr}) + \frac{1}{T} \operatorname{rank}(\bar{P}_2 - P_2^{tr}).
\end{aligned}$$

Note that $\tilde{\varepsilon} - \bar{\varepsilon} = \mathbb{E}\tilde{\varepsilon}$ and that $\operatorname{rank}(\mathbb{E}\tilde{\varepsilon}) = 1$ by the i.i.d assumption. Therefore, we have

$$\begin{aligned}
& \frac{1}{T} \operatorname{rank}(\bar{P}_1 - P_1^{tr}) \\
\le \;& \frac{1}{T} \operatorname{rank}\left\{\frac{1}{T} M_l U' M_l \tilde{\varepsilon}' \left[\left(\frac{1}{T} \bar{\varepsilon} M_l U M_l U' M_l \bar{\varepsilon}'\right)^{-1} - \left(\frac{1}{T} \tilde{\varepsilon} M_l U M_l U' M_l \tilde{\varepsilon}'\right)^{-1}\right] \tilde{\varepsilon} M_l U M_l\right\} \qquad (21) \\
& + \frac{1}{T} \operatorname{rank}\left\{\frac{1}{T} M_l U' M_l (\tilde{\varepsilon} - \bar{\varepsilon})' \left(\frac{1}{T} \bar{\varepsilon} M_l U M_l U' M_l \bar{\varepsilon}'\right)^{-1} \bar{\varepsilon} M_l U M_l\right\} \\
& + \frac{1}{T} \operatorname{rank}\left\{\frac{1}{T} M_l U' M_l \tilde{\varepsilon}' \left(\frac{1}{T} \bar{\varepsilon} M_l U M_l U' M_l \bar{\varepsilon}'\right)^{-1} (\tilde{\varepsilon} - \bar{\varepsilon}) M_l U M_l\right\}.
\end{aligned}$$

The latter two ranks are no larger than $\operatorname{rank}(\mathbb{E}\tilde{\varepsilon}) = 1$. Since

$$\begin{aligned}
& \left(\frac{1}{T} \bar{\varepsilon} M_l U M_l U' M_l \bar{\varepsilon}'\right)^{-1} - \left(\frac{1}{T} \tilde{\varepsilon} M_l U M_l U' M_l \tilde{\varepsilon}'\right)^{-1} \\
= \;& \left(\frac{1}{T} \tilde{\varepsilon} M_l U M_l U' M_l \tilde{\varepsilon}'\right)^{-1} \left(\frac{1}{T} \tilde{\varepsilon} M_l U M_l U' M_l \tilde{\varepsilon}' - \frac{1}{T} \bar{\varepsilon} M_l U M_l U' M_l \bar{\varepsilon}'\right) \left(\frac{1}{T} \bar{\varepsilon} M_l U M_l U' M_l \bar{\varepsilon}'\right)^{-1},
\end{aligned}$$

the first rank on the right hand side of (21) is no larger than $2\,\mathrm{rank}(\mathbb{E}\tilde{\varepsilon}) = 2$.

To summarize, $\frac{1}{T}\,\mathrm{rank}(\bar{P}_1 - P_1^{tr}) \to 0$ and similarly, $\frac{1}{T}\,\mathrm{rank}(\bar{P}_2 - P_2^{tr}) \to 0$ as $p, T \to_c \infty$. Therefore $\mathcal{L}\left(\bar{F}_p, F_p^{tr}\right) \overset{\text{a.s.}}{\to} 0$, which, taken together with (20), yields

$$\mathcal{L}\left(\bar{F}_p, \hat{F}_p\right) \overset{\text{a.s.}}{\to} 0.$$

Since $\bar{P}_1\bar{P}_2$ is invariant to rescaling of $\bar{\varepsilon}$, and since $Var\left(\bar{\varepsilon}_{ij}\right) \to 1$ as $p, T \to_c \infty$, we may and will assume that $\bar{\varepsilon}_{ij}$ are i.i.d. with zero mean, unit variance, and satisfy $|\varepsilon_{ij}| \leq \bar{\delta}_T\sqrt{T}$, where $\bar{\delta}_T \to 0$ and $\bar{\delta}_T\sqrt{T} \to \infty$ as $p, T \to_c \infty$.

Step 3: perturbing the projections. Matrices U, $\left(\bar{\varepsilon}M_l\bar{\varepsilon}'\right)^{-1}$, and $\left(\bar{\varepsilon}M_lUM_lU'M_l\bar{\varepsilon}'\right)^{-1}$ involved in the definitions of projections $\bar{P}_1$ and $\bar{P}_2$ may have large norms, asymptotically. Therefore, we winsorize U and replace the two other matrices by matrices of bounded norms.

The winsorization of U is done as follows. Let

$$M_lU'M_l = \sum_{j=1}^{T}\sigma_j u_j v_j'$$

be a singular value decomposition of $M_lU'M_l$, where $\sigma_1 \geq ... \geq \sigma_T$ are the singular values. Throughout this note, we will assume that T is an odd integer. The case of even T can be analyzed similarly, and we omit the corresponding analysis. From (10), $\sigma_{2j-1} = \sigma_{2j} = \left(2 - 2\cos\omega_j\right)^{-1/2}$ for $j = 1, ..., (T-1)/2$, and $\sigma_T = 0$. Let

$$U_t' = \sum_{j=1}^{\rceil tp\lceil -1}\sigma_{\rceil tp\lceil}u_jv_j' + \sum_{j=\rceil tp\lceil}^{T-1}\sigma_ju_jv_j' + \frac{1}{2}u_Tv_T',$$

where t is a small positive number and $\rceil x\lceil$ denotes the smallest even integer larger than or equal to x.

For future reference, note that the norm of $M_lU'M_l$ is of order T^2, whereas

$$\|U_t\| = \sigma_{\rceil tp\lceil} = \left(2 - 2\cos\omega_{\rceil tp\lceil/2}\right)^{-1/2} \leq \sqrt{2}T/\left(\pi tp\right), \tag{22}$$

where the latter inequality uses the fact that $1 - \cos x \geq x^2/4$ for $x \in [0, \pi/2]$. For any fixed $t > 0$, the right hand side of (22) remains bounded as $p, T \to_c \infty$.

Let

$$\tilde{P}_{1t} = \frac{1}{T}U_t'\bar{\varepsilon}'\left(\frac{1}{T}\bar{\varepsilon}U_tU_t'\bar{\varepsilon}'\right)^{-1}\bar{\varepsilon}U_t \text{ and } \tilde{P}_2 = \frac{1}{T}\bar{\varepsilon}'\left(\frac{1}{T}\bar{\varepsilon}\bar{\varepsilon}'\right)^{-1}\bar{\varepsilon},$$

and let $\tilde{F}_{pt}\left(\lambda\right)$ be the empirical d.f. of the eigenvalues of $\tilde{P}_{1t}\tilde{P}_2$.

Lemma 5 *Let Y_1 and Y_2 be $n \times m$ matrices and let P_{Y_1} and P_{Y_2} be projections on the spaces spanned by the columns of Y_1 and Y_2, respectively. If $\mathrm{rank}\left(Y_1 - Y_2\right) = k$, then there exist $n \times k$ matrices y_1 and y_2 such that $P_{Y_1} - P_{Y_2} = P_{y_1} - P_{y_2}$, where P_{y_1} and P_{y_2} are projections on the spaces spanned by the columns of y_1 and y_2, respectively. In particular, $\mathrm{rank}\left(P_{Y_1} - P_{Y_2}\right) \leq 2k$.*

Proof. Assume that $Y_1 - Y_2 = ab$, where a is $n \times k$ and $b = \begin{pmatrix} 0, & I_k \end{pmatrix}$. This assumption does not lead to a loss of generality because P_{Y_1} and P_{Y_2} are invariant with respect to multiplication of Y_1 and Y_2 from the right by arbitrary invertible $m \times m$ matrices. Let us partition Y_1 and Y_2 as $[Y_{11}, Y_{12}]$ and $[Y_{21}, Y_{22}]$, where Y_{12} and Y_{22} are the last k columns of Y_1 and Y_2, respectively. We have $Y_{21} = Y_{11}$ and $Y_{22} + a = Y_{12}$. Denote $I_m - P_{Y_{21}}$ as M_1, where $P_{Y_{21}}$ is the projection on the space spanned by the columns of Y_{21}, and let $y_2 = M_1Y_{22}$. Note that

$$P_{Y_2} = P_{[Y_{21}, y_2]} = P_{Y_{21}} + P_{y_2},$$

where the second equality holds because Y_{21} is orthogonal to y_2. Similarly, we have

$$P_{Y_1} = P_{Y_{11}} + P_{y_1} = P_{Y_{21}} + P_{y_1},$$

where $y_1 = M_1 Y_{12}$. Therefore, $P_{Y_1} - P_{Y_2} = P_{y_1} - P_{y_2}$.□

By rank inequality (Bai and Silverstein (2010, thm. A.43)) and Lemma 5,

$$\begin{aligned} \mathcal{L}\left(\bar{F}_p, \tilde{F}_{pt}\right) &\leq \frac{1}{T}\operatorname{rank}(\bar{P}_1\bar{P}_2 - \tilde{P}_{1t}\tilde{P}_2) \\ &\leq \frac{1}{T}\operatorname{rank}(\bar{P}_1 - \tilde{P}_{1t}) + \frac{1}{T}\operatorname{rank}(\bar{P}_2 - \tilde{P}_2) \leq \frac{2\rceil tp\lceil + 2}{T}, \end{aligned} \tag{23}$$

which converges to zero as $p, T \to_c \infty$ and $t \to 0$. A similar inequality holds for the Gaussian analogues, $\hat{F}_{Gp}(\lambda), \tilde{F}_{Gpt}(\lambda)$, of d.f.'s $\bar{F}_p(\lambda), \tilde{F}_{pt}(\lambda)$.

Next, let

$$\bar{P}_{1t} = \frac{1}{T}U_t'\bar{\varepsilon}'\left(\frac{1}{T}\bar{\varepsilon}U_tU_t'\bar{\varepsilon}' + tI_p\right)^{-1}\bar{\varepsilon}U_t \text{ and } \bar{P}_{2t} = \frac{1}{T}\bar{\varepsilon}'\left(\frac{1}{T}\bar{\varepsilon}\bar{\varepsilon}' + tI_p\right)^{-1}\bar{\varepsilon},$$

and let $\bar{F}_{pt}(\lambda)$ be the empirical d.f. of the eigenvalues of $\bar{P}_{1t}\bar{P}_{2t}$ (for later use, we denote the Gaussian analogue of $\bar{F}_{pt}(\lambda)$ as $\hat{F}_{Gpt}(\lambda)$). Since the eigenvalues of $\bar{P}_{1t}\bar{P}_{2t}$ and $\bar{P}_{2t}\bar{P}_{1t}\bar{P}_{2t}$ coincide, and the eigenvalues of $\tilde{P}_{1t}\tilde{P}_2$ and $\tilde{P}_2\tilde{P}_{1t}\tilde{P}_2$ coincide, Corollary A.41 of Bai and Silverstein (2010) yields

$$\begin{aligned} \mathcal{L}^3\left(\bar{F}_{pt}, \tilde{F}_{pt}\right) &\leq \frac{1}{T}\operatorname{tr}\left(\bar{P}_{2t}\bar{P}_{1t}\bar{P}_{2t} - \tilde{P}_2\tilde{P}_{1t}\tilde{P}_2\right)^2 \\ &\leq \frac{K}{T}\operatorname{tr}\left(\bar{P}_{1t} - \tilde{P}_{1t}\right)^2 + \frac{K}{T}\operatorname{tr}\left(\bar{P}_{2t} - \tilde{P}_2\right)^2 \end{aligned} \tag{24}$$

for an absolute constant K.

On the other hand,

$$\begin{aligned} \frac{1}{T}\operatorname{tr}\left(\bar{P}_{1t} - \tilde{P}_{1t}\right)^2 &= \frac{1}{T}\operatorname{tr}\left(\frac{1}{T}\bar{\varepsilon}U_tU_t'\bar{\varepsilon}'\left[\left(\frac{1}{T}\bar{\varepsilon}U_tU_t'\bar{\varepsilon}' + tI_p\right)^{-1} - \left(\frac{1}{T}\bar{\varepsilon}U_tU_t'\bar{\varepsilon}'\right)^{-1}\right]\right)^2 \\ &= \frac{t^2}{T}\operatorname{tr}\left(\frac{1}{T}\bar{\varepsilon}U_tU_t'\bar{\varepsilon}' + tI_p\right)^{-2} \leq \frac{t^2}{T}\operatorname{tr}\left(\frac{1}{2T}\bar{\varepsilon}\bar{\varepsilon}' + tI_p\right)^{-2}, \end{aligned}$$

where the last inequality follows from the fact that, by construction, $U_tU_t' \geq I_T/2$. By Theorem 3.6 in Bai and Silverstein (2010), the empirical d.f. of the eigenvalues of $\frac{1}{T}\bar{\varepsilon}\bar{\varepsilon}'$ a.s. converges to the Marchenko-Pastur distribution. Hence, for any fixed $t > 0$,

$$\frac{t^2}{T}\operatorname{tr}\left(\frac{1}{2T}\bar{\varepsilon}\bar{\varepsilon}' + tI_p\right)^{-2} \stackrel{\text{a.s.}}{\to} 4ct^2\int_a^b \frac{\sqrt{(b-\lambda)(\lambda-a)}}{2\pi\lambda(\lambda+2t)^2}\mathrm{d}\lambda,$$

where $b = (1+\sqrt{c})^2$ and $a = (1-\sqrt{c})^2$.

For $c \in (0,1)$, the above a.s. limit is bounded by Kt^2, where K is a constant that depends only on c. For $c = 1$, we have

$$4t^2\int_0^4 \frac{\sqrt{(4-\lambda)\lambda}\mathrm{d}\lambda}{2\pi\lambda(\lambda+2t)^2} < t^2\int_0^\infty \frac{4\mathrm{d}\lambda}{\pi\lambda^{1/2}(\lambda+2t)^2} = \sqrt{t/2}.$$

In any case, the limit converges to zero as $t \to 0$. Similarly,

$$\frac{1}{T}\operatorname{tr}\left(\bar{P}_{2t} - \tilde{P}_2\right)^2 = \frac{t^2}{T}\operatorname{tr}\left(\frac{1}{T}\bar{\varepsilon}\bar{\varepsilon}' + tI_p\right)^{-2}$$

with the a.s. limit of the latter expression converging to zero as $t \to 0$.

Using (24), we arrive at the following result. With probability one, for any $\delta > 0$, there exists $t_\delta > 0$ s.t. for any $t \in (0, t_\delta)$

$$\limsup \mathcal{L}\left(\bar{F}_{pt}, \tilde{F}_{pt}\right) < \delta \tag{25}$$

as $p, T \to_c \infty$. A similar result holds for $\mathcal{L}\left(\hat{F}_{Gpt}, \tilde{F}_{Gpt}\right)$. Combining (23), (25), and similar inequalities for the Gaussian case, we conclude that the asymptotic equivalence $\bar{F}_p \stackrel{\text{P}}{\sim} \hat{F}_{Gp}$ would follow from $\bar{F}_{pt} \stackrel{\text{P}}{\sim} \hat{F}_{Gpt}$ for all fixed $t > 0$.

Step 4: using the Lindeberg principle. Let $\bar{m}_{pt}(z)$ and $m_{pt}(z)$ be the Stieltjes transforms of $\bar{F}_{pt}$ and $\hat{F}_{Gpt}$, respectively. By Theorem 4, the equivalence $\bar{F}_{pt} \overset{\mathrm{P}}{\sim} \hat{F}_{Gpt}$ would follow from the pointwise in $z \in \mathbb{C}^+$ convergence $|\bar{m}_{pt}(z) - m_{pt}(z)| \overset{\mathrm{P}}{\to} 0$. In this section, we establish the latter convergence using Chatterjee's (2006) extension of the Lindeberg principle.

The Lindeberg principle is a method of establishing the convergence in distribution of sums of independent random variables to a normal one by showing the closeness of the expectations of three-times differentiable functions of the original sums and sums of independent normals. The method is concisely described in Bentkus et al (2000). It has been extended by Chatterjee (2006) beyond sums, to cover nonlinear functions of random variables. We will use Chatterjee's Theorem 1.1 that we reproduce here for the reader's convenience.

Theorem 6 *(Chatterjee (2006)) Suppose X and Y are random vector in $\mathbb{R}^n$ with Y having independent components. For $1 \le i \le n$, let*

$$\begin{aligned} A_i &= \mathbb{E}\left|\mathbb{E}\left(X_i \left| X_1, ..., X_{i-1}\right| - \mathbb{E}\left(Y_i\right)\right)\right|, \\ B_i &= \mathbb{E}\left|\mathbb{E}\left(X_i^2 \left| X_1, ..., X_{i-1}\right| - \mathbb{E}\left(Y_i^2\right)\right)\right|. \end{aligned}$$

Let M_3 be a bound on $\max_i \left(\mathbb{E}\left|X_i\right|^3 + \mathbb{E}\left|Y_i\right|^3\right)$. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a thrice continuously differentiable function, and for $r = 1, 2, 3$, let $L_r(f)$ be a finite constant such that $|\partial_i^r f(x)| \le L_r(f)$ for each i and x, where ∂_i^r denotes the r-fold derivative in the i-th coordinate. Then,

$$|\mathbb{E}f(X) - \mathbb{E}f(Y)| \le \sum_{i=1}^{n} \left(A_i L_1(f) + \frac{1}{2} B_i L_2(f)\right) + \frac{1}{6} n L_3(f) M_3.$$

Let $n = pT$, $X = \operatorname{vec}(\bar{\varepsilon})$, and $Y = \operatorname{vec}(\eta)$ so that $A_i = B_i = 0$. Since $|X_i|^3 \le |X_i|^2 \bar{\delta}_T \sqrt{T}$ and Y_i are standard normals, we have

$$M_3 \le \bar{\delta}_T T^{1/2} + 2^{3/2}/\pi^{1/2}.$$

Further, let $g : \mathbb{R} \to \mathbb{R}$ be a thrice continuously differentiable function with bounded derivatives up to the third order. Finally, let $f(X) = g(\Re \bar{m}_{pt}(z))$ and $f(Y) = g(\Re m_{pt}(z))$. The following lemma is proven at the end of this section.

Lemma 7 *For any $z \in \mathbb{C}^+$ and $t > 0$, there exists $K > 0$ such that*

$$\left|\frac{\partial m_{pt}(z)}{\partial \eta_{ij}}\right| \le K T^{-3/2}, \left|\frac{\partial^2 m_{pt}(z)}{\partial \eta_{ij}^2}\right| \le K T^{-2}, \text{ and } \left|\frac{\partial^3 m_{pt}(z)}{\partial \eta_{ij}^3}\right| \le K T^{-5/2}.$$

Similar inequalities hold for the derivatives of $\bar{m}_{pt}(z)$ with respect to $\bar{\varepsilon}_{ij}$.

The lemma and Theorem 6 yield

$$|\mathbb{E}f(X) - \mathbb{E}f(Y)| \le K \frac{p}{T} \bar{\delta}_T$$

for some $K > 0$. This implies $|\mathbb{E}f(X) - \mathbb{E}f(Y)| \to 0$ as $p, T \to_c \infty$ because $\bar{\delta}_T \to 0$ and $p/T \to c$. Furthermore, since g can be *any* thrice continuously differentiable function with bounded derivatives, we have

$$|\Re \bar{m}_{pt}(z) - \Re m_{pt}(z)| \overset{\mathrm{P}}{\to} 0.$$

Similarly, setting $f(X) = g(\Im \bar{m}_{pt}(z))$ and $f(Y) = g(\Im m_{pt}(z))$, and using Lemma 7 and Theorem 6, we obtain $|\Im \bar{m}_{pt}(z) - \Im m_{pt}(z)| \overset{\mathrm{P}}{\to} 0$, and hence,

$$|\bar{m}_{pt}(z) - m_{pt}(z)| \overset{\mathrm{P}}{\to} 0,$$

as required.

It remains to prove Lemma 7. Recall that

$$m_{pt}(z) = \frac{1}{T}\operatorname{tr} B^{-1} \equiv \frac{1}{T}\operatorname{tr}\left(P_{1t}P_{2t} - zI_T\right)^{-1}$$

with

$$\begin{aligned} P_{1t} &= \frac{1}{T}U_t'\eta'\left(\frac{1}{T}\eta U_t U_t'\eta' + tI_p\right)^{-1}\eta U_t \equiv \frac{1}{T}U_t'\eta' D_t^{-1}\eta U_t, \\ P_{2t} &= \frac{1}{T}\eta'\left(\frac{1}{T}\eta\eta' + tI_p\right)^{-1}\eta \equiv \frac{1}{T}\eta' A_t^{-1}\eta. \end{aligned}$$

Note that

$$\begin{aligned} \frac{\partial \eta}{\partial \eta_{ij}} &= e_i e_j', \; \frac{\partial A_t^{-1}}{\partial \eta_{ij}} = -\frac{1}{T}A_t^{-1}\left(e_i e_j'\eta' + \eta e_j e_i'\right)A_t^{-1}, \text{ and} \\ \frac{\partial D_t^{-1}}{\partial \eta_{ij}} &= -\frac{1}{T}D_t^{-1}\left(e_i e_j' U_t U_t'\eta' + \eta U_t U_t' e_j e_i'\right)D_t^{-1}. \end{aligned}$$

Therefore, the chain rule for the derivative $\partial B/\partial \eta_{ij}$ yields a sum of terms, each of which has form

$$\pm\frac{1}{\sqrt{T}}M_L^{(1)}e_L e_R' M_R^{(1)}$$

where $e_L e_R'$ equals either $e_i e_j'$ or $e_j e_i'$ (the superscript '(1)' reminds us that the terms correspond to the first-order derivative of B). The "left" matrix $M_L^{(1)}$ belongs to the set

$$M_L^{(1)} \in \{U_t', P_{1t}, P_{1t}U_t', H_1, P_{1t}H_2, P_{1t}P_{2t}\} \tag{26}$$

with

$$H_1 = \frac{1}{\sqrt{T}}U_t'\eta' D_t^{-1} \text{ and } H_2 = \frac{1}{\sqrt{T}}\eta' A_t^{-1}.$$

The "right" matrix $M_R^{(1)}$ belongs to the set

$$M_R^{(1)} \in \{I_T, P_{2t}, U_t P_{2t}, H_2', H_1' P_{2t}, U_t P_{1t} P_{2t}\}. \tag{27}$$

For some constant $K > 0$ that depends on t, we have

$$\left\|M_L^{(1)}\right\| < K \text{ and } \left\|M_R^{(1)}\right\| < K.$$

To see this, note that $\|U_t\|, \|P_{1t}\|$, and $\|P_{2t}\|$ are clearly bounded, whereas

$$\begin{aligned} \|H_1\| &\leq \left\|\frac{1}{\sqrt{T}}U_t'\eta' D_t^{-1/2}\right\|\left\|D_t^{-1/2}\right\| = \|P_{1t}\|^{1/2}\left\|D_t^{-1/2}\right\| < K, \\ \|H_2\| &\leq \left\|\frac{1}{\sqrt{T}}\eta' A_t^{-1/2}\right\|\left\|A_t^{-1/2}\right\| = \|P_{2t}\|^{1/2}\left\|A_t^{-1/2}\right\| < K. \end{aligned}$$

Since

$$\frac{\partial m_{pt}(z)}{\partial \eta_{ij}} = -\frac{1}{T}\operatorname{tr}\left(B^{-1}\frac{\partial B}{\partial \eta_{ij}}B^{-1}\right)$$

and $\left\|B^{-1}\right\| \leq 1/\Im z$, we conclude that

$$\left|\frac{\partial m_{pt}(z)}{\partial \eta_{ij}}\right| \leq KT^{-3/2} \tag{28}$$

for some constant K that depends on t and z.

To obtain the second derivative $\partial^2 B/\partial\eta_{ij}^2$, we need to differentiate each term $\pm\frac{1}{\sqrt{T}}M_L^{(1)}e_Le_R'M_R^{(1)}$ in the expansion of the first derivative separately. This amounts to obtaining either $\partial M_L^{(1)}/\partial\eta_{ij}$ or $\partial M_R^{(1)}/\partial\eta_{ij}$. Given (26) and (27), we have

$$\frac{\partial M_L^{(1)}}{\partial\eta_{ij}} \in \left\{0, \frac{\partial P_{1t}}{\partial\eta_{ij}}, \frac{\partial P_{1t}}{\partial\eta_{ij}}U_t', \frac{\partial H_1}{\partial\eta_{ij}}, \frac{\partial P_{1t}}{\partial\eta_{ij}}H_2, P_{1t}\frac{\partial H_2}{\partial\eta_{ij}}, \frac{\partial\left(P_{1t}P_{2t}\right)}{\partial\eta_{ij}}\right\},$$

and

$$\frac{\partial M_R^{(1)}}{\partial\eta_{ij}} \in \left\{0, \frac{\partial P_{2t}}{\partial\eta_{ij}}, U_t\frac{\partial P_{2t}}{\partial\eta_{ij}}, \frac{\partial H_2'}{\partial\eta_{ij}}, \frac{\partial H_1'}{\partial\eta_{ij}}P_{2t}, H_1'\frac{\partial P_{2t}}{\partial\eta_{ij}}, U_t\frac{\partial\left(P_{1t}P_{2t}\right)}{\partial\eta_{ij}}\right\}.$$

Whatever the specific values of $\partial M_L^{(1)}/\partial\eta_{ij}$ and $\partial M_L^{(1)}/\partial\eta_{ij}$ are, these derivatives can be represented as sums of terms of the form

$$\pm\frac{1}{\sqrt{T}}M_Le_Le_R'M_R \tag{29}$$

where M_L and M_R may be different from $M_L^{(1)}$ and $M_R^{(1)}$, but must satisfy

$$\|M_L\| < K \text{ and } \|M_R\| < K.$$

To see this, it is sufficient to verify that any of the matrices

$$\frac{\partial P_{1t}}{\partial\eta_{ij}}, \frac{\partial P_{2t}}{\partial\eta_{ij}}, \frac{\partial H_1}{\partial\eta_{ij}}, \frac{\partial H_2}{\partial\eta_{ij}}, \text{ and } \frac{\partial\left(P_{1t}P_{2t}\right)}{\partial\eta_{ij}}$$

can be represented as a sum of terms of the form (29). For $\partial\left(P_{1t}P_{2t}\right)/\partial\eta_{ij}$, we have established this fact above. For the other matrices, we have

$$\begin{aligned}
\frac{\partial P_{1t}}{\partial\eta_{ij}} &= \frac{1}{\sqrt{T}}U_t'e_je_i'H_1' + \frac{1}{\sqrt{T}}H_1e_ie_j'U_t - \frac{1}{\sqrt{T}}H_1e_ie_j'U_tP_{1t} - \frac{1}{\sqrt{T}}P_{1t}U_t'e_je_i'H_1',\\
\frac{\partial P_{2t}}{\partial\eta_{ij}} &= \frac{1}{\sqrt{T}}e_je_i'H_2' + \frac{1}{\sqrt{T}}H_2e_ie_j - \frac{1}{\sqrt{T}}H_2e_ie_j'P_{2t} - \frac{1}{\sqrt{T}}P_2e_je_i'H_2',\\
\frac{\partial H_1}{\partial\eta_{ij}} &= \frac{1}{\sqrt{T}}U_t'e_je_i'D_t^{-1} - \frac{1}{\sqrt{T}}H_1e_ie_j'U_tH_1 - \frac{1}{\sqrt{T}}P_{1t}U_t'e_je_i'D_t^{-1}, \text{ and}\\
\frac{\partial H_2}{\partial\eta_{ij}} &= \frac{1}{\sqrt{T}}e_je_i'A_t^{-1} - \frac{1}{\sqrt{T}}H_2e_ie_j'H_2 - \frac{1}{\sqrt{T}}P_{2t}e_je_i'A_t^{-1}.
\end{aligned}$$

Hence, indeed, these matrices can be represented as sums of terms of the form (29). To summarize, the second derivative $\partial^2 B/\partial\varepsilon_{ij}^2$ can be represented as a sum of terms of the form

$$\pm\frac{1}{T}M_L^{(2)}e_Le_{ML}'M_M^{(2)}e_{MR}e_R'M_R^{(2)}$$

where the "left", "middle", and "right" matrices $M_L^{(2)}$, $M_M^{(2)}$, $M_R^{(2)}$ are products of constant matrices of bounded norm and terms of the form

$$P_{1t}, P_{2t}, H_1, H_1', H_2, H_2', P_{1t}P_{2t}, D_t^{-1}, A_t^{-1}. \tag{30}$$

In particular,

$$\left\|M_L^{(2)}\right\| < K, \left\|M_M^{(2)}\right\| < K, \left\|M_R^{(2)}\right\| < K.$$

On the other hand,

$$\frac{\partial^2 m_{pt}(z)}{\partial\eta_{ij}^2} = \frac{2}{T}\operatorname{tr}\left(B^{-1}\frac{\partial B}{\partial\eta_{ij}}B^{-1}\frac{\partial B}{\partial\eta_{ij}}B^{-1}\right) - \frac{1}{T}\operatorname{tr}\left(B^{-1}\frac{\partial^2 B}{\partial\eta_{ij}^2}B^{-1}\right).$$

Therefore,

$$\left|\frac{\partial^2 m_{pt}(z)}{\partial \eta_{ij}^2}\right| \leq KT^{-2}. \tag{31}$$

For the third derivative, the same logic implies that $\partial^3 B/\partial\eta_{ij}^3$ can be represented as a sum of terms of the form

$$\pm\frac{1}{T^{3/2}} M_L^{(3)} e_L e_{ML}' M_{ML}^{(3)} e_{LM} e_{RM}' M_{MR}^{(3)} e_{MR} e_R' M_R^{(3)},$$

where "left", "middle-left", "middle-right", and "right" matrices $M_L^{(3)}$, $M_{ML}^{(3)}$, $M_{MR}^{(3)}$, $M_R^{(2)}$ satisfy

$$\left\|M_L^{(3)}\right\| < K, \left\|M_{ML}^{(3)}\right\| < K, \left\|M_{MR}^{(3)}\right\| < K, \left\|M_R^{(3)}\right\| < K.$$

The arguments used to establish this fact remain the same as above. The only additional fact that needs to be established is that $\partial D_t^{-1}/\partial\eta_{ij}$ and $\partial A_t^{-1}/\partial\eta_{ij}$ can be represented as sums of terms of the form (29). The reason we need this is that D_t^{-1} and A_t^{-1} enter products defining $M_L^{(2)}$, $M_M^{(2)}$, $M_R^{(2)}$, in addition to matrices $P_{1t}, P_{2t}, H_1, H_1'$, H_2, and H_2' (see (30)). We have

$$\begin{aligned}\frac{\partial D_t^{-1}}{\partial\eta_{ij}} &= -\frac{1}{\sqrt{T}} D_t^{-1} e_i e_j' U_t H_1 - \frac{1}{\sqrt{T}} H_1' U_t' e_j e_i' D_t^{-1}, \\ \frac{\partial A_t^{-1}}{\partial\eta_{ij}} &= -\frac{1}{\sqrt{T}} A_t^{-1} e_i e_j' H_2 - \frac{1}{\sqrt{T}} H_2' e_j e_i' A_t^{-1}.\end{aligned}$$

Hence, indeed, $\partial D_t^{-1}/\partial\eta_{ij}$ and $\partial A_t^{-1}/\partial\eta_{ij}$ are sums of terms $\pm\frac{1}{\sqrt{T}} M_L e_L e_R' M_R$ with $\|M_L\| < K$ and $\|M_R\| < K$.

A straightforward calculation shows that

$$\begin{aligned}\frac{\partial^3 m_{pt}(z)}{\partial\eta_{ij}^3} &= -\frac{6}{T}\operatorname{tr}\left(B^{-1}\frac{\partial B}{\partial\eta_{ij}}B^{-1}\frac{\partial B}{\partial\eta_{ij}}B^{-1}\frac{\partial B}{\partial\eta_{ij}}B^{-1}\right) \\ &\quad +\frac{3}{T}\operatorname{tr}\left(B^{-1}\frac{\partial^2 B}{\partial\eta_{ij}^2}B^{-1}\frac{\partial B}{\partial\eta_{ij}}B^{-1}\right) \\ &\quad +\frac{3}{T}\operatorname{tr}\left(B^{-1}\frac{\partial B}{\partial\eta_{ij}}B^{-1}\frac{\partial^2 B}{\partial\eta_{ij}^2}B^{-1}\right) - \frac{1}{T}\operatorname{tr}\left(B^{-1}\frac{\partial^3 B}{\partial\eta_{ij}^3}B^{-1}\right).\end{aligned}$$

The facts that $\left\|B^{-1}\right\|$ is bounded, and that $\partial B/\partial\varepsilon_{ij}$, $\partial^2 B/\partial^2\varepsilon_{ij}$ and $\partial^3 B/\partial^3\varepsilon_{ij}$ can be represented as sum of terms, respectively,

$$\begin{aligned}&\pm\frac{1}{\sqrt{T}} M_L^{(1)} e_L e_R' M_R^{(1)}, \\ &\pm\frac{1}{T} M_L^{(2)} e_L e_{ML}' M_M^{(2)} e_{MR} e_R' M_R^{(2)}, \text{ and} \\ &\pm\frac{1}{T^{3/2}} M_L^{(3)} e_L e_{ML}' M_{ML}^{(3)} e_{LM} e_{RM}' M_{MR}^{(3)} e_{MR} e_R' M_R^{(3)}\end{aligned}$$

with norm-bounded matrices $M_j^{(1)}, M_j^{(2)}$, and $M_j^{(3)}$, implies that

$$\left|\frac{\partial^3 m_{pt}(z)}{\partial\eta_{ij}^3}\right| < KT^{-5/2}. \tag{32}$$

The proof of inequalities (28), (31), (32) for $\bar{m}_{pt}(z)$ is exactly the same as above, after η is replaced by $\bar{\varepsilon}$. □

2.1.4 Identities for Stieltjes transform

For the rest of the proof of Theorem OW1, we assume that the data are generated by

$$X_t = X_{t-1} + \eta_t,\ t = 1, ..., T$$

with i.i.d. $\eta_t \sim N(0, I_p)$ and $X_0 = 0$. Replacing ε by η in the definition (1) of S_{ij}, we obtain

$$S_{00} = \frac{1}{T}\eta M_l \eta',\ S_{01} = \frac{1}{T}\eta M_l U' M_l \eta',\ \text{and } S_{11} = \frac{1}{T}\eta M_l U M_l U' M_l \eta'.$$

Further, for the rest of the proof, we will assume that

$$c \in (0, 1). \tag{33}$$

For $c = 1$, the Wachter limit W_c equals the distribution having mass one at unity. On the other hand, matrix $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$ (with S_{11}^{-1} and S_{00}^{-1} being Moore-Penrose generalized inverses) equals I_p plus a matrix of rank converging to zero as $p, T \to_1 \infty$. Hence, Theorem OW1 holds.

Let $\hat{\eta} = \eta\mathcal{F}^*$ be a $p \times T$ matrix whose rows are the discrete Fourier transforms at frequencies $0, \omega_1, ..., \omega_{\bar{T}}$ of the rows of η. Here and in the rest of the proof, $\bar{T} = T - 1$. The discrete Fourier transform matrix $\mathcal{F}$ is as defined in (9). Further, let $\hat{\eta}_{-0}$ be the $p \times \bar{T}$ matrix obtained from $\hat{\eta}$ by removing its first column, corresponding to zero frequency. The diagonalization equations for M_l, $M_l U' M_l$, and $M_l U M_l U' M_l$, given in (10), yield

$$S_{00} = \frac{1}{T}\hat{\eta}_{-0}\hat{\eta}_{-0}^*,\ S_{01} = \frac{1}{T}\hat{\eta}_{-0}\hat{\nabla}\hat{\eta}_{-0}^*,\ \text{and } S_{11} = \frac{1}{T}\hat{\eta}_{-0}\hat{\nabla}^*\hat{\nabla}\hat{\eta}_{-0}^*.$$

Below we will work with real-valued sin and cos Fourier transforms of η. In addition, we will interchange the order of frequencies so that ω_{s_1} and ω_{s_2} with $s_1 + s_2 = T$ become adjacent pairs. Specifically, recall that T is assumed to be odd so that $\bar{T}$ is even. Let $P = \{p_{st}\}$ be a $\bar{T} \times \bar{T}$ permutation matrix with elements

$$p_{st} = \begin{cases} 1 \text{ if } s = 1, ..., \bar{T}/2 \text{ and } t = 2s - 1 \\ 1 \text{ if } s = \bar{T}/2 + 1, ..., \bar{T} \text{ and } t = 2\left(\bar{T} - s + 1\right) \\ 0 \text{ otherwise} \end{cases}, \tag{34}$$

and let

$$W = \frac{1}{\sqrt{2}} I_{\bar{T}/2} \otimes \begin{pmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{pmatrix}, \tag{35}$$

where $\otimes$ denotes the Kronecker product, and $\mathrm{i} = \sqrt{-1}$ is the imaginary unit. Further, let $\epsilon = \hat{\eta}_{-0}PW^*/\sqrt{\bar{T}T}$ and $\nabla = \operatorname{diag}\left\{\nabla_1, ..., \nabla_{\bar{T}/2}\right\}$ with

$$\nabla_j = -\frac{1}{2}\begin{pmatrix} 1 & -\cot(\omega_j/2) \\ \cot(\omega_j/2) & 1 \end{pmatrix}. \tag{36}$$

A direct calculation shows that

$$\nabla\nabla' = \nabla'\nabla = \operatorname{diag}\left\{r_1^{-1}I_2, ..., r_{\bar{T}/2}^{-1}I_2\right\} \text{ with } r_j = 4\sin^2(\omega_j/2). \tag{37}$$

Lemma 8 *The columns of ϵ are i.i.d. $N\left(0, I_p/\bar{T}\right)$ vectors. Matrix $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$ equals $CD^{-1}C'A^{-1}$ where*

$$C = \epsilon\nabla'\epsilon', D = \epsilon\nabla\nabla'\epsilon', \text{ and } A = \epsilon\epsilon'.$$

Proof. Let $P_+ = \operatorname{diag}\{1, P\}$ and $W_+ = \operatorname{diag}\{1, W\}$. Note that P_+ is an orthogonal matrix and W_+ is a unitary matrix. In particular, $P_+W_+^*W_+P_+' = I_T$. The statement about ϵ follows from the rotational invariance of Gaussian distribution and from the fact that $\mathcal{F}^*P_+W_+^*/\sqrt{T}$ is an orthogonal matrix. The statement about $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$ follows from a direct verification of the identity $WP'\hat{\nabla}PW^* = \nabla'$.□

The convenience of the $CD^{-1}C'A^{-1}$ representation of $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$ stems from the block-diagonality of ∇ and the diagonality of $\nabla\nabla'$. Let $\epsilon_{(j)}$ be a $p \times 2$ matrix that consists of the $(2j-1)$-th and the $2j$-th columns of ϵ. In particular, $\epsilon = \left[\epsilon_{(1)}, ..., \epsilon_{(\bar{T}/2)}\right]$. Then C, D, A can be represented as sums of independent components of rank two. Specifically,

$$C = \sum \epsilon_{(j)} \nabla_j' \epsilon_{(j)}', D = \sum r_j^{-1} \epsilon_{(j)} \epsilon_{(j)}', \text{ and } A = \sum \epsilon_{(j)} \epsilon_{(j)}'.$$

Below, we exploit these representations to derive the following identities that involve the Stieltjes transform $m_p(z)$ of the empirical d.f. of the eigenvalues of $CD^{-1}C'A^{-1}$.

$$m_p(z) = \frac{\bar{T}}{p}\frac{1}{1-z} - \frac{1}{p}\sum_{j=1}^{\bar{T}/2} \frac{1}{(1-z)^2} \operatorname{tr}\left(\left[I_2, r_j \nabla_j'\right] \Omega_j^{(q)} \left[I_2, r_j \nabla_j'\right]'\right), \quad (38)$$

$$\frac{\bar{T}}{p} + z m_p(z) = \frac{\bar{T}}{p}\frac{1}{1-z} - \frac{1}{p}\sum_{j=1}^{\bar{T}/2} \frac{1}{(1-z)^2} \operatorname{tr}\left(\left[I_2, r_j z \nabla_j'\right] \Omega_j^{(q)} \left[I_2, z r_j \nabla_j'\right]'\right), \quad (39)$$

$$1 + z m_p(z) = \frac{\bar{T}}{p}\frac{1}{1-z} - \frac{1}{p}\sum_{j=1}^{\bar{T}/2} \frac{1}{(1-z)^2} \operatorname{tr}\left(\left[I_2, r_j z \nabla_j'\right] \Omega_j^{(q)} \left[I_2, r_j \nabla_j'\right]'\right), \quad (40)$$

$$0 = \frac{1}{p}\sum_{j=1}^{\bar{T}/2} \frac{1}{1-z} \operatorname{tr}\left(\left[0, I_2\right] \Omega_j^{(q)} \left[I_2, r_j \nabla_j'\right]'\right), \quad (41)$$

where

$$\Omega_j^{(q)} \equiv \Omega_{pj}^{(q)}(z) = \begin{pmatrix} \frac{1}{1-z} I_2 + v_j^{(q)}(z) & \frac{r_j}{1-z}\nabla_j' + u_j^{(q)\prime}(z) \\ \frac{r_j}{1-z}\nabla_j + u_j^{(q)}(z) & \frac{r_j z}{1-z} I_2 + z\tilde{v}_j^{(q)}(z) \end{pmatrix}^{-1}, \quad (42)$$

and the 2×2 matrices $v_j^{(q)} \equiv v_j^{(q)}(z)$, $u_j^{(q)} \equiv u_j^{(q)}(z)$, and $\tilde{v}_j^{(q)} \equiv \tilde{v}_j^{(q)}(z)$ are defined as follows. Let

$$\begin{aligned} A_j &= A - \epsilon_{(j)}\epsilon_{(j)}', C_j = C - \epsilon_{(j)} \nabla_j' \epsilon_{(j)}', \ D_j = D - r_j^{-1} \epsilon_{(j)} \epsilon_{(j)}', \\ M_j &= C_j D_j^{-1} C_j' - z A_j, \text{ and } \tilde{M}_j = C_j' A_j^{-1} C_j - z D_j. \end{aligned}$$

Then,

$$v_j^{(q)} = \epsilon_{(j)}' M_j^{-1} \epsilon_{(j)}, \ u_j^{(q)} = \epsilon_{(j)}' D_j^{-1} C_j' M_j^{-1} \epsilon_{(j)}, \text{ and } \tilde{v}_j^{(q)} = \epsilon_{(j)}' \tilde{M}_j^{-1} \epsilon_{(j)}.$$

The entries of these matrices are quadratic forms in the columns of $\epsilon_{(j)}$. In what follows, we use superscript '(q)' to denote matrices that involve quadratic forms in the columns of $\epsilon_{(j)}$ to distinguish them from related matrices that do not involve such quadratic forms.

First, we establish the following lemma. Let

$$\begin{aligned} w_j^{(q)} &= \epsilon_{(j)}' D_j^{-1} C_j' M_j^{-1} C_j' D_j^{-1} \epsilon_{(j)}, \ \tilde{w}_j^{(q)} = \epsilon_{(j)}' A_j^{-1} C_j \tilde{M}_j^{-1} C_j A_j^{-1} \epsilon_{(j)}, \\ \tilde{u}_j^{(q)} &= \epsilon_{(j)}' A_j^{-1} C_j \tilde{M}_j^{-1} \epsilon_{(j)}, \ s_j^{(q)} = \epsilon_{(j)}' D_j^{-1} \epsilon_{(j)}, \text{ and } \tilde{s}_j^{(q)} = \epsilon_{(j)}' A_j^{-1} \epsilon_{(j)}. \end{aligned}$$

Further, let

$$\begin{aligned} & v_p \equiv v_p(z) = \operatorname{tr}\left(M^{-1}\right)/\bar{T}, && \tilde{v}_p \equiv \tilde{v}_p(z) = \operatorname{tr}\left(\tilde{M}^{-1}\right)/\bar{T}, \\ & u_p \equiv u_p(z) = \operatorname{tr}\left(D^{-1}C'M^{-1}\right)/\bar{T}, && \tilde{u}_p \equiv \tilde{u}_p(z) = \operatorname{tr}\left(A^{-1}C\tilde{M}^{-1}\right)/\bar{T}, \\ & w_p \equiv w_p(z) = \operatorname{tr}\left(D^{-1}C'M^{-1}C'D^{-1}\right)/\bar{T}, && \tilde{w}_p \equiv \tilde{w}_p(z) = \operatorname{tr}\left(A^{-1}C\tilde{M}^{-1}CA^{-1}\right)/\bar{T}, \\ & s_p = \operatorname{tr} D^{-1}/\bar{T}, \text{ and} && \tilde{s}_p = \operatorname{tr} A^{-1}/\bar{T}, \end{aligned}$$

where $M = CD^{-1}C' - zA$ and $\tilde{M} = C'A^{-1}C - zD$. For the reader's convenience, Table 1 lists definitions of matrices and scalars used in our proofs below.

Lemma 9 *The following identities hold*

$$u_j^{(q)} = \tilde{u}_j^{(q)\prime}, \ z\tilde{v}_j^{(q)} = w_j^{(q)} - s_j^{(q)}, \text{ and } z v_j^{(q)} = \tilde{w}_j^{(q)} - \tilde{s}_j^{(q)}. \quad (43)$$

Similarly,

$$u_p = \tilde{u}_p, \ z\tilde{v}_p = w_p - s_p, \text{ and } z v_p = \tilde{w}_p - \tilde{s}_p. \quad (44)$$

Table 1: Definitions of matrices, quadratic forms and traces that are used in the derivations below. Notations used in this table suppress the dependence of various quantities, such as M, v_p, u_p, w_p, etc., on z.

$p\times p$ matrices	2×2 matrices	scalars
$M = CD^{-1}C' - zA$	$v_j^{(q)} = \varepsilon'_{(j)} M_j^{-1} \varepsilon_{(j)}$	$v_p = \frac{1}{T}\operatorname{tr}\left\{M^{-1}\right\}$
$\tilde{M} = C'A^{-1}C - zD$	$u_j^{(q)} = \varepsilon'_{(j)} D_j^{-1} C'_j M_j^{-1} \varepsilon_{(j)}$	$u_p = \frac{1}{T}\operatorname{tr}\left\{D^{-1}C'M^{-1}\right\}$
$C_j = C - \varepsilon_{(j)}\nabla'_j\varepsilon'_{(j)},$	$w_j^{(q)} = \varepsilon'_{(j)} D_j^{-1} C'_j M_j^{-1} C'_j D_j^{-1} \varepsilon_{(j)}$	$w_p = \frac{1}{T}\operatorname{tr}\left\{D^{-1}C'M^{-1}CD^{-1}\right\}$
$D_j = D - r_j^{-1}\varepsilon_{(j)}\varepsilon'_{(j)},$	$s_j^{(q)} = \varepsilon'_{(j)} D_j^{-1} \varepsilon_{(j)}$	$s_p = \frac{1}{T}\operatorname{tr}\left\{D^{-1}\right\}$
$A_j = A - \varepsilon_{(j)}\varepsilon'_{(j)},$	$\tilde{v}_j^{(q)} = \varepsilon'_{(j)} \tilde{M}_j^{-1} \varepsilon_{(j)}$	$\tilde{v}_p = \frac{1}{T}\operatorname{tr}\left\{\tilde{M}^{-1}\right\}$
$M_j = C_j D_j^{-1} C'_j - zA_j$	$\tilde{u}_j^{(q)} = \varepsilon'_{(j)} A_j^{-1} C_j \tilde{M}_j^{-1} \varepsilon_{(j)}$	$\tilde{u}_p = \frac{1}{T}\operatorname{tr}\left\{A^{-1}C\tilde{M}^{-1}\right\}$
$\tilde{M}_j = C'_j A_j^{-1} C_j - zD_j$	$\tilde{w}_j^{(q)} = \varepsilon'_{(j)} A_j^{-1} C_j \tilde{M}_j^{-1} C_j A_j^{-1} \varepsilon_{(j)}$	$\tilde{w}_p = \frac{1}{T}\operatorname{tr}\left\{A^{-1}C\tilde{M}^{-1}C'A^{-1}\right\}$
	$\tilde{s}_j^{(q)} = \varepsilon'_{(j)} A_j^{-1} \varepsilon_{(j)}$	$\tilde{s}_p = \frac{1}{T}\operatorname{tr}\left\{A^{-1}\right\}$
		$m_p = \frac{1}{p}\operatorname{tr}\left\{\left(CD^{-1}C'A^{-1} - zI_p\right)^{-1}\right\}$

Proof. The identity $u_j^{(q)} = \tilde{u}_j^{(q)\prime}$ is established by the following sequence of equalities

$$\begin{aligned}
u_j^{(q)} &= \epsilon'_{(j)} D_j^{-1} C'_j M_j^{-1} \epsilon_{(j)} = \epsilon'_{(j)} D_j^{-1} C'_j \left(C_j D_j^{-1} C'_j - zA_j\right)^{-1} \epsilon_{(j)} \\
&= \epsilon'_{(j)} \left(C_j - zA_j \left(C'_j\right)^{-1} D_j\right)^{-1} \epsilon_{(j)} = \left(\epsilon'_{(j)} \left(C'_j - zD_j \left(C_j\right)^{-1} A_j\right)^{-1} \epsilon_{(j)}\right)' \\
&= \left(\epsilon'_{(j)} A_j^{-1} C_j \left(C'_j A_j^{-1} C_j - zD_j\right)^{-1} \epsilon_{(j)}\right)' = \left(\epsilon'_{(j)} A_j^{-1} C_j \tilde{M}_j^{-1} \epsilon_{(j)}\right)' = \tilde{u}_j^{(q)\prime}.
\end{aligned}$$

The relationship $z\tilde{v}_j^{(q)} = w_j^{(q)} - s_j^{(q)}$ is obtained as follows

$$\begin{aligned}
z\tilde{v}_j^{(q)} + s_j^{(q)} &= \epsilon'_{(j)} \left(z\tilde{M}_j^{-1} + D_j^{-1}\right) \epsilon_{(j)} = \epsilon'_{(j)} D_j^{-1} \left(zI_p \left(C'_j A_j^{-1} C_j D_j^{-1} - zI_p\right)^{-1} + I_p\right) \epsilon_{(j)} \\
&= \epsilon'_{(j)} D_j^{-1} \left(-I_p + C'_j A_j^{-1} C_j D_j^{-1} \left(C'_j A_j^{-1} C_j D_j^{-1} - zI_p\right)^{-1} + I_p\right) \epsilon_{(j)} \\
&= \epsilon'_{(j)} D_j^{-1} C'_j \left(C'_j - zD_j C_j^{-1} A_j\right)^{-1} \epsilon_{(j)} = \epsilon'_{(j)} D_j^{-1} C'_j \left(D_j^{-1} C'_j - zC_j^{-1} A_j\right)^{-1} D_j^{-1} \epsilon_{(j)} \\
&= \epsilon'_{(j)} D_j^{-1} C'_j \left(C_j D_j^{-1} C'_j - zA_j\right)^{-1} C_j D_j^{-1} \epsilon_{(j)} = w_j^{(q)}.
\end{aligned}$$

The relationship $zv_j^{(q)} = \tilde{w}_j^{(q)} - \tilde{s}_j^{(q)}$ is obtained as follows

$$\begin{aligned}
zv_j^{(q)} + \tilde{s}_j^{(q)} &= \epsilon'_{(j)} \left(zM_j^{-1} + A_j^{-1}\right) \epsilon_{(j)} = \epsilon'_{(j)} A_j^{-1} \left(zI_p \left(C_j D_j^{-1} C'_j A_j^{-1} - zI_p\right)^{-1} + I_p\right) \epsilon_{(j)} \\
&= \epsilon'_{(j)} A_j^{-1} \left(-I_p + C_j D_j^{-1} C'_j A_j^{-1} \left(C_j D_j^{-1} C'_j A_j^{-1} - zI_p\right)^{-1} + I_p\right) \epsilon_{(j)} \\
&= \epsilon'_{(j)} A_j^{-1} C_j \left(C_j - zA_j C_j'^{-1} D_j\right)^{-1} \epsilon_{(j)} = \epsilon'_{(j)} A_j^{-1} C_j \left(A_j^{-1} C_j - zC_j'^{-1} D_j\right)^{-1} A_j^{-1} \epsilon_{(j)} \\
&= \epsilon'_{(j)} A_j^{-1} C_j \left(C'_j A_j^{-1} C_j - zD_j\right)^{-1} C'_j A_j^{-1} \epsilon_{(j)} = \tilde{w}_j^{(q)}.
\end{aligned}$$

Identities (44) are established similarly. The only differences are that the matrices involved are not indexed by j, and instead of the quadratic forms in the columns of $\epsilon_{(j)}$ we work with traces.□

Derivation of (38) Applying the Sherman-Morrison-Woodbury formula

$$\left(V + XWY\right)^{-1} = V^{-1} - V^{-1}X\left(W^{-1} + YV^{-1}X\right)^{-1} YV^{-1} \tag{45}$$

to the right hand side of

$$D^{-1} = \left(D_j + r_j^{-1}\epsilon_{(j)}\epsilon'_{(j)}\right)^{-1},$$

we obtain

$$D^{-1} = D_j^{-1} - D_j^{-1}\epsilon_{(j)}\left(r_jI_2 + s_j^{(q)}\right)^{-1}\epsilon'_{(j)}D_j^{-1}. \quad (46)$$

Using this and the identity

$$C = C_j + \epsilon_{(j)}\nabla'_j\epsilon'_{(j)}, \quad (47)$$

we expand $CD^{-1}C'$ in the following form

$$\begin{aligned}
&C_jD_j^{-1}C'_j + \epsilon_{(j)}\nabla'_j\epsilon'_{(j)}D_j^{-1}C'_j - C_jD_j^{-1}\epsilon_{(j)}\left(r_jI_2 + s_j^{(q)}\right)^{-1}\epsilon'_{(j)}D_j^{-1}C'_j + C_jD_j^{-1}\epsilon_{(j)}\nabla_j\epsilon'_{(j)} \\
&-\epsilon_{(j)}\nabla'_js_j^{(q)}\left(r_jI_2 + s_j^{(q)}\right)^{-1}\epsilon'_{(j)}D_j^{-1}C'_j - C_jD_j^{-1}\epsilon_{(j)}\left(r_jI_2 + s_j^{(q)}\right)^{-1}s_j^{(q)}\nabla_j\epsilon'_{(j)} + \epsilon_{(j)}\nabla'_js_j^{(q)}\nabla_j\epsilon'_{(j)} \\
&-\epsilon_{(j)}\nabla'_js_j^{(q)}\left(r_jI_2 + s_j^{(q)}\right)^{-1}s_j^{(q)}\nabla_j\epsilon'_{(j)}.
\end{aligned}$$

Simplifying this expression yields

$$\begin{aligned}
CD^{-1}C' \;=\;& C_jD_j^{-1}C'_j - C_jD_j^{-1}\epsilon_{(j)}\left(r_jI_2 + s_j^{(q)}\right)^{-1}\epsilon'_{(j)}D_j^{-1}C'_j + \epsilon_{(j)}\nabla'_jr_j\left(r_jI_2 + s_j^{(q)}\right)^{-1}\epsilon'_{(j)}D_j^{-1}C'_j \\
&+C_jD_j^{-1}\epsilon_{(j)}\left(r_jI_2 + s_j^{(q)}\right)^{-1}r_j\nabla_j\epsilon'_{(j)} + \epsilon_{(j)}\nabla'_js_j^{(q)}\left(r_jI_2 + s_j^{(q)}\right)^{-1}r_j\nabla_j\epsilon'_{(j)}.
\end{aligned}$$

Since $M = CD^{-1}C' - zA$ and $A = A_j + \epsilon_{(j)}\epsilon'_{(j)}$, it follows that

$$M^{-1} = \left(M_j + \alpha_jK_j\alpha'_j\right)^{-1}, \quad (48)$$

where

$$\alpha_j = [\epsilon_{(j)}, C_jD_j^{-1}\epsilon_{(j)}]$$

and

$$K_j = \begin{pmatrix} \nabla'_js_j^{(q)}\left(r_jI_2 + s_j^{(q)}\right)^{-1}r_j\nabla_j - zI_2 & \nabla'_jr_j\left(r_jI_2 + s_j^{(q)}\right)^{-1} \\ \left(r_jI_2 + s_j^{(q)}\right)^{-1}r_j\nabla_j & -\left(r_jI_2 + s_j^{(q)}\right)^{-1} \end{pmatrix}.$$

Applying (45) to the right hand side of (48), we obtain

$$M^{-1} = M_j^{-1} - M_j^{-1}\alpha_j\left(K_j^{-1} + \alpha'_jM_j^{-1}\alpha_j\right)^{-1}\alpha'_jM_j^{-1}. \quad (49)$$

The identity $\nabla'_j\nabla_j = r_j^{-1}I_2$ yields

$$K_j = \begin{pmatrix} \nabla'_j & 0 \\ 0 & I_2 \end{pmatrix}\begin{pmatrix} s_j^{(q)}\left(r_jI_2 + s_j^{(q)}\right)^{-1}r_j - zr_jI_2 & \left(r_jI_2 + s_j^{(q)}\right)^{-1}r_j \\ \left(r_jI_2 + s_j^{(q)}\right)^{-1}r_j & -\left(r_jI_2 + s_j^{(q)}\right)^{-1} \end{pmatrix}\begin{pmatrix} \nabla_j & 0 \\ 0 & I_2 \end{pmatrix},$$

which implies that

$$K_j^{-1} = \frac{1}{1-z}\begin{pmatrix} \nabla_j^{-1} & 0 \\ 0 & I_2 \end{pmatrix}\begin{pmatrix} r_j^{-1}I_2 & I_2 \\ I_2 & z\left(r_jI_2 + s_j^{(q)}\right) - s_j^{(q)} \end{pmatrix}\begin{pmatrix} \nabla_j'^{-1} & 0 \\ 0 & I_2 \end{pmatrix},$$

and therefore, using $\nabla'_j\nabla_j = r_j^{-1}I_2$ again, we obtain

$$K_j^{-1} = \begin{pmatrix} \frac{1}{1-z}I_2 & \frac{1}{1-z}r_j\nabla'_j \\ \frac{1}{1-z}r_j\nabla_j & \frac{z}{1-z}r_jI_2 - s_j^{(q)} \end{pmatrix}. \quad (50)$$

Further, the definitions of $v_j^{(q)}$, $u_j^{(q)}$ and $w_j^{(q)}$ yield

$$\alpha_j' M_j^{-1} \alpha_j = \begin{pmatrix} v_j^{(q)} & u_j^{(q)\prime} \\ u_j^{(q)} & w_j^{(q)} \end{pmatrix}. \tag{51}$$

Using (50) and (51) in (49), we obtain

$$M^{-1} = M_j^{-1} - M_j^{-1} \alpha_j \Omega_j^{(q)} \alpha_j' M_j^{-1}, \tag{52}$$

where

$$\Omega_j^{(q)} = \begin{pmatrix} \frac{1}{1-z} I_2 + v_j^{(q)} & \frac{1}{1-z} r_j \nabla_j' + u_j^{(q)\prime} \\ \frac{1}{1-z} r_j \nabla_j + u_j^{(q)} & \frac{z}{1-z} r_j I_2 - s_j^{(q)} + w_j^{(q)} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{1-z} I_2 + v_j^{(q)} & \frac{1}{1-z} r_j \nabla_j' + u_j^{(q)\prime} \\ \frac{1}{1-z} r_j \nabla_j + u_j^{(q)} & \frac{z}{1-z} r_j I_2 + z \tilde{v}_j^{(q)} \end{pmatrix}^{-1}, \tag{53}$$

and the latter equality holds by Lemma 9.

Equation (52) yields

$$\epsilon_{(j)}' M^{-1} \epsilon_{(j)} = v_j^{(q)} - \left[v_j^{(q)}, u_j^{(q)\prime} \right] \Omega_j^{(q)} \left[v_j^{(q)}, u_j^{(q)\prime} \right]'. \tag{54}$$

Note that

$$v_j^{(q)} = \left[v_j^{(q)}, u_j^{(q)\prime} \right] \Omega_j^{(q)} (\Omega_j^{(q)})^{-1} \left[I_2, 0 \right]' = \left[v_j^{(q)}, u_j^{(q)\prime} \right] \Omega_j^{(q)} \left(\frac{1}{1-z} \left[I_2, r_j \nabla_j' \right]' + \left[v_j^{(q)}, u_j^{(q)\prime} \right]' \right),$$

and thus, (54) can be rewritten as

$$\begin{aligned} \epsilon_{(j)}' M^{-1} \epsilon_{(j)} &= \frac{1}{1-z} \left[v_j^{(q)}, u_j^{(q)\prime} \right] \Omega_j^{(q)} \left[I_2, r_j \nabla_j' \right]' \\ &= \frac{1}{1-z} \left(\left[\frac{1}{1-z} I_2 + v_j^{(q)}, \frac{1}{1-z} r_j \nabla_j' + u_j^{(q)\prime} \right] - \left[\frac{1}{1-z} I_2, \frac{1}{1-z} r_j \nabla_j' \right] \right) \Omega_j^{(q)} \left[I_2, r_j \nabla_j' \right]' \\ &= \frac{1}{1-z} \left(\left[I_2, 0 \right] \left[I_2, r_j \nabla_j' \right]' - \left[\frac{1}{1-z} I_2, \frac{1}{1-z} r_j \nabla_j' \right] \Omega_j^{(q)} \left[I_2, r_j \nabla_j' \right]' \right) \\ &= \frac{1}{1-z} I_2 - \frac{1}{(1-z)^2} \left[I_2, r_j \nabla_j' \right] \Omega_j^{(q)} \left[I_2, r_j \nabla_{2j}' \right]'. \end{aligned}$$

To summarize, we have the following identity

$$\epsilon_{(j)}' M^{-1} \epsilon_{(j)} = \frac{1}{1-z} I_2 - \frac{1}{(1-z)^2} \left[I_2, r_j \nabla_j' \right] \Omega_j^{(q)} \left[I_2, r_j \nabla_j' \right]'. \tag{55}$$

Recall that by definition,

$$m_p(z) = \frac{1}{p} \operatorname{tr} \left[\left(C D^{-1} C' A^{-1} - z I_p \right)^{-1} \right] = \frac{1}{p} \operatorname{tr} \left[A M^{-1} \right] = \frac{1}{p} \sum_{j=1}^{\bar{T}/2} \operatorname{tr} \left[\epsilon_{(j)}' M^{-1} \epsilon_{(j)} \right].$$

This equation and representation (55) yield identity (38)

$$m_p(z) = \frac{\bar{T}}{p} \frac{1}{1-z} - \frac{1}{p} \sum_{j=1}^{\bar{T}/2} \frac{1}{(1-z)^2} \operatorname{tr} \left(\left[I_2, r_j \nabla_j' \right] \Omega_j^{(q)} \left[I_2, r_j \nabla_j' \right]' \right).$$

Derivation of identity (39) Since the eigenvalues of $CD^{-1}C'A^{-1}$ coincide with those of $C'A^{-1}CD^{-1}$, we have

$$m_p(z) = \frac{1}{p}\operatorname{tr}\left[\left(C'A^{-1}CD^{-1} - zI_p\right)^{-1}\right] = \frac{1}{p}\operatorname{tr}\left[D\tilde{M}^{-1}\right] = \frac{1}{p}\sum_{j=1}^{\bar{T}/2}\operatorname{tr}\left[r_j^{-1}\epsilon'_{(j)}\tilde{M}^{-1}\epsilon_{(j)}\right]. \tag{56}$$

Note that matrix $\tilde{M}$ can be obtained from M by swapping A for D and C for C'. Performing such a swap in the above derivations of (55) yields

$$\epsilon'_{(j)}\tilde{M}^{-1}\epsilon_{(j)} = \frac{r_j}{1-z}I_2 - \frac{r_j^2}{(1-z)^2}\left[I_2, \nabla_j\right]\tilde{\Omega}_j^{(q)}\left[I_2, \nabla_j\right]', \tag{57}$$

where

$$\tilde{\Omega}_j^{(q)} = \begin{pmatrix} \frac{r_j}{1-z}I_2 + \tilde{v}_j^{(q)} & \frac{r_j}{1-z}\nabla_j + \tilde{u}_j^{(q)\prime} \\ \frac{r_j}{1-z}\nabla'_j + \tilde{u}_j^{(q)} & \frac{z}{1-z}I_2 - \tilde{s}_j^{(q)} + \tilde{w}_j^{(q)} \end{pmatrix}^{-1}.$$

Lemma 9 implies that

$$\tilde{\Omega}_j^{(q)} = \begin{pmatrix} \frac{r_j}{1-z}I_2 + z^{-1}\left(w_j^{(q)} - s_j^{(q)}\right) & \frac{r_j}{1-z}\nabla_j + u_j^{(q)} \\ \frac{r_j}{1-z}\nabla'_j + u_j^{(q)\prime} & \frac{z}{1-z}I_2 + zv_j^{(q)} \end{pmatrix}^{-1} = \begin{pmatrix} 0 & zI_2 \\ I_2 & 0 \end{pmatrix}\Omega_j^{(q)}\begin{pmatrix} 0 & z^{-1}I_2 \\ I_2 & 0 \end{pmatrix},$$

so that (57) yields

$$\epsilon'_{(j)}\tilde{M}^{-1}\epsilon_{(j)} = \frac{r_j}{1-z}I_2 - \frac{zr_j^2}{(1-z)^2}\left[z^{-1}\nabla_j, I_2\right]\Omega_j^{(q)}\left[z^{-1}\nabla_j, I_2\right]'. \tag{58}$$

Combining this with (56) gives us

$$m_p(z) = \frac{\bar{T}}{p}\frac{1}{1-z} - \frac{1}{p}\sum_{j=1}^{\bar{T}/2}\operatorname{tr}\left[\frac{zr_j}{(1-z)^2}\left[z^{-1}\nabla_j, I_2\right]\Omega_j^{(q)}\left[z^{-1}\nabla_j, I_2\right]'\right].$$

Further, since $r_j\nabla_j\nabla'_j = I_2$, we have

$$\frac{zr_j}{(1-z)^2}\operatorname{tr}\left[\left[z^{-1}\nabla_j, I_2\right]\Omega_j^{(q)}\left[z^{-1}\nabla_j, I_2\right]'\right] = \frac{zr_j}{(1-z)^2}\operatorname{tr}\left[r_j\nabla_j\nabla'_j\left[z^{-1}\nabla_j, I_2\right]\Omega_j^{(q)}\left[z^{-1}\nabla_j, I_2\right]'\right]$$

$$= \frac{zr_j}{(1-z)^2}\operatorname{tr}\left[r_j\left[z^{-1}\nabla'_j\nabla_j, \nabla'_j\right]\Omega_j^{(q)}\left[z^{-1}\nabla_j\nabla'_j, \nabla'_j\right]'\right] = \frac{z^{-1}}{(1-z)^2}\operatorname{tr}\left(\left[I_2, r_jz\nabla'_j\right]\Omega_j^{(q)}\left[I_2, zr_j\nabla'_j\right]'\right),$$

and therefore,

$$m_p(z) = \frac{\bar{T}}{p}\frac{1}{1-z} - \frac{1}{p}\sum_{j=1}^{\bar{T}/2}\frac{z^{-1}}{(1-z)^2}\operatorname{tr}\left(\left[I_2, r_jz\nabla'_j\right]\Omega_j^{(q)}\left[I_2, zr_j\nabla'_j\right]'\right),$$

which is equivalent to identity (39),

$$\frac{\bar{T}}{p} + zm_p(z) = \frac{\bar{T}}{p}\frac{1}{1-z} - \frac{1}{p}\sum_{j=1}^{\bar{T}/2}\frac{1}{(1-z)^2}\operatorname{tr}\left(\left[I_2, r_jz\nabla'_j\right]\Omega_j^{(q)}\left[I_2, zr_j\nabla'_j\right]'\right).$$

Derivation of identity (40) Multiplying both sides of the identity

$$MA^{-1} = CD^{-1}C'A^{-1} - zI_p$$

by AM^{-1}, taking trace, dividing by p, and rearranging yields

$$1+zm_p(z)=\frac{1}{p}\sum_{j=1}^{\bar{T}/2}\operatorname{tr}\left[\nabla_j'\epsilon_{(j)}'D^{-1}C'M^{-1}\epsilon_{(j)}\right]. \tag{59}$$

Equations (46), (47), and (52) imply that

$$\begin{aligned}D^{-1}C'M^{-1} &= \left(D_j^{-1}-D_j^{-1}\epsilon_{(j)}\left(r_jI_2+s_j^{(q)}\right)^{-1}\epsilon_{(j)}'D_j^{-1}\right)\left(C_j'+\epsilon_{(j)}\nabla_j\epsilon_{(j)}'\right)\\ &\quad\times\left(M_j^{-1}-M_j^{-1}\alpha_j\Omega_j^{(q)}\alpha_j'M_j^{-1}\right).\end{aligned}$$

Opening up brackets, we obtain

$$\begin{aligned}&D^{-1}C'M^{-1}\\ =\;&D_j^{-1}C_j'M_j^{-1}-D_j^{-1}\epsilon_{(j)}\left(r_jI_2+s_j^{(q)}\right)^{-1}\epsilon_{(j)}'D_j^{-1}C_j'M_j^{-1}+D_j^{-1}\epsilon_{(j)}\nabla_j\epsilon_{(j)}'M_j^{-1}\\ &-D_j^{-1}C_j'M_j^{-1}\alpha_j\Omega_j^{(q)}\alpha_j'M_j^{-1}-D_j^{-1}\epsilon_{(j)}\left(r_jI_2+s_j^{(q)}\right)^{-1}\epsilon_{(j)}'D_j^{-1}\epsilon_{(j)}\nabla_j\epsilon_{(j)}'M_j^{-1}\\ &+D_j^{-1}\epsilon_{(j)}\left(r_jI_2+s_j^{(q)}\right)^{-1}\epsilon_{(j)}'D_j^{-1}C_j'M_j^{-1}\alpha_j\Omega_j^{(q)}\alpha_j'M_j^{-1}-D_j^{-1}\epsilon_{(j)}\nabla_j\epsilon_{(j)}'M_j^{-1}\alpha_j\Omega_j^{(q)}\alpha_j'M_j^{-1}\\ &+D_j^{-1}\epsilon_{(j)}\left(r_jI_2+s_j^{(q)}\right)^{-1}\epsilon_{(j)}'D_j^{-1}\epsilon_{(j)}\nabla_j\epsilon_{(j)}'M_j^{-1}\alpha_j\Omega_j^{(q)}\alpha_j'M_j^{-1}.\end{aligned}$$

Multiplying from the left by $\epsilon_{(j)}'$ and from the right by $\epsilon_{(j)}$, and using the definitions of $u_j^{(q)}, v_j^{(q)}, s_j^{(q)}$, and $w_j^{(q)}$, we obtain

$$\begin{aligned}&\epsilon_{(j)}'D^{-1}C'M^{-1}\epsilon_{(j)}\\ =\;&u_j^{(q)}-s_j^{(q)}\left(r_jI_2+s_j^{(q)}\right)^{-1}u_j^{(q)}+s_j^{(q)}\nabla_jv_j^{(q)}-\left[u_j^{(q)},w_j^{(q)}\right]\Omega_j^{(q)}\left[v_j^{(q)},u_j^{(q)\prime}\right]'\\ &-s_j^{(q)}\left(r_jI_2+s_j^{(q)}\right)^{-1}s_j^{(q)}\nabla_jv_j^{(q)}+s_j^{(q)}\left(r_jI_2+s_j^{(q)}\right)^{-1}\left[u_j^{(q)},w_j^{(q)}\right]\Omega_j^{(q)}\left[v_j^{(q)},u_j^{(q)\prime}\right]'\\ &-s_j^{(q)}\nabla_j\left[v_j^{(q)},u_j^{(q)\prime}\right]\Omega_j^{(q)}\left[v_j^{(q)},u_j^{(q)\prime}\right]'+s_j^{(q)}\left(r_jI_2+s_j^{(q)}\right)^{-1}s_j^{(q)}\nabla_j\left[v_j^{(q)},u_j^{(q)\prime}\right]\Omega_j^{(q)}\left[v_j^{(q)},u_j^{(q)\prime}\right]'.\end{aligned}$$

Rearranging terms and simplifying gives us

$$\begin{aligned}\epsilon_{(j)}'D^{-1}C'M^{-1}\epsilon_{(j)} &= r_j\left(r_jI_2+s_j^{(q)}\right)^{-1}s_j^{(q)}\nabla_j\left(v_j^{(q)}-\left[v_j^{(q)},u_j^{(q)\prime}\right]\Omega_j^{(q)}\left[v_j^{(q)},u_j^{(q)\prime}\right]'\right)\\ &\quad+r_j\left(r_jI_2+s_j^{(q)}\right)^{-1}\left(u_j^{(q)}-\left[u_j^{(q)},w_j^{(q)}\right]\Omega_j^{(q)}\left[v_j^{(q)},u_j^{(q)\prime}\right]'\right).\end{aligned} \tag{60}$$

As follows from (54) and (55)

$$v_j^{(q)}-\left[v_j^{(q)},u_j^{(q)\prime}\right]\Omega_j^{(q)}\left[v_j^{(q)},u_j^{(q)\prime}\right]'=\frac{1}{1-z}I_2-\frac{1}{(1-z)^2}\left[I_2,r_j\nabla_j'\right]\Omega_j^{(q)}\left[I_2,r_j\nabla_j'\right]'. \tag{61}$$

Further,

$$\begin{aligned}u_j^{(q)}-\left[u_j^{(q)},w_j^{(q)}\right]\Omega_j^{(q)}\left[v_j^{(q)},u_j^{(q)\prime}\right]' &= \left[u_j^{(q)},w_j^{(q)}\right]\Omega_j^{(q)}(\Omega_j^{(q)})^{-1}\left[I_2,0\right]'-\left[u_j^{(q)},w_j^{(q)}\right]\Omega_j^{(q)}\left[v_j^{(q)},u_j^{(q)\prime}\right]'\\ &= \frac{1}{1-z}\left[u_j^{(q)},w_j^{(q)}\right]\Omega_j^{(q)}\left[I_2,r_j\nabla_j'\right]'\end{aligned}$$

Note that

$$
\begin{aligned}
& \frac{1}{1-z}\left[u_j^{(q)}, w_j^{(q)}\right] \Omega_j^{(q)}\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime} \\
= & \frac{1}{1-z}\left(\left[\frac{r_j}{1-z} \nabla_j+u_j^{(q)}, \frac{r_j z}{1-z} I_2+w_j^{(q)}-s_j^{(q)}\right]-\left[\frac{r_j}{1-z} \nabla_j, \frac{r_j z}{1-z} I_2-s_j^{(q)}\right]\right) \Omega_j^{(q)}\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime} \\
= & \frac{1}{1-z}\left(\left[0, I_2\right]\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime}-\frac{r_j z}{1-z}\left[z^{-1} \nabla_j, I_2\right] \Omega_j^{(q)}\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime}+\left[0, s_j^{(q)}\right] \Omega_j^{(q)}\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime}\right) \\
= & \frac{1}{1-z} r_j \nabla_j-\frac{r_j z}{(1-z)^2}\left[z^{-1} \nabla_j, I_2\right] \Omega_j^{(q)}\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime}+\frac{1}{1-z}\left[0, s_j^{(q)}\right] \Omega_j^{(q)}\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime} .
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
u_j^{(q)}-\left[u_j^{(q)}, w_j^{(q)}\right] \Omega_j^{(q)}\left[v_j^{(q)}, u_j^{(q)\prime}\right]^{\prime} = & \frac{1}{1-z} r_j \nabla_j-\frac{r_j z}{(1-z)^2}\left[z^{-1} \nabla_j, I_2\right] \Omega_j^{(q)}\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime} \\
& +\frac{1}{1-z}\left[0, s_j^{(q)}\right] \Omega_j^{(q)}\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime} .
\end{aligned}
$$

Using this and (61) in (60), we obtain

$$
\begin{aligned}
& \epsilon_{(j)}^{\prime} D^{-1} C^{\prime} M^{-1} \epsilon_{(j)} \\
= & r_j\left(r_j I_2+s_j^{(q)}\right)^{-1} s_j^{(q)} \nabla_j\left(\frac{1}{1-z} I_2-\frac{1}{(1-z)^2}\left[I_2, r_j \nabla_j^{\prime}\right] \Omega_j^{(q)}\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime}\right)+r_j\left(r_j I_2+s_j^{(q)}\right)^{-1} \\
& \times\left(\frac{1}{1-z} r_j \nabla_j-\frac{r_j z}{(1-z)^2}\left[z^{-1} \nabla_j, I_2\right] \Omega_j^{(q)}\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime}+\frac{1}{1-z}\left[0, s_j^{(q)}\right] \Omega_j^{(q)}\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime}\right) \\
= & \frac{1}{1-z} r_j \nabla_j-\frac{r_j\left(r_j I_2+s_j^{(q)}\right)^{-1}}{(1-z)^2}\left[s_j^{(q)} \nabla_j+r_j \nabla_j, z s_j^{(q)}+z r_j I_2\right] \Omega_j^{(q)}\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime} \\
= & \frac{r_j}{1-z} \nabla_j-\frac{z r_j}{(1-z)^2}\left[z^{-1} \nabla_j, I_2\right] \Omega_j^{(q)}\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime},
\end{aligned}
$$

that is,

$$
\epsilon_{(j)}^{\prime} D^{-1} C^{\prime} M^{-1} \epsilon_{(j)}=\frac{r_j}{1-z} \nabla_j-\frac{z r_j}{(1-z)^2}\left[z^{-1} \nabla_j, I_2\right] \Omega_j^{(q)}\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime} . \tag{62}
$$

This identity together with (59) yield

$$
\begin{aligned}
1+z m_p(z) & =\frac{1}{p} \sum_{j=1}^{\bar{T} / 2} \operatorname{tr}\left[\nabla_j^{\prime}\left(\frac{r_j}{1-z} \nabla_j-\frac{z r_j}{(1-z)^2}\left[z^{-1} \nabla_j, I_2\right] \Omega_j^{(q)}\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime}\right)\right] \\
& =\frac{1}{p} \sum_{j=1}^{\bar{T} / 2} \operatorname{tr}\left[\left(\frac{1}{1-z} I_2-\frac{1}{(1-z)^2}\left[I_2, r_j z \nabla_j^{\prime}\right] \Omega_j^{(q)}\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime}\right)\right],
\end{aligned}
$$

which is equivalent to identity (40),

$$
1+z m_p(z)=\frac{\bar{T}}{p} \frac{1}{1-z}-\frac{1}{p} \sum_{j=1}^{\bar{T} / 2} \frac{1}{(1-z)^2} \operatorname{tr}\left(\left[I_2, r_j z \nabla_j^{\prime}\right] \Omega_j^{(q)}\left[I_2, r_j \nabla_j^{\prime}\right]^{\prime}\right) .
$$

Derivation of identity (41) An obvious identity

$$
\frac{1}{p} \operatorname{tr}\left[C^{\prime} M^{-1}\right]=\frac{1}{p} \operatorname{tr}\left[D D^{-1} C^{\prime} M^{-1}\right]
$$

and representations $C' = \sum_{j=1}^{\bar{T}/2} \epsilon_{(j)} \nabla_j \epsilon'_{(j)}$ and $D = \sum_{j=1}^{\bar{T}/2} r_j^{-1} \epsilon_{(j)} \epsilon'_{(j)}$ yield

$$\frac{1}{p}\sum_{j=1}^{\bar{T}/2} \operatorname{tr}\left[\nabla_j \epsilon'_{(j)} M^{-1} \epsilon_{(j)}\right] = \frac{1}{p}\sum_{j=1}^{\bar{T}/2} \operatorname{tr}\left[r_j^{-1} \epsilon'_{(j)} D^{-1} C' M^{-1} \epsilon_{(j)}\right].$$

Using (62) and (55) in this equation, we obtain

$$\begin{aligned} & \frac{1}{p}\sum_{j=1}^{\bar{T}/2} \operatorname{tr}\left[\nabla_j \left(\frac{1}{1-z} I_2 - \frac{1}{(1-z)^2}\left[I_2, r_j \nabla'_j\right] \Omega_j^{(q)} \left[I_2, r_j \nabla'_j\right]'\right)\right] \\ = \; & \frac{1}{p}\sum_{j=1}^{\bar{T}/2} \operatorname{tr}\left[r_j^{-1} \left(\frac{r_j}{1-z} \nabla_j - \frac{z r_j}{(1-z)^2}\left[z^{-1}\nabla_j, I_2\right] \Omega_j^{(q)} \left[I_2, r_j \nabla'_j\right]'\right)\right]. \end{aligned}$$

Equivalently,

$$\begin{aligned} 0 &= \frac{1}{p}\sum_{j=1}^{\bar{T}/2} \operatorname{tr}\left[\left(\frac{1}{(1-z)^2}\left[\nabla_j, I_2\right] \Omega_j^{(q)} \left[I_2, r_j \nabla'_j\right]' - \frac{1}{(1-z)^2}\left[\nabla_j, z I_2\right] \Omega_j^{(q)} \left[I_2, r_j \nabla'_j\right]'\right)\right] \\ &= \frac{1}{p}\sum_{j=1}^{\bar{T}/2} \frac{1}{1-z} \operatorname{tr}\left(\left[0, I_2\right] \Omega_j^{(q)} \left[I_2, r_j \nabla'_j\right]'\right), \end{aligned}$$

which is the same as identity (41).

2.1.5 From identities to a system of approximate equations

As we will see below, matrices $v_p I_2, \tilde{v}_p I_2, u_p I_2,$ etc. are close to $v_j^{(q)}, \tilde{v}_j^{(q)}, u_j^{(q)}$, etc., uniformly in $j = 1, ..., \bar{T}/2$ (see Table 1 for definitions of $v_p, \tilde{v}_p, v_j^{(q)}, \tilde{v}_j^{(q)}$, etc.) Therefore, matrices $\Omega_j^{(q)}$, defined in (53), can be well approximated by

$$\begin{aligned} \Omega_j &= \begin{pmatrix} \frac{1}{1-z} I_2 + v_p I_2 & \frac{1}{1-z} r_j \nabla'_j + u_p I_2 \\ \frac{1}{1-z} r_j \nabla_j + u_p I_2 & \frac{z}{1-z} r_j I_2 + (w_p - s_p) I_2 \end{pmatrix}^{-1} \\ &= \frac{1-z}{\delta_{pj}} \begin{pmatrix} \frac{z}{1-z} r_j I_2 + z \tilde{v}_p I_2 & -\frac{1}{1-z} r_j \nabla'_j - u_p I_2 \\ -\frac{1}{1-z} r_j \nabla_j - u_p I_2 & \frac{1}{1-z} I_2 + v_p I_2 \end{pmatrix}, \end{aligned} \tag{63}$$

where the latter equality follows from (44), and

$$\delta_{pj} = (1-z)\left(z\tilde{v}_p v_p - u_p^2\right) + z\tilde{v}_p + r_j\left(u_p + z v_p - 1\right). \tag{64}$$

Approximating $\Omega_j^{(q)}$ by Ω_j, sums by integrals, and $p/\bar{T}$ by c in equations (38-41), we obtain the following system of "approximate equations"

$$\begin{cases} m_p(z) = \frac{1}{2\pi c}\int_0^{2\pi} \delta_p^{-1}(\varphi)\left(z\tilde{v}_p v_p - u_p^2 - 4 v_p \sin^2\varphi\right) \mathrm{d}\varphi + e_{1p}(z) \\ m_p(z) = \frac{1}{2\pi c}\int_0^{2\pi} \delta_p^{-1}(\varphi)\left(z v_p \tilde{v}_p - \tilde{v}_p - u_p^2\right) \mathrm{d}\varphi + e_{2p}(z) \\ 1 + z m_p(z) = \frac{1}{2\pi c}\int_0^{2\pi} \delta_p^{-1}(\varphi)\left(2 u_p \sin^2\varphi + z\tilde{v}_p v_p - u_p^2\right) \mathrm{d}\varphi + e_{3p}(z) \\ 0 = \frac{1}{2\pi c}\int_0^{2\pi} \delta_p^{-1}(\varphi)\left(4 v_p \sin^2\varphi + 2 u_p\right) \mathrm{d}\varphi + e_{4p}(z) \end{cases}, \tag{65}$$

where

$$\delta_p(\varphi) = (1-z)\left(z\tilde{v}_p v_p - u_p^2\right) + z\tilde{v}_p + 4\sin^2\varphi\left(z v_p + u_p - 1\right), \tag{66}$$

and $e_{kp}(z)$ are the approximation errors. Of course, system (65) can be viewed simply as the definition of $e_{kp}(z)$, $k = 1, ..., 4$. That these quantities are indeed the errors of the above mentioned approximations will be clear from the proof of the following lemma.

Lemma 10 *There exists $\zeta > 0$ such that, for any z with $\Re z = 0$ and $\Im z > \zeta$,*

$$e_{kp}(z) \overset{\text{a.s.}}{\to} 0$$

for $k = 1, ..., 4$ as $p, T \to_c \infty$.

Proof. Consider a decomposition $e_{kp}(z) = e_{kp}^{(1)}(z) + e_{kp}^{(2)}(z)$, where $e_{kp}^{(1)}(z)$ is the error due to the replacement of $\Omega_j^{(q)}$ by Ω_j and $e_{kp}^{(2)}(z)$ is due to the replacement of sums by integrals, and $p/\bar{T}$ by c. For example, for $k = 1$, we have

$$e_{1p}^{(1)}(z) = \frac{1}{p}\sum_{j=1}^{\bar{T}/2} \frac{1}{(1-z)^2} \operatorname{tr}\left(\left[I_2, r_j\nabla_j'\right]\left(\Omega_j - \Omega_j^{(q)}\right)\left[I_2, r_j\nabla_j'\right]'\right) \tag{67}$$

and

$$\begin{aligned} e_{1p}^{(2)}(z) &= \frac{\bar{T}}{p}\frac{1}{1-z} - \frac{1}{p}\sum_{j=1}^{\bar{T}/2}\frac{1}{(1-z)^2}\operatorname{tr}\left(\left[I_2, r_j\nabla_j'\right]\Omega_j\left[I_2, r_j\nabla_j'\right]'\right) \\ &\quad -\frac{1}{2\pi c}\int_0^{2\pi}\delta_p^{-1}(\varphi)\left(z\tilde{v}_p v_p - u_p^2 - 4v_p\sin^2\varphi\right)\mathrm{d}\varphi. \end{aligned} \tag{68}$$

Using equations (53) and (63), we obtain

$$\Omega_j - \Omega_j^{(q)} = \Omega_j\begin{pmatrix} v_j^{(q)} - v_p I_2 & u_j^{(q)\prime} - u_p I_2 \\ u_j^{(q)} - u_p I_2 & w_j^{(q)} - s_j^{(q)} - (w_p - s_p) I_2 \end{pmatrix}\Omega_j^{(q)}.$$

Therefore, for any $z \in \mathbb{C}^+$, the convergences $e_{kp}^{(1)}(z) \overset{\text{a.s.}}{\to} 0$, $k = 1, ..., 4$, would follow from the a.s. uniform in j convergence to zero of all the elements of the matrix sandwiched between Ω_j and $\Omega_j^{(q)}$ in the above display, and from the a.s. uniform in j boundedness of $\|\Omega_j\|$, $\left\|\Omega_j^{(q)}\right\|$, and $\|r_j\nabla_j'\|$. The uniform convergences of the matrix elements are established in Lemma 14 below. The uniform boundedness of $\|r_j\nabla_j'\|$ immediately follows from the definitions of r_j and ∇_j. The uniform boundedness of $\|\Omega_j\|$ follows from Lemmas 15 and 16. Finally, the uniform boundedness of $\left\|\Omega_j^{(q)}\right\|$ follows from that of $\|\Omega_j\|$ and from Lemma 14.

For $e_{1p}^{(2)}(z)$, using the explicit expression (63) for matrix Ω_j and the identities $r_j\nabla_j'\nabla_j = I_2$ and $\operatorname{tr}\nabla_j = -1$, we obtain

$$\begin{aligned} &\frac{\bar{T}}{p}\frac{1}{1-z} - \frac{1}{p}\sum_{j=1}^{\bar{T}/2}\frac{1}{(1-z)^2}\operatorname{tr}\left(\left[I_2, r_j\nabla_j'\right]\Omega_j\left[I_2, r_j\nabla_j'\right]'\right) \\ &= \frac{\bar{T}}{p}\frac{1}{1-z} - \frac{1}{p}\sum_{j=1}^{\bar{T}/2}\frac{2}{1-z}\frac{z\tilde{v}_p + r_j(u_p + v_p - 1)}{\delta_{pj}} = \frac{\bar{T}}{p}\frac{2}{\bar{T}}\sum_{j=1}^{\bar{T}/2}\delta_{pj}^{-1}\left(z\tilde{v}_p v_p - u_p^2 - r_j v_p\right) \end{aligned}$$

Since $r_j = 4\sin^2(\pi j/T)$ and $\bar{T} = T - 1$, the latter expression can be interpreted as a Riemann sum approximation for the integral

$$\frac{1}{2\pi c}\int_0^{2\pi}\delta_p^{-1}(\varphi)\left(z\tilde{v}_p v_p - u_p^2 - 4v_p\sin^2\varphi\right)\mathrm{d}\varphi.$$

The derivative of the integrand with respect to φ equals

$$-4\delta_p^{-2}(\varphi)\sin 2\varphi\left((v_p + u_p - 1)\left(z\tilde{v}_p v_p - u_p^2\right) + z v_p\tilde{v}_p\right). \tag{69}$$

As follows from Lemmas 15 and 16, there exists $\zeta > 0$ s.t. for any z with $\Im z > \zeta$, expression (69) a.s. remains uniformly bounded in φ as $p, T \to_c \infty$. Therefore, $e_{1p}^{(2)}(z) \overset{\text{a.s.}}{\to} 0$. The convergences $e_{kp}^{(2)}(z) \overset{\text{a.s.}}{\to} 0$, $k = 2, .., 4$, are established similarly, and we omit the corresponding details.□

In the remaining part of this section, we formulate and prove Lemmas 14, 15, 16, referred to in the above proof. We start from the following three auxiliary results.

Lemma 11 *Let Ω be a $p \times p$ deterministic complex matrix, and $\xi \sim N_p\left(0, I_p/\bar{T}\right)$. Then, for any $\rho \geq 2$*

$$\mathbb{E}\left|\xi'\Omega\xi - \operatorname{tr}\Omega/T\right|^\rho \leq K p^{\rho/2}\bar{T}^{-\rho}\left\|\Omega\right\|^\rho,$$

where K depends only on ρ.

Proof. The lemma is a straightforward corollary of Lemma 2.7 in Bai and Silverstein (1998).□

Lemma 12 *As $p, T \to_c \infty$,*

$$s_p \overset{\text{a.s.}}{\to} 4c^2/\left(1 - c^2\right). \tag{70}$$

Proof. By definition, $s_p = \operatorname{tr} D^{-1}/\bar{T} = \operatorname{tr}\left(\epsilon\nabla\nabla'\epsilon'\right)^{-1}/\bar{T}$. Let $F_D(x)$ denote the empirical distribution of the eigenvalues of D, and let

$$\hat{m}_D(z) = \int \frac{1}{x - z} \mathrm{d}F_D(x)$$

be its Stieltjes transform. Then by Theorem 1.1 of Silverstein and Bai (1995), for any $z \in \mathbb{C}^+$, $\hat{m}_D(z) \overset{\text{a.s.}}{\to} m_D(z)$ with

$$z = -\frac{1}{m_D(z)} + \int \frac{\tau \mathrm{d}H(\tau)}{1 + \tau c m_D(z)},$$

where $H(\tau)$ is the limit of the empirical distribution of the diagonal elements of $\nabla\nabla'$, r_j^{-1}, $j = 1, ..., \bar{T}/2$. Recall that

$$r_j = 4\sin^2(\omega_j/2) = 2 - 2\cos\omega_j = 2\left(1 - \cos\left(2\pi j/T\right)\right).$$

Therefore, $H(\tau)$ is the cumulative distribution function of the random variable $\left[2\left(1 - \cos U\right)\right]^{-1}$, where U is distributed uniformly on the interval $[0, \pi]$. This fact implies that

$$\begin{aligned} z &= -\frac{1}{m_D(z)} + \frac{1}{\pi}\int_0^\pi \frac{\mathrm{d}u}{2\left(1 - \cos u\right) + c m_D(z)} \\ &= -\frac{1}{m_D(z)} + \frac{1}{2\pi\mathrm{i}} \oint_{|s|=1} \frac{\mathrm{d}s}{s\left(2\left(1 - \frac{s + s^{-1}}{2}\right) + c m_D(z)\right)} \\ &= -\frac{1}{m_D(z)} - \frac{1}{2\pi\mathrm{i}} \oint_{|s|=1} \frac{\mathrm{d}s}{\left(s^2 - \left(2 + c m_D(z)\right)s + 1\right)}. \end{aligned}$$

The integrand has two poles at

$$s_{1,2} = \frac{c m_D(z) + 2 \pm \sqrt{c^2 m_D^2(z) + 4 c m_D(z)}}{2}.$$

Note that $s_1 s_2 = 1$, which implies that one of them is inside the contour and the other is outside. Therefore, we have

$$\begin{aligned} z &= -\frac{1}{m_D(z)} \pm \frac{1}{s_1 - s_2} \\ &= -\frac{1}{m_D(z)} \pm \frac{1}{\sqrt{c^2 m_D^2(z) + 4 c m_D(z)}} \end{aligned} \tag{71}$$

where the choice of $+$ or $-$ sign depends on which of $s_{1,2}$ is inside the contour. Squaring and rearranging, we obtain

$$c\left(zm_D(z)+1\right)^2\left(cm_D(z)+4\right)-m_D(z)=0. \tag{72}$$

Further, since $\min_{j=1,\ldots,\bar{T}/2} r_j^{-1}\geq 1/4$, denoting the k-th largest eigenvalue of a symmetric matrix S as $\lambda_k(S)$, we have (see e.g. Bai and Yin (1993))

$$\lambda_p(D)=\lambda_p\left(\epsilon\nabla\nabla'\epsilon'\right)\geq\frac{\lambda_p(\epsilon\epsilon')}{4}\overset{\text{a.s.}}{\to}\frac{\left(1-\sqrt{c}\right)^2}{4}.$$

Therefore, $m_D(z)$ is analytic at $z=0$, $\hat{m}_D(0)\overset{\text{a.s.}}{\to} m_D(0)$, and $m_D(0)$ satisfies equation (72) with $z=0$. That is,

$$\hat{m}_D(0)\overset{\text{a.s.}}{\to} m_D(0)=\frac{4c}{1-c^2}.$$

But $s_p=\frac{p}{T}\hat{m}_D(0)$. Hence, we have (70).□

Lemma 13 *Let $\mu_{\min,j},\mu_{\max,j}$ and $\mu_{\min,0},\mu_{\max,0}$ be the smallest and largest eigenvalues of A_j and of A, respectively. Then,*

$$\begin{aligned}\left\|M_j^{-1}\right\| &\leq 1/\left((\Im z)\,\mu_{\min,j}\right),\ \left\|D_j^{-1}\right\|\leq 4/\mu_{\min,j},\ \left\|D_j^{-1}C_j'\right\|^2\leq 4\mu_{\max,j}/\mu_{\min,j},\\ \left\|M^{-1}\right\| &\leq 1/\left((\Im z)\,\mu_{\min,0}\right),\ \left\|D^{-1}\right\|\leq 4/\mu_{\min,0},\ \text{and}\ \left\|D^{-1}C'\right\|^2\leq 4\mu_{\max,0}/\mu_{\min,0}.\end{aligned}$$

Further,

$$\left|\operatorname{tr}\left(M_j^{-1}-M^{-1}\right)\right|\leq 8/\left((\Im z)\,\mu_{\min,j}\right),\ \left|\operatorname{tr}\left(D_j^{-1}C_j'M_j^{-1}-D^{-1}C'M^{-1}\right)\right|\leq 32\mu_{\max,0}^{1/2}/\left((\Im z)\,\mu_{\min,j}^{3/2}\right),$$

and

$$\left|\operatorname{tr}\left(D_j^{-1}C_j'M_j^{-1}C_jD_j^{-1}-D^{-1}C'M^{-1}CD^{-1}\right)\right|\leq 96\mu_{\max,0}/\left((\Im z)\,\mu_{\min,j}^2\right).$$

Proof. By definition of M_j, we have

$$\left\|M_j^{-1}\right\|=\left\|A_j^{-1/2}\left(A_j^{-1/2}C_jD_j^{-1}C_j'A_j^{-1/2}-zI_p\right)^{-1}A_j^{-1/2}\right\|\leq\left\|A_j^{-1}\right\|\left\|\left(A_j^{-1/2}C_jD_j^{-1}C_j'A_j^{-1/2}-zI_p\right)^{-1}\right\|.$$

On the other hand, $\left\|A_j^{-1}\right\|=\mu_{\min,j}^{-1}$ and $\left\|\left(A_j^{-1/2}C_jD_j^{-1}C_j'A_j^{-1/2}-zI_p\right)^{-1}\right\|\leq 1/(\Im z)$. Therefore,

$$\left\|M_j^{-1}\right\|\leq 1/\left((\Im z)\,\mu_{\min,j}\right). \tag{73}$$

The required bound for $\left\|M^{-1}\right\|$ is established similarly.

Further, denoting the k-th largest eigenvalue of a symmetric matrix S as $\lambda_k(S)$, we have

$$\left\|D_j^{-1}\right\|=1/\lambda_p(D_j)\leq 1/\left(\lambda_p\left(\nabla\nabla'\right)\mu_{\min,j}\right)\leq 4/\mu_{\min,j}. \tag{74}$$

The required bound on $\left\|D^{-1}\right\|$ is established similarly. Next,

$$\left\|D_j^{-1}C_j'\right\|^2=\left\|D_j^{-1}C_j'C_jD_j^{-1}\right\|=\left\|D_j^{-1}\varepsilon_{-(j)}\nabla_{-j}\varepsilon_{-(j)}'\varepsilon_{-(j)}\nabla_{-j}'\varepsilon_{-(j)}'D_j^{-1}\right\|,$$

where ∇_{-j} is the block-diagonal matrix obtained from ∇ by removing its j-th 2×2 block, and $\varepsilon_{-(j)}$ is obtained from ε by removing the $2j-1$-th and $2j$-th columns. On the other hand,

$$\begin{aligned}\left\|D_j^{-1}\varepsilon_{-(j)}\nabla_{-j}\varepsilon_{-(j)}'\varepsilon_{-(j)}\nabla_{-j}'\varepsilon_{-(j)}'D_j^{-1}\right\| &\leq \mu_{\max,j}\left\|D_j^{-1}\varepsilon_{-(j)}\nabla_{-j}\nabla_{-j}'\varepsilon_{-(j)}'D_j^{-1}\right\|\\ &= \mu_{\max,j}\left\|D_j^{-1}D_jD_j^{-1}\right\|=\mu_{\max,j}\left\|D_j^{-1}\right\|.\end{aligned}$$

Using (74), we obtain

$$\left\|D_j^{-1}C_j'\right\|^2 \leq 4\mu_{\max,j}/\mu_{\min,j}. \tag{75}$$

The required bound for $\left\|D^{-1}C'\right\|$ is established similarly.

Now let us establish the bounds on the differences of traces. As follows from (52), M_j^{-1} differs from M^{-1} by a matrix of rank no larger than 4. Therefore,

$$\left|\mathrm{tr}\left(M_j^{-1} - M^{-1}\right)\right| \leq 4\left\|M_j^{-1} - M^{-1}\right\| \leq 4\left(\left\|M_j^{-1}\right\| + \left\|M^{-1}\right\|\right), \tag{76}$$

so that

$$\left|\mathrm{tr}\left(M_j^{-1} - M^{-1}\right)\right| \leq 4/\left(\left(\Im z\right)\mu_{\min,j}\right) + 4/\left(\left(\Im z\right)\mu_{\min,0}\right) \leq 8/\left(\left(\Im z\right)\mu_{\min,j}\right), \tag{77}$$

where the last inequality holds because $A - A_j$ is a positive-semidefinite matrix and hence $\mu_{\min,j} \leq \mu_{\min,0}$.

Similarly, $D_j^{-1}C_j'M_j^{-1}$ differs from $D^{-1}C'M^{-1}$ by a matrix with rank no larger than 8. It is because

$$D_j^{-1}C_j'M_j^{-1} - D^{-1}C'M^{-1} = D_j^{-1}C_j'\left(M_j^{-1} - M^{-1}\right) + D_j^{-1}\left(C_j' - C'\right)M^{-1} + \left(D_j^{-1} - D^{-1}\right)C'M^{-1},$$

where the rank of $M_j^{-1} - M^{-1}$ is no larger than 4, and the ranks of $C_j' - C'$ and $D_j^{-1} - D^{-1}$ are no larger than 2 each. Therefore,

$$\left|\mathrm{tr}\left(D_j^{-1}C_j'M_j^{-1} - D^{-1}C'M^{-1}\right)\right| \leq 8\left(\left\|D_j^{-1}C_j'\right\|\left\|M_j^{-1}\right\| + \left\|D^{-1}C'\right\|\left\|M^{-1}\right\|\right) \leq 32\mu_{\max,0}^{1/2}/\left(\left(\Im z\right)\mu_{\min,j}^{3/2}\right),$$

where we used (73) and (75). Finally, $D_j^{-1}C_j'M_j^{-1}C_jD_j^{-1}$ differs from $D^{-1}C'M^{-1}CD^{-1}$ by a matrix with rank no larger than 12. Therefore,

$$\left|\mathrm{tr}\left(D_j^{-1}C_j'M_j^{-1}C_jD_j^{-1} - D^{-1}C'M^{-1}CD^{-1}\right)\right| \leq 96\mu_{\max,0}/\left(\left(\Im z\right)\mu_{\min,j}^2\right).\square$$

Now, we are ready to formulate and prove Lemmas 14, 15, and 16.

Lemma 14 *For any pair* $\left(x_j^{(q)}, x_p\right) \in \left\{\left(s_j^{(q)}, s_p\right), \left(u_j^{(q)}, u_p\right), \left(v_j^{(q)}, v_p\right), \left(w_j^{(q)}, w_p\right)\right\}$ *and any* $z \in \mathbb{C}^+$, *as* $p, T \to_c \infty$, *we have*

$$\max_{j=1,\ldots,\bar{T}/2}\left\|x_j^{(q)} - x_pI_2\right\| \overset{\text{a.s.}}{\to} 0.$$

Proof. First, let us prove the convergence

$$\max_{j=1,\ldots,\bar{T}/2}\left\|s_j^{(q)} - s_pI_2\right\| \overset{\text{a.s.}}{\to} 0. \tag{78}$$

Since the square of the spectral norm is no larger than the sum of the squared elements of the matrix, it is sufficient to prove the element-wise convergences. Take, for example, the element in the second row and the second column of $s_j^{(q)} - s_pI_2$. We need to show that

$$\max_{j=1,\ldots,\bar{T}/2}\left|\epsilon_{2j}'D_j^{-1}\epsilon_{2j} - s_p\right| \overset{\text{a.s.}}{\to} 0, \tag{79}$$

or, equivalently, that for any $\tau > 0$

$$\Pr\left(\max_{j=1,\ldots,\bar{T}/2}\left|\epsilon_{2j}'D_j^{-1}\epsilon_{2j} - s_p\right| > \tau \text{ i.o.}\right) = 0, \tag{80}$$

where "i.o." stands for "infinitely often".

Let $\mathcal{B}_j$ be the indicator function of the event $\mu_{\min,j} > \underline{\mu}$, where $\underline{\mu}$ is a positive number smaller than $(1-\sqrt{c})^2$ (recall that $c < 1$ without loss of generality). Theorem II.13 of Davidson and Szarek (2001) implies that the probability of $\mathcal{B}_j = 0$, is decaying exponentially fast in p. Hence, by the Borel-Cantelli lemma,

$$\Pr\left(\cup_{j=1}^{\bar{T}/2}(\mathcal{B}_j = 0) \text{ i.o.}\right) = 0.$$

On the other hand,

$$\begin{aligned} &\Pr\left(\max_{j=1,\ldots,\bar{T}/2}\left|\epsilon_{2j}' D_j^{-1}\epsilon_{2j} - s_p\right| > \tau \text{ i.o.}\right) \\ \leq\; &\Pr\left(\max_{j=1,\ldots,\bar{T}/2}\left|\epsilon_{2j}' D_j^{-1}\epsilon_{2j} - s_p\right| > \tau \cap_{j=1}^{\bar{T}/2}(\mathcal{B}_j = 1) \text{ i.o.}\right) + \Pr\left(\cup_{j=1}^{\bar{T}/2}(\mathcal{B}_j = 0) \text{ i.o.}\right). \end{aligned}$$

Therefore,

$$\begin{aligned} &\Pr\left(\max_{j=1,\ldots,\bar{T}/2}\left|\epsilon_{2j}' D_j^{-1}\epsilon_{2j} - s_p\right| > \tau \text{ i.o.}\right) \\ \leq\; &\Pr\left(\max_{j=1,\ldots,\bar{T}/2}\left|\epsilon_{2j}' D_j^{-1}\epsilon_{2j} - s_p\right| > \tau \cap_{j=1}^{T/2}(\mathcal{B}_j = 1) \text{ i.o.}\right) \\ \leq\; &\Pr\left(\max_{j=1,\ldots,\bar{T}/2}\left|\mathcal{B}_j\epsilon_{2j}' D_j^{-1}\epsilon_{2j} - \mathcal{B}_j s_p\right| > \tau \text{ i.o.}\right) \\ \leq\; &\Pr\left(\max_{j=1,\ldots,\bar{T}/2}\left|\epsilon_{2j}' D_j^{-1}\mathcal{B}_j\epsilon_{2j} - \operatorname{tr}\left[D_j^{-1}\mathcal{B}_j\right]/\bar{T}\right| > \tau/2 \text{ i.o.}\right) \\ &+ \Pr\left(\max_{j=1,\ldots,\bar{T}/2}\left|\operatorname{tr}\left[\left(D^{-1} - D_j^{-1}\right)\mathcal{B}_j\right]/\bar{T}\right| > \tau/2 \text{ i.o.}\right). \end{aligned}$$

By Lemma 11, for any $\rho \geq 2$,

$$\mathbb{E}\left|\epsilon_{2j}' D_j^{-1}\mathcal{B}_j\epsilon_{2j} - \operatorname{tr}\left[D_j^{-1}\mathcal{B}_j\right]/\bar{T}\right|^{\rho} \leq K p^{\rho/2}\bar{T}^{-\rho}\mathbb{E}\left\|D_j^{-1}\mathcal{B}_j\right\|^{\rho} \tag{81}$$

On the other hand, by Lemma 13, $\left\|D_j^{-1}\right\| \leq 4/\mu_{\min,j}$ and thus,

$$\left\|D_j^{-1}\mathcal{B}_j\right\| \leq 4/\underline{\mu}, \text{ and } \mathbb{E}\left\|D_j^{-1}\mathcal{B}_j\right\|^{\rho} \leq \left(4/\underline{\mu}\right)^{\rho}.$$

Combining this with (81), we obtain

$$\mathbb{E}\left|\epsilon_{2j}' D_j^{-1}\mathcal{B}_j\epsilon_{2j} - \operatorname{tr}\left[D_j^{-1}\mathcal{B}_j\right]/\bar{T}\right|^{\rho} \leq K p^{\rho/2}\bar{T}^{-\rho},$$

where K depends only on ρ and c. By Markov's inequality

$$\Pr\left(\max_{j=1,\ldots,\bar{T}/2}\left|\epsilon_{2j}' D_j^{-1}\mathcal{B}_j\epsilon_{2j} - \operatorname{tr}\left[D_j^{-1}\mathcal{B}_j\right]/\bar{T}\right| > \tau/2\right) \leq \frac{\bar{T}}{2}\frac{K p^{\rho/2}\bar{T}^{-\rho}}{(\tau/2)^{\rho}}.$$

The right hand side of the latter inequality is summable over p for sufficiently large ρ, and therefore,

$$\Pr\left(\max_{j=1,\ldots,\bar{T}/2}\left|\epsilon_{2j}' D_j^{-1}\mathcal{B}_j\epsilon_{2j} - \operatorname{tr}\left[D_j^{-1}\mathcal{B}_j\right]/\bar{T}\right| > \tau/2 \text{ i.o.}\right) = 0. \tag{82}$$

Finally, since the rank of the positive semi-definite matrix $D_j^{-1} - D^{-1}$ is no larger than two, we have by Weyl's theorem (see Theorem 4.3.6 in Horn and Johnson (1985))

$$\left|\operatorname{tr}\left[\left(D^{-1} - D_j^{-1}\right)\mathcal{B}_j\right]\right| \leq 2\left\|D_j^{-1}\mathcal{B}_j\right\| \leq 8/\underline{\mu},$$

and

$$\Pr\left(\max_{j=1,\ldots,\bar{T}/2}\left|\operatorname{tr}\left[\left(D^{-1} - D_j^{-1}\right)\mathcal{B}_j\right]/\bar{T}\right| > \tau/2 \text{ i.o.}\right) = 0. \tag{83}$$

Equalities (82) and (83) imply that (80) holds.

The convergence of the element in the first row and the first column of $s_j^{(q)} - s_p I_2$ can be shown similarly to (79). For the off-diagonal elements, note that

$$\epsilon'_{2j-1} D_j^{-1} \epsilon_{2j} = \frac{1}{2} \left(\epsilon'_{2j-1}, \epsilon'_{2j} \right) \begin{pmatrix} 0 & D_j^{-1} \\ D_j^{-1} & 0 \end{pmatrix} \begin{pmatrix} \epsilon_{2j-1} \\ \epsilon_{2j} \end{pmatrix}.$$

Hence, we again can use Lemma 11 and the Borel-Cantelli lemma to obtain desired results.

The a.s. convergences of the maxima over $j = 1, ..., \bar{T}/2$ of $\left\| u_j^{(q)} - u_p I_2 \right\|, \left\| v_j^{(q)} - v_p I_2 \right\|$ and $\left\| w_j^{(q)} - w_p I_2 \right\|$ can be proven by closely following the strategy of the above proof of (78). We omit details. The only two new aspects of the remaining proofs are related to the need for bounds on the spectral norms of $D_j^{-1} C_j' M_j^{-1}$, M_j^{-1}, and $D_j^{-1} C_j' M_j^{-1} C_j D_j^{-1}$, and on the differences between the traces of these matrices and the traces of $D^{-1} C' M^{-1}$, M^{-1}, and $D^{-1} C' M^{-1} C D^{-1}$, respectively. Such bounds are provided by Lemma 13.□

Lemma 15 *For any $z \in \mathbb{C}^+$, quantities $u_p \equiv u_p(z)$, $v_p \equiv v_p(z)$, $w_p \equiv w_p(z)$, and $s_p \equiv s_p$ almost surely remain bounded as $p, T \to_c \infty$.*

Proof. By definition

$$|u_p| \leq \frac{p}{T} \left\| D^{-1} C' M^{-1} \right\|, |v_p| \leq \frac{p}{T} \left\| M^{-1} \right\|, |w_p| \leq \frac{p}{T} \left\| D^{-1} C' M^{-1} C D^{-1} \right\|, \text{ and } |s_p| \leq \frac{p}{T} \left\| D^{-1} \right\|.$$

Therefore, Lemma 15 follows from Lemma 13 and the convergences (see e.g. Bai and Yin (1993)) $\mu_{\min,0} \overset{\text{a.s.}}{\to} (1 - \sqrt{c})^2$ and $\mu_{\max,0} \overset{\text{a.s.}}{\to} (1 + \sqrt{c})^2$.□

Lemma 16 *There exists $\zeta > 0$ such that, for any z with $\Re z = 0$ and $\Im z > \zeta$, a.s.,*

$$\liminf_{p,T \to_c \infty} \max_{j=1,...,T/2} |\delta_{pj}| > c^2 / \left(1 - c^2\right) \text{ and } \liminf_{p,T \to_c \infty} \sup_{\varphi} |\delta_p(\varphi)| > c^2 / \left(1 - c^2\right).$$

Proof. By Lemma 9, $z\tilde{v}_p = w_p - s_p$. Elementary algebra then yields the following representation $\delta_{pj} = \delta_{pj}^{(1)} + \frac{1}{z} \delta_{pj}^{(2)}$, where

$$\delta_{pj}^{(1)} = (r_j + s_p)(zv_p - 1),$$

and

$$\delta_{pj}^{(2)} = (zw_p)(1 + v_p - zv_p) - s_p(zv_p) + r_j(zu_p) - \frac{1 - z}{z} (zu_p)^2$$

Note that for $z \in \mathbb{C}^+$, $v_p \in \mathbb{C}^+$. Hence, for $z \in \mathbb{C}^+$ such that $\Re z = 0$, we have $\Re(zv_p) < 0$ and

$$|zv_p - 1| > 1. \tag{84}$$

This inequality and Lemma 12 imply that, for any $z \in \mathbb{C}^+$ such that $\Re z = 0$, $\left| \delta_{pj}^{(1)} \right| > \frac{2c^2}{1-c^2}$ for sufficiently large p, T as $p, T \to_c \infty$, a.s.

Further, Lemma 13 implies that $|zu_p|, |zv_p|$, and $|zw_p|$ (as well as $|u_p|, |v_p|$, and $|w_p|$) remain bounded for sufficiently large p, T as $p, T \to_c \infty$, a.s. Moreover, the presence of the imaginary part of z in the denominator of the bound on $\left\| M^{-1} \right\|$ in Lemma 13 imply that, for $z \in \mathbb{C}^+$ such that $\Re z = 0$, the value of the bound on $|zu_p|, |zv_p|$, and $|zw_p|$ does not depend on z. In particular, for any such z, $\left| \delta_{pj}^{(2)} \right|$ is bounded for sufficiently large p, T as $p, T \to_c \infty$, a.s., uniformly in j, with the value of the bound independent from z with $\Im z > \zeta$.[3] Hence, by choosing ζ sufficiently large, we can ensure that, for any z with $\Re z = 0$ and $\Im z > \zeta$, $\left| \frac{1}{z} \delta_{pj}^{(2)} \right| < \frac{1}{2} \left| \delta_{pj}^{(1)} \right|$, and therefore

$$|\delta_{pj}| > c^2 / \left(1 - c^2\right).$$

A proof of the a.s. uniform over φ bound on $\delta_p(\varphi)$ is almost identical to the above proof, and therefore we omit details.□

[3] For such z, there exist bounds on $|u_p|, |v_p|$, and $|w_p|$ that depend only on ζ.

2.1.6 Solving the system

By definition, $|m_p(z)|$ is bounded by $(\Im z)^{-1}$. Further, by Lemmas 9 and 15, $v_p \equiv v_p(z), \tilde{v}_p \equiv \tilde{v}_p(z)$, and $u_p \equiv u_p(z)$ are a.s. bounded by absolute value. Therefore, with probability one, there exists a subsequence of p, T along which $m_p(z), v_p(z), \tilde{v}_p(z)$, and $u_p(z)$ converge to some limits $m, v, \tilde{v}$, and u.

These limits must satisfy a limiting version of system (65), that is equations

$$\begin{cases} m = \frac{1}{2\pi c}\int_0^{2\pi} \delta^{-1}(\varphi)\left(z\tilde{v}v - u^2 - 4v\sin^2\varphi\right)\mathrm{d}\varphi \\ m = \frac{1}{2\pi c}\int_0^{2\pi} \delta^{-1}(\varphi)\left(zv\tilde{v} - \tilde{v} - u^2\right)\mathrm{d}\varphi \\ 1 + zm = \frac{1}{2\pi c}\int_0^{2\pi} \delta^{-1}(\varphi)\left(2u\sin^2\varphi + zv\tilde{v} - u^2\right)\mathrm{d}\varphi \\ 0 = \frac{1}{2\pi c}\int_0^{2\pi} \delta^{-1}(\varphi)\left(2v\sin^2\varphi + u\right)\mathrm{d}\varphi \end{cases}, \tag{85}$$

where

$$\delta(\varphi) = (1 - z)\left(z\tilde{v}v - u^2\right) + z\tilde{v} + 4\sin^2\varphi\,(zv + u - 1). \tag{86}$$

Let us consider, until further notice, only such z that $\Re z = 0$ and $\Im z > \zeta$, for some $\zeta > 0$. Let us solve system (85) for m. Adding two times the last equation to the first one, and subtracting the second equation we obtain

$$0 = \frac{1}{2\pi c}\int_0^{2\pi} \delta^{-1}(\varphi)\,(2u + \tilde{v})\,\mathrm{d}\varphi. \tag{87}$$

Note that $\int_0^{2\pi} \delta^{-1}(\varphi)\,\mathrm{d}\varphi \neq 0$. Otherwise, from the second equation of (85), we have $m = 0$, which cannot be true. Indeed, for any $0 \leq \lambda \leq 1$ and $z \in \mathbb{C}^+$ with $\Re z = 0$,

$$\Im\left(\frac{1}{\lambda - z}\right) = \frac{\Im z}{\lambda^2 + (\Im z)^2} \geq \frac{\Im z}{1 + (\Im z)^2}.$$

Therefore, $\Im m_p(z) \geq \Im z / \left(1 + (\Im z)^2\right)$, and $m_p(z)$ cannot converge to $m = 0$.

Since $\int_0^{2\pi} \delta^{-1}(\varphi)\,\mathrm{d}\varphi \neq 0$, (87) yields

$$\tilde{v} + 2u = 0 \tag{88}$$

with $\tilde{v} \neq 0$ and $u \neq 0$ (if one of them equals zero, the other equals zero too, and $m = 0$ by the second equation of (85), which is impossible). Since $u \neq 0$, the last equation implies that $v \neq 0$ as well.

Further, subtracting from the third equation the sum of z times the second and u/v times the last equation, and using (88), we obtain

$$1 = \frac{1}{2\pi c}\int_0^{2\pi} \delta^{-1}(\varphi)\,\frac{u}{v}\,(2zv + u)(zv - v - 1)\,\mathrm{d}\varphi. \tag{89}$$

This equation, together with (88) and the second equation of (85) yield

$$m = \frac{v\,(2zv + u - 2)}{(1 + v - zv)(2zv + u)}. \tag{90}$$

Next, for the integrand in the last equation of (85), it is straightforward to verify using (66) and (88) that

$$\delta^{-1}(\varphi)\left(2v\sin^2\varphi + u\right) = \frac{1}{2}\frac{v}{zv + u - 1} + \delta^{-1}(\varphi)\,\frac{u}{2}\left(\frac{(1 - z)\,v\,(2zv + u) + 2\,(2zv + u - 1)}{zv + u - 1}\right). \tag{91}$$

This assumes that

$$zv + u - 1 \neq 0, \tag{92}$$

which must hold because otherwise,

$$\delta(\varphi) = (1 - z)\left(z\tilde{v}v - u^2\right) + z\tilde{v}$$

would not depend on φ and the last equation of (85) would imply that $u+v=0$. The latter equation and the equality $zv+u-1=0$ would yield $v=-(1-z)^{-1}$, which, when combined with the second equation of (85), would give us $m=-c^{-1}(1-z)^{-1}$. This cannot be true because m, being a limit of $m_p(z)$, must satisfy $\Im m\geq 0$ for $\Im z>0$.

Equations (89), (91), and the last equation of (85) imply that

$$u=\frac{2c}{2c-1-(1-z)\,v\,(1-c)}-2zv. \tag{93}$$

Combining this with (90) yields

$$m=v\frac{1-c}{c}. \tag{94}$$

Finally, elementary calculations given at the end of this section show that

$$\left(\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{x+2\sin^2\varphi}\mathrm{d}\varphi\right)^2=\frac{1}{x\,(x+2)}, \tag{95}$$

where $x\in\mathbb{C}\backslash[-2,0]$. Using (95), (89), and the definition of $\delta(\varphi)$, we obtain the following relationship

$$\left(\frac{2cv\,(zv+u-1)}{u\,(2zv+u)\,(zv-v-1)}\right)^2=\frac{4\,(zv+u-1)^2}{u\,((1-z)\,(-2zv-u)-2z)\,(-u+uz+2)\,(u+2vz-2)}, \tag{96}$$

that holds as long as

$$\frac{u\,((1-z)\,(-2zv-u)-2z)}{2\,(zv+u-1)}\in\mathbb{C}\backslash\left[-2,0\right].$$

The latter inclusion holds because otherwise $\delta(\varphi)$ is not a bounded function of φ, which would contradict Lemma 16.

Using (93) in (96), and simplifying, we find that there exist only three possibilities. Either

$$v=-\frac{1}{1-z}, \tag{97}$$

or

$$1-(c+cz-1)\,v+z\,(1-z)\,(1-c)\,v^2=0, \tag{98}$$

or

$$\frac{c}{1-c}-(c+cz-z)\,v+z\,(1-z)\,(1-c)\,v^2=0. \tag{99}$$

Equation (97) cannot hold because otherwise, (94) would imply that $\Im m<0$, which is impossible as argued above. Equation (98) taken together with (93) implies that

$$u+zv-1=0,$$

which was ruled out above. This leaves us with (99), so that, using (94), we get

$$m=\frac{-(z-c-cz)\pm\sqrt{(z-c-cz)^2-4c\,(1-z)\,z}}{2z\,(1-z)\,c}. \tag{100}$$

For $z\in\mathbb{C}^+$ with $\Re z=0$, the imaginary part of the right hand side of (100) is negative when '$-$' is used in front of the square root. Here we choose the branch of the square root, with the cut along the positive real semi-axis, which has positive imaginary part. Since $\Im m$ cannot be negative, we conclude that

$$m=\frac{-(z-c-cz)+\sqrt{(z-c-cz)^2-4c\,(1-z)\,z}}{2z\,(1-z)\,c}. \tag{101}$$

But the right hand side of the above equality is the value of the limit of the Stieltjes transforms of the eigenvalues of the multivariate beta matrix $B_p\left(p,(T-p)/2\right)$ as $p,T\to_c\infty$. This can be verified directly by using the formula for such a limit, given for example in Theorem 1.6 of Bai, Hu, Pan and Zhou (2015). As follows from Wachter (1980), the weak limit of the empirical distribution of the eigenvalues of the multivariate beta matrix $B_p\left(p,(T-p)/2\right)$ as $p,T\to_c\infty$ equals $W\left(\lambda;c/\left(1+c\right),2c/\left(1+c\right)\right)$.

Equation (101) shows that, for z with $\Re z=0$ and $\Im z>\zeta$, with probability one, any converging subsequence of $m_p(z)$ converges to the same limit. Hence, $m_p(z)$ a.s. converges for all z with $\Re z=0$ and $\Im z>\zeta$. Note that $m_p(z)$ is a sequence of bounded analytic functions in the domain $\{z:\Im z>d\}$, where d is an arbitrary positive number. Therefore, by Vitaly's convergence theorem (see Titchmarsh (1939), p.168) $m_p(z)$ a.s. converges to m, described by (101), for any $z\in\mathbb{C}^+$. The a.s. convergence of $F_p\left(\lambda\right)$ to the Wachter distribution follows from the Continuity Theorem for the Stieltjes transforms (see, for example, Corollary 1 in Geronimo and Hill (2003)).

Note that we have just proven the a.s. weak convergence $F_p\left(\lambda\right)\Rightarrow W_c\left(\lambda\right)$ for Gaussian data. Since, as has been shown above, the Gaussian and non-Gaussian versions of $F_p\left(\lambda\right)$ are asymptotically equivalent in probability, we conclude that the weak convergence $F_p\left(\lambda\right)\Rightarrow W_c\left(\lambda\right)$ takes place for non-Gaussian data too, albeit, possibly only in probability.

Proof of (95). Consider

$$\mathcal{I}=\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{x+2\sin^2\varphi}\mathrm{d}\varphi,$$

where $x\in\mathbb{C}\backslash\left[-2,0\right]$. Changing the variable of integration to $z=\exp\left\{\mathrm{i}\varphi\right\}$, we obtain

$$\mathcal{I}=\frac{1}{2\pi\mathrm{i}}\oint\limits_{|z|=1}\frac{1}{x-\left(z-z^{-1}\right)^2/2}\frac{\mathrm{d}z}{z}=-\frac{1}{2\pi\mathrm{i}}\oint\limits_{|z|=1}\frac{2z}{\left(z^2-x_1\right)\left(z^2-x_2\right)}\mathrm{d}z,$$

where

$$x_{1,2}=x+1\pm\sqrt{x\left(x+2\right)}.$$

Since $x_1x_2=1$, whereas $|x_1|\neq1$ and $|x_2|\neq1$, there are only two poles of the integrand that are situated inside the unit circle. They are either $x_1^{1/2},-x_1^{1/2}$, which we shall call case 1, or $x_2^{1/2},-x_2^{1/2}$, which we shall call case 2. By Cauchy's residue theorem,

$$\mathcal{I}=\mp\frac{2}{x_1-x_2},$$

with "$-$" corresponding to case 1 and "$+$" corresponding to case 2. Whatever the case, we have

$$\mathcal{I}^2=\frac{4}{\left(x_1-x_2\right)^2}=\frac{1}{x\left(x+2\right)}.$$

2.2 Proof of Corollary OW2

Since the probability limits of $PB_{r,p,T}/p^2$ and $PB_{0,p,T}/p^2$ are the same as long as $r/p\to0$, we shall only compute the latter limit. As $p,T\to_c\infty$, we have

$$PB_{0,p,T}/p^2\xrightarrow{\mathrm{P}}c^{-1}\int\lambda\mathrm{d}W_c(\lambda).$$

Using the explicit formula for the density of the Wachter distribution OW(9), we obtain,

$$\int\lambda\mathrm{d}W_c(\lambda)=\frac{1+c}{2\pi c}\int_{b_-}^{b_+}\frac{\sqrt{\left(b_+-\lambda\right)\left(\lambda-b_-\right)}}{1-\lambda}\mathrm{d}\lambda+\max\left\{0,2-1/c\right\},\qquad(102)$$

where

$$b_\pm=c\left(\sqrt{2}\mp\sqrt{1-c}\right)^{-2}.$$

Denote $\int \lambda \mathrm{d}W_c(\lambda) - \max\{0, 2-1/c\}$ as $\mathcal{I}$. Let

$$x = (\lambda - b_-)\,/\,(b_+ - b_-)$$

so that $\lambda = b_- + (b_+ - b_-)\,x$. Then

$$\mathcal{I} = \frac{1+c}{2c\pi}\int_0^1 \frac{(b_+ - b_-)^2\sqrt{(1-x)\,x}}{1 - b_- - (b_+ - b_-)\,x}\mathrm{d}x.$$

Changing variables to θ where $x = (1-\cos\theta)/2$ so that $\mathrm{d}x = \frac{1}{2}\sin\theta\mathrm{d}\theta$, we obtain

$$\mathcal{I} = \frac{1+c}{8c\pi}\int_0^\pi \frac{(b_+ - b_-)^2\sin^2\theta}{\frac{2-b_+-b_-}{2} + \frac{b_+-b_-}{2}\cos\theta}\mathrm{d}\theta.$$

Further, letting $z = \cos\theta + \mathrm{i}\sin\theta$ so that

$$\cos\theta = \frac{z+z^{-1}}{2},\ \sin\theta = \frac{z-z^{-1}}{2\mathrm{i}},\ \text{and } \mathrm{d}\theta = \frac{\mathrm{d}z}{\mathrm{i}z},$$

we obtain

$$\mathcal{I} = -\frac{1+c}{16c\pi\mathrm{i}}\oint\limits_{|z|=1} \frac{(b_+ - b_-)^2\left(\frac{z-z^{-1}}{2}\right)^2}{\frac{2-b_+-b_-}{2} + \frac{b_+-b_-}{2}\frac{z+z^{-1}}{2}}\frac{\mathrm{d}z}{z},$$

where the contour integral is taken over the unit circle in the complex plane. Noting that

$$\frac{2-b_+-b_-}{2} + \frac{b_+-b_-}{2}\frac{z+z^{-1}}{2} = (a+bz)\left(a+bz^{-1}\right),$$

where

$$a = \frac{\sqrt{1-b_-}+\sqrt{1-b_+}}{2},\ b = \frac{\sqrt{1-b_-}-\sqrt{1-b_+}}{2},$$

we represent $\mathcal{I}$ in the following form

$$\mathcal{I} = -\frac{1+c}{64c\pi\mathrm{i}}\oint\limits_{|z|=1} \frac{(b_+ - b_-)^2\left(z^2-1\right)^2}{a\,(a+bz)\,(z+b/a)}\frac{\mathrm{d}z}{z^2}.$$

Since $a > b > 0$, the integrand has poles at 0 and $-b/a$. The corresponding residues are

$$r_0 = \frac{1+c}{2c}\left(a^2+b^2\right),$$

and

$$r_{-b/a} = -\frac{1+c}{2c}\left(a^2-b^2\right),$$

so that

$$\mathcal{I} = \frac{1+c}{2c}\left(a^2+b^2\right) - \frac{1+c}{2c}\left(a^2-b^2\right).$$

Noting that

$$a^2+b^2 = \frac{-c+2c^2+1}{(c+1)^2} \text{ and } a^2-b^2 = \frac{1-2c}{1+c},$$

we further simplify the above expression for $\mathcal{I}$ to obtain

$$\mathcal{I} = \frac{2c}{c+1}.$$

Therefore,

$$\int \lambda \mathrm{d}W_c(\lambda) = \frac{2c}{c+1} + \max\{0, 2-1/c\},$$

and

$$PB_{0,p,T}/p^2 \xrightarrow{\mathrm{P}} \frac{2}{c+1} + \frac{1}{c}\max\{0, 2-1/c\}.$$

As follows from Theorem OW1, in cases where ε_t are Gaussian, the convergence is a.s.

2.3 Proof of Corollary OW3

As explained in OW, for $c < 1/2$, we have

$$LR_{r,p,T}/p^2 \geq -\left(T/p^2\right) \sum_{j=r+1}^{p} \overline{\log}(1-\lambda_j).$$

Therefore,

$$LR_{r,p,T}/p^2 \geq -\left(T/p\right) \int \overline{\log}\left(1-\lambda\right) \mathrm{d}F_p\left(\lambda\right) + \left(T/p^2\right) \sum_{j=1}^{r} \overline{\log}(1-\lambda_j).$$

If $r/p \to 0$ as $p, T \to_c \infty$, the second term on the right hand side of the above display converges to zero. Therefore, by Theorem OW1, for any $\delta > 0$,

$$\Pr\left(LR_{r,p,T}/p^2 < -c^{-1} \int \overline{\log}\left(1-\lambda\right) \mathrm{d}W_c(\lambda) - \delta\right) \to 0,$$

and in cases of Gaussian ε_t, almost surely,

$$\liminf LR_{r,p,T}/p^2 \geq -c^{-1} \int \overline{\log}\left(1-\lambda\right) \mathrm{d}W_c(\lambda).$$

By definition of $\overline{\log}$,

$$-c^{-1} \int \overline{\log}\left(1-\lambda\right) \mathrm{d}W_c(\lambda) = -c^{-1} \int \log\left(1-\lambda\right) \mathrm{d}W_c(\lambda) \equiv \underline{LR}_c.$$

Using the explicit formula for the density of the Wachter distribution OW(9), we obtain

$$\underline{LR}_c = -\frac{1+c}{2\pi c^2} \int_{b_-}^{b_+} \log\left(1-\lambda\right) \frac{\sqrt{\left(b_+ - \lambda\right)\left(\lambda - b_-\right)}}{\lambda\left(1-\lambda\right)} \mathrm{d}\lambda,$$

where

$$b_\pm = c\left(\sqrt{2} \mp \sqrt{1-c}\right)^{-2}.$$

Let $x = \left(\lambda - b_-\right) / \left(b_+ - b_-\right)$ so that $\lambda = b_- + \left(b_+ - b_-\right) x$. Then

$$\underline{LR}_c = -\frac{1+c}{2c^2\pi} \int_0^1 \frac{\log\left(1 - b_- - \left(b_+ - b_-\right) x\right) \sqrt{\left(1-x\right) x} \left(b_+ - b_-\right)^2}{\left(\left(b_+ - b_-\right) x + b_-\right)\left(1 - b_- - \left(b_+ - b_-\right) x\right)} \mathrm{d}x.$$

Changing variables to θ where $x = \left(1 - \cos\theta\right)/2$ so that $\mathrm{d}x = \frac{1}{2}\sin\theta \mathrm{d}\theta$, we obtain

$$\underline{LR}_c = -\frac{1+c}{8c^2\pi} \int_0^\pi \frac{\log\left(\frac{2-b_+-b_-}{2} + \frac{b_+-b_-}{2}\cos\theta\right)\left(b_+ - b_-\right)^2 \sin^2\theta}{\left(\frac{b_++b_-}{2} - \frac{b_+-b_-}{2}\cos\theta\right)\left(\frac{2-b_+-b_-}{2} + \frac{b_+-b_-}{2}\cos\theta\right)} \mathrm{d}\theta.$$

Further, letting $z = \cos\theta + \mathrm{i}\sin\theta$ so that

$$\cos\theta = \frac{z + z^{-1}}{2},\ \sin\theta = \frac{z - z^{-1}}{2\mathrm{i}},\ \text{and } \mathrm{d}\theta = \frac{\mathrm{d}z}{\mathrm{i}z},$$

we obtain

$$\underline{LR}_c = \frac{1+c}{16c^2\pi\mathrm{i}} \oint_{|z|=1} \frac{\log\left(\frac{2-b_+-b_-}{2} + \frac{b_+-b_-}{2}\frac{z+z^{-1}}{2}\right)\left(b_+ - b_-\right)^2 \left(\frac{z-z^{-1}}{2}\right)^2}{\left(\frac{b_++b_-}{2} - \frac{b_+-b_-}{2}\frac{z+z^{-1}}{2}\right)\left(\frac{2-b_+-b_-}{2} + \frac{b_+-b_-}{2}\frac{z+z^{-1}}{2}\right)} \frac{\mathrm{d}z}{z},$$

where the contour integral is taken over the unit circle in the complex plane. Noting that

$$\frac{2-b_+-b_-}{2}+\frac{b_+-b_-}{2}\frac{z+z^{-1}}{2}=(a+bz)\left(a+bz^{-1}\right),$$

where

$$a=\frac{\sqrt{1-b_-}+\sqrt{1-b_+}}{2},\ b=\frac{\sqrt{1-b_-}-\sqrt{1-b_+}}{2},$$

and that

$$\frac{b_++b_-}{2}-\frac{b_+-b_-}{2}\frac{z+z^{-1}}{2}=(e-dz)\left(e-dz^{-1}\right),$$

where

$$e=\frac{\sqrt{b_+}+\sqrt{b_-}}{2}, d=\frac{\sqrt{b_+}-\sqrt{b_-}}{2},$$

we represent $\underline{LR}_c$ in the following form

$$\begin{aligned}\underline{LR}_c &= \frac{1+c}{4c^2\pi\mathrm{i}}\oint\limits_{|z|=1}\frac{\log\left((a+bz)\left(a+bz^{-1}\right)\right)(ed)^2\left(z^2-1\right)^2}{(e-dz)(ez-d)(a+bz)(az+b)}\frac{\mathrm{d}z}{z}\\ &= \frac{1+c}{4c^2\pi\mathrm{i}}\oint\limits_{|z|=1}\frac{\log\left((a+bz)\left(a+bz^{-1}\right)\right)ed^2\left(z^2-1\right)^2}{a(e-dz)\left(z-\frac{d}{e}\right)(a+bz)\left(z+\frac{b}{a}\right)}\frac{\mathrm{d}z}{z}.\end{aligned}$$

The integral has form $\underline{LR}_c=\oint\limits_{|z|=1}\log\left(q(z)q\left(z^{-1}\right)\right)H(z)z^{-1}\mathrm{d}z$ with $H(z)=H\left(z^{-1}\right)$. Hence, expanding the logarithm yields two identical terms, so that

$$\underline{LR}_c=\frac{1+c}{2c^2\pi\mathrm{i}}\oint\limits_{|z|=1}\frac{\log(a+bz)ed^2\left(z^2-1\right)^2}{a(e-dz)\left(z-\frac{d}{e}\right)(a+bz)\left(z+\frac{b}{a}\right)}\frac{\mathrm{d}z}{z}.$$

Since $a>b>0$ and $e>d>0$, $\log(a+bz)$ is analytic inside the unit circle and the integrand has three simple poles there: 0, $-b/a$, and d/e. The corresponding residues are

$$r_0=-\frac{1+c}{c^2}\frac{ed}{ab}\log a=-\frac{1+c}{c^2}\log a,$$

$$\begin{aligned}r_{-b/a} &= \frac{1+c}{c^2}\frac{\log\left(a-\frac{b^2}{a}\right)e^2d^2\left(a^2-b^2\right)}{ab(ae+db)(be+ad)}\\ &= \frac{1+c}{c^2}\sqrt{1-b_-}\sqrt{1-b_+}\log\left(a-\frac{b^2}{a}\right)\\ &= \frac{1-2c}{c^2}\log\left(a-\frac{b^2}{a}\right),\end{aligned}$$

and

$$\begin{aligned}r_{d/e} &= \frac{1+c}{c^2}\frac{\log\left(a+\frac{bd}{e}\right)ed\left(e^2-d^2\right)}{(ae+bd)(be+ad)}\\ &= \frac{1+c}{c^2}\sqrt{b_+}\sqrt{b_-}\log\left(a+\frac{bd}{e}\right)\\ &= \frac{1}{c}\log\left(a+\frac{bd}{e}\right).\end{aligned}$$

Summing up, we obtain

$$\underline{LR}_c = -\frac{1+c}{c^2}\log a + \frac{1-2c}{c^2}\log\left(a - \frac{b^2}{a}\right) + \frac{1}{c}\log\left(a + \frac{bd}{e}\right).$$

Noting that

$$a = \frac{\sqrt{1-c}}{1+c}, a^2 - b^2 = \frac{1-2c}{1+c}, e = \frac{\sqrt{2c}}{1+c}, \text{ and } ae + bd = \frac{\sqrt{2c\left(1-c\right)}}{1+c},$$

we further simplify the above expression for $\underline{LR}_c$ to obtain

$$\underline{LR}_c = \frac{1+c}{c^2}\log\left(1+c\right) - \frac{1-c}{c^2}\log\left(1-c\right) + \frac{1-2c}{c^2}\log\left(1-2c\right).$$

3 Sequential asymptotics and over-rejection

3.1 Proof of Theorem OW4

First, let us show that the weak limit $F_0\left(\lambda\right)$ of $F_\gamma\left(\lambda\right)$ as $\gamma \to 0$ exists and equals the continuous part of the Marchenko-Pastur distribution with density OW(20). By definition and Theorem OW1, $F_\gamma\left(\lambda\right)$ is the (scaled) Wachter d.f. $W\left(\gamma\lambda; \gamma/\left(1+\gamma\right), 2\gamma/\left(1+\gamma\right)\right)$ with density $f_\gamma(\lambda)$ and support $[\hat{b}_-, \hat{b}_+]$ given by

$$f_\gamma(\lambda) = \frac{1+\gamma}{2\pi}\frac{\sqrt{(\hat{b}_+ - \lambda)(\lambda - \hat{b}_-)}}{\lambda\left(1-\gamma\lambda\right)}, \text{ and } \hat{b}_\pm = \left(\sqrt{2} \mp \sqrt{1-\gamma}\right)^{-2}.$$

As $\gamma \to 0$, $\hat{b}_\pm \to a_\pm$, where $a_\pm = \left(1 \pm \sqrt{2}\right)^2$ as in OW(19), and $f_\gamma(\lambda)$ converges to the density given by OW(20). This implies the weak convergence of $F_\gamma\left(\lambda\right)$ to $F_0\left(\lambda\right)$ with F_0 supported on $[a_-, a_+]$ and having density OW(20).

Next, recall that matrix

$$\frac{1}{p}\int_0^1 \left(\mathrm{d}B\right)F'\left(\int_0^1 FF'\mathrm{d}u\right)^{-1}\int_0^1 F\left(\mathrm{d}B\right)' \tag{103}$$

has been derived by Johansen (1991) as the limit in distribution of matrix $\frac{T}{p}\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{01}'\tilde{S}_{00}^{-1}$ as $T \to \infty$, where

$$\tilde{S}_{00} = \frac{1}{T}\varepsilon M_l\varepsilon',\ \tilde{S}_{01} = \frac{1}{T}\varepsilon M_l U'\varepsilon',\text{ and } \tilde{S}_{11} = \frac{1}{T}\varepsilon U M_l U'\varepsilon'. \tag{104}$$

In that derivation, the distribution of the data generating process is inessential. Only the i.i.d.-ness of innovations and the existence of their second moments are of importance. Therefore, for the purpose of proving Theorem OW4, we may and will assume that ε is Gaussian.

In addition, and again without loss of generality, we will assume that ε and (103) are defined on the common probability space so that the convergence of $\frac{T}{p}\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{01}'\tilde{S}_{00}^{-1}$ to (103) is in probability. We denote the empirical d.f. of the eigenvalues of $\frac{T}{p}\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{01}'\tilde{S}_{00}^{-1}$ as $\tilde{F}_{p,T}\left(\lambda\right)$, and note that, $\mathcal{L}\left(\tilde{F}_{p,T}, F_{p,0}\right) \to 0$ as $T \to \infty$ while p is held fixed. Here $\mathcal{L}\left(\cdot,\cdot\right)$ is the Lévy distance.

To establish Theorem OW4, we need to show that $F_{p,0}(\lambda) \Rightarrow F_0(\lambda)$ in probability as $p \to \infty$. It is sufficient to show that for any $\delta > 0$ and all sufficiently large p,

$$\Pr\left(\mathcal{L}\left(F_0, F_{p,0}\right) < \delta\right) > 1 - \delta. \tag{105}$$

We shall split $\mathcal{L}\left(F_0, F_{p,0}\right)$ into several parts, and show that each of them is small with high probability.

Let

$$C = \epsilon\nabla'\epsilon', D = \epsilon\nabla\nabla'\epsilon', \text{ and } A = \epsilon\epsilon',$$

as in Lemma 8. For any p, let T_γ be the smallest integer satisfying $p/T_\gamma \leq \gamma$ and let $T > T_\gamma$. Let $F_{p,T}$ denote the empirical distributions of eigenvalues of

$$\frac{T}{p}CD^{-1}C'A^{-1}, \tag{106}$$

and let F_{p,T_γ} be its counterpart when T is replaced by T_γ.

By the triangle inequality,

$$\mathcal{L}(F_0, F_{p,0}) \leq \mathcal{L}(F_0, F_\gamma) + \mathcal{L}\left(F_\gamma, F_{p,T_\gamma}\right) + \mathcal{L}\left(F_{p,T_\gamma}, F_{p,T}\right) + \mathcal{L}\left(F_{p,T}, \tilde{F}_{p,T}\right) + \mathcal{L}\left(\tilde{F}_{p,T}, F_{p,0}\right). \tag{107}$$

We can choose $\gamma > 0$ so small that the first term on the right hand side of (107) satisfies

$$\mathcal{L}(F_0, F_\gamma) < \delta/4. \tag{108}$$

By Theorem OW1, the second term a.s. converges to zero as $p \to \infty$. Therefore, for all sufficiently large p, we have

$$\Pr\left(\mathcal{L}\left(F_\gamma, F_{p,T_\gamma}\right) < \delta/4\right) > 1 - \delta/4. \tag{109}$$

Further, for any p and T, the rank of the difference between $\frac{T}{p}CD^{-1}C'A^{-1}$ and $\frac{T}{p}\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{01}'\tilde{S}_{00}^{-1}$ remains below a fixed positive integer, say m. Indeed, by Lemma 8, $\frac{T}{p}CD^{-1}C'A^{-1}$ equals $\frac{T}{p}S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$, where S_{ij} are defined by (1). On the other hand, comparing (1) with (104), we see that the rank of the difference between $\frac{T}{p}S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$ and $\frac{T}{p}\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{01}'\tilde{S}_{00}^{-1}$ is bounded by a fixed positive integer. Therefore, for the fourth term on the right hand side of (107) we have

$$\mathcal{L}\left(F_{p,T}, \tilde{F}_{p,T}\right) < m/p. \tag{110}$$

As was mentioned above, as $T \to \infty$ while p is fixed, the fifth term satisfies

$$\mathcal{L}\left(\tilde{F}_{p,T}, F_{p,0}\right) \to 0. \tag{111}$$

To establish (105), it remains to show that the third term on the right hand side of (107), $\mathcal{L}\left(F_{p,T_\gamma}, F_{p,T}\right)$, is small with high probability for sufficiently small γ, all large p, and all $T > \tilde{T}$, where $\tilde{T}$ may depend on p. Hence, the following lemma completes the proof of Theorem OW4.

Lemma 17 *For any $\delta > 0$, there exists $\gamma_\delta > 0$ such that for any $\gamma \in (0, \gamma_\delta)$, all sufficiently large p, and all $T > \tilde{T}$, where $\tilde{T}$ may depend on p, we have*

$$\Pr\left(\mathcal{L}\left(F_{p,T_\gamma}, F_{p,T}\right) < \delta/4\right) > 1 - \delta/2. \tag{112}$$

3.1.1 Proof of Lemma 17 (Lévy distance between F_{p,T_γ} and $F_{p,T}$ is small)

Below, whenever we need to say "for any $\delta > 0$, there exists $\gamma_\delta > 0$ such that for any $\gamma \in (0, \gamma_\delta)$, all sufficiently large p, and all $T > \tilde{T}$, where $\tilde{T}$ may depend on p", we will abbreviate this statement by "under conditions of Lemma 17".

Let $\xi = \sqrt{\bar{T}}\epsilon$, where $\bar{T} = T - 1$. Then the elements of ξ are i.i.d. $N(0,1)$. Some of the arguments below are more conveniently expressed in terms of ξ rather than ϵ. Consider

$$M_{p,T} = (T/p)\left(\xi\xi'/T\right)^{-1/2}\left(\xi\nabla'\xi'/T\right)\left(\xi\nabla\nabla'\xi'/T\right)^{-1}\left(\xi\nabla\xi'/T\right)\left(\xi\xi'/T\right)^{-1/2}.$$

Note that $M_{p,T}$ is identical to the real symmetric matrix $A^{-1/2}C\left(pD/T\right)^{-1}C'A^{-1/2}$, and thus, F_{p,T_γ} and $F_{p,T}$ are the empirical distributions of eigenvalues of M_{p,T_γ} and $M_{p,T}$, respectively. By Theorem A.45 (norm inequality) of Bai and Silverstein (2010),

$$\mathcal{L}\left(F_{p,T_\gamma}, F_{p,T}\right) \leq \left\|M_{p,T_\gamma} - M_{p,T}\right\|. \tag{113}$$

Hence, it is sufficient to prove that under conditions of Lemma 17,

$$\Pr\left(\left\|M_{p,T_\gamma} - M_{p,T}\right\| < \delta/4\right) > 1 - \delta/2. \tag{114}$$

Let us introduce some new notation. Let $\bar{T}_\gamma = T_\gamma - 1$, and let ∇_γ be defined similarly to ∇ with T replaced by T_γ. Consider a partition $\xi = \left[\xi_\gamma, \xi_{-\gamma}\right]$, where ξ_γ and $\xi_{-\gamma}$ are $p \times \bar{T}_\gamma$ and $p \times (\bar{T} - \bar{T}_\gamma)$, respectively. Define

$$A_\gamma = \xi_\gamma \xi_\gamma'/T_\gamma, D_\gamma = \xi_\gamma \nabla_\gamma \nabla_\gamma' \xi_\gamma'/T_\gamma,\ C_\gamma = \xi_\gamma \nabla_\gamma' \xi_\gamma'/T_\gamma,$$

and

$$A_0 = \xi\xi'/T, D_0 = \xi\nabla\nabla'\xi'/T,\ C_0 = \xi\nabla'\xi'/T.$$

Then

$$M_{p,T_\gamma} = A_\gamma^{-1/2} C_\gamma \left(pD_\gamma/T_\gamma\right)^{-1} C_\gamma' A_\gamma^{-1/2} \text{ and } M_{p,T} = A_0^{-1/2} C_0 \left(pD_0/T\right)^{-1} C_0' A_0^{-1/2}$$

By Theorem 1 of Onatski and Wang (2017a), when $p \to \infty$,

$$\left\|M_{p,T_\gamma}\right\| \overset{\text{a.s.}}{\to} \left(\sqrt{2} - \sqrt{1-\gamma}\right)^{-2}. \tag{115}$$

In particular, $\left\|M_{p,T_\gamma}\right\|$ a.s. remains bounded by an absolute constant. Convergences (115) and

$$\left\|A_\gamma^{-1/2}\right\| \overset{\text{a.s.}}{\to} (1-\sqrt{\gamma})^{-1},\ \ \left\|A_\gamma^{1/2}\right\| \overset{\text{a.s.}}{\to} (1+\sqrt{\gamma}) \tag{116}$$

(see e.g. Geman (1980) and Silverstein (1985)) imply that there exists an absolute constant K such that, for any sufficiently small γ,

$$\limsup \left\|\left(pD_\gamma/T_\gamma\right)^{-1/2} C_\gamma'\right\| < K \tag{117}$$

with probability one. In addition, for any sufficiently small γ, (116) implies

$$\limsup \left\|A_\gamma^{-1/2} - I_p\right\| \leq 3\sqrt{\gamma}. \tag{118}$$

Further, it is also true that, for any sufficiently small γ,

$$\limsup \left\|A_0^{-1/2} - I_p\right\| \leq 3\sqrt{\gamma} \tag{119}$$

with probability one. To see this, consider a $p_\gamma \times \bar{T}$ matrix η with $p_\gamma = \lfloor \gamma \bar{T} \rfloor \geq p$ (here $\lfloor \cdot \rfloor$ denotes the integer part of a real number), such that the upper $p \times \bar{T}$ block of η coincides with ξ, and the remaining part of η consists of i.i.d. $N(0,1)$ variables independent from ξ. Note that A_0 can be viewed as a $p \times p$ principal submatrix of $A_{\gamma 0} \equiv \eta\eta'/T$. By Theorem 4.3.15 of Horn and Johnson (1985),

$$\lambda_{\min}\left(A_{\gamma 0}\right) \leq \lambda_{\min}\left(A_0\right) \leq \lambda_{\max}\left(A_0\right) \leq \lambda_{\max}\left(A_{\gamma 0}\right), \tag{120}$$

where $\lambda_{\min}(S)$ and $\lambda_{\max}(S)$ denote the smallest and the largest eigenvalues of a symmetric matrix S. Since $A_{\gamma 0}$ is the sample covariance matrix with $p_\gamma = \lfloor \gamma \bar{T} \rfloor$, its largest and smallest eigenvalues a.s. converge to the same limits as those of A_γ, and thus, (120) yields (119).

Inequalities (118) and (119), and convergence (115) imply that to establish Lemma 17, it is sufficient to show that under conditions of Lemma 17,

$$\Pr\left(\left\|\tilde{M}_{p,T_\gamma} - \tilde{M}_{p,T}\right\| < \delta/4\right) > 1 - \delta/2, \tag{121}$$

where

$$\tilde{M}_{p,T_\gamma} = C_\gamma \left(pD_\gamma/T_\gamma\right)^{-1} C_\gamma',\ \text{and}\ \tilde{M}_{p,T} = C_0 \left(pD_0/T\right)^{-1} C_0'.$$

Let

$$\alpha_D = \left(pD_\gamma/T_\gamma\right)^{1/2} \left(pD_0/T\right)^{-1} \left(pD_\gamma/T_\gamma\right)^{1/2} - I_p \text{ and } \alpha_C = C_0 - C_\gamma.$$

Using the identity

$$C_0 \left(pD_0/T\right)^{-1} C_0' = \left(C_\gamma + \alpha_C\right) \left(pD_\gamma/T_\gamma\right)^{-1/2} \left(I_p + \alpha_D\right) \left(pD_\gamma/T_\gamma\right)^{-1/2} \left(C_\gamma' + \alpha_C'\right),$$

it is straightforward to verify that

$$\tilde{M}_{p,T} - \tilde{M}_{p,T_\gamma} = \left(\beta_1 + \beta_1'\right) + \beta_2 + \left(\beta_3 + \beta_3'\right) + \beta_4, \tag{122}$$

where

$$\begin{aligned}
\beta_1 &= \alpha_C \left(pD_\gamma/T_\gamma\right)^{-1} C_\gamma', \\
\beta_2 &= C_\gamma \left(pD_\gamma/T_\gamma\right)^{-1/2} \alpha_D \left(pD_\gamma/T_\gamma\right)^{-1/2} C_\gamma', \\
\beta_3 &= \alpha_C \left(pD_\gamma/T_\gamma\right)^{-1/2} \alpha_D \left(pD_\gamma/T_\gamma\right)^{-1/2} C_\gamma', \text{ and} \\
\beta_4 &= \alpha_C \left(pD_\gamma/T_\gamma\right)^{-1/2} \left(I_p + \alpha_D\right) \left(pD_\gamma/T_\gamma\right)^{-1/2} \alpha_C'.
\end{aligned}$$

The following two lemmas are proven in the next two subsections of this note.

Lemma 18 *There exists γ_0 such that for any $\gamma \in (0, \gamma_0)$, the smallest eigenvalue of pD_γ/T_γ a.s. converges as $p \to \infty$ to a number larger than 1/17.*

Lemma 19 *Under conditions of Lemma 17,*

$$\Pr\left(\|\alpha_D\| \le K\gamma \text{ and } \|\alpha_C\| \le K\sqrt{\gamma}\right) > 1 - \delta/2,$$

where K is an absolute constant.

These lemmas together with inequality (117) and the decomposition (122) guarantee that under conditions of Lemma 17,

$$\Pr\left(\left\|\tilde{M}_{p,T_\gamma} - \tilde{M}_{p,T}\right\| \le K\sqrt{\gamma}\right) > 1 - \delta/2.$$

This implies (121), which yields Lemma 17.

3.1.2 Proof of Lemma 18 (lower bound on the smallest eigenvalue of pD_γ/T_γ)

Without loss of generality, assume that T_γ is odd so that $\bar{T}_\gamma \equiv T_\gamma - 1$ is even. For any $\delta > 0$, define $D_{\gamma\delta} = \xi_\gamma \Delta_{\gamma\delta} \xi_\gamma' / T_\gamma$, where

$$\Delta_{\gamma\delta} = \left\{ \frac{1}{2\left(1 + \delta - \cos\omega_1\right)} I_2, ..., \frac{1}{2\left(1 + \delta - \cos\omega_{\bar{T}_\gamma/2}\right)} I_2 \right\}$$

with $\omega_j = 2\pi j / T_\gamma$ (cf. (37)). Note that $\|\Delta_{\gamma\delta}\| < 1/(2\delta)$. Denote the Stieltjes transform of the empirical d.f. of the eigenvalues of $D_{\gamma\delta}$ as $m_{\gamma\delta,p}(z)$. By Silverstein and Bai (1995), for any $z \in \mathbb{C}^+$, $m_{\gamma\delta,p}(z) \overset{\text{a.s.}}{\to} m_{\gamma\delta}(z)$ as $p \to \infty$, where $m \equiv m_{\gamma\delta}(z)$ satisfies

$$z = -\frac{1}{m} + \frac{1}{2\pi} \int_0^{2\pi} \frac{\mathrm{d}u}{2\left(1 + \delta - \cos u\right) + \gamma m}. \tag{123}$$

Let $F_{\gamma\delta}(\lambda)$ be the d.f. whose Stieltjes transform is $m_{\gamma\delta}(z)$, and let b be the lower boundary of its support. By Theorem 1.1 of Bai and Silverstein (1998), the smallest eigenvalue of $D_{\gamma\delta}$ a.s. converges to b. Since $D_{\gamma\delta} \le D_\gamma$, it remains to show that for any sufficiently small γ, $b > 1/(17\gamma)$.

Silverstein and Choi (1995) show that the support of $F_{\gamma\delta}(\lambda)$ can be found as follows. Find $S \subset \mathbb{R}$, such that for any $m \in S$, z is well defined by (123) as a function of m and has positive derivative at m. Then the complement of the support of $F_{\gamma\delta}(\lambda)$ coincides with $z(S)$.

Clearly, $z(m)$ is well defined for (i): $m < -\gamma^{-1}(2\delta + 4)$, (ii): $m \in \left(-\gamma^{-1} 2\delta, 0\right)$, and (iii): $m > 0$. For case (i), we have

$$\mathrm{d}z/\mathrm{d}m = m^{-2} - \frac{1}{2\pi} \int_0^{2\pi} \frac{\gamma \mathrm{d}u}{\left(2\left(1 + \delta - \cos u\right) + \gamma m\right)^2} \le m^{-2} - \gamma^{-1} m^{-2} < 0.$$

For case (ii), we have

$$\mathrm{d}^2 z/\mathrm{d}m^2 = -2m^{-3} + \frac{1}{2\pi}\int_0^{2\pi} \frac{2\gamma^2 \mathrm{d}u}{\left(2\left(1+\delta-\cos u\right)+\gamma m\right)^3} > 0$$

so that $\mathrm{d}z/\mathrm{d}m$ is strictly increasing. Furthermore, as $m \uparrow 0$, $\mathrm{d}z/\mathrm{d}m \to +\infty$ and $z \to +\infty$. Let $(\bar{m}, 0) \subseteq \left(-\gamma^{-1}2\delta, 0\right)$ be the interval where $\mathrm{d}z/\mathrm{d}m > 0$. Then, according to Silverstein and Choi's (1995) result, $(z(\bar{m}), +\infty)$ is outside the support of $F_{\gamma\delta}(\lambda)$ and hence, any point in the support of $F_{\gamma\delta}(\lambda)$ is no larger than $z(\bar{m})$.

For case (iii), we need to use an explicit form of $z(m)$. Changing the variable of integration in (123) from u to $s = e^{\mathrm{i}u}$, we obtain

$$z(m) = -\frac{1}{m} + \frac{1}{2\pi\mathrm{i}} \oint_{|s|=1} \frac{\mathrm{d}s}{s\left(2+2\delta - s - s^{-1} + \gamma m\right)}.$$

Using Cauchy's theorem,

$$z(m) = -m^{-1} + \left(\left(2+2\delta+\gamma m\right)^2 - 4\right)^{-1/2},$$

and hence

$$\mathrm{d}z/\mathrm{d}m = \frac{1}{m^2} - \frac{\gamma\left(2+2\delta+\gamma m\right)}{\left(\left(2+2\delta+\gamma m\right)^2-4\right)^{3/2}}.$$

Let $x = 2+2\delta+\gamma m$. Then, $\mathrm{d}z/\mathrm{d}m > 0$ at $m > 0$ if and only if

$$g(x) \equiv \gamma^2\left(x^2-4\right)^3 - \left(x-2-2\delta\right)^4 x^2 > 0$$

at $x > 2+2\delta$. When $\gamma = \delta = 0$, $g(x)$ has two roots at zero and four roots at two. By continuity, for small γ and δ, there are two roots in a neighborhood of zero, and four in a neighborhood of two. Now, for $\gamma < 1$, $g(x) \to -\infty$ as $x \to +\infty$. Furthermore, $g(2) < 0$ and $g(2+2\delta) > 0$. Therefore, there can be either one or three roots of $g(x)$ that satisfy $x > 2+2\delta$.

Subcase (1): There is only one root $x_1 > 2+2\delta$. We then have $\mathrm{d}z/\mathrm{d}m > 0$ for $m \in \left(0, \frac{x_1-2-2\delta}{\gamma}\right)$, and thus, according to Silverstein and Choi's (1995) result, the lower boundary of support of $F_{\gamma\delta}(\lambda)$ equals

$$b = -\frac{\gamma}{x_1-2-2\delta} + \frac{1}{\sqrt{x_1^2-4}}. \tag{124}$$

Let $\delta = o(\gamma^2)$. Then, writing $x_1 = 2 + a_1\gamma + a_2\gamma^2 + o\left(\gamma^2\right)$, and substituting this to $g(x_1) = 0$, we find that

$$x_1 = 2 + 16\gamma^2 + o\left(\gamma^2\right).$$

Using this in (124), we obtain $b = 1/\left(16\gamma\right) + o\left(\gamma^{-1}\right)$. Hence, $b > 1/(17\gamma)$ for sufficiently small γ.

Subcase (2): There are three roots $x_1 \le x_2 \le x_3$, each of which is larger than $2+2\delta$. Then, $\mathrm{d}z/\mathrm{d}m > 0$ for $m \in \left(0, \frac{x_1-2-2\delta}{\gamma}\right)$ and for $m \in \left(\frac{x_2-2-2\delta}{\gamma}, \frac{x_3-2-2\delta}{\gamma}\right)$. Note that as m goes from 0 to $\frac{x_1-2-2\delta}{\gamma}$, $z(m)$ goes from $-\infty$ to b defined by (124). Therefore, the lower boundary of the support of $F_{\gamma\delta}(\lambda)$ cannot be smaller than b. Repeating arguments used for subcase (1), we again find that $b > 1/(17\gamma)$ for sufficiently small γ.

3.1.3 Proof of Lemma 19 (α_D and α_C are small)

First, let us focus on α_D. Define

$$\alpha \equiv \left(I_p + \alpha_D\right)^{-1} - I_p = \left(pD_\gamma/T_\gamma\right)^{-1/2}\left(pD_0/T - pD_\gamma/T_\gamma\right)\left(pD_\gamma/T_\gamma\right)^{-1/2}.$$

We would like to show that under conditions of Lemma 17,

$$\Pr\left(\|\alpha\| \le K\gamma\right) > 1 - \delta/4. \tag{125}$$

Consider partition $\nabla\nabla' = \operatorname{diag}\{\Delta_\gamma, \Delta_{-\gamma}\}$, where Δ_γ is $\bar{T}_\gamma \times \bar{T}_\gamma$. Then,

$$\alpha = (pD_\gamma/T_\gamma)^{-1/2} \left(p\xi_\gamma \Delta_\gamma \xi_\gamma'/T^2 - pD_\gamma/T_\gamma + p\xi_{-\gamma}\Delta_{-\gamma}\xi_{-\gamma}'/T^2\right) (pD_\gamma/T_\gamma)^{-1/2}. \tag{126}$$

By Lemma 18, almost surely, as $p \to \infty$,

$$\limsup \left\| (pD_\gamma/T_\gamma)^{-1/2} \right\| < \sqrt{17} < 5. \tag{127}$$

Further,

$$p\xi_\gamma \Delta_\gamma \xi_\gamma'/T^2 - pD_\gamma/T_\gamma = p\xi_\gamma \left(\Delta_\gamma/T^2 - \nabla_\gamma \nabla_\gamma'/T_\gamma^2\right) \xi_\gamma'.$$

Recall that the diagonal elements of Δ_γ have form $\frac{1}{2}(1-\cos 2\pi j/T)^{-1}$ with $j \le \bar{T}_\gamma/2$. The diagonal elements of $\nabla_\gamma \nabla_\gamma'$ have a similar form with T replaced by T_γ. Since

$$\cos x = 1 - \frac{1}{2}x^2 + \frac{1}{4!}x^4 \cos t$$

for some $t \in [0, x]$, we have

$$\frac{1}{2T^2}(1-\cos 2\pi j/T)^{-1} = \frac{1}{(2\pi j)^2}\left(1 - \frac{\cos t}{12}\frac{(2\pi j)^2}{T^2}\right)^{-1}$$

for some $t \in [0, \pi]$, and hence

$$\frac{1}{2T^2}(1-\cos 2\pi j/T)^{-1} - \frac{1}{(2\pi j)^2} = \frac{\cos t}{12T^2}\left(1 - \frac{\cos t}{12}\frac{(2\pi j)^2}{T^2}\right)^{-1}.$$

Since $j \le \bar{T}_\gamma/2$ and $T > T_\gamma$, we have

$$1 - \frac{\cos t}{12}\frac{(2\pi j)^2}{T^2} > 1 - \pi^2/12 > 1/12,$$

and thus

$$\left| \frac{1}{2T^2}(1-\cos 2\pi j/T)^{-1} - \frac{1}{(2\pi j)^2} \right| < 1/T^2. \tag{128}$$

A similar inequality holds for the elements of $\nabla_\gamma \nabla_\gamma'$:

$$\left| \frac{1}{2T_\gamma^2}(1-\cos 2\pi j/T_\gamma)^{-1} - \frac{1}{(2\pi j)^2} \right| < 1/T_\gamma^2. \tag{129}$$

Therefore,

$$\left| \frac{1}{2T^2}(1-\cos 2\pi j/T)^{-1} - \frac{1}{2T_\gamma^2}(1-\cos 2\pi j/T_\gamma)^{-1} \right| < 2/T_\gamma^2,$$

and hence,

$$\left\| p\xi_\gamma \Delta_\gamma \xi_\gamma'/T^2 - pD_\gamma/T_\gamma \right\| < \left\| 2\,(p/T_\gamma)\,\xi_\gamma \xi_\gamma'/T_\gamma \right\| < 4\gamma \tag{130}$$

with high probability for any sufficiently small γ, sufficiently large p, and all $T > T_\gamma$. To obtain the last inequality in (130), we used the fact that the largest eigenvalue of $\xi_\gamma \xi_\gamma'/T_\gamma$ a.s. converges to $\left(1+\sqrt{\gamma}\right)^2$.

Consider now the component $p\xi_{-\gamma}\Delta_{-\gamma}\xi_{-\gamma}'/T^2$ of (126). Since $1 - \cos x \ge x^2/6$ for $x \in [0, \pi]$, we have

$$2T^2(1-\cos 2\pi j/T) > (2\pi j)^2/3. \tag{131}$$

Let us represent $\Delta_{-\gamma}$ as $\operatorname{diag}\{\Delta_{-\gamma,1}, ..., \Delta_{-\gamma,k}\}$, where

$$k = \left(\bar{T} - \bar{T}_\gamma\right)/\bar{T}_\gamma$$

and each block $\Delta_{-\gamma,i}$ is $\bar{T}_\gamma$-dimensional. We can choose T so that k is an integer, so such a representation is possible. Using the fact that the diagonal elements of $\Delta_{-\gamma,i}/T^2$ have form

$$\left(2T^2\left(1-\cos 2\pi j/T\right)\right)^{-1} \text{ with } j = i\bar{T}_\gamma/2+1, ..., (i+1)\,\bar{T}_\gamma/2,$$

we find that the upper bound on the diagonal elements of $\Delta_{-\gamma,i}/T^2$ equals $\left(2T^2\left(1-\cos i\bar{T}_\gamma\pi/T\right)\right)^{-1}$. By (131), this is no larger than $3/\left(i\pi\bar{T}_\gamma\right)^2$.

Let us partition $\xi_{-\gamma}$ conformably with $\Delta_{-\gamma}$ so that $\xi_{-\gamma} = \left[\xi_{-\gamma,1}, ..., \xi_{-\gamma,k}\right]$. Then, from the above, we have

$$\left\|p\xi_{-\gamma}\Delta_{-\gamma}\xi_{-\gamma}'/T^2\right\| \le 3p/\left(\pi^2\bar{T}_\gamma\right)\sum_{i=1}^{k} i^{-2}\left\|\xi_{-\gamma,i}\xi_{-\gamma,i}'/\bar{T}_\gamma\right\|.$$

The Gaussian concentration inequality for the singular values of a rectangular matrix with i.i.d. Gaussian entries (see Theorem II.13 of Davidson and Szarek (2001)) implies that, for any $t>0$,

$$\Pr\left(\left\|\xi_{-\gamma,i}\xi_{-\gamma,i}'/\bar{T}_\gamma\right\| \ge \left(1+\sqrt{p/\bar{T}_\gamma}+t\right)^2\right) < \exp\left\{-\bar{T}_\gamma t^2/2\right\}.$$

Take $t = i^{1/4}$. Then,

$$\sum_{i=1}^{k}\Pr\left(\left\|\xi_{-\gamma,i}\xi_{-\gamma,i}'/\bar{T}_\gamma\right\| \ge \left(1+\sqrt{p/\bar{T}_\gamma}+i^{1/4}\right)^2\right) < \sum_{i=1}^{\infty}\exp\left\{-\bar{T}_\gamma i^{1/2}/2\right\}.$$

Clearly, the right hand side of the above inequality can be made arbitrarily small by choosing sufficiently large T_γ. Therefore, with large probability, for sufficiently large T_γ, all $\left\|\xi_{-\gamma,i}\xi_{-\gamma,i}'/\bar{T}_\gamma\right\|$ are smaller than $\left(1+\sqrt{p/\bar{T}_\gamma}+i^{1/4}\right)^2$ and

$$\left\|p\xi_{-\gamma}\Delta_{-\gamma}\xi_{-\gamma}'/T^2\right\| \le 3p/\left(\pi^2\bar{T}_\gamma\right)\sum_{i=1}^{k} i^{-2}\left(1+\sqrt{p/\bar{T}_\gamma}+i^{1/4}\right)^2 \le K\gamma \tag{132}$$

for some constant K that does not depend on γ. Using the definition of α, (130), (132), and (127), we conclude that inequality (125) does take place under conditions of Lemma 17.

Let us now consider α_C. Partition ∇ as $\operatorname{diag}\{\nabla_{1\gamma}, \nabla_{-\gamma}\}$, where $\nabla_{1\gamma}$ is $\bar{T}_\gamma\times\bar{T}_\gamma$.Write α_C in the following form

$$\alpha_C = \left(-\xi\xi'/(2T)+\xi_\gamma\left(\nabla_{1\gamma}'+I_{\bar{T}_\gamma}/2\right)\xi_\gamma'/T - C_\gamma\right)+\xi_{-\gamma}\left(\nabla_{-\gamma}'+I_{\bar{T}-\bar{T}_\gamma}/2\right)\xi_{-\gamma}'/T.$$

Let us denote $\nabla_{1\gamma}'+I_{\bar{T}_\gamma}/2$ as $\hat{\nabla}_{1\gamma}'$ and $\nabla_\gamma'+I_{\bar{T}_\gamma}/2$ as $\hat{\nabla}_\gamma'$. Then

$$-\xi\xi'/(2T)+\xi_\gamma\left(\nabla_\gamma'+I_{\bar{T}_\gamma}/2\right)\xi_\gamma'/T - C_\gamma = -\xi\xi'/(2T)+\xi_\gamma\left(\hat{\nabla}_{1\gamma}'/T-\hat{\nabla}_\gamma'/T_\gamma\right)\xi_\gamma'+\xi_\gamma\xi_\gamma'/(2T_\gamma).$$

By definition, the block-diagonal elements of $\hat{\nabla}_{1\gamma}'$ have form (cf. (36))

$$\begin{pmatrix} 0 & -\frac{1}{2}\frac{\sin 2\pi j/T}{1-\cos 2\pi j/T} \\ \frac{1}{2}\frac{\sin 2\pi j/T}{1-\cos 2\pi j/T} & 0 \end{pmatrix}$$

The block-diagonal elements of $\hat{\nabla}_\gamma'$ have a similar form with T replaced by T_γ. Now,

$$\sin x = x - \frac{\cos t_1}{3!}x^3 \text{ and } \cos x = 1-\frac{1}{2}x^2+\frac{\cos t_2}{4!}x^4$$

for some $t_1, t_2 \in [0, x]$. Therefore, we have

$$\frac{1}{2}\frac{\sin 2\pi j/T}{1-\cos 2\pi j/T} = \frac{2\pi j/T - \frac{\cos t_1}{6}\left(2\pi j/T\right)^3}{\left(2\pi j/T\right)^2-\frac{\cos t_2}{12}\left(2\pi j/T\right)^4} = \frac{1}{2\pi j/T}\frac{1-\frac{\cos t_1}{6}\left(2\pi j/T\right)^2}{1-\frac{\cos t_2}{12}\left(2\pi j/T\right)^2},$$

so that

$$\frac{1}{2T}\frac{\sin 2\pi j/T}{1-\cos 2\pi j/T}-\frac{1}{2\pi j}=\frac{(2\pi j/T)^2}{2\pi j}\frac{\frac{\cos t_2}{12}-\frac{\cos t_1}{6}}{1-\frac{\cos t_2}{12}(2\pi j/T)^2}$$

and thus,

$$\left|\frac{1}{2T}\frac{\sin 2\pi j/T}{1-\cos 2\pi j/T}-\frac{1}{2\pi j}\right|<6\pi j/T^2. \tag{133}$$

Similarly,

$$\left|\frac{1}{2T_\gamma}\frac{\sin 2\pi j/T_\gamma}{1-\cos 2\pi j/T_\gamma}-\frac{1}{2\pi j}\right|<6\pi j/T_\gamma^2. \tag{134}$$

Let $\xi_{\gamma 1}$ and $\xi_{\gamma 2}$ be $p\times\bar{T}_\gamma/2$ matrices that consists of the odd and even columns of ξ_γ, respectively. Then, the latter two inequalities and the fact that $j\leq\bar{T}_\gamma/2$ imply that

$$\left\|\xi_\gamma\left(\hat{\nabla}'_{1\gamma}/T-\hat{\nabla}'_\gamma/T_\gamma\right)\xi'_\gamma\right\|\leq 2\left\|\xi_{\gamma 1}\Gamma\xi'_{\gamma 2}\right\|, \tag{135}$$

where Γ is a diagonal matrix with diagonal elements smaller than $3\pi/T_\gamma$ by absolute value. On the other hand, $\left\|\xi_{\gamma 1}\Gamma\xi'_{\gamma 2}\right\|$ is the square root of the largest eigenvalue of $\xi_{\gamma 1}\Gamma\xi'_{\gamma 2}\xi_{\gamma 2}\Gamma\xi'_{\gamma 1}$.

Note that the rank of $T_\gamma\Gamma\xi'_{\gamma 2}\xi_{\gamma 2}\Gamma$ is no larger than p, and there exists an orthogonal transformation R such that $RT_\gamma\Gamma\xi'_{\gamma 2}\xi_{\gamma 2}\Gamma R'$ is diagonal with only the first p diagonal elements potentially non-zero. Furthermore, these non-zero diagonal elements coincide with the eigenvalues of $T_\gamma\xi_{\gamma 2}\Gamma^2\xi'_{\gamma 2}$. But

$$T_\gamma\xi_{\gamma 2}\Gamma^2\xi'_{\gamma 2}\leq\frac{(3\pi)^2}{2}\frac{\xi_{\gamma 2}\xi'_{\gamma 2}}{T_\gamma/2}.$$

With high probability, for sufficiently any small γ and large p,

$$\left\|\xi_{\gamma 2}\xi'_{\gamma 2}/\left(T_\gamma/2\right)\right\|<2.$$

Hence, the only p potentially non-zero diagonal elements of $RT_\gamma\Gamma\xi'_{\gamma 2}\xi_{\gamma 2}\Gamma R'$ are smaller than $(3\pi)^2$.

Let $\xi_{\gamma 11}$ be the $p\times p$ matrix that consists of the first p columns of $\xi_{\gamma 1}R'$. Note that the entries of $\xi_{\gamma 11}$ are i.i.d. standard normals. Then, we have

$$\xi_{\gamma 1}\Gamma\xi'_{\gamma 2}\xi_{\gamma 2}\Gamma\xi'_{\gamma 1}\leq(3\pi)^2\,\xi_{\gamma 11}\xi'_{\gamma 11}/T_\gamma.$$

Since the norm of $\xi_{\gamma 11}\xi'_{\gamma 11}/p$ is smaller than 5 with high probability for sufficiently large p, it must be that

$$\left\|\xi_{\gamma 1}\Gamma\xi'_{\gamma 2}\xi_{\gamma 2}\Gamma\xi'_{\gamma 1}\right\|\leq(9\pi)^2\,\gamma$$

with high probability for any sufficiently small γ, large p, and all $T>\tilde{T}$, where $\tilde{T}$ may depend on p. Combining this with (135), we obtain

$$\left\|\xi_\gamma\left(\hat{\nabla}'_{1\gamma}/T-\hat{\nabla}'_\gamma/T_\gamma\right)\xi'_\gamma\right\|\leq 18\pi\sqrt{\gamma}. \tag{136}$$

Further,

$$\left\|-\xi\xi'/\left(2T\right)+\xi_\gamma\xi'_\gamma/\left(2T_\gamma\right)\right\|\leq\frac{1}{2}\left\|I_p-\xi\xi'/T\right\|+\frac{1}{2}\left\|I_p-\xi_\gamma\xi'_\gamma/T_\gamma\right\|\leq 4\sqrt{\gamma}$$

with high probability, for sufficiently small γ, large p, and $T>T_\gamma$. Combining this with (136), we obtain

$$\left\|-\xi\xi'/\left(2T\right)+\xi_\gamma\left(\nabla'_{1\gamma}+I_{\bar{T}_\gamma}/2\right)\xi'_\gamma/T-C_\gamma\right\|\leq 20\pi\sqrt{\gamma}. \tag{137}$$

Next, consider $\xi_{-\gamma}\left(\nabla'_{-\gamma}+I_{\bar{T}-\bar{T}_\gamma}/2\right)\xi'_{-\gamma}/T$ part of α_C. Let $\xi_{-\gamma,1}$ and $\xi_{-\gamma,2}$ be $p\times\left(\bar{T}-\bar{T}_\gamma\right)/2$ matrices that consist of odd and even columns of $\xi_{-\gamma}$. Then,

$$\left\|\xi_{-\gamma}\left(\nabla'_{-\gamma}+I_{\bar{T}-\bar{T}_\gamma}/2\right)\xi'_{-\gamma}/T\right\|^2\leq 4\left\|\xi_{-\gamma,2}\Upsilon\xi'_{-\gamma,1}\right\|^2=4\left\|\xi_{-\gamma,2}\Upsilon\xi'_{-\gamma,1}\xi_{-\gamma,1}\Upsilon\xi'_{-\gamma,2}\right\|,$$

where

$$\Upsilon=\operatorname{diag}\left\{\frac{1}{2T}\frac{\sin 2\pi j/T}{1-\cos 2\pi j/T}\right\}_{j=\bar{T}_{\gamma}/2+1}^{\bar{T}/2}.$$

Let R be the orthogonal matrix such that $R\Upsilon\xi_{-\gamma,1}'\xi_{-\gamma,1}\Upsilon R'$ is diagonal. Note that the rank of $\Upsilon\xi_{-\gamma,1}'\xi_{-\gamma,1}\Upsilon$ is no larger than p. Therefore, there are only p potentially non-zero elements on the diagonal of $R\Upsilon\xi_{-\gamma,1}'\xi_{-\gamma,1}\Upsilon R'$. Without loss of generality, these are the first p elements. Let $\xi_{-\gamma,21}$ be the first p columns of $\xi_{-\gamma,2}R'$. Then, we have

$$\begin{aligned}\left\|\xi_{-\gamma}\left(\nabla_{-\gamma}'+I_{\bar{T}-\bar{T}_{\gamma}}/2\right)\xi_{-\gamma}'/T\right\|^{2} &\leq 4\left\|\xi_{-\gamma,21}\xi_{-\gamma,21}'\right\|\left\|R\Upsilon\xi_{-\gamma,1}'\xi_{-\gamma,1}\Upsilon R'\right\| \\ &= 4\left\|\xi_{-\gamma,21}\xi_{-\gamma,21}'\right\|\left\|\xi_{-\gamma,1}\Upsilon^{2}\xi_{-\gamma,1}'\right\|.\end{aligned} \tag{138}$$

Consider the partition $\xi_{-\gamma,1}=\left[\xi_{-\gamma,11},...,\xi_{-\gamma,1k}\right]$, and note that $\Upsilon^{2}=\operatorname{diag}\left\{\Upsilon_{1}^{2},...,\Upsilon_{k}^{2}\right\}$ with

$$\Upsilon_{i}=\operatorname{diag}\left\{\frac{1}{2T}\frac{\sin\left(\left(i\frac{\bar{T}_{\gamma}}{2}+1\right)\frac{2\pi}{T}\right)}{1-\cos\left(\left(i\frac{\bar{T}_{\gamma}}{2}+1\right)\frac{2\pi}{T}\right)},...,\frac{1}{2T}\frac{\sin\left((i+1)\frac{\bar{T}_{\gamma}}{2}\frac{2\pi}{T}\right)}{1-\cos\left((i+1)\frac{\bar{T}_{\gamma}}{2}\frac{2\pi}{T}\right)}\right\}.$$

We have

$$\left(\frac{1}{2T}\frac{\sin\left(\left(i\frac{\bar{T}_{\gamma}}{2}+k\right)\frac{2\pi}{T}\right)}{1-\cos\left(\left(i\frac{\bar{T}_{\gamma}}{2}+k\right)\frac{2\pi}{T}\right)}\right)^{2}=\frac{1}{2T^{2}}\frac{\cos^{2}\left(i\frac{\bar{T}_{\gamma}}{2}+k\right)\frac{\pi}{T}}{1-\cos\left(i\frac{\bar{T}_{\gamma}}{2}+k\right)\frac{2\pi}{T}}\leq\frac{1}{2T^{2}}\frac{1}{1-\cos\frac{i\bar{T}_{\gamma}\pi}{T}}<\frac{3}{\bar{T}_{\gamma}^{2}\pi^{2}}\frac{1}{i^{2}}.$$

Therefore,

$$\left\|\xi_{-\gamma,1}\Upsilon^{2}\xi_{-\gamma,1}'\right\|\leq\sum_{i=1}^{k}\left\|\xi_{-\gamma,1i}\Upsilon_{i}^{2}\xi_{-\gamma,1i}'\right\|\leq\sum_{i=1}^{k}\frac{3}{2\bar{T}_{\gamma}\pi^{2}}\frac{1}{i^{2}}\left\|\frac{\xi_{-\gamma,1i}\xi_{-\gamma,1i}'}{\bar{T}_{\gamma}/2}\right\|.$$

Using the large deviation inequality argument as above, we conclude that with high probability,

$$\left\|\xi_{-\gamma,1}\Upsilon^{2}\xi_{-\gamma,1}'\right\|\leq K/\bar{T}_{\gamma},$$

where K is an absolute constant. This and (138) yield

$$\left\|\xi_{-\gamma}\left(\nabla_{-\gamma}'+I_{\bar{T}-\bar{T}_{\gamma}}/2\right)\xi_{-\gamma}'/T\right\|^{2}\leq 4K\left(p/\bar{T}_{\gamma}\right)\left\|\xi_{-\gamma,21}\xi_{-\gamma,21}'/p\right\|\leq K_{1}\gamma,$$

where K_1 is an absolute constant. Hence, with high probability, for any sufficiently small γ, large p, and all $T>\tilde{T}$, where $\tilde{T}$ may depend on p,

$$\left\|\xi_{-\gamma}\left(\nabla_{-\gamma}'+I_{\bar{T}-\bar{T}_{\gamma}}/2\right)\xi_{-\gamma}'/T\right\|\leq K\sqrt{\gamma}$$

for some absolute constant K. Combining this with (137), we conclude that under conditions of Lemma 17,

$$\Pr\left(\|\alpha_{C}\|\leq K\sqrt{\gamma}\right)>1-\delta/4$$

for some absolute constant K.

3.2 Proof of Theorem OW5

The plan of our proof is as follows. First, we show that $\int\lambda \mathrm{d}F_{0}(\lambda)=2$, as stated by the theorem. Next, we prove that, for any $\delta>0$,

$$\Pr\left(\left|\int\left(\lambda-\min\left(a_{+}+\delta,\lambda\right)\right)\mathrm{d}F_{p,0}(\lambda)\right|>\delta\right)\to 0 \tag{139}$$

as $p \to \infty$. For this, we establish the convergence $\lambda_{2,0} \xrightarrow{\mathrm{P}} a_+$, and show that $\lambda_{1,0} = o_{\mathrm{P}}(p)$ as $p \to \infty$. Since

$$\int (\lambda - \min(a_+ + \delta, \lambda))\, \mathrm{d}F_{p,0}(\lambda) = \frac{1}{p}\sum_{j=1}^{p} (\lambda_{j,0} - \min(a_+ + \delta, \lambda_{j,0})),$$

such an asymptotic behavior of $\lambda_{2,0}$ and $\lambda_{1,0}$ implies (139). Finally, by Theorem OW4,

$$\Pr\left(\left|\int \min(a_+ + \delta, \lambda)\, \mathrm{d}F_{p,0}(\lambda) - \int \lambda \mathrm{d}F_0(\lambda)\right| > \delta\right) \to 0.$$

This convergence and (139) yield Theorem OW5.

Throughout the proof, we interpret matrix

$$\frac{1}{p}\int_0^1 (\mathrm{d}B)\, F' \left(\int_0^1 FF'\mathrm{d}u\right)^{-1} \int_0^1 F\, (\mathrm{d}B)' \tag{140}$$

as the probability limit of $\frac{T}{p}\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{01}'\tilde{S}_{00}^{-1}$ when $T \to \infty$ (see Johansen (1991)), where

$$\tilde{S}_{00} = \frac{1}{T}\varepsilon M_l \varepsilon', \tilde{S}_{01} = \frac{1}{T}\varepsilon M_l U'\varepsilon', \text{ and } \tilde{S}_{11} = \frac{1}{T}\varepsilon U M_l U'\varepsilon'. \tag{141}$$

The same argument as in the proof of Theorem OW4 allows us to assume, without loss of generality, that ε is Gaussian.

3.2.1 Showing that $\int \lambda \mathrm{d}F_0(\lambda) = 2$

By OW(20),

$$\int \lambda \mathrm{d}F_0(\lambda) = \int_{a_-}^{a_+} \frac{\lambda}{2\pi} \frac{\sqrt{(a_+ - \lambda)(\lambda - a_-)}}{\lambda}\mathrm{d}\lambda,$$

where $a_\pm = \left(1 \pm \sqrt{2}\right)^2$. Let $x = (\lambda - a_-)/(a_+ - a_-)$ so that $\lambda = a_- + (a_+ - a_-)x$. Then

$$\int \lambda \mathrm{d}F_0(\lambda) = \frac{(a_+ - a_-)^2}{2\pi}\int_0^1 \sqrt{(1-x)x}\mathrm{d}x = \frac{(a_+ - a_-)^2}{2\pi}\frac{\pi}{8} = 2.$$

3.2.2 Convergence of the second largest eigenvalue

In this subsection, we would like to show that $\lambda_{2,0} \xrightarrow{\mathrm{P}} a_+$ as $p \to \infty$. Suppose this is not true. Then, for some $\delta > 0$ and any $p_0 > 0$, there exists $p > p_0$ such that

$$\Pr(|\lambda_{2,0} - a_+| > \delta) > 2\delta.$$

Denote the eigenvalues of $\frac{T}{p}\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{01}'\tilde{S}_{00}^{-1}$ as $\tilde{\lambda}_{1,pT} \geq \ldots \geq \tilde{\lambda}_{p,pT}$. Since the latter matrix converges in probability to (140) as $T \to \infty$, it must be true that for any p and all sufficiently large T,

$$\Pr\left(\left|\lambda_{2,0} - \tilde{\lambda}_{2,pT}\right| > \delta/2\right) < \delta.$$

Hence, to prove the convergence $\lambda_{2,0} \xrightarrow{\mathrm{P}} a_+$, it is sufficient to show that for any $p_0 > 0$, there exists $p > p_0$ such that for all sufficiently large T

$$\Pr\left(\left|\tilde{\lambda}_{2,pT} - a_+\right| > \delta/2\right) < \delta. \tag{142}$$

We can interpret $p\tilde{\lambda}_{j,pT}/T$ as the j-th largest eigenvalue of the product of projections $P_2\tilde{P}_1P_2$, where

$$\tilde{P}_1 = M_l U'\varepsilon' \left(\varepsilon U M_l U'\varepsilon'\right)^{-1} \varepsilon U M_l, \text{ and } P_2 = M_l\varepsilon' \left(\varepsilon M_l \varepsilon'\right)^{-1} \varepsilon M_l.$$

Let

$$P_1 = M_l U' M_l \varepsilon' \left(\varepsilon M_l U M_l U' M_l \varepsilon'\right)^{-1} \varepsilon M_l U M_l.$$

Then, by Lemma 5, there exist projections $P_{\tilde{y}}$ and P_y on one-dimensional subspaces of $\mathbb{R}^T$, such that

$$P_2 \tilde{P}_1 P_2 - P_2 P_1 P_2 = P_2 P_{\tilde{y}} P_2 - P_2 P_y P_2.$$

Since $P_2 P_{\tilde{y}} P_2$ and $P_2 P_y P_2$ are positive semi-definite, by interlacing inequalities (see Theorem 4.3.4 in Horn and Johnson (1985)),

$$\lambda_{3,pT} \leq \tilde{\lambda}_{2,pT} \leq \lambda_{1,pT}, \tag{143}$$

where $p\lambda_{j,pT}/T$ is the j-th largest eigenvalue of $P_2 P_1 P_2$.

Note that $\lambda_{j,pT}$ equal the eigenvalues of $\frac{T}{p} S_{01} S_{11}^{-1} S_{01}' S_{00}^{-1}$, where S_{ij} are as defined in (1), that is

$$S_{00} = \frac{1}{T} \varepsilon M_l \varepsilon', \;\; S_{01} = \frac{1}{T} \varepsilon M_l U' M_l \varepsilon', \text{ and } S_{11} = \frac{1}{T} \varepsilon M_l U M_l U' M_l \varepsilon'.$$

By Lemma 8,

$$\frac{T}{p} S_{01} S_{11}^{-1} S_{01}' S_{00}^{-1} = \frac{T}{p} C D^{-1} C' A^{-1},$$

where

$$C = \epsilon \nabla' \epsilon', D = \epsilon \nabla \nabla' \epsilon', \;\; A = \epsilon \epsilon',$$

and ϵ is a $p \times \bar{T}$ matrix ($\bar{T} \equiv T - 1$) with i.i.d. $N\left(0, I_p/\bar{T}\right)$ entries.

For any p, let T_γ be the smallest integer satisfying $p/T_\gamma \leq \gamma$. Onatski and Wang's (2017a) Theorem 1 and Theorem OW1 imply that $\lambda_{1,pT_\gamma} \overset{\text{a.s.}}{\to} b_+$ and $\lambda_{3,pT_\gamma} \overset{\text{a.s.}}{\to} b_+$ as $p \to \infty$, where

$$b_+ = \left(\sqrt{2} - \sqrt{1-\gamma}\right)^{-2}.$$

As $\gamma \to 0$, we have $|b_+ - a_+| \to 0$. Therefore, for sufficiently small γ and large p,

$$\Pr\left(\left|\lambda_{1,pT_\gamma} - a_+\right| \vee \left|\lambda_{3,pT_\gamma} - a_+\right| > \delta/4\right) < \delta/2.$$

This inequality, together with (143) and inequality (114) imply (142).

3.2.3 Asymptotic behavior of the largest eigenvalue

In this subsection, we would like to show that $\lambda_{1,0} = o_{\mathrm{P}}(p)$ as $p \to \infty$. Since (140) is the probability limit of $\frac{T}{p} \tilde{S}_{01} \tilde{S}_{11}^{-1} \tilde{S}_{01}' \tilde{S}_{00}^{-1}$ as $T \to \infty$, it is sufficient to show that for any $\delta > 0$, all sufficiently large p, and all $T > \tilde{T}$, where $\tilde{T}$ may depend on p,

$$\Pr\left(\tilde{\lambda}_{1,pT} > \delta p\right) < \delta. \tag{144}$$

We start from reinterpreting the eigenvalue $\tilde{\lambda}_{1,pT}$ as the largest eigenvalue of matrix $\frac{T}{p} \tilde{A}^{-1/2} \tilde{C} \tilde{D}^{-1} \tilde{C}' \tilde{A}^{-1/2}$, defined below. Recall definitions (9), (34), and (35) of the discrete Fourier transform matrix $\mathcal{F}$, the permutation matrix P, and the unitary matrix W. Let $P_+ = \operatorname{diag}\{1, P\}$, $W_+ = \operatorname{diag}\{1, W\}$, and let a $p \times 1$ vector η_0 and a $p \times \bar{T}$ matrix η_{-0} (with $\bar{T} = T - 1$) be defined as

$$\left[\eta_0, \eta_{-0}\right] \equiv \varepsilon \mathcal{F}^* P_+ W_+^* / \sqrt{T}.$$

Since $\mathcal{F}^* P_+ W_+^* / \sqrt{T}$ is an orthogonal matrix, the entries of $\left[\eta_0, \eta_{-0}\right]$ are i.i.d. $N(0,1)$.

Using (10) and (141), we obtain after some algebra

$$\tilde{S}_{00} = \eta_{-0} \eta_{-0}' / T,$$

$$\tilde{S}_{01} = \varepsilon M_l U' M_l \varepsilon' / T + \varepsilon M_l U' l l' \varepsilon' / T^2 = \eta_{-0} \nabla' \eta_{-0}' / T + \eta_{-0} y \eta_0' / T,$$

where $y' = \left(y_1', ..., y_{\bar{T}/2}'\right)$ with

$$y_j' = -\frac{1}{\sqrt{2}}\left(1, \frac{\sin\omega_j}{1-\cos\omega_j}\right),$$

and

$$\tilde{S}_{11} = \eta_{-0}\nabla\nabla'\eta_{-0}'/T + \eta_{-0}x\eta_0'/T + \eta_0 x'\eta_{-0}'/T + \left(T^2-1\right)\eta_0\eta_0'/(12T),$$

where $x = \nabla y$, so that $x' = \left(x_1', ..., x_{\bar{T}/2}'\right)$ with

$$x_j' = -\frac{1}{\sqrt{2}}\left(\frac{\cos\omega_j}{1-\cos\omega_j}, \frac{-\sin\omega_j}{1-\cos\omega_j}\right).$$

Next, let R_0 be a random orthogonal $p \times p$ matrix, independent from η_{-0}, and such that $R_0\eta_0$ equals $\|\eta_0\| e_{1p}$, where e_{1p} is the first column of I_p. Denote $\|\eta_0\|$ as χ_p, and $R_0\eta_{-0}$ as ξ, and let

$$\begin{aligned}
\tilde{A} &= R_0\tilde{S}_{00}R_0' = \xi\xi'/T, \\
\tilde{C} &= R_0\tilde{S}_{01}R_0' = \xi\nabla'\xi'/T + \chi_p\xi y e_{1p}'/T, \text{ and} \\
\tilde{D} &= R_0\tilde{S}_{11}R_0' = \xi\nabla\nabla'\xi'/T + \chi_p\xi x e_{1p}'/T + \chi_p e_{1p}x'\xi'/T + \chi_p^2\left(T^2-1\right)e_{1p}{e_{1p}}'/(12T).
\end{aligned}$$

Since R_0 is an invertible matrix, the eigenvalues of $\frac{T}{p}\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{01}'\tilde{S}_{00}^{-1}$ and $\frac{T}{p}\tilde{A}^{-1/2}\tilde{C}\tilde{D}^{-1}\tilde{C}'\tilde{A}^{-1/2}$ coincide, and in particular, $\tilde{\lambda}_{1,pT}$ is the largest eigenvalue of $\frac{T}{p}\tilde{A}^{-1/2}\tilde{C}\tilde{D}^{-1}\tilde{C}'\tilde{A}^{-1/2}$ as claimed above.

Now recall some notation from Section 3.1.1. Specifically, recall that T_γ denotes the smallest integer satisfying $p/T_\gamma \le \gamma$, $\bar{T}_\gamma = T_\gamma - 1$, and ∇_γ denotes a $\bar{T}_\gamma \times \bar{T}_\gamma$ matrix defined similarly to ∇ with T replaced by T_γ. Further,

$$A_\gamma = \xi_\gamma\xi_\gamma'/T_\gamma,\ C_\gamma = \xi_\gamma\nabla_\gamma'\xi_\gamma'/T_\gamma,\ D_\gamma = \xi_\gamma\nabla_\gamma\nabla_\gamma'\xi_\gamma'/T_\gamma,$$

where ξ_γ is a $p \times \bar{T}_\gamma$ matrix from the partition $\xi = \left[\xi_\gamma, \xi_{-\gamma}\right]$.

Let us define

$$\begin{aligned}
\tilde{A}_\gamma &= A_\gamma,\ \tilde{C}_\gamma = C_\gamma + \chi_p\xi_\gamma y_\gamma e_{1p}'/T_\gamma, \text{ and} \\
\tilde{D}_\gamma &= D_\gamma + \chi_p\xi_\gamma x_\gamma e_{1p}'/T_\gamma + \chi_p e_{1p}x_\gamma'\xi_\gamma'/T_\gamma + \chi_p^2\left(T_\gamma^2-1\right)e_{1p}e_{1p}'/(12T_\gamma),
\end{aligned}$$

where y_γ and x_γ defined as y and x after T is replaced by T_γ, and let $\tilde{\lambda}_{1,pT_\gamma}$ be the largest eigenvalue of $\frac{T_\gamma}{p}\tilde{A}_\gamma^{-1/2}\tilde{C}_\gamma\tilde{D}_\gamma^{-1}\tilde{C}_\gamma'\tilde{A}_\gamma^{-1/2}$. The following two lemmas are established in the next two sections of this note.

Lemma 20 *Under conditions of Lemma 17,*

$$\Pr\left(\tilde{\lambda}_{1,pT} > K\gamma p\left(1+\tilde{\lambda}_{1,pT_\gamma}\right)\right) < \delta,$$

where K is an absolute constant.

Lemma 21 *For any $\delta > 0$, there exists $\gamma_\delta > 0$ such that for any $\gamma \in (0, \gamma_\delta)$ and all sufficiently large p,*

$$\Pr\left(\tilde{\lambda}_{1,pT_\gamma} > K\gamma^{-3/4}\right) < \delta,$$

where K is an absolute constant.

Choosing γ sufficiently small and using Lemmas 20 and 21, we obtain (144).

3.2.4 Proof of Lemma 20 (lower bound on $\tilde{\lambda}_{1,pT}$ in terms of $\tilde{\lambda}_{1,pT_\gamma}$)

Let

$$\tilde{\alpha}_C = \tilde{C} - \tilde{C}_\gamma \text{ and } \tilde{\alpha}_D = \left(p\tilde{D}_\gamma/T_\gamma\right)^{1/2}\left(p\tilde{D}/T\right)^{-1}\left(p\tilde{D}_\gamma/T_\gamma\right)^{1/2} - I_p.$$

Then

$$\begin{aligned}\tilde{\lambda}_{1,pT} &= \left\|\tilde{A}^{-1/2}\left(\tilde{C}_\gamma + \tilde{\alpha}_C\right)\left(p\tilde{D}/T\right)^{-1}\left(\tilde{C}_\gamma + \tilde{\alpha}_C\right)'\tilde{A}^{-1/2}\right\| \\ &\leq 2\left\|\tilde{A}^{-1}\right\|\left(\left\|\tilde{C}_\gamma\left(p\tilde{D}/T\right)^{-1}\tilde{C}_\gamma'\right\| + \left\|\tilde{\alpha}_C\left(p\tilde{D}/T\right)^{-1}\tilde{\alpha}_C'\right\|\right).\end{aligned}$$

Since

$$\begin{aligned}\tilde{C}_\gamma\left(p\tilde{D}/T\right)^{-1}\tilde{C}_\gamma' &= \tilde{C}_\gamma\left(p\tilde{D}_\gamma/T_\gamma\right)^{-1/2}(I_p + \tilde{\alpha}_D)\left(p\tilde{D}_\gamma/T_\gamma\right)^{-1/2}\tilde{C}_\gamma' \\ &= \tilde{C}_\gamma\left(p\tilde{D}_\gamma/T_\gamma\right)^{-1}\tilde{C}_\gamma' + \tilde{C}_\gamma\left(p\tilde{D}_\gamma/T_\gamma\right)^{-1/2}\tilde{\alpha}_D\left(p\tilde{D}_\gamma/T_\gamma\right)^{-1/2}\tilde{C}_\gamma',\end{aligned}$$

we have

$$\begin{aligned}\tilde{\lambda}_{1,pT} &\leq 2\left\|\tilde{A}^{-1}\right\|\left(\left\|\tilde{C}_\gamma\left(p\tilde{D}_\gamma/T_\gamma\right)^{-1}\tilde{C}_\gamma'\right\|(1 + \|\tilde{\alpha}_D\|) + \left\|\left(p\tilde{D}/T\right)^{-1}\right\|\|\tilde{\alpha}_C\|^2\right) \\ &\leq 2\left\|\tilde{A}^{-1}\right\|\left(\tilde{\lambda}_{1,pT_\gamma}\left\|\tilde{A}_\gamma\right\|(1 + \|\tilde{\alpha}_D\|) + \left\|\left(p\tilde{D}/T\right)^{-1}\right\|\|\tilde{\alpha}_C\|^2\right). \qquad (145)\end{aligned}$$

We will now establish bounds on various terms in the latter expression.

Bounds on $\left\|\tilde{A}_\gamma\right\|$ and $\left\|\tilde{A}^{-1}\right\|$. Since $\left\|\tilde{A}_\gamma\right\| \overset{\text{a.s.}}{\to} \left(1+\sqrt{\gamma}\right)^2$ and $\left\|\tilde{A}_\gamma^{-1}\right\| \overset{\text{a.s.}}{\to} \left(1-\sqrt{\gamma}\right)^{-2}$ as $p \to \infty$, we have

$$\left\|\tilde{A}_\gamma\right\| \leq 4 \text{ and } \left\|\tilde{A}^{-1}\right\| \leq \left\|\tilde{A}_\gamma^{-1}\right\| \leq 4 \qquad (146)$$

for any $\gamma \in (0, 1/4)$, with probability arbitrarily close to one, for all sufficiently large p, and all $T > T_\gamma$.

Bound on $\left\|\left(p\tilde{D}/T\right)^{-1}\right\|$. This norm equals the inverse of the smallest eigenvalue of $p\tilde{S}_{11}/T$. Recall that (see (141) and (1))

$$p\tilde{S}_{11}/T = p\varepsilon UM_lU'\varepsilon'/T^2 \text{ and } pS_{11}/T = p\varepsilon M_lUM_lU'M_l\varepsilon'/T^2$$

Note that the eigenvalues of UM_lU' are the same as those of $M_lU'UM_l$. Further,

$$M_lU'UM_l = M_lU'M_lUM_l + M_lU'll'UM_l/T = M_lUM_lU'M_l + M_lU'll'UM_l/T.$$

Hence, the j-th largest eigenvalue of UM_lU' is no smaller than the j-th largest eigenvalue of $M_lUM_lU'M_l$. Therefore, the probability that the smallest eigenvalue of $p\tilde{S}_{11}/T$ is below some number, say $\delta > 0$, is no larger than the probability that the smallest eigenvalue of pS_{11}/T is below δ.

On the other hand, in notation of Section 3.1.1, $pS_{11}/T = pD_0/T$. Therefore, by Lemmas 18 and 19, for any $\delta > 0$, all sufficiently large p, and all $T > \tilde{T}$, where $\tilde{T}$ may depend on p,

$$\Pr\left(\left\|(pS_{11}/T)^{-1}\right\| > K\right) < \delta/4,$$

where K is an absolute constant. The same inequality must hold for $\left\|\left(p\tilde{S}_{11}/T\right)^{-1}\right\|$ and thus, for any $\delta > 0$, all sufficiently large p, and all $T > \tilde{T}$, where $\tilde{T}$ may depend on p,

$$\Pr\left(\left\|\left(p\tilde{D}/T\right)^{-1}\right\| > K\right) < \delta/4. \qquad (147)$$

Bound on $\|\tilde{\alpha}_C\|$. Consider the following decomposition

$$\begin{aligned}\tilde{\alpha}_C &= C_0 - C_\gamma + \chi_p \xi y e_{1p}'/T - \chi_p \xi_\gamma y_\gamma e_{1p}'/T_\gamma \\ &= C_0 - C_\gamma + \chi_p \xi_\gamma \left(y_{1:\bar{T}_\gamma}/T - y_\gamma/T_\gamma \right) e_{1p}' + \chi_p \xi_{-\gamma} y_{T_\gamma:\bar{T}} e_{1p}'/T,\end{aligned}$$

where $y_{s:t}$ denotes the sub-vector of y that consists of all entries of y starting from entry s and ending with entry t. By Lemma 19, under conditions of Lemma 17,

$$\Pr\left(\|C_0 - C_\gamma\| \leq K\sqrt{\gamma}\right) > 1 - \delta/8. \tag{148}$$

Next, by definition of y,

$$y_{1:\bar{T}_\gamma}'/T - y_\gamma'/T_\gamma = \left(y_1'/T - y_{\gamma,1}'/T_\gamma, ..., y_{\bar{T}_\gamma/2}'/T - y_{\gamma,\bar{T}_\gamma/2}'/T_\gamma \right),$$

where

$$y_j'/T - y_{\gamma,j}'/T_\gamma = -\frac{1}{\sqrt{2}} \left(T^{-1} - T_\gamma^{-1}, \frac{1}{T} \frac{\sin \omega_j}{1 - \cos \omega_j} - \frac{1}{T_\gamma} \frac{\sin \omega_{j\gamma}}{1 - \cos \omega_{j\gamma}} \right).$$

Using (133) and (134), we conclude that for some absolute constant K and all $T > T_\gamma$,

$$\left\| y_{1:T_\gamma}/T - y_\gamma/T_\gamma \right\| \leq K T_\gamma^{-1/2} \leq K\sqrt{\gamma/p}.$$

This inequality and the fact that χ_p^2 is independent from ξ and has the chi-squared distribution with p degrees of freedom imply that

$$\left\| \chi_p \xi_\gamma \left(y_{1:\bar{T}_\gamma}/T - y_\gamma/T_\gamma \right) e_{1p}' \right\| \leq K\sqrt{p\gamma} \tag{149}$$

with probability arbitrarily close to one, for all sufficiently large p, and all $T > T_\gamma$.

Finally, since by definition of y and by (131),

$$\left\| y_{T_\gamma:\bar{T}}/T \right\|^2 \leq 1/(4T) + \sum_{s=\bar{T}_\gamma/2+1} s^{-2} \leq K/T_\gamma \leq K\gamma/p,$$

we have

$$\left\| \chi_p \xi_{-\gamma} y_{T_\gamma:\bar{T}} e_{1p}'/T \right\| \leq K\sqrt{p\gamma} \tag{150}$$

with probability arbitrarily close to one, for all sufficiently large p, and all $T > T_\gamma$.

Combining (148-150), we obtain that under conditions of Lemma 17,

$$\Pr\left(\|\tilde{\alpha}_C\| \leq K\sqrt{p\gamma}\right) > 1 - \delta/4. \tag{151}$$

Bound on $\|\tilde{\alpha}_D\|$. Using the definition of $\tilde{\alpha}_D$, we have

$$\begin{aligned}\|\tilde{\alpha}_D\| &\leq 1 + \left\| \left(p\tilde{D}_\gamma/T_\gamma\right)^{1/2} \left(p\tilde{D}/T\right)^{-1} \left(p\tilde{D}_\gamma/T_\gamma\right)^{1/2} \right\| = 1 + \left\| \left(p\tilde{D}/T\right)^{-1/2} \left(p\tilde{D}_\gamma/T_\gamma\right) \left(p\tilde{D}/T\right)^{-1/2} \right\| \\ &\leq 2 + \left\| \left(p\tilde{D}/T\right)^{-1/2} \left(p\tilde{D}/T - p\tilde{D}_\gamma/T_\gamma\right) \left(p\tilde{D}/T\right)^{-1/2} \right\| \leq 2 + \left\| \left(p\tilde{D}/T\right)^{-1} \right\| \left\| p\tilde{D}/T - p\tilde{D}_\gamma/T_\gamma \right\|.\end{aligned}$$

By (147), we have with high probability

$$\|\tilde{\alpha}_D\| \leq 2 + K \left\| p\tilde{D}/T - p\tilde{D}_\gamma/T_\gamma \right\|$$

for some absolute constant K, all sufficiently large p, and all $T > \tilde{T}$, where $\tilde{T}$ may depend on p.

Now, consider $p\tilde{D}/T - p\tilde{D}_\gamma/T_\gamma$. By definition, we have

$$\begin{aligned} p\tilde{D}/T - p\tilde{D}_\gamma/T_\gamma &= pD_0/T - pD_\gamma/T_\gamma + \frac{1}{12}p\chi_p^2 e_{1p}e_{1p}'\left(T_\gamma^{-2} - T^{-2}\right) \\ &\quad + p\chi_p\left(\xi_\gamma\left(x_{1:\bar{T}_\gamma}/T^2 - x_\gamma/T_\gamma^2\right)e_{1p}' + e_{1p}\left(x_{1:\bar{T}_\gamma}'/T^2 - x_\gamma'/T_\gamma^2\right)\xi_\gamma'\right) \\ &\quad + p\chi_p\left(\xi_{-\gamma}x_{T_\gamma:\bar{T}}e_{1p}' + e_{1p}x_{T_\gamma:\bar{T}}'\xi_{-\gamma}'\right)/T^2 \end{aligned}$$

The decomposition

$$D_0 = \xi_\gamma\Delta_\gamma\xi_\gamma'/T + \xi_{-\gamma}\Delta_{-\gamma}\xi_{-\gamma}'/T$$

and inequalities (130) and (132) imply that, under conditions of Lemma 17,

$$\Pr\left(\|pD_0/T - pD_\gamma/T_\gamma\| \le K\gamma\right) > 1 - \delta/8. \tag{152}$$

Further, with probability arbitrarily close to one, for sufficiently large p and all $T > T_\gamma$,

$$\left\|\frac{1}{12}p\chi_p^2 e_{1p}e_{1p}'\left(T_\gamma^{-2} - T^{-2}\right)\right\| \le K\gamma^2. \tag{153}$$

Next, by definition of x,

$$x_{1:\bar{T}_\gamma}'/T^2 - x_\gamma'/T_\gamma^2 = \left(x_1'/T^2 - x_{\gamma,1}'/T_\gamma^2, ..., x_{\bar{T}_\gamma/2}'/T^2 - x_{\gamma,\bar{T}_\gamma/2}'/T_\gamma^2\right),$$

where

$$x_j'/T^2 - x_{\gamma,j}'/T_\gamma^2 = -\frac{1}{\sqrt{2}}\left(\frac{1}{T^2}\frac{\cos\omega_j}{1-\cos\omega_j} - \frac{1}{T_\gamma^2}\frac{\cos\omega_{j\gamma}}{1-\cos\omega_{j\gamma}}, \frac{1}{T^2}\frac{-\sin\omega_j}{1-\cos\omega_j} - \frac{1}{T_\gamma^2}\frac{-\sin\omega_{j\gamma}}{1-\cos\omega_{j\gamma}}\right).$$

Using (128-129) and (133-134), we conclude that for some absolute constant K and all $T > T_\gamma$,

$$\left\|x_{1:\bar{T}_\gamma}/T^2 - x_\gamma/T_\gamma^2\right\| \le KT_\gamma^{-1} \le K\gamma/p.$$

Therefore, with probability arbitrarily close to one, for all sufficiently large p, and all $T > T_\gamma$,

$$\left\|p\chi_p\left(\xi_\gamma\left(x_{1:\bar{T}_\gamma}/T^2 - x_\gamma/T_\gamma^2\right)e_{1p}' + e_{1p}\left(x_{1:\bar{T}_\gamma}'/T^2 - x_\gamma'/T_\gamma^2\right)\xi_\gamma'\right)\right\| \le K\gamma p. \tag{154}$$

Finally, by definition and by (131)

$$\left\|x_{T_\gamma:\bar{T}}/T\right\|^2 \le K\gamma/p.$$

Therefore, with probability arbitrarily close to one, for all sufficiently large p, and all $T > T_\gamma$,

$$\left\|p\chi_p\left(\xi_{-\gamma}x_{T_\gamma:\bar{T}}e_{1p}' + e_{1p}x_{T_\gamma:\bar{T}}'\xi_{-\gamma}'\right)/T^2\right\| \le K\gamma^{3/2}p^{1/2}. \tag{155}$$

Combining (152-155), we obtain that under conditions of Lemma 17,

$$\Pr\left(\|\tilde{\alpha}_D\| \le Kp\gamma\right) > 1 - \delta/4. \tag{156}$$

The established bounds on $\left\|\tilde{A}_\gamma\right\|$, $\left\|\tilde{A}^{-1}\right\|$, $\left\|\left(p\tilde{D}/T\right)^{-1}\right\|$, $\|\tilde{\alpha}_C\|$, and $\|\tilde{\alpha}_D\|$ together with (145) yield Lemma 20.

3.2.5 Proof of Lemma 21 (lower bound on $\tilde{\lambda}_{1,pT_\gamma}$)

The largest eigenvalue, $p\tilde{\lambda}_{1,pT_\gamma}/T_\gamma$, of $\tilde{C}_\gamma \tilde{D}_\gamma^{-1}\tilde{C}_\gamma' \tilde{A}_\gamma^{-1}$ equals that of the product of the projections on the column spaces of matrices $\nabla_\gamma'\xi_\gamma' + \chi_p y_\gamma e_{1p}'$ and ξ_γ'. Since the former space is spanned by the columns of $\left[y_\gamma, \nabla_\gamma'\xi_\gamma'\right]$, $p\tilde{\lambda}_{1,pT_\gamma}/T_\gamma$ cannot be larger than $\left\|\tilde{\mathcal{P}}_1\mathcal{P}_2\right\|^2$, where $\tilde{\mathcal{P}}_1$ and $\mathcal{P}_2$ are the projections on the column spaces of $\left[y_\gamma, \nabla_\gamma'\xi_\gamma'\right]$ and ξ_γ'.

Denote the projection on the column space of $\nabla_\gamma'\xi_\gamma'$ as $\mathcal{P}_1$, and that on the orthogonal space as $\mathcal{M}_1$. Then

$$\tilde{\mathcal{P}}_1\mathcal{P}_2 = \mathcal{P}_1\mathcal{P}_2 + \mathcal{P}_{\mathcal{M}_1 y_\gamma}\mathcal{P}_2,$$

where $\mathcal{P}_{\mathcal{M}_1 y_\gamma}$ is the projection on the space generated by the vector $\mathcal{M}_1 y_\gamma$. This yields

$$p\tilde{\lambda}_{1,pT_\gamma}/T_\gamma \le 2\left\|\mathcal{P}_1\mathcal{P}_2\right\|^2 + 2\left\|\mathcal{P}_{\mathcal{M}_1 y_\gamma}\mathcal{P}_2\right\|^2 = 2\left\|\mathcal{P}_1\mathcal{P}_2\right\|^2 + 2\frac{y_\gamma'\mathcal{M}_1\mathcal{P}_2\mathcal{M}_1 y_\gamma}{y_\gamma'\mathcal{M}_1 y_\gamma}.$$

Onatski and Wang's (2017a) Theorem 1 implies that, as $p \to \infty$,

$$\left\|\mathcal{P}_1\mathcal{P}_2\right\|^2 \overset{\text{a.s.}}{\to} \gamma\left(\sqrt{2} - \sqrt{1-\gamma}\right)^{-2}. \tag{157}$$

Therefore, to establish Lemma 21, it is sufficient to show that for any $\delta > 0$, there exists $\gamma_\delta > 0$ such that for any $\gamma \in (0, \gamma_\delta)$ and all sufficiently large p,

$$\Pr\left(\frac{y_\gamma'\mathcal{M}_1\mathcal{P}_2\mathcal{M}_1 y_\gamma}{y_\gamma'\mathcal{M}_1 y_\gamma} > K\gamma^{1/4}\right) < \delta, \tag{158}$$

where K is an absolute constant. Below, we will denote absolute constants that may take different values from one appearance to another as K, K_1, and K_2. We will denote constants that depend on the value of γ as K_γ.

To simplify notation, we will omit the subscript γ from y_γ, ∇_γ, and ξ_γ. However, we will keep notation T_γ and $\bar{T}_\gamma = T_\gamma - 1$ to remind the reader that p/T_γ is close to γ. Our plan is to derive bounds on $y'\mathcal{M}_1 y$ and $y'\mathcal{M}_1\mathcal{P}_2\mathcal{M}_1 y$, and then combine these bounds to obtain (158).

By definition,

$$y'\mathcal{M}_1 y = y'y - x'\xi'\left(\xi\nabla\nabla'\xi'\right)^{-1}\xi x,$$

where

$$x = \nabla y = -\frac{1}{\sqrt{2}}\left(\frac{\cos\omega_1}{1-\cos\omega_1}, \frac{-\sin\omega_1}{1-\cos\omega_1}, \frac{\cos\omega_2}{1-\cos\omega_2}, \frac{-\sin\omega_2}{1-\cos\omega_2}, \ldots\right)'.$$

Therefore,

$$y'\mathcal{M}_1 y = x'\Delta^{-1}x - \sum_{i,j=1}^{\bar{T}_\gamma} x_i x_j \xi_i'\left(\xi\Delta\xi'\right)^{-1}\xi_j.$$

where x_j and ξ_j are, respectively, the j-th element of x and the j-th column of ξ, and

$$\Delta = \nabla'\nabla = \operatorname{diag}\left\{(2 - 2\cos\omega_s)^{-1} I_2\right\}_{s=1}^{\bar{T}_\gamma/2}. \tag{159}$$

Denoting matrix $p\xi\Delta\xi'/T_\gamma^2$ as W, we obtain

$$y'\mathcal{M}_1 y = \sum_{i,j} x_i x_j\left(\Delta_{ii}^{-1}\delta_{i=j} - p\xi_i' W^{-1}\xi_j/T_\gamma^2\right), \tag{160}$$

where $\delta_{i=j}$ is the Kronecker delta. First, we are going to analyze the part of the sum corresponding to $i = j$. We call this part the diagonal component of $y'\mathcal{M}_1 y$.

The diagonal component of $y'\mathcal{M}_1 y$. Using the Sherman-Morrison-Woodbury (SMW) formula (see (45)), we obtain

$$W^{-1} = W_{-i}^{-1} - \frac{W_{-i}^{-1}\xi_i\xi_i' W_{-i}^{-1}/p}{\Delta_{ii}^{-1}T_\gamma^2/p^2 + \xi_i' W_{-i}^{-1}\xi_i/p}, \tag{161}$$

where $W_{-i} = \left(p/T_\gamma^2\right)\sum_{j\neq i}\Delta_{jj}\xi_j\xi_j'$. Therefore,

$$\Delta_{ii}^{-1} - p\xi_i' W^{-1}\xi_i/T_\gamma^2 = \frac{\Delta_{ii}^{-2}T_\gamma^2/p^2}{\Delta_{ii}^{-1}T_\gamma^2/p^2 + \xi_i' W_{-i}^{-1}\xi_i/p}.$$

Since $x_i^2 = 2s_i^2\Delta_{ii}^2$, where $s_i^2 = \cos^2\omega_{\lceil i/2\rceil}$ for odd i and $s_i^2 = \sin^2\omega_{\lceil i/2\rceil}$ for even i, we have

$$x_i^2\left(\Delta_{ii}^{-1} - p\xi_i' W^{-1}\xi_i/T_\gamma^2\right) = \frac{2s_i^2 T_\gamma^2/p^2}{b_{pi} + \xi_i' W_{-i}^{-1}\xi_i/p} \tag{162}$$

where $b_{pi} = \Delta_{ii}^{-1}T_\gamma^2/p^2$.

Bounds on $\xi_i' W_{-i}^{-1}\xi_i/p$. In this subsection, we show that there exist $K_2 > K_1 > 0$ such that $\xi_i' W_{-i}^{-1}\xi_i/p \in [K_1, K_2]$ with overwhelming probability as $p\to\infty$.

Definition 22 *(Tao and Vu (2011)) Let $\mathcal{E}$ be an event depending on p. Then $\mathcal{E}$ holds with overwhelming probability (w.ow.p.) if $\Pr(\mathcal{E}) \geq 1 - O_K\left(p^{-K}\right)$ for every constant $K > 0$. Here $O_K\left(p^{-K}\right)$ denotes a quantity that is smaller than Bp^{-K} with constant B that may depend on K.*

Assuming that $\gamma < 1/2$, we have

$$W_{-i} \equiv \left(p/T_\gamma^2\right)\sum_{j\neq i}\Delta_{jj}\xi_j\xi_j' \geq \left(p/T_\gamma^2\right)\sum_{j\neq i, j\leq 2p}\Delta_{jj}\xi_j\xi_j',$$

and, by (159), $\min_{j\leq 2p}\Delta_{jj} \geq \left(2\pi p/T_\gamma\right)^{-2}$, so that

$$\left\|W_{-i}^{-1}\right\| \leq \left\|\left(\left(2\pi\right)^{-2}p^{-1}\sum_{j\neq i, j\leq 2p}\xi_j\xi_j'\right)^{-1}\right\| = 2\pi^2\lambda_{\min}^{-1}, \tag{163}$$

where $\lambda_{\min}$ is the smallest eigenvalue of the Wishart matrix $\sum_{j\neq i, j\leq 2p}\xi_j\xi_j'/\left(2p\right)$. Therefore,

$$\xi_i' W_{-i}^{-1}\xi_i/p \leq 2\pi^2\lambda_{\min}^{-1}\xi_i'\xi_i/p.$$

Gaussian concentration inequalities for $\chi^2(p)$ and $\lambda_{\min}$ (see e.g. Theorem II.13 of Davidson and Szarek (2001)) imply that there exist $K_2 > 0$ and $K > 0$ such that

$$\Pr\left(\xi_i' W_{-i}^{-1}\xi_i/p \geq K_2\right) \leq e^{-Kp}. \tag{164}$$

Now, let us establish a lower bound. The following inequality follows from the tail inequality for linear combinations of χ^2 (see Laurent and Massart (2000), Lemma 1)

$$\Pr\left(\xi_i' W_{-i}^{-1}\xi_i/p \leq \operatorname{tr} W_{-i}^{-1}/p - (2/p)\sqrt{t\operatorname{tr} W_{-i}^{-2}}\right) \leq e^{-t},$$

where t is any positive number. Setting $t = \sqrt{p}/2$ and using

$$\frac{2}{p}\sqrt{t\operatorname{tr} W_{-i}^{-2}} \leq 2\sqrt{\frac{2}{p}t\left\|W_{-i}^{-1}\right\|\frac{1}{2p}\operatorname{tr} W_{-i}^{-1}} \leq \frac{2}{p}t\left\|W_{-i}^{-1}\right\| + \frac{1}{2p}\operatorname{tr} W_{-i}^{-1},$$

we obtain

$$\Pr\left(\xi_i' W_{-i}^{-1}\xi_i/p \le \operatorname{tr} W_{-i}^{-1}/(2p) - \left\|W_{-i}^{-1}\right\|/\sqrt{p}\right) \le e^{-\sqrt{p}/2}. \tag{165}$$

To analyze the term $\operatorname{tr} W_{-i}^{-1}/(2p)$ in (165), consider the decomposition

$$W_{-i} = \left(p/T_\gamma^2\right)\sum_{j\neq i, j<p/2}\Delta_{jj}\xi_j\xi_j' + \left(p/T_\gamma^2\right)\sum_{j\neq i, j\geq p/2}\Delta_{jj}\xi_j\xi_j'.$$

Since the rank of the first term on the right hand side is smaller than $p/2$, we have by Weyl's inequalities for eigenvalues, for any $j \le p/2$,

$$\lambda_j\left(W_{-i}^{-1}\right) > \lambda_{j+p/2}\left(\left(\left(p/T_\gamma^2\right)\sum_{j\neq i, j\geq p/2}\Delta_{jj}\xi_j\xi_j'\right)^{-1}\right) > \left\|\left(p/T_\gamma^2\right)\sum_{j\neq i, j\geq p/2}\Delta_{jj}\xi_j\xi_j'\right\|^{-1}.$$

where $\lambda_j(M)$ denotes the j-th largest eigenvalue of symmetric matrix M. Therefore,

$$\operatorname{tr} W_{-i}^{-1}/(2p) > \frac{1}{4}\left\|\left(p/T_\gamma^2\right)\sum_{j\neq i, j\geq p/2}\Delta_{jj}\xi_j\xi_j'\right\|^{-1}.$$

Since $\Delta_{jj} = \left(2 - 2\cos\frac{2\pi\lceil j/2\rceil}{T}\right)^{-1}$ and $1-\cos x > x^2/6$ for $x \in (0,\pi]$,

$$\Delta_{jj} < \frac{6}{2\left(2\pi\lceil j/2\rceil/T_\gamma\right)^2} \le 3T_\gamma^2/\left(\pi^2 j^2\right),$$

and

$$\operatorname{tr} W_{-i}^{-1}/(2p) > \frac{\pi^2}{12}\left\|\sum_{j\neq i, j\geq p/2} p j^{-2}\xi_j\xi_j'\right\|^{-1}. \tag{166}$$

For simplicity, assume that $i < p/2$ (the other case can be analyzed similarly with only minor changes). Let $\psi = \left(\xi_{p/2}', ..., \xi_{\bar{T}_\gamma}'\right)'$ and

$$f(\psi) = \sqrt{p}\left\|\left[\frac{1}{p/2}\xi_{p/2}, ..., \frac{1}{\bar{T}_\gamma}\xi_{\bar{T}_\gamma}\right]\right\|$$

Then, $f(\cdot)$ is a $2/\sqrt{p}$-Lipschitz function. Indeed, let $\delta \equiv \left(\delta_{p/2}', ..., \delta_{\bar{T}_\gamma}'\right)' \in \mathbb{R}^{(T_\gamma - p/2)p}$. Then

$$|f(\psi+\delta) - f(\psi)| \le \sqrt{p}\left\|\left[\frac{1}{p/2}\delta_{p/2}, ..., \frac{1}{\bar{T}_\gamma}\delta_{\bar{T}_\gamma}\right]\right\| \le \frac{2}{\sqrt{p}}\|\delta\|_2,$$

where $\|\delta\|_2$ is the Euclidean norm of δ. Therefore, by Gaussian concentration inequality (see Ledoux (2000) prop. 2.18), for every $t \ge 0$,

$$\Pr\left(f(\psi) \ge \mathbb{E}f(\psi) + t\right) \le e^{-pt^2/8}. \tag{167}$$

On the other hand, by Latala's (2004) Theorem 1, there exists K such that

$$\mathbb{E}f(\psi) \le \sqrt{p}K\left(\sqrt{\sum_{j\geq p/2} j^{-2}} + \sqrt{p/(p/2)^2} + \sqrt[4]{\sum_{j\geq p/2} p j^{-4}}\right).$$

The right hand side of the above inequality is smaller than some other absolute constant K, so that

$$\mathbb{E}f(\psi) \le K, \tag{168}$$

and hence, there exist $K_1 > 0$ and $K_2 > 0$ such that

$$\Pr\left(f(\psi) \geq K_1\right) \leq e^{-K_2 p}. \tag{169}$$

Combining (166) and (169) yields

$$\begin{aligned} \Pr\left\{\operatorname{tr} W_{-i}^{-1}/(2p) \leq \pi^2/\left(12K_1^2\right)\right\} &\leq \Pr\left(\left\|\sum_{j\geq p/2} pj^{-2}\xi_j\xi_j'\right\|^{-1/2} \leq K_1^{-1}\right) \\ &= \Pr\left(f(\psi)^{-1} \leq K_1^{-1}\right) = \Pr\left(f(\psi) \geq K_1\right) \leq e^{-K_2 p}. \end{aligned}$$

Hence, for some absolute constants $K_1 > 0$ and $K_2 > 0$, we have

$$\Pr\left(\operatorname{tr} W_{-i}^{-1}/(2p) \leq K_1\right) \leq e^{-K_2 p}. \tag{170}$$

Combining (165) and (170), and recalling that $\left\|W_{-i}^{-1}\right\|$ is bounded from above with probability approaching one exponentially fast (see (163)), we conclude that there exists an absolute constant $K_1 > 0$ such that

$$\Pr\left(\xi_i' W_{-i}^{-1}\xi_i/p \leq K_1\right) < 2e^{-\frac{1}{2}\sqrt{p}}. \tag{171}$$

This and (164) yield the following lemma.

Lemma 23 *For some absolute constants $K_1 > 0$ and $K_2 > 0$, $\xi_i' W_{-i}^{-1}\xi_i/p \in [K_1, K_2]$ w.ow.p.*

The order of the diagonal component. Lemma 23 and equation (162) imply that

$$\sum_i \frac{2s_i^2 T_\gamma^2/p^2}{b_{pi} + K_2} \leq \sum_i x_i^2\left(\Delta_{ii}^{-1} - p\xi_i' W^{-1}\xi_i/T_\gamma^2\right) \leq \sum_i \frac{2s_i^2 T_\gamma^2/p^2}{b_{pi} + K_1} \tag{172}$$

w.ow.p. On the other hand, since $s_i^2 + s_{i+1}^2 = 1$ and $b_{pi} = b_{p,i+1}$ for odd i, we have, for any $K > 0$,

$$\sum_{i=1}^{\bar{T}_\gamma} \frac{2s_i^2 T_\gamma^2/p^2}{b_{pi} + K} = \sum_{j=1}^{\bar{T}_\gamma/2} \frac{1}{k - \cos(2\pi j/T_\gamma)} = \frac{T_\gamma}{2\pi}\int_0^\pi \frac{\mathrm{d}\varphi}{k - \cos\varphi} + O(1)$$

as $p \to \infty$, where $k = 1 + \frac{1}{2}Kp^2/T_\gamma^2$. According to Gradshteyn and Ryzhik (2000, formula (2.553)),

$$\frac{T_\gamma}{2\pi}\int_0^\pi \frac{\mathrm{d}\varphi}{k - \cos\varphi} = \frac{T_\gamma}{2\sqrt{k^2 - 1}}.$$

Hence, there exist $K_1 > 0$ and $K_2 > 0$ such that, w.ow.p.,

$$p/\left(K_2\gamma^2\right) \leq \sum_i x_i^2\left(\Delta_{ii}^{-1} - p\xi_i' W^{-1}\xi_j/T_\gamma^2\right) \leq p/\left(K_1\gamma^2\right). \tag{173}$$

The off-diagonal component of $y'\mathcal{M}_1 y$. We will now establish a bound on the second moment of the off-diagonal component of $y'\mathcal{M}_1 y$. The square of this component consists of three parts

$$\left(-\frac{p^2}{T_\gamma^2}\sum_{i\neq j} x_i x_j \frac{1}{p}\xi_i' W^{-1}\xi_j\right)^2 = (p/T_\gamma)^4\left(M_1 + M_2 + M_3\right),$$

where

$$\begin{aligned} M_1 &= 2\sum_{i\neq j} x_i^2 x_j^2\left(\xi_i' W^{-1}\xi_j/p\right)^2, \\ M_2 &= 4\sum_{i\neq j\neq t} x_i x_j^2 x_t\left(\xi_i' W^{-1}\xi_j/p\right)\left(\xi_j' W^{-1}\xi_t/p\right), \text{ and} \\ M_3 &= \sum_{i\neq j\neq s\neq t} x_i x_j x_s x_t\left(\xi_i' W^{-1}\xi_j/p\right)\left(\xi_s' W^{-1}\xi_t/p\right). \end{aligned}$$

The above sums run over ordered pairs, triples, and quadruples of unequal indices (no repeated indices in any of these sets). Multiplier 2 in the definition of M_1 takes into account the fact that the term, say, $x_1^2 x_2^2 \left(\xi_1' W^{-1}\xi_2/p\right)^2$ appears in $\left(\sum_{i\neq j} x_i x_j \xi_i' W^{-1}\xi_j/p\right)^2$ four times (corresponding to $x_1x_2x_1x_2$, $x_1x_2x_2x_1$, $x_2x_1x_1x_2$, and $x_2x_1x_2x_1$), whereas it appears in $\sum_{i\neq j} x_i^2 x_j^2 \left(\xi_i' W^{-1}\xi_j/p\right)^2$ only two times (corresponding to $x_1^2x_2^2$ and $x_2^2x_1^2$). Multiplier 4 in the definition of M_2 has a similar justification.

Analysis of M_1. By (161),

$$\xi_i' W^{-1}\xi_j/p = \xi_i'\left(W_{-i}^{-1} - \frac{W_{-i}^{-1}\xi_i\xi_i' W_{-i}^{-1}/p}{b_{pi} + \xi_i' W_{-i}^{-1}\xi_i/p}\right)\xi_j/p = \frac{b_{pi}\left(\xi_i' W_{-i}^{-1}\xi_j/p\right)}{b_{pi} + \xi_i' W_{-i}^{-1}\xi_i/p}. \tag{174}$$

Similarly,

$$\xi_i' W_{-i}^{-1}\xi_j/p = \frac{b_{pj}\left(\xi_i' W_{-i,j}^{-1}\xi_j/p\right)}{b_{pj} + \xi_j' W_{-i,j}^{-1}\xi_j/p},$$

where $W_{-i,j} = \left(p/T_\gamma^2\right)\sum_{s\neq i,j}\Delta_{ss}\xi_s\xi_s'$, so that

$$x_i^2 x_j^2\left(\xi_i' W^{-1}\xi_j/p\right)^2 = \left(\xi_i' W_{-i,j}^{-1}\xi_j/p\right)^2\left(\frac{x_j b_{pj}}{b_{pj} + \xi_j' W_{-i,j}^{-1}\xi_j/p}\frac{x_i b_{pi}}{b_{pi} + \xi_i' W_{-i}^{-1}\xi_i/p}\right)^2. \tag{175}$$

Lemma 24 *For any $i \neq j$, $x_i^2 x_j^2\left(\xi_i' W^{-1}\xi_j/p\right)^2 \leq K_\gamma p^4$ for some constant K_γ that depends on γ.*

Proof. Similarly to (174), we get

$$\xi_i' W_{-i}^{-1}\xi_i/p = \xi_i' W_{-i,j}^{-1}\xi_i/p - \frac{\left(\xi_i' W_{-i,j}^{-1}\xi_j/p\right)^2}{b_{pj} + \xi_j' W_{-i,j}^{-1}\xi_j/p}.$$

Using this in (175), we obtain

$$x_i^2 x_j^2\left(\xi_i' W^{-1}\xi_j/p\right)^2 = \frac{\left(x_j b_{pj}\right)^2\left(x_i b_{pi}\right)^2\left(\xi_i' W_{-i,j}^{-1}\xi_j/p\right)^2}{\left(\left(b_{pi} + \xi_i' W_{-i,j}^{-1}\xi_i/p\right)\left(b_{pj} + \xi_j' W_{-i,j}^{-1}\xi_j/p\right) - \left(\xi_i' W_{-i,j}^{-1}\xi_j/p\right)^2\right)^2}.$$

By the Cauchy-Schwarz inequality,

$$\left(\xi_j' W_{-i,j}^{-1}\xi_j/p\right)\left(\xi_i' W_{-i,j}^{-1}\xi_i/p\right) - \left(\xi_i' W_{-i,j}^{-1}\xi_j/p\right)^2 \geq 0.$$

Therefore,

$$\begin{aligned} x_i^2 x_j^2\left(\xi_i' W^{-1}\xi_j/p\right)^2 &\leq \frac{\left(x_j b_{pj}\right)^2\left(x_i b_{pi}\right)^2\left(\xi_i' W_{-i,j}^{-1}\xi_j/p\right)^2}{\left(b_{pi}\xi_j' W_{-i,j}^{-1}\xi_j/p + b_{pj}\xi_i' W_{-i,j}^{-1}\xi_i/p\right)^2} \\ &\leq \frac{\left(x_j b_{pj}\right)^2\left(x_i b_{pi}\right)^2}{2b_{pi}b_{pj}} = 2s_j^2 s_i^2\Delta_{ii}\Delta_{jj}\left(T_\gamma/p\right)^4 \leq K_\gamma p^4. \square \end{aligned}$$

The fact that

$$x_j^2 b_{pj}^2 = 2s_j^2\left(T_\gamma/p\right)^4, \tag{176}$$

together with Lemma 23 and an analogous result for $\xi_j' W_{-i,j}^{-1}\xi_j/p$ imply that the last squared term in (175) is bounded by K_γ w.ow.p. This and Lemma 24 guarantee that, for any integer k,

$$\mathbb{E}x_i^2 x_j^2\left(\xi_i' W^{-1}\xi_j/p\right)^2 \leq K_\gamma\mathbb{E}\left(\xi_i' W_{-i,j}^{-1}\xi_j/p\right)^2 + o(p^{-k}) \tag{177}$$

as $p \to \infty$.

But

$$\mathbb{E}\left(\xi_i' W_{-i,j}^{-1} \xi_j / p\right)^2 = \mathbb{E}\left(\operatorname{tr} W_{-i,j}^{-2} / p\right) / p \leq \mathbb{E}\mu_{\min}^{-2} / p,$$

where $\mu_{\min}$ is the smallest eigenvalue of a Wishart matrix $\sum_{s \neq i,j, s \leq 2p} \xi_s \xi_s' / (2p)$ (see (163) for a similar inequality derived above). Lemma 25 below implies that $\mathbb{E}\mu_{\min}^{-2}$ is bounded by an absolute constant for all sufficiently large p, so that

$$\mathbb{E}\left(\xi_i' W_{-i,j}^{-1} \xi_j / p\right)^2 \leq K/p.$$

This inequality, inequality (177) and the fact that there are less than T_γ^2 terms in the sum defining M_1 implies that

$$\mathbb{E} M_1 \leq K_\gamma p \tag{178}$$

for all sufficiently large p. How large p needs to be may depend on γ.

Lemma 25 *Let η be a $p \times m$ matrix with i.i.d. $N(0,1)$ entries, where $p \leq (1+\tau) m/2$ with $\tau \in (0,1)$. Let $\mu_{\min}$ be the smallest eigenvalue of $\eta\eta'/m$. Then, for any $\rho > 0$, there exists $K_{\rho,\tau} > 0$, which may depend on ρ and τ, such that $\mathbb{E}\mu_{\min}^{-\rho} \leq K_{\rho,\tau}$ for all sufficiently large p and m along a sequence $p, m \to_\tau \infty$.*

Proof: It follows from Chen and Dongarra (2005, p. 610) that

$$\Pr\left(\mu_{\min} \leq \mu\right) < m^{m-p+1} \mu^{\frac{m-p+1}{2}} / \Gamma\left(m-p+2\right).$$

Their $\lambda_{\min}, n$, and m equal $m\mu_{\min}, m$, and p in our notation, respectively. By Stirling's formula (see e.g. 6.1.38 in Abramowitz and Stegun (1970)),

$$\Gamma\left(m-p+2\right) \geq \sqrt{2\pi}\left(m-p+1\right)^{m-p+3/2} e^{-(m-p+1)}.$$

Further, for $p \leq (1+\tau) m/2$, we have $(1-\tau) m/2 \leq m-p+1$. Therefore, for all $N > 0$ we have

$$\Pr\left(\mu_{\min}^{-1} > N\right) < m^{m-p+1} N^{-\frac{m-p+1}{2}} / \Gamma\left(m-p+2\right) \leq \left(2\pi\left(m-p+1\right)\right)^{-1/2} \left(N\left(\frac{1-\tau}{2e}\right)^2\right)^{-\frac{m-p+1}{2}}.$$

Hence, for any $q > 0$ and sufficiently large p, m along a sequence $p, m \to_\tau \infty$, we have

$$\Pr\left(\mu_{\min}^{-1} > N\right) \leq N^{-q}\left(\frac{1-\tau}{2e}\right)^{-2q}.$$

On the other hand, according to Lemma 2.6 of Bai and Silverstein (1998), if for all $N > 0$, $\Pr\left(\mu_{\min}^{-1} > N\right) \leq N^{-q} K$ for some positive q and K, then, for any positive $\rho < q$,

$$\mathbb{E}\mu_{\min}^{-\rho} \leq K^{\rho/q} \frac{q}{q-\rho}. \square$$

Analysis of M_2. Similarly to (175), we have

$$x_i x_j^2 x_t \left(\xi_i' W^{-1} \xi_j / p\right)\left(\xi_j' W^{-1} \xi_t / p\right) = \left(\xi_i' W_{-i,j}^{-1} \xi_j / p\right)\left(\xi_j' W_{-j,t}^{-1} \xi_t / p\right)$$
$$\times \frac{x_i b_{pi}}{b_{pi} + \xi_i' W_{-i,j}^{-1} \xi_i / p} \frac{x_t b_{pt}}{b_{pt} + \xi_t' W_{-j,t}^{-1} \xi_t / p} \frac{x_j^2 b_{pj}^2}{\left(b_{pj} + \xi_j' W_{-j}^{-1} \xi_j / p\right)^2}$$

Further,

$$\xi_i' W_{-i,j}^{-1} \xi_j / p = \xi_i' W_{-i,j,t}^{-1} \xi_j / p - \frac{\left(\xi_i' W_{-i,j,t}^{-1} \xi_t / p\right)\left(\xi_t' W_{-i,j,t}^{-1} \xi_j / p\right)}{b_{pt} + \xi_t' W_{-i,j,t}^{-1} \xi_t / p}$$

and

$$\xi_j' W_{-j,t}^{-1}\xi_t/p = \xi_j' W_{-i,j,t}^{-1}\xi_t/p - \frac{\left(\xi_j' W_{-i,j,t}^{-1}\xi_i/p\right)\left(\xi_i' W_{-i,j,t}^{-1}\xi_t/p\right)}{b_{pi} + \xi_i' W_{-i,j,t}^{-1}\xi_i/p}.$$

To shorten notation, denote

$$\xi_i' W_{-i,j,t}^{-1}\xi_j/p = a_{ij}, \xi_i' W_{-i,j,t}^{-1}\xi_t/p = a_{it}, \text{ etc.}$$

Further, let $a_{ii}^{(+t)} = \xi_i' W_{-i,j}^{-1}\xi_i/p$, $a_{tt}^{(+i)} = \xi_t' W_{-j,t}^{-1}\xi_t/p$, and $a_{jj}^{(+i,t)} = \xi_j' W_{-j}^{-1}\xi_j/p$. Then

$$\begin{aligned} x_i x_j^2 x_t \left(\xi_i' W^{-1}\xi_j/p\right)\left(\xi_j' W^{-1}\xi_t/p\right) &= \left(a_{ij} - \frac{a_{it}a_{jt}}{b_{pt} + a_{tt}}\right)\left(a_{jt} - \frac{a_{it}a_{ij}}{b_{pi} + a_{ii}}\right) \\ &\quad \times \frac{x_i b_{pi}}{b_{pi} + a_{ii}^{(+t)}} \frac{x_t b_{pt}}{b_{pt} + a_{tt}^{(+i)}} \frac{x_j^2 b_{pj}^2}{\left(b_{pj} + a_{jj}^{(+i,t)}\right)^2} \end{aligned} \tag{179}$$

Using the identity

$$\frac{1}{b+x} = \frac{1}{b+\alpha} - \frac{x-\alpha}{(b+x)(b+\alpha)},$$

we expand the right hand side of (179) as follows

$$\begin{aligned} &\left(a_{ij} - \frac{a_{it}a_{jt}}{b_{pt} + a_{tt}}\right)\left(a_{jt} - \frac{a_{it}a_{ij}}{b_{pi} + a_{ii}}\right) \\ &\times \left(\frac{x_i b_{pi}}{b_{pi} + \mathbb{E}a_{ii}^{(+t)}} - \frac{x_i b_{pi}\left(a_{ii}^{(+t)} - \mathbb{E}a_{ii}^{(+t)}\right)}{\left(b_{pi} + a_{ii}^{(+t)}\right)\left(b_{pi} + \mathbb{E}a_{ii}^{(+t)}\right)}\right) \\ &\times \left(\frac{x_t b_{pt}}{b_{pt} + \mathbb{E}a_{tt}^{(+i)}} - \frac{x_t b_{pt}\left(a_{tt}^{(+i)} - \mathbb{E}a_{tt}^{(+i)}\right)}{\left(b_{pt} + a_{tt}^{(+i)}\right)\left(b_{pt} + \mathbb{E}a_{tt}^{(+i)}\right)}\right) \\ &\times \left(\frac{x_j b_{pj}}{b_{pj} + \mathbb{E}a_{jj}^{(+i,t)}} - \frac{x_j b_{pj}\left(a_{jj}^{(+i,t)} - \mathbb{E}a_{jj}^{(+i,t)}\right)}{\left(b_{pj} + a_{jj}^{(+i,t)}\right)\left(b_{pj} + \mathbb{E}a_{jj}^{(+i,t)}\right)}\right)^2 \end{aligned}$$

It is straightforward to verify that $\mathbb{E}(a_{ij}a_{jt}) = 0$. Therefore, opening up brackets in the above expression and taking expectation, we obtain a sum of terms each of which is proportional to a monomial in a_{ij}, a_{it}, a_{jt}, $a_{ii}^{(+t)} - \mathbb{E}a_{ii}^{(+t)}$, $a_{tt}^{(+i)} - \mathbb{E}a_{tt}^{(+i)}$, and $a_{jj}^{(+i,t)} - \mathbb{E}a_{jj}^{(+i,t)}$ of degree no less than three. Moreover, the coefficients of proportionality are smaller by absolute value than a quantity which depends only on γ, w.ow.p.

The validity of the last statement follows from Lemma 23 (cf. discussion immediately below (176)) and from the fact that $\mathbb{E}a_{ii}^{(+t)}$, $\mathbb{E}a_{tt}^{(+i)}$ and $\mathbb{E}a_{jj}^{(+i,t)}$ are bounded from below by a positive absolute constant. We establish this fact for $\mathbb{E}a_{jj}^{(+i,t)}$ (a proof for the other expectations is similar). By definition of $a_{jj}^{(+i,t)}$ and by (166),

$$\mathbb{E}a_{jj}^{(+i,t)} = \mathbb{E}\,\mathrm{tr}\, W_{-j}^{-1}/p \geq \frac{\pi^2}{6}\mathbb{E}\left\|\sum_{s\neq j, s\geq p/2} ps^{-2}\xi_s\xi_s'\right\|^{-1}$$

By Jensen's inequality

$$\mathbb{E}\left\|\sum_{s\neq j, s\geq p/2} ps^{-2}\xi_s\xi_s'\right\|^{-1} \geq \left(\mathbb{E}\left\|\sum_{s\neq j, s\geq p/2} ps^{-2}\xi_s\xi_s'\right\|^{1/2}\right)^{-2} \geq K$$

for some absolute constant K, with the last inequality following from (168).

Lemma 24 and the boundedness of the coefficients of proportionality w.ow.p. imply that an upper bound on the expected value of the right hand side of (179) would follow from upper bounds on the expected value of monomials in a_{ij}, a_{it}, a_{jt}, $a_{ii}^{(+t)} - \mathbb{E}a_{ii}^{(+t)}$, $a_{tt}^{(+i)} - \mathbb{E}a_{tt}^{(+i)}$, and $a_{jj}^{(+i,t)} - \mathbb{E}a_{jj}^{(+i,t)}$ of degree no less than three.

We will use Hölder's inequality. Take, for example, the monomial $a_{ij}a_{it}\left(a_{ii}^{(+t)} - \mathbb{E}a_{ii}^{(+t)}\right)$. We have

$$\mathbb{E}\left|a_{ij}a_{it}\left(a_{ii}^{(+t)} - \mathbb{E}a_{ii}^{(+t)}\right)\right| \leq \left(\mathbb{E}a_{ij}^2\right)^{1/2}\left(\mathbb{E}a_{it}^4\right)^{1/4}\left(\mathbb{E}\left(a_{ii}^{(+t)} - \mathbb{E}a_{ii}^{(+t)}\right)^4\right)^{1/4}. \tag{180}$$

On the other hand, for any $k \geq 2$

$$\mathbb{E}\left|a_{ij}\right|^k \equiv \mathbb{E}\left|\xi_i' W_{-i,j,t}^{-1}\xi_j/p\right|^k = \mathbb{E}\left|\frac{1}{2p}\left[\xi_i', \xi_j'\right]\begin{pmatrix} 0 & W_{-i,j,t}^{-1} \\ W_{-i,j,t}^{-1} & 0 \end{pmatrix}\begin{bmatrix} \xi_i \\ \xi_j \end{bmatrix}\right|^k.$$

Therefore, by Lemma 2.7 of Bai and Silverstein (1998),

$$\mathbb{E}\left|a_{ij}\right|^k \leq Kp^{-k/2}\left|\frac{1}{p}\mathbb{E}\operatorname{tr}W_{-i,j,t}^{-2}\right|^{k/2} \leq Kp^{-k/2},$$

where the latter inequality can be established using a slightly modified version of (163) and from Lemma 25. A similar inequality holds for $\mathbb{E}\left|a_{it}\right|^k$. Finally,

$$\mathbb{E}\left|a_{ii}^{(+t)} - \mathbb{E}a_{ii}^{(+t)}\right|^k \leq K\mathbb{E}\left|a_{ii}^{(+t)} - \operatorname{tr}W_{-i,j}^{-1}/p\right|^k + K\mathbb{E}\left|\operatorname{tr}W_{-i,j}^{-1}/p - \mathbb{E}\operatorname{tr}W_{-i,j}^{-1}/p\right|^k.$$

The first term on the right hand side of the above inequality is bounded by $Kp^{-k/2}$ similarly to $\mathbb{E}\left|a_{ij}\right|^k$. To bound the second term, we use the following decomposition

$$\operatorname{tr}W_{-i,j}^{-1}/p - \mathbb{E}\operatorname{tr}W_{-i,j}^{-1}/p = \sum_{s\neq i,j}\left(\mathbb{E}_s - \mathbb{E}_{s-1}\right)\operatorname{tr}\left(W_{-i,j}^{-1} - W_{-i,j,s}^{-1}\right)/p,$$

where $\mathbb{E}_s$ denotes the expectation conditional on $\varepsilon_{(1)}, ..., \varepsilon_{(s)}$, and $\mathbb{E}_0 \equiv \mathbb{E}$ denotes the unconditional expectation. SMW formula yields

$$\operatorname{tr}\left(W_{-i,j}^{-1} - W_{-i,j,s}^{-1}\right) = -\frac{\xi_s' W_{-i,j,s}^{-2}\xi_s/p}{b_{ps} + \xi_s' W_{-i,j,s}^{-1}\xi_s/p}.$$

Hence,

$$\mathbb{E}\left|\operatorname{tr}\left(W_{-i,j}^{-1} - W_{-i,j,s}^{-1}\right)\right|^k = \mathbb{E}\frac{\left|\xi_s' W_{-i,j,s}^{-2}\xi_s/p\right|^k}{\left|b_{ps} + \xi_s' W_{-i,j,s}^{-1}\xi_s/p\right|^k} \leq \mathbb{E}\left\|W_{-i,j,s}^{-1}\right\|^k \leq K,$$

where the last inequality follows from an inequality similar to (163) and from Lemma 25.

Now using the Burkholder inequality (see Bai and Silverstein (1998), Lemma 2.2), we get

$$\mathbb{E}\left|\operatorname{tr}W_{-i,j}^{-1}/p - \mathbb{E}\operatorname{tr}W_{-i,j}^{-1}/p\right|^k \leq KT_\gamma^{k/2}p^{-k},$$

and hence

$$\mathbb{E}\left|a_{ii}^{(+t)} - \mathbb{E}a_{ii}^{(+t)}\right|^k \leq K_\gamma p^{-k/2}.$$

Recalling (180), we obtain

$$\mathbb{E}\left|a_{ij}a_{it}\left(a_{ii}^{(+t)} - \mathbb{E}a_{ii}^{(+t)}\right)\right| \leq K_\gamma p^{-3/2}.$$

Similarly, expected value of monomials of order r are bounded by $K_\gamma p^{-r/2}$. We conclude that, for any $k > 0$,

$$\mathbb{E}\left|x_i x_j^2 x_t \frac{1}{p^2}\xi_i' W^{-1}\xi_j\xi_j' W^{-1}\xi_t\right| \leq K_\gamma p^{-3/2} + o(p^{-k})$$

as $p \to \infty$. Since there are less than T_γ^3 such terms in the sum defining M_2, we have

$$\mathbb{E}M_2 \leq K_\gamma p^{3/2} \tag{181}$$

for all sufficiently large p.

Analysis of M_3. By (175)

$$\begin{aligned}
& x_i x_j x_s x_t \frac{1}{p^2}\xi_i' W^{-1}\xi_j \xi_s' W^{-1}\xi_t \\
= & \left(\frac{1}{p}\xi_i' W_{-i,j}^{-1}\xi_j \frac{x_j b_{pj}}{b_{pj}+\frac{1}{p}\xi_j' W_{-i,j}^{-1}\xi_j}\frac{x_i b_{pi}}{b_{pi}+\frac{1}{p}\xi_i' W_{-i}^{-1}\xi_i}\right) \\
& \times \left(\frac{1}{p}\xi_s' W_{-s,t}^{-1}\xi_t \frac{x_s b_{ps}}{b_{ps}+\frac{1}{p}\xi_s' W_{-s,t}^{-1}\xi_s}\frac{x_t b_{pt}}{b_{pt}+\frac{1}{p}\xi_t' W_{-t}^{-1}\xi_t}\right).
\end{aligned}$$

Further,

$$\begin{aligned}
\frac{1}{p}\xi_i' W_{-i,j}^{-1}\xi_j &= \frac{1}{p}\xi_i' W_{-i,j,t}^{-1}\xi_j - \frac{\frac{1}{p}\xi_i' W_{-i,j,t}^{-1}\xi_t \frac{1}{p}\xi_t' W_{-i,j,t}^{-1}\xi_j}{b_{pt}+\frac{1}{p}\xi_t' W_{-i,j,t}^{-1}\xi_t}, \\
\frac{1}{p}\xi_s' W_{-s,t}^{-1}\xi_t &= \frac{1}{p}\xi_s' W_{-i,s,t}^{-1}\xi_t - \frac{\frac{1}{p}\xi_s' W_{-i,s,t}^{-1}\xi_i \frac{1}{p}\xi_i' W_{-i,s,t}^{-1}\xi_t}{b_{pi}+\frac{1}{p}\xi_i' W_{-i,s,t}^{-1}\xi_i}.
\end{aligned}$$

Denoting $\frac{1}{p}\xi_i' W_{-i,j,t}^{-1}\xi_j$ as a_{ij}^{+s}, $\frac{1}{p}\xi_t' W_{-t}^{-1}\xi_t$ as $a_{tt}^{+i,j,s}$, etc., we have

$$x_i x_j x_s x_t \frac{1}{p^2}\xi_i' W^{-1}\xi_j \xi_s' W^{-1}\xi_t = \prod_{r=1}^{6} M_{3r}$$

where

$$\begin{aligned}
M_{31} &= a_{ij}^{+s} - \frac{a_{it}^{+s}a_{jt}^{+s}}{b_{pt}+a_{tt}^{+s}}, M_{32} = a_{st}^{+j} - \frac{a_{is}^{+j}a_{it}^{+j}}{b_{pi}+a_{ii}^{+j}}, \\
M_{33} &= \frac{x_j b_{pj}}{b_{pj}+a_{jj}^{+s,t}}, M_{34} = \frac{x_i b_{pi}}{b_{pi}+a_{ii}^{+j,s,t}}, \text{ and} \\
M_{35} &= \frac{x_s b_{ps}}{b_{ps}+a_{ss}^{+i,j}}, M_{36} = \frac{x_t b_{pt}}{b_{pt}+a_{tt}^{+i,j,s}}.
\end{aligned}$$

Consider now the following identities

$$M_{31} = a_{ij}^{+s} - \frac{a_{it}^{+s}a_{jt}^{+s}}{b_{pt}+\mathbb{E}a_{tt}^{+s}} + \frac{a_{it}^{+s}a_{jt}^{+s}\left(a_{tt}^{+s}-\mathbb{E}a_{tt}^{+s}\right)}{\left(b_{pt}+\mathbb{E}a_{tt}^{+s}\right)^2} - \frac{a_{it}^{+s}a_{jt}^{+s}\left(a_{tt}^{+s}-\mathbb{E}a_{tt}^{+s}\right)^2}{\left(b_{pt}+\mathbb{E}a_{tt}^{+s}\right)^2\left(b_{pt}+a_{tt}^{+s}\right)}, \tag{182}$$

$$M_{32} = a_{st}^{+j} - \frac{a_{is}^{+j}a_{it}^{+j}}{b_{pi}+\mathbb{E}a_{ii}^{+j}} + \frac{a_{is}^{+j}a_{it}^{+j}\left(a_{ii}^{+j}-\mathbb{E}a_{ii}^{+j}\right)}{\left(b_{pi}+\mathbb{E}a_{ii}^{+j}\right)^2} - \frac{a_{is}^{+j}a_{it}^{+j}\left(a_{ii}^{+j}-\mathbb{E}a_{ii}^{+j}\right)^2}{\left(b_{pi}+\mathbb{E}a_{ii}^{+j}\right)^2\left(b_{pi}+a_{ii}^{+j}\right)}, \tag{183}$$

$$M_{33} = \sum_{r=0}^{2}\frac{(-1)^r x_j b_{pj}\left(a_{jj}^{+s,t}-\mathbb{E}a_{jj}^{+s,t}\right)^r}{\left(b_{pj}+\mathbb{E}a_{jj}^{+s,t}\right)^{1+r}} - \frac{x_j b_{pj}\left(a_{jj}^{+s,t}-\mathbb{E}a_{jj}^{+s,t}\right)^3}{\left(b_{pj}+\mathbb{E}a_{jj}^{+s,t}\right)^3\left(b_{pj}+a_{jj}^{+s,t}\right)}, \tag{184}$$

$$M_{34} = \sum_{r=0}^{2}\frac{(-1)^r x_i b_{pi}\left(a_{ii}^{+j,s,t}-\mathbb{E}a_{ii}^{+j,s,t}\right)^r}{\left(b_{pi}+\mathbb{E}a_{ii}^{+j,s,t}\right)^{1+r}} - \frac{x_i b_{pi}\left(a_{ii}^{+j,s,t}-\mathbb{E}a_{ii}^{+j,s,t}\right)^3}{\left(b_{pi}+\mathbb{E}a_{ii}^{+j,s,t}\right)^3\left(b_{pi}+a_{ii}^{+j,s,t}\right)}, \tag{185}$$

$$M_{35} = \sum_{r=0}^{2}\frac{(-1)^r x_s b_{ps}\left(a_{ss}^{+i,j}-\mathbb{E}a_{ss}^{+i,j}\right)^r}{\left(b_{ps}+\mathbb{E}a_{ss}^{+i,j}\right)^{1+r}} - \frac{x_s b_{ps}\left(a_{ss}^{+i,j}-\mathbb{E}a_{ss}^{+i,j}\right)^3}{\left(b_{ps}+\mathbb{E}a_{ss}^{+i,j}\right)^3\left(b_{ps}+a_{ss}^{+i,j}\right)}, \tag{186}$$

and

$$M_{36} = \sum_{r=0}^{2} \frac{(-1)^r x_t b_{pt} \left(a_{tt}^{+i,j,s} - \mathbb{E}a_{tt}^{+i,j,s}\right)^r}{\left(b_{pt} + \mathbb{E}a_{tt}^{+i,j,s}\right)^{1+r}} - \frac{x_t b_{pt} \left(a_{tt}^{+i,j,s} - \mathbb{E}a_{tt}^{+i,j,s}\right)^3}{\left(b_{pt} + \mathbb{E}a_{tt}^{+i,j,s}\right)^3 \left(b_{pt} + a_{tt}^{+i,j,s}\right)}. \tag{187}$$

Further, note that

$$a_{ss}^{+i,j} = a_{ss}^{+j} - \frac{\left(a_{is}^{+j}\right)^2}{b_{pi} + a_{ii}^{+j}} = a_{ss}^{+j} - \frac{\left(a_{is}^{+j}\right)^2}{b_{pi} + \mathbb{E}a_{ii}^{+j}} + \frac{\left(a_{is}^{+j}\right)^2 \left(a_{ii}^{+j} - \mathbb{E}a_{ii}^{+j}\right)}{\left(b_{pi} + \mathbb{E}a_{ii}^{+j}\right)\left(b_{pi} + a_{ii}^{+j}\right)} \tag{188}$$

and

$$a_{tt}^{+i,j,s} = a_{tt}^{+j,s} - \frac{\left(a_{it}^{+j,s}\right)^2}{b_{pi} + a_{ii}^{+j,s}} = a_{tt}^{+j,s} - \frac{\left(a_{it}^{+j,s}\right)^2}{b_{pi} + \mathbb{E}a_{ii}^{+j,s}} + \frac{\left(a_{it}^{+j,s}\right)^2 \left(a_{ii}^{+j,s} - \mathbb{E}a_{ii}^{+j,s}\right)}{\left(b_{pi} + \mathbb{E}a_{ii}^{+j,s}\right)\left(b_{pi} + a_{ii}^{+j,s}\right)}. \tag{189}$$

Using (188) and (189) in the terms of (186) and (187) corresponding to $r = 1$, we obtain

$$\begin{aligned} M_{35} \quad = \quad & \frac{x_s b_{ps}}{b_{ps} + \mathbb{E}a_{ss}^{+i,j}} - \frac{x_s b_{ps} \left(\mathbb{E}a_{ss}^{+j} - \mathbb{E}a_{ss}^{+i,j}\right)}{\left(b_{ps} + \mathbb{E}a_{ss}^{+i,j}\right)^2} - \frac{x_s b_{ps} \left(a_{ss}^{+j} - \mathbb{E}a_{ss}^{+j}\right)}{\left(b_{ps} + \mathbb{E}a_{ss}^{+i,j}\right)^2} \\ & + \frac{x_s b_{ps} \left(a_{is}^{+j}\right)^2}{\left(b_{ps} + \mathbb{E}a_{ss}^{+i,j}\right)^2 \left(b_{pi} + \mathbb{E}a_{ii}^{+j}\right)} + \frac{x_s b_{ps} \left(a_{ss}^{+i,j} - \mathbb{E}a_{ss}^{+i,j}\right)^2}{\left(b_{ps} + \mathbb{E}a_{ss}^{+i,j}\right)^3} \\ & - \frac{x_s b_{ps} \left(a_{is}^{+j}\right)^2 \left(a_{ii}^{+j} - \mathbb{E}a_{ii}^{+j}\right)}{\left(b_{ps} + \mathbb{E}a_{ss}^{+i,j}\right)^2 \left(b_{pi} + \mathbb{E}a_{ii}^{+j}\right)\left(b_{pi} + a_{ii}^{+j}\right)} - \frac{x_s b_{ps} \left(a_{ss}^{+i,j} - \mathbb{E}a_{ss}^{+i,j}\right)^3}{\left(b_{ps} + \mathbb{E}a_{ss}^{+i,j}\right)^3 \left(b_{ps} + a_{ss}^{+i,j}\right)} \end{aligned} \tag{190}$$

and

$$\begin{aligned} M_{36} \quad = \quad & \frac{x_t b_{pt}}{b_{pt} + \mathbb{E}a_{tt}^{+i,j,s}} - \frac{x_t b_{pt} \left(\mathbb{E}a_{tt}^{+j,s} - \mathbb{E}a_{tt}^{+i,j,s}\right)}{\left(b_{pt} + \mathbb{E}a_{tt}^{+i,j,s}\right)^2} - \frac{x_t b_{pt} \left(a_{tt}^{+j,s} - \mathbb{E}a_{tt}^{+j,s}\right)}{\left(b_{pt} + \mathbb{E}a_{tt}^{+i,j,s}\right)^2} \\ & + \frac{x_t b_{pt} \left(a_{it}^{+j,s}\right)^2}{\left(b_{pt} + \mathbb{E}a_{tt}^{+i,j,s}\right)^2 \left(b_{pi} + \mathbb{E}a_{ii}^{+j,s}\right)} + \frac{x_t b_{pt} \left(a_{tt}^{+i,j,s} - \mathbb{E}a_{tt}^{+i,j,s}\right)^2}{\left(b_{pt} + \mathbb{E}a_{tt}^{+i,j,s}\right)^3} \\ & - \frac{x_t b_{pt} \left(a_{it}^{+j,s}\right)^2 \left(a_{ii}^{+j,s} - \mathbb{E}a_{ii}^{+j,s}\right)}{\left(b_{pt} + \mathbb{E}a_{tt}^{+i,j,s}\right)^2 \left(b_{pi} + \mathbb{E}a_{ii}^{+j,s}\right)\left(b_{pi} + a_{ii}^{+j,s}\right)} - \frac{x_t b_{pt} \left(a_{tt}^{+i,j,s} - \mathbb{E}a_{tt}^{+i,j,s}\right)^3}{\left(b_{pt} + \mathbb{E}a_{tt}^{+i,j,s}\right)^3 \left(b_{pt} + a_{tt}^{+i,j,s}\right)}. \end{aligned} \tag{191}$$

Using identities (182-185) and (190-191), we represent the product $\prod_{r=1}^{6} M_{3r}$ in the form of a weighted sum of monomials in a_{ij}^{+s}, a_{it}^{+s}, a_{jt}^{+s}, $a_{tt}^{+s} - \mathbb{E}a_{tt}^{+s}$, a_{st}^{+j}, a_{is}^{+j}, a_{it}^{+j}, $a_{ii}^{+j} - \mathbb{E}a_{ii}^{+j}$, etc. A somewhat lengthy but straightforward inspection reveals that the expectation of all the monomial terms of degree less than five is zero. Take, for example the monomial term

$$\begin{aligned} & a_{ij}^{+s} a_{st}^{+j} \frac{x_j b_{pj} \left(a_{jj}^{+s,t} - \mathbb{E}a_{jj}^{+s,t}\right)}{\left(b_{pj} + \mathbb{E}a_{jj}^{+s,t}\right)^2} \frac{x_i b_{pi} \left(a_{ii}^{+j,s,t} - \mathbb{E}a_{ii}^{+j,s,t}\right)}{\left(b_{pi} + \mathbb{E}a_{ii}^{+j,s,t}\right)^2} \\ & \times \left(\frac{1}{b_{ps} + \mathbb{E}a_{ss}^{+i,j}} - \frac{x_s b_{ps} \left(\mathbb{E}a_{ss}^{+j} - \mathbb{E}a_{ss}^{+i,j}\right)}{\left(b_{ps} + \mathbb{E}a_{ss}^{+i,j}\right)^2} \right) \left(\frac{x_t b_{pt}}{b_{pt} + \mathbb{E}a_{tt}^{+i,j,s}} - \frac{x_t b_{pt} \left(\mathbb{E}a_{tt}^{+j,s} - \mathbb{E}a_{tt}^{+i,j,s}\right)}{\left(b_{pt} + \mathbb{E}a_{tt}^{+i,j,s}\right)^2} \right), \end{aligned}$$

which is obtained by taking the product of the first order terms in (182-185) and zeroth order terms in (190-191). We have

$$\mathbb{E}\left[a_{ij}^{+s}a_{st}^{+j}\left(a_{jj}^{+s,t}-\mathbb{E}a_{jj}^{+s,t}\right)\left(a_{ii}^{+j,s,t}-\mathbb{E}a_{ii}^{+j,s,t}\right)\right]=0$$

because the expression under the expectation can be represented in the form of a weighted sum of monomials in the components of vector ξ_i of order three and one only. The expectation of such monomials, conditional on $\xi_j,\xi_s,\xi_t,W_{-i,j,t}^{-1},W_{-i,s,t}^{-1},W_{-i,j}^{-1}$, and $W_{-i,j}^{-1}$ is zero.

By the same logic as in the above subsection, the expectation of the monomial terms of order five and more in the expansion of $\prod_{r=1}^{6}M_{3r}$ is bounded by $K_\gamma p^{-5/2}$. Since there are no more than T_γ^4 such terms, we have

$$\mathbb{E}M_3\leq K_\gamma p^{3/2} \tag{192}$$

for all sufficiently large p.

Combining (178), (181), and (192) yields

$$\mathbb{E}\left(\frac{p^2}{T_\gamma^2}\sum_{i\neq j}x_ix_j\frac{1}{p}\xi_i'W^{-1}\xi_j\right)^2\leq K_\gamma p^{3/2}$$

or all sufficiently large p. By Markov's inequality,

$$\Pr\left(\left|\frac{p^2}{T_\gamma^2}\sum_{i\neq j}x_ix_j\frac{1}{p}\xi_i'W^{-1}\xi_j\right|\geq p^{4/5}\right)\leq K_\gamma p^{-1/10}. \tag{193}$$

This inequality and (173) imply that there exist absolute constants $K_1>0$ and $K_2>0$ such that

$$\Pr\left(\frac{p}{K_2\gamma^2}\leq y'\mathcal{M}_1y\leq\frac{p}{K_1\gamma^2}\right)\to 1$$

as $p\to\infty$.

Comparison to $y'\mathcal{M}_1\mathcal{P}_2\mathcal{M}_1y$. Represent y as $y=y_t+y_{-t}$, where

$$y_{tj}=\begin{cases}y_j \text{ for } j\leq t\\ 0 \text{ otherwise}\end{cases}.$$

We will choose the value of the integer t later.

We have

$$y'\mathcal{M}_1\mathcal{P}_2\mathcal{M}_1y\leq 2y_t'\mathcal{M}_1\mathcal{P}_2\mathcal{M}_1y_t+2y_{-t}'\mathcal{M}_1\mathcal{P}_2\mathcal{M}_1y_{-t}$$

Now, recall that $\mathcal{P}_1=I_T-\mathcal{M}_1$. Therefore,

$$y_{-t}'\mathcal{M}_1\mathcal{P}_2\mathcal{M}_1y_{-t}=y_{-t}'\mathcal{P}_2y_{-t}-2y_{-t}'\mathcal{P}_1\mathcal{P}_2y_{-t}+y_{-t}'\mathcal{P}_1\mathcal{P}_2\mathcal{P}_1y_{-t},$$

and hence,

$$y_{-t}'\mathcal{M}_1\mathcal{P}_2\mathcal{M}_1y_{-t}\leq y_{-t}'\mathcal{P}_2y_{-t}+3\left\|\mathcal{P}_1\mathcal{P}_2\right\|\left\|y_{-t}\right\|^2. \tag{194}$$

Further, clearly

$$y_t'\mathcal{M}_1\mathcal{P}_2\mathcal{M}_1y_t\leq y_t'\mathcal{M}_1y_t. \tag{195}$$

Analysis of $y_t'\mathcal{M}_1 y_t$. By definition

$$y_t'\mathcal{M}_1 y_t = \sum_{i=1}^{t} x_i^2 \left(\Delta_{ii}^{-1} - p\xi_i' W^{-1}\xi_i / T_\gamma^2\right) - (p/T_\gamma)^2 \sum_{i\neq j}^{t} x_i x_j \xi_i' W^{-1}\xi_j / p.$$

Similarly to (172), we have w.ow.p.,

$$\sum_{i=1}^{t} \frac{2s_i^2 T_\gamma^2/p^2}{b_{pi}+K_2} \leq \sum_{i=1}^{t} x_i^2 \left(\Delta_{ii}^{-1} - p\xi_i' W^{-1}\xi_i / T_\gamma^2\right) \leq \sum_{i=1}^{t} \frac{2s_i^2 T_\gamma^2/p^2}{b_{pi}+K_1}. \tag{196}$$

Let t be an even integer. Then,

$$\sum_{i=1}^{t} \frac{2s_i^2 T_\gamma^2/p^2}{b_{pi}+K} = \sum_{j=1}^{t/2} \frac{1}{k_K - \cos\frac{2\pi j}{T_\gamma}} = \frac{T_\gamma}{2\pi}\int_0^{\pi t/T_\gamma} \frac{\mathrm{d}\varphi}{k_K - \cos\varphi} + O(1)$$

as $p \to \infty$, where $k_K = 1 + \frac{1}{2}Kp^2/T_\gamma^2$. According to Gradshteyn and Ryzhik (2000, 2.553 (3)),

$$\frac{T_\gamma}{2\pi}\int_0^{\pi t/T_\gamma} \frac{\mathrm{d}\varphi}{k_K - \cos\varphi} = \frac{T_\gamma}{\pi\sqrt{k_K^2-1}} \arctan\left(\frac{2\sqrt{k_K^2-1}}{Kp^2/T_\gamma^2}\tan\frac{\pi t}{2T_\gamma}\right)$$

Choosing t to be the even integer closest to $p\,(p/T_\gamma)^{1/4}$, we obtain (using linear approximations of tan and arctan around zero)

$$\lim_{p/T_\gamma\to 0} \frac{K\,(p/T_\gamma)^{3/4}}{T_\gamma}\left(\frac{T_\gamma}{2\pi}\int_0^{\pi t/T_\gamma} \frac{\mathrm{d}\varphi}{k_K - \cos\varphi}\right) = 1,$$

so that for any $\delta > 0$, there exists $\gamma_\delta > 0$ such that for all $\gamma < \gamma_\delta$, we have

$$\sum_{i=1}^{t} \frac{2s_i^2 T_\gamma^2/p^2}{b_{pi}+K} \leq (1+\delta)\frac{T_\gamma}{K\,(p/T_\gamma)^{3/4}}.$$

Combining this with (196) yields

$$\sum_{i=1}^{t} x_i^2 \left(\Delta_{ii}^{-1} - \frac{p}{T_\gamma^2}\xi_i' W^{-1}\xi_i\right) \leq \frac{T_\gamma}{K\,(p/T_\gamma)^{3/4}}$$

for an absolute constant $K > 0$ and all sufficiently small γ, w.ow.p. Since, similarly to (193), we have

$$\Pr\left(\left|\frac{p^2}{T_\gamma^2}\sum_{i\neq j}^{t} x_i x_j \frac{1}{p}\xi_i' W^{-1}\xi_j\right| \geq p^{4/5}\right) \leq K_\gamma p^{-1/10},$$

we conclude that, for all sufficiently small γ,

$$\Pr\left(y_t'\mathcal{M}_1 y_t \leq Kp/\gamma^{7/4}\right) \to 1 \tag{197}$$

as $p \to \infty$.

Analysis of $y_{-t}'\mathcal{P}_2 y_{-t} + 3\,\|\mathcal{P}_1\mathcal{P}_2\|\,\|y_{-t}\|^2$. By definition, we have $\|y_{-t}\|^2 = \sum_{j=t+1}^{\bar{T}_\gamma} y_j^2$. Recall that

$$y = -\frac{1}{\sqrt{2}}\left(1, \frac{\sin\omega_1}{1-\cos\omega_1}, 1, \frac{\sin\omega_2}{1-\cos\omega_2}, ...\right)'.$$

Therefore,

$$\sum_{j=t+1}^{\bar{T}_\gamma} y_j^2 = \frac{1}{4}\left(\bar{T}_\gamma - t\right) + \frac{1}{2}\sum_{j=t/2+1}^{\bar{T}_\gamma/2}\left(\frac{\sin\omega_j}{1-\cos\omega_j}\right)^2 \le \frac{1}{4}\bar{T}_\gamma + \sum_{j=t/2+1}^{\bar{T}_\gamma/2}\frac{1}{1-\cos\frac{2\pi j}{T_\gamma}}.$$

For t which is the even integer closest to $p\,(p/T_\gamma)^{1/4}$, we have

$$\sum_{j=t/2+1}^{\bar{T}_\gamma/2}\frac{1}{1-\cos\frac{2\pi j}{T_\gamma}} = \frac{T_\gamma}{2\pi}\int_{\pi(p/T_\gamma)^{5/4}}^{\pi}\frac{\mathrm{d}x}{1-\cos x} + O(1) = \frac{T_\gamma}{2\pi}\cot\left[\frac{1}{2}\pi\,(p/T_\gamma)^{5/4}\right] + O(1)$$

as $p\to\infty$, so that

$$\|y_{-t}\|^2 \le KT_\gamma/\,(p/T_\gamma)^{5/4}$$

for all sufficiently small γ and some absolute constant $K>0$, as $p\to\infty$.

Since, by Theorem 1 of Onatski and Wang (2017a), $\|\mathcal{P}_1\mathcal{P}_2\| \le K\sqrt{\gamma}$ as $p\to\infty$, we have

$$3\,\|\mathcal{P}_1\mathcal{P}_2\|\,\|y_{-t}\|^2 \le KT_\gamma/\,(p/T_\gamma)^{3/4}\,.$$

On the other hand, since $\mathcal{P}_2$ is a projection on a random p-dimensional subspace of $\mathbb{R}^{\bar{T}_\gamma}$,

$$y_{-t}'\mathcal{P}_2 y_{-t} \le K\,\|y_{-t}\|^2\,\frac{p}{T_\gamma} \le KT_\gamma/\,(p/T_\gamma)^{1/4}$$

with high probability as $p\to\infty$. Hence,

$$y_{-t}'\mathcal{P}_2 y_{-t} + 3\,\|\mathcal{P}_1\mathcal{P}_2\|\,\|y_{-t}\|^2 \le KT_\gamma/\,(p/T_\gamma)^{3/4} \qquad (198)$$

with high probability as $p\to\infty$.

Combining (197) and (198) yields

$$y'\mathcal{M}_1\mathcal{P}_2\mathcal{M}_1 y \le Kp/\gamma^{7/4}$$

with high probability for all sufficiently small γ as $p\to\infty$. To summarize,

$$\frac{y'\mathcal{M}_1\mathcal{P}_2\mathcal{M}_1 y}{y'\mathcal{M}_1 y} \le \frac{Kp/\gamma^{7/4}}{p/\gamma^2} \le K\gamma^{1/4}$$

with high probability, for all sufficiently small γ as $p\to\infty$.

4 Monte Carlo

In this section, we explore the sensitivity of the empirical distribution of the squared canonical correlations to the nuisance parameters Ψ and Γ. All figures given below show the 5-th and 95-th percentiles of the MC distributions of the squared canonical correlations, λ_{p+1-i}, (solid lines) plotted against the $100(i-1/2)/p$ quantiles of the corresponding Wachter limit. The figures correspond to the entries of ε having Student's $t(3)$ distribution. The results for Gaussian and centered $\chi^2(1)$ distributions are very similar.

4.1 Sensitivity to Ψ

We simulate the data generating process

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1}\Gamma_i\Delta X_{t-i} + \Psi F_t + \varepsilon_t,$$

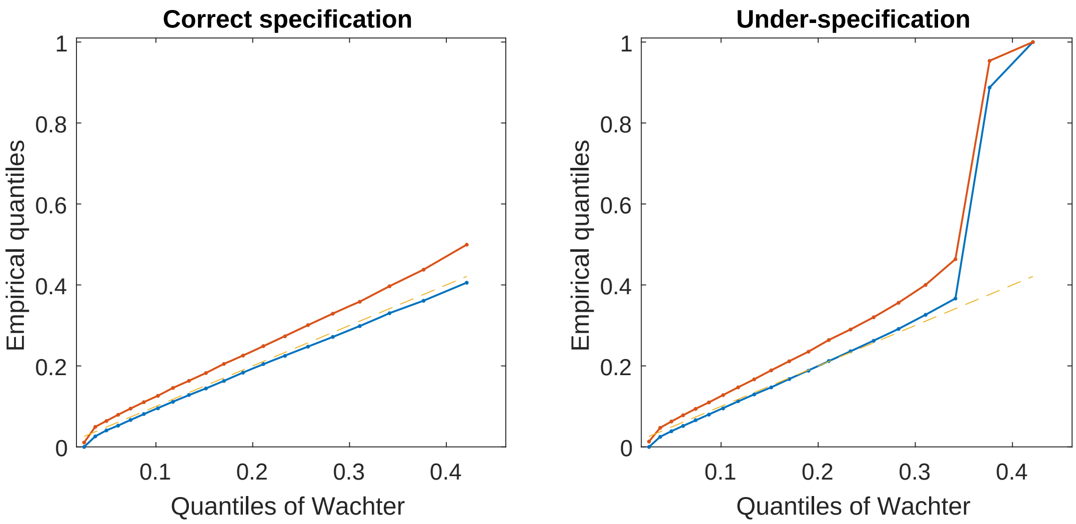

Figure 1: The 5-th and 95-th quantiles of the MC distribution of λ_{p+1-i} plotted against $100(i-1/2)/p$ quantiles of $W_{1/10}(\lambda)$. $(p,T)=(20,200)$. The data generating process has a linear deterministic trend with i.i.d. $N(0,1)$ coefficients collected in matrix Ψ (the first column - intercept, the second column - slope). Right panel corresponds to the case where the trend is omitted from the econometrician's model.

where Ψ is a $p\times 2$ matrix with i.i.d. $N(0,\sigma^2)$ entries, and $F_t=(1,t)'$. Matrices Π and Γ_i are set to be zero. That is, the data components are random walks with heterogeneous linear time trends. The initial values are zero, and the sample size is $(p,T)=(20,200)$.

The left panel of Figure 1 corresponds to the case where the econometrician's model

$$\Delta X_t=\Pi X_{t-1}+\Phi D_t+\varepsilon_t \tag{199}$$

is correctly specified. That is, $D_t=(1,t)'$ and Φ is not constrained to be zero. The right panel corresponds to an under-specification, where the deterministic terms are mistakenly omitted from (199). Parameter σ^2 is set to one.

We see that omitting two deterministic terms (the constant and the time index) leads to a deviation of the two largest squared canonical correlations from the 45° line. This is, perhaps, not surprising because under the mis-specification the canonical correlations are based on changes and levels of the raw data, as opposed to the residuals from the regressions on the deterministic terms. Therefore, the changes and levels contain two deterministic components resulting in the two largest canonical correlations being large.

The degree of the deviation of the two largest squared canonical correlations depends on the value of σ. When σ decreases, the deviation becomes smaller, and entirely disappears when $\sigma=0$. When we increase σ to 1.9, the time trend starts to dominate the data so much that matrix S_{11} becomes very poorly scaled under the mis-specification, and the numerical results become inaccurate.

When the model is correctly specified, the MC quantiles of Wachter plots lie close to the 45° line. However, in contrast to Figure OW4, the line is closer to the 5-th and is further away from the 95-th MC quantile. This phenomenon does not disappear even when $\sigma=0$.

4.2 Sensitivity to Γ

First, we generate data with $\Gamma_1=\sigma vv'$, where v is a p-dimensional vector uniformly distributed on the unit sphere, and $\sigma\in(0,1)$ so that the generated process does not become $I(2)$. We set Π, Ψ, and the initial values to zero. The samples size is $(p,T)=(20,200)$. The econometrician's model is (199) with $\Phi=0$.

Figure 2 reports results for $\sigma=0.1,0.4,0.7$, and 0.9. As σ increases, the MC distribution of the largest squared canonical correlation shifts upwards and away from the corresponding quantile of the Wachter limit.

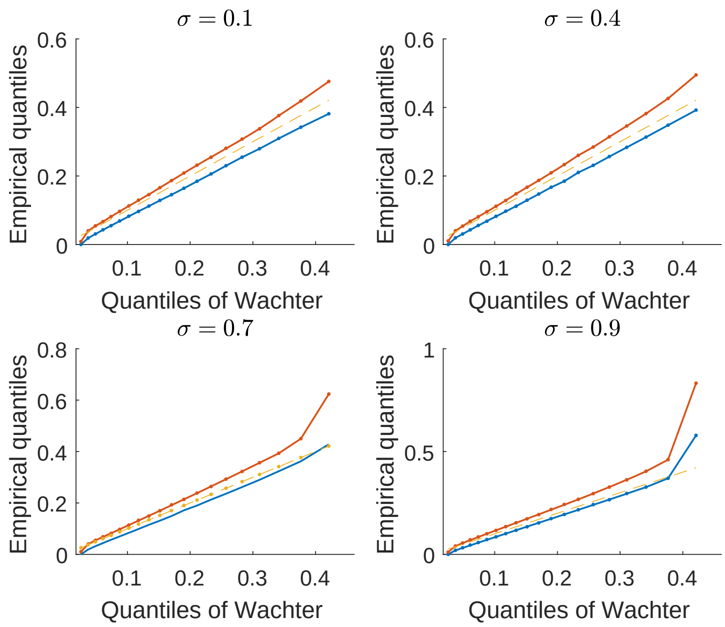

Figure 2: The 5-th and 95-th quantiles of the MC distribution of λ_{p+1-i} plotted against $100(i-1/2)/p$ quantiles of $W_{1/10}(\lambda)$. $(p,T) = (20,200)$. The data generating process has $\Gamma_1 = \sigma vv'$, where v is uniformly distributed on the unit sphere.

The deviation becomes clearly noticeable for $\sigma = 0.7$. The other squared canonical correlations remain close to the Wachter limit.

We repeat this MC experiment with $\Gamma_1 = \sigma vv'$, where v is a $p \times 2$ matrix distributed as the first two columns of a random orthogonal matrix (uniformly distributed over the orthogonal group). The results are reported in Figure 3. Now the two largest squared canonical correlations deviate from the corresponding quantiles of the Wachter limit for relatively large σ. The reason is the presence in ΔX_t and X_{t-1} of two persistent and related stochastic components, $v'\Delta X_t$ and $v'X_{t-1}$.

Note that the econometrician's model is still (199). Hence, it omits the lag ΔX_{t-1}, and the canonical correlations are computed using the raw data, not regressed on ΔX_{t-1}.

Overall, we see that making Ψ or Γ non-zero does influence the empirical distribution of the squared canonical correlations when the econometrician's model is misspecified. However, this influence is mostly confined to a few of the largest squared canonical correlations. In particular, for low-rank non-zero Ψ or Γ, the empirical distribution of the squared canonical correlations remains close to the Wachter limit in terms of the Lévy distance.

References

[1] Abramowitz, M. and I.A. Stegun, eds., (1970) *Handbook of Mathematical Functions,* Dover, New York.

[2] Bai, Z.D., J. Hu, G. Pan, and W. Zhou (2015) "Convergence of the Empirical Spectral Distribution Function of Beta Matrices," *Bernoulli* 21, 1538-1574.

[3] Bai, Z. D. and J. W. Silverstein, (1998) "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices," *The Annals of Probability* 26, 316-345.

[4] Bai, Z. D. and J. W. Silverstein, (2010) *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer Verlag, New York.

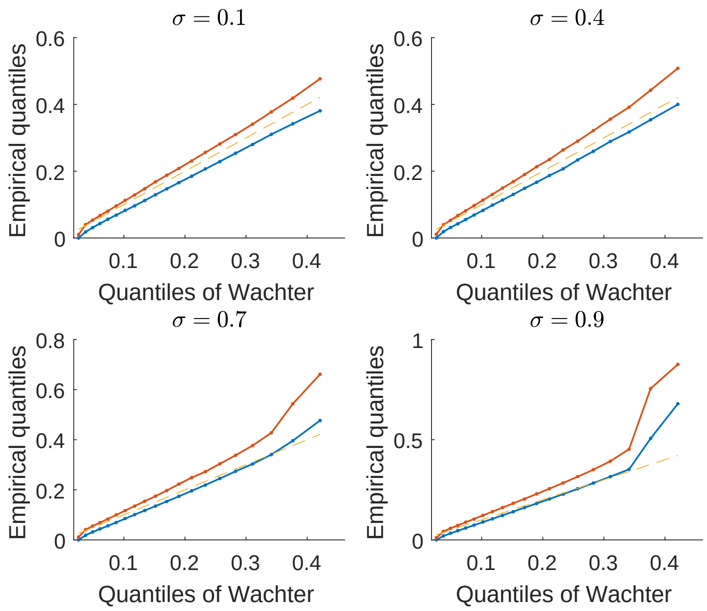

Figure 3: Same as Figure 2, except $\Gamma_1 = \sigma v v'$, where v is a $p \times 2$ matrix distributed as the first two columns of a random orthogonal matrix.

[5] Bai, Z. D. and Yin, Y. Q. (1993) "Limit of the smallest eigenvalue of a large dimensional sample covariance matrix," *Annals of Probability* 21, 1275-1294.

[6] Bentkus, V., Götze, F., Paulauskas, V., and Račkauskas, A. (2000) "The accuracy of Gaussian Approximation in Banach Spaces," in Prohorov and Statulevičius (eds.) *Limit Theorems of Probability Theory,* Springer.

[7] Chen, Z. and J. J. Dongarra, (2005) "Condition number of Gaussian random matrices," *SIAM J. Matrix. Anal. Appl.* 27, 603-620.

[8] Chatterjee, S. (2006) "A Generalization of the Lindeberg Principle," *The Annals of Probability* 34, 2061-76.

[9] Davidson, K.R. and S.J. Szarek (2001) Local operator theory, random matrices and Banach spaces, in: W.B. Johnson and J. Lindenstrauss (Eds.), Handbook of the Geometry of Banach Spaces, Vol. I, North-Holland, Amsterdam, 2001, 317–366.

[10] Geman, S. (1980) "A limit theorem for the norm of random matrices," *Annals of Probability* 8, 252–261.

[11] Geronimo, J.S. and T.P. Hill (2003) "Necessary and sufficient condition that the limit of Stieltjes transforms is a Stieltjes transform," *Journal of Approximation Theory* 121, 54-60.

[12] Golub, G.H. and C. F. Van Loan (1996) *Matrix Computations*, The John Hopkins University Press.

[13] Gradshteyn, I.S., and Ryzhik, I.M. (2000) *Table of Integrals, Series, and Products.* Sixth edition. Academic Press.

[14] Horn, R.A. and C.R. Johnson (1985) *Matrix Analysis*, Cambridge University press.

[15] Johansen, S. (1991) "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica* 59, 1551-1580.

[16] Johansen, S. (1995) *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press.

[17] Latala, R. (2004) "Some Estimates of Norms of Random Matrices," *Proceedings of the American Mathematical Society* 133, 1273-1282.

[18] Laurent, B. and Massart, P. (2000) "Adaptive Estimation of a Quadratic Functional by Model Selection," *The Annals of Statistics* 28, 1302-1338.

[19] Ledoux, M. (2000) *The Concentration of Measure Phenomenon*, the American Mathematical Society.

[20] Olver, F.W.J. (1997) *Asymptotics and Special Functions*, A K Peters.

[21] Onatski, A. and Wang, C. (2017) "Alternative asymptotics for cointegration tests in large VARs," manuscript, University of Cambridge.

[22] Onatski, A. and Wang, C. (2017a) "Extreme canonical correlations and high-dimensional cointegration analysis," manuscript, University of Cambridge.

[23] Silverstein, J.W. (1985) "The smallest eigenvalue of a large dimensional Wishart matrix," *Annals of Probability* 13, 1364-1368.

[24] Silverstein, J.W. and Z.D. Bai (1995) "On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices," *Journal of Multivariate Analysis* 54, 175–192.

[25] Silverstein, J.W., and Choi, S.I. (1995) "Analysis of the limiting spectral distribution of large dimensional random matrices," Journal of Multivariate Analysis 54, 295–309.

[26] Tao, T. and Vu, V. (2011) "Random Matrices: Universality of Local Eigenvalue Statistics," *Acta Mathematica* 206, 127–204.

[27] Titchmarsh, E. C. (1939). *The Theory of Functions*, second edition. Oxford University Press, London.

[28] Wachter, K. W. (1980). "The limiting empirical measure of multiple discriminant ratios," *Annals of Statistics* 8, 937–957.

[29] Yang, Y. and G. Pan (2012) "The convergence of the empirical distribution of canonical correlation coefficients," *Electronic Journal of Probability* 64, 1-13.