Supplementary Material: Computationally efficient banding of large covariance matrices for ordered data and connections to banding the inverse Cholesky factor

Y. Wang
wang.yanpin@scrippshealth.org
Scripps Health
La Jolla, CA 92037

M. J. Daniels
mjdaniels@austin.utexas.edu
University of Texas at Austin
Austin, TX 78712

1 STATEMENT OF SOME ESSENTIAL RESULTS

Result S1 (T.W.Anderson, 1984): Let

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

be a square matrix, where Σ_{11}, Σ_{22} are square submatrices with non-zero determinants. We denote $\Sigma_{11\bullet 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ and $\Sigma_{22\bullet 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$.
Then

a)
$$det(\Sigma) = det(\Sigma_{11})det(\Sigma_{22\bullet 1}),$$

b)
$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{11\bullet 2}^{-1} & -\Sigma_{11\bullet 2}^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11\bullet 2}^{-1} & \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11\bullet 2}^{-1}\Sigma_{12}\Sigma_{22}^{-1} \end{pmatrix},$$

with similar results for $\Sigma_{11\bullet 2}$ and Σ_{22}.

Result S2 (Joe, 2006): For $-1 < \pi_{j,k} < 1$, $j, k \in \{1, 2, ..., p\}$ and $j \neq k$, the corresponding correlation matrix $R = (\rho_{j,k})$ is positive-definite.

Result S3 (Daniels and Pourahmadi, 2009): For $-1 < \pi_{j,k} < 1$, for $j, k \in \{1, 2, ..., p\}$, $j \neq k$, and R be the corresponding correlation matrix constructed from the elements of Π, we have

$$det(R) = \prod_{t=2}^{p}\prod_{j=1}^{t-1}(1 - \pi_{j,t}^2).$$

Result S4: For a k-band partial autocorrelation matrix of a multivariate normal random vector, the exponential part of the multivariate normal density is only a function of partial autocorrelations of lags $\{1, \ldots, k\}$.
Proof: Using Result 1 from the main paper, under the condition of a k_0 band partial autocorrelation matrix,

$$h(\hat{\Pi}_{k_0}, \hat{\sigma}_0) = \exp(-\frac{1}{2}trace(D^{-1}R^{-1}D^{-1}S)) \tag{1}$$

is only related to $(k_0+1)\times(k_0+1)$ principle sub-matrices of S, i.e., it is only affected by sample partial autocorrelations with lag not greater than k_0.

Result S5: Let $R_{p\times p} = (\rho_{j,k})$ be a correlation matrix, and $\Pi_{p\times p} = (\pi_{j,k})$ be the corresponding partial autocorrelation matrix of $R_{p\times p}$. Then,

$$det(R_{p\times p}) = [\det(R_2)(1-\nu_2^T R_2^{-1}\nu_2)(1-r_2^T R_2^{-1} r_2)] \bullet (1-\pi_{1,p}^2).$$

where $r_2 = R[2:p-1,p]$, $\nu_2 = R_{p\times p}^T[1,2:p-1]$ and $R_2 = R_{p\times p}[2:p-2,2:p-2]$.

Proof: First, we partition $R_{p\times p}$ as follows, $R_{p\times p} = \begin{pmatrix} R_{p-1\times p-1} & r_1 \\ r_1^T & \rho_{p,p} \end{pmatrix}$,
where $R_{p-1p-1} = R_{[1:p-11:p-1]}$ and $r_1 = R_{[1,1:p-1]}^T = (\rho_{1,p}, r_2^T)^T$. By Result S1 and since $\rho_{p,p} = 1$, $\det(R_{p\times p}) = \det(R_{p-1\times p-1})(1-r_1^T R_{p-1\times p-1}^{-1} r_1)$.
Furthermore, we can partition $R_{p-1\times p-1}$ in the following way,

$$R_{(p-1)\times(p-1)} = \begin{pmatrix} \rho_{1,1} & \nu_2^T \\ \nu_2 & R_2 \end{pmatrix},$$

So, $\det(R_{(p-1)\times(p-1)}) = (1-\nu_2^T R_2^{-1}\nu_2)\det(R_2)$.

Now, let $A_{11\bullet 2} = 1-\nu_2^T R_2^{-1}\nu_2$, then,

$$R_{(p-1)\times(p-1)}^{-1} = \begin{pmatrix} A_{11\bullet 2}^{-1} & -A_{11\bullet 2}^{-1}\nu_2^T R_2^{-1} \\ -R_2^{-1}\nu_2 A_{11\bullet 2}^{-1} & R_2^{-1}\nu_2 A_{11\bullet 2}^{-1}\nu_2^T R_2^{-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & R_2^{-1} \end{pmatrix}$$

$$= A_{11\bullet 2}^{-1}\begin{pmatrix} 1 & -\nu_2^T R_2^{-1} \\ -R_2^{-1}\nu_2 & R_2^{-1}\nu_2\nu_2^T R_2^{-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & R_2^{-1} \end{pmatrix}.$$

We now show

$$\begin{aligned} & r_1^T R_{(p-1)\times(p-1)}^{-1} r_1 \\ &= \begin{pmatrix} \rho_{1,p} & r_2^T \end{pmatrix} \cdot \left[A_{11\bullet 2}^{-1}\begin{pmatrix} 1 & -\nu_2^T R_2^{-1} \\ -R_2^{-1}\nu_2 & R_2^{-1}\nu_2\nu_2^T R_2^{-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & R_2^{-1} \end{pmatrix} \right] \cdot \begin{pmatrix} \rho_{1,p} \\ r_2 \end{pmatrix} \\ &= A_{11\bullet 2}^{-1} \cdot [\rho_{1,p}^2 - 2\rho_{1,p}\cdot r_2^T R_2^{-1}\nu_2 + (r_2^T R_2^{-1}\nu_2)^2] + r_2 R_2^{-1} r_2^T \\ &= A_{11\bullet 2}^{-1}\cdot(\rho_{1,p}^2 - \nu_2^T R_2^{-1} r_2)^2 + r_2^T R_2^{-1} r_2. \end{aligned}$$

Therefore,

$$\begin{aligned} & \det(R_{p\times p}) \\ &= (1-\nu_2^T R_2^{-1}\nu_2)\det(R_2)\cdot\{1 - r_2^T R_2^{-1} r_2 - A_{11\bullet 2}^{-1}\cdot(\rho_{1,p}^2 - \nu_2^T R_2^{-1} r_2)^2\} \\ &= (1-\nu_2^T R_2^{-1} r_2)(1-r_2^T R_2^{-1} r_2)\det(R_2)\cdot\left[1 - \left(\frac{\rho_{1,p}^2 - \nu_2^T R_2^{-1} r_2}{\sqrt{(1-\nu_2^T R_2^{-1}\nu_2)(1-r_2^T R_2^{-1} r_2)}}\right)^2\right] \\ &= (1-\nu_2^T R_2^{-1}\nu_2)(1-r_2^T R_2^{-1} r_2)\det(R_2)\cdot(1-\pi_{1,p}^2). \end{aligned}$$

2 PROOF OF RESULTS AND THEOREMS FROM SECTION 3.1

Proof of Result 1
First, let $R = (\rho_{j,k})_{p\times p}$ be a correlation matrix and denote $R_i = R_{[i:p-1,i:p-1]}$, $\nu_i = R^T_{[1,i:p-1]}$, and $r_i = R_{[i:p-1,p]}$.
We partition R as

$$R = \left(\begin{array}{ccccc} 1 & \rho_{1,2} & \dots & \rho_{1,p-1} & \rho_{1,p} \\ \rho_{2,1} & 1 & \dots & \rho_{2,p-1} & \rho_{2,p} \\ \dots\dots & & & & \\ \rho_{p-1,1} & \rho_{p-1,2} & \dots & 1 & \rho_{p-1,p} \\ \rho_{p,1} & \rho_{p,2} & \dots & \rho_{p,p-1} & 1 \end{array}\right) = \left(\begin{array}{cc} R_1 & r_1 \\ r_1^T & 1_{1\times 1} \end{array}\right),$$

where $r_1^T = \left(\begin{array}{cccc} \rho_{1,p} & \rho_{2,p} & \dots & \rho_{p-1,p} \end{array}\right)^T = \left(\begin{array}{cc} \rho_{1,p} & r_2^T \end{array}\right)^T$.

Thus, by Result S1

$$R^{-1} = A^{-1}_{22\bullet 1} \times \left(\begin{array}{cc} R_1^{-1} r_1 r_1^T R_1^{-1} & -R_1^{-1} r_1 \\ -r_1^T R_1^{-1} & 1_{1\times 1} \end{array}\right) + \left(\begin{array}{cc} R_1 & 0_{(p-1)\times 1} \\ 0_{1\times(p-1)} & 0_{1\times 1} \end{array}\right),$$

where $A_{22\bullet 1} = 1 - r_1^T R_1^{-1} r_1 = \prod_{j=1}^{p-1}(1-\pi^2_{j,p})$, and $det(R) = A_{22\bullet 1} det(R_1)$.
Furthermore, partition R_1 as

$$R_1 = \left(\begin{array}{cc} 1_{1\times 1} & \nu_2^T \\ \nu_2 & R_2 \end{array}\right).$$

We can then show that

$$R_1^{-1} = A^{-1}_{11\bullet 2} \left(\begin{array}{cc} 1_{1\times 1} & -\nu_2^T R_2^{-1} \\ -R_2^{-1}\nu_2 & R_2^{-1}\nu_2\nu_2^T R_2^{-1} \end{array}\right) + \left(\begin{array}{cc} 0_{1\times 1} & 0_{1\times(p-2)} \\ 0_{(p-2)\times 1} & R_2^{-1} \end{array}\right),$$

$$\begin{aligned} R_1^{-1} r_1 &= \frac{\rho_{1,p} - r_2^T R_2^{-1} r_2}{1-\nu_2^T R_2^{-1}\nu_2} \left(\begin{array}{c} 1_{1\times(p-2)} \\ -R_2^{-1}\nu_2 \end{array}\right) + \left(\begin{array}{c} 0_{1\times(p-2)} \\ R_2^{-1} r_2 \end{array}\right) \\ &= \pi_{1,p}\sqrt{\frac{1-r_2^T R_2^{-1} r_2}{1-\nu_2^T R_2^{-1}\nu_2}} \left(\begin{array}{c} 1_{1\times(p-2)} \\ -R_2^{-1}\nu_2 \end{array}\right) + \left(\begin{array}{c} 0_{1\times(p-2)} \\ R_2^{-1} r_2 \end{array}\right) \end{aligned}$$

$$\begin{aligned} & R_1^{-1} r_1 r_1^T R_1^{-1} \\ &= \pi^2_{1,p} \bullet \frac{1-r_2^T R_2^{-1} r_2}{1-\nu_2^T R_2^{-1}\nu_2} \left(\begin{array}{cc} 1_{1\times 1} & -\nu_2^T R_2^{-1} \\ -R_2^{-1}\nu_2 & R_2^{-1}\nu_2\nu_2^T R_2^{-1} \end{array}\right) \\ &+ \pi_{1,p} \bullet \sqrt{\frac{1-r_2^T R_2^{-1} r_2}{1-\nu_2^T R_2^{-1}\nu_2}} \left(\begin{array}{cc} 0_{1\times 1} & r_2^T R_2^{-1} \\ R_2^{-1} r_2 & -R_2^{-1}(\nu_2 r_2^T + r_2\nu_2^T) R_2^{-1} \end{array}\right) + \left(\begin{array}{cc} 0_{1\times 1} & 0_{1\times(p-2)} \\ 0_{(p-2)\times 1} & R_2^{-1} r_2 r_2^T R_2^{-1} \end{array}\right). \end{aligned}$$

Therefore,

$$
\begin{aligned}
&\begin{pmatrix} R_1^{-1} r_1 r_1^T R_1^{-1} & -R_1^{-1} r_1 \\ -r_1^T R_1^{-1} & 1_{1\times 1} \end{pmatrix} \\
&= \pi_{1,p}^2 \bullet \frac{\prod_{j=2}^{p-1}(1-\pi_{j,p}^2)}{\prod_{k=2}^{p-1}(1-\pi_{1,k}^2)} \begin{pmatrix} 1_{1\times 1} & -\nu_2^T R_2^{-1} & 0_{1\times 1} \\ -R_2^{-1}\nu_2 & R_2^{-1}\nu_2\nu_2^T R_2^{-1} & 0_{(p-2)\times 1} \\ 0_{1\times 1} & 0_{1\times(p-2)} & 0_{1\times 1} \end{pmatrix} \\
&+ \pi_{1,p} \bullet \sqrt{\frac{\prod_{j=2}^{p-1}(1-\pi_{j,p}^2)}{\prod_{k=2}^{p-1}(1-\pi_{1,k}^2)}} \begin{pmatrix} 0_{1\times 1} & r_2^T R_2^{-1} & -1_{1\times 1} \\ R_2^{-1} r_2 & -R_2^{-1}(\nu_2 r_2^T + r_2 \nu_2^T) R_2^{-1} & R_2^{-1}\nu_2 \\ -1_{1\times 1} & \nu_2^T R^{-1} & 0_{1\times 1} \end{pmatrix} \\
&+ \begin{pmatrix} 0_{1\times 1} & 0_{1\times(p-2)} & 0_{1\times 1} \\ 0_{(p-2)\times 1} & R_2^{-1} r_2 r_2^T R_2^{-1} & -R_2^{-1} r_2 \\ 0_{1\times 1} & -r_2^T R_2^{-1} & 1_{1\times 1} \end{pmatrix},
\end{aligned}
$$

since the partial autocorrelation matrix $\Pi = (\pi_{j,k})_{p\times p}$ has k_0 bands, $\pi_{j,j+k} = 0$ for $k > k_0$. Hence,

$$
\begin{pmatrix} R_1^{-1} r_1 r_1^T R_1^{-1} & -R_1^{-1} r_1 \\ -r_1^T R_1^{-1} & 1 \end{pmatrix} = \begin{pmatrix} 0_{1\times 1} & 0_{1\times(p-2)} & 0_{1\times 1} \\ 0_{(p-2)\times 1} & R_2^{-1} r_2 r_2^T R_2^{-1} & -R_2^{-1} r_2 \\ 0_{1\times 1} & -r_2^T R_2^{-1} & 1_{1\times 1} \end{pmatrix}, \text{ (for } p-1 > k_0).
$$

Similarly, for $p - i > k_0$, we obtain

$$
\begin{pmatrix} R_i^{-1} r_i r_i^T R_i^{-1} & -R_i^{-1} r_i \\ -r_i^T R_i^{-1} & 1_{1\times 1} \end{pmatrix} = \begin{pmatrix} 0_{1\times 1} & 0_{1\times(p-i-1)} & 0_{1\times 1} \\ 0_{(p-i-1)\times 1} & R_{i+1}^{-1} r_{i+1} r_{i+1}^T R_{i+1}^{-1} & -R_{i+1}^{-1} r_{i+1} \\ 0_{1\times 1} & -r_{i+1}^T R_{i+1}^{-1} & 1_{1\times 1} \end{pmatrix}.
$$

Therefore,

$$
\begin{pmatrix} R_1^{-1} r_1 r_1^T R_1^{-1} & -R_1^{-1} r_1 \\ -r_1^T R_1^{-1} & 1_{1\times 1} \end{pmatrix} = \begin{pmatrix} 0_{(p-k_0-1)\times(p-k_0-1)} & 0_{(p-k_0-1)\times k_0} & 0_{(p-k_0-1)\times 1} \\ 0_{k_0\times(p-k_0-1)} & R_{p-k_0}^{-1} r_{p-k_0} r_{p-k_0}^T R_{p-k_0}^{-1} & -R_{p-k_0}^{-1} r_{p-k_0} \\ 0_{1\times(p-k_0-1)} & -r_{p-k_0}^T R_{p-k_0}^{-1} & 1_{1\times 1} \end{pmatrix},
$$

where $R_{p-k_0} = R_{[p-k_0:p-1,p-k_0:p-1]}$ is a principal submatrix of R with rows and columns from $(p-k_0)$ to $(p-1)$, and $r_{p-k_0} = R_{[p-k_0:p-1,p]}$.
Therefore, under the assumption of k_0 bands, we obtain

$$
\begin{aligned}
&R^{-1} \\
&= A_{22\cdot 1}^{-1} \begin{pmatrix} 0_{(p-k_0-1)\times(p-k_0-1)} & 0_{(p-k_0-1)\times k_0} & 0_{(p-k_0-1)\times 1} \\ 0_{k_0\times(p-k_0-1)} & R_{p-k_0}^{-1} r_{p-k_0}^T r_{p-k_0} R_{p-k_0}^{-1} & -R_{p-k_0}^{-1} r_{p-k_0} \\ 0_{1\times(p-k_0-1)} & -r_{p-k_0}^T R_{p-k_0}^{-1} & 1_{1\times 1} \end{pmatrix} + \begin{pmatrix} R_1^{-1} & 0_{(p-1)\times 1} \\ 0_{1\times(p-1)} & 0_{1\times 1} \end{pmatrix} \\
&= \prod_{j=1}^{p-1}(1-\pi_{j,p}^2)^{-1} \begin{pmatrix} 0_{(p-k_0-1)\times(p-k_0-1)} & 0_{(p-k_0-1)\times k_0} & 0_{(p-k_0-1)\times 1} \\ 0_{1\times(p-k_0-1)} & R_{p-k_0}^{-1} r_{p-k_0} r_{p-k_0}^T R_{p-k_0}^{-1} & -R_{p-k_0}^{-1} r_{p-k_0} \\ 0_{1\times(p-k_0-1)} & -r_{p-k_0}^T R_{p-k_0}^{-1} & 1_{1\times 1} \end{pmatrix} + \begin{pmatrix} R_1^{-1} & 0_{(p-1)\times 1} \\ 0_{1\times(p-1)} & 0_{1\times 1} \end{pmatrix} \\
&= \prod_{j=p-k_0}^{p-1}(1-\pi_{j,p}^2)^{-1} \begin{pmatrix} 0_{(p-k_0-1)\times(p-k_0-1)} & 0_{(p-k_0-1)\times k_0} & 0_{(p-k_0-1)\times 1} \\ 0_{1\times(p-k_0-1)} & R_{p-k_0}^{-1} r_{p-k_0} r_{p-k_0}^T R_{p-k_0}^{-1} & -R_{p-k_0}^{-1} r_{p-k_0} \\ 0_{1\times(p-k_0-1)} & -r_{p-k_0}^T R_{p-k_0}^{-1} & 1_{1\times 1} \end{pmatrix} + \begin{pmatrix} R_1^{-1} & 0_{(p-1)\times 1} \\ 0_{1\times(p-1)} & 0_{1\times 1} \end{pmatrix}.
\end{aligned}
$$

Now, we re-write R_1 similar to R to obtain

$$
\begin{aligned}
& R^{-1} \\
= \; & \prod_{j=p-k_0}^{p-1}(1-\pi_{j,p}^2)^{-1}\begin{pmatrix} 0_{(p-k_0-1)\times(p-k_0-1)} & 0_{(p-k_0-1)\times k_0} & 0_{(p-k_0-1)\times 1} \\ 0_{k_0\times(p-k_0-1)} & M_{jp}M_{jp}^T & -M_{jp} \\ 0_{1\times(p-k_0-1)} & -M_{jp}^T & 1_{1\times 1} \end{pmatrix} \\
& +\ldots\ldots+ \\
& +\prod_{j=s-k_0}^{s-1}(1-\pi_{j,s}^2)^{-1}\begin{pmatrix} 0_{(p-k_0-s-1)\times(p-k_0-s-1)} & 0_{(p-k_0-s-1)\times k_0} & 0_{(p-k_0-s-1)\times 1} & 0_{(p-k_0-s-1)\times s} \\ 0_{k_0\times(p-k_0-s-1))} & M_{js}M_{js}^T & -M_{js} & 0_{k_0\times s} \\ 0_{1\times(p-k_0-s-1)} & -M_{js}^T & 1_{1\times 1} & 0_{1\times s} \\ 0_{s\times(p-k_0-s-1)} & 0_{s\times k_0} & 0_{s\times 1} & 0_{s\times s} \end{pmatrix} \\
& +\ldots\ldots+ \\
& +\prod_{j=1}^{k_0}(1-\pi_{j,k_0+1}^2)^{-1}\begin{pmatrix} 0_{1\times 1} & 0_{1\times k_0} & 0_{1\times 1} & 0_{1\times(p-k_0-2)} \\ 0_{k_0\times 1} & M_{jk_0+1}M_{jk_0+1}^T & -M_{jk_0+1} & 0_{1\times(p-k_0-2)} \\ 0_{1\times 1} & -M_{jk_0+1}^T & 1_{1\times 1} & 0_{1\times(p-k_0-2)} \\ 0_{(p-k_0-2)\times 1} & 0_{(p-k_0-2)\times k_0} & 0_{(p-k_0-2)\times 1} & 0_{(p-k_0-2)\times(p-k_0-2)} \end{pmatrix} \\
& +\begin{pmatrix} R^{-1}_{[1:k_0 1:k_0]} & 0_{k_0\times(p-k_0)} \\ 0_{(p-k_0)\times k_0} & 0_{(p-k_0)\times(p-k_0)} \end{pmatrix},
\end{aligned}
$$

where $M_{jp}=R^{-1}_{[p-k_0:p-1,p-k_0:p-1]}R_{[p-k_0:p-1,p]}$, $M_{js}=R^{-1}_{[s-k_0:s-1,s-k_0:s-1]}R_{[s-k_0:s-1,s]}$, and $M_{jk_0+1}=R^{-1}_{[2:k_0+1,2:k_0+1]}R_{[2:k_0+1,k_0+2]}$. This shows that R^{-1} is a sum of $p\times p$ matrices including only $(k_0+1)\times(k_0+1)$ non-zero principal sub-matrices. As a result, we only need to invert (k_0-1)-dimensional matrices to move from a k_0 band Π to R^{-1}. Furthermore, both R^{-1} and the precision matrix, $\Sigma^{-1}=D^{-1}R^{-1}D^{-1}$ are k_0-band matrices.

Proof of Theorem 1:

Assume $\hat{\pi}=(\hat{\pi}_{1,2},\hat{\pi}_{2,3},...,\hat{\pi}_{p-1,p},...,\hat{\pi}_{1,p-1},\hat{\pi}_{2,p},\hat{\pi}_{1,p})$ is the mle of multivariate normal likelihood function, $\widetilde{\pi}^{1,p}=(\widetilde{\pi}_{1,2},\widetilde{\pi}_{2,3},...\widetilde{\pi}_{p-1,p},...,\widetilde{\pi}_{1,p-1},\widetilde{\pi}_{2,p},\widetilde{\pi}_{1p})$ as in $\tilde{\pi}_{j,j+l}=\max_{\pi_{j,j+l}}G^\star(\pi_{j,j+l}: j=1,\ldots,p-l)$, (maximizer of the objective function). We are going to prove $\hat{\pi}$ is same as $\widetilde{\pi}$. Let $\widetilde{\rho}^{1,p}=(\widetilde{\rho}_{1,2},\widetilde{\rho}_{2,3},...,\widetilde{\rho}_{p-1,p},...,\widetilde{\rho}_{1,p-1},\widetilde{\rho}_{2,p},\widetilde{\rho}_{1,p})$ be the corresponding estimators of the correlation coefficients. We use an induction argument.

1) For $k=1$, $\widetilde{\pi}_{j,j+1}$, the maximizer of $G(\pi_{j,j+1})$, is also the mle of the corresponding correlation coefficients $\rho_{j,j+1}$ since $\rho_{j,j+1}=\pi_{j,j+1}$.

2) For $k=t\in\{1,\ldots.p-2\}$, assume

$$\widetilde{\pi}^{j,j+t}=(\widetilde{\pi}_{j,j+1},\widetilde{\pi}_{j+1,j+2},...\widetilde{\pi}_{j+t-1,j+t},...,\widetilde{\pi}_{j,j+t-1},\widetilde{\pi}_{j+1,j+t},\widetilde{\pi}_{j,j+t})$$

is the maximizer of the multivariate normal likelihood with $j=1,2,...,p-t$, i.e., $\widetilde{\pi}^{jj+t}=\hat{\pi}^{jj+t}$. Then, the corresponding correlation coefficient estimators

$$\widetilde{\rho}^{j,j+t}=(\widetilde{\rho}_{j,j+1},\widetilde{\rho}_{j+1,j+2},...,\widetilde{\rho}_{j+t-1,j+t},...,\widetilde{\rho}_{j,j+t-1},\widetilde{\rho}_{j+1,j+t},\widetilde{\rho}_{j,j+t})$$

are the maximizers of the multivariate normal likelihood on the ρ scale, i.e., $\widetilde{\rho}^{j,j+t}=\hat{\rho}^{j,j+t}$.

3) for k=t+1, let $\hat{\Pi}_{j,j+t+1\backslash j+t+1}=\{\hat{\pi}_{j,j+1},...,\hat{\pi}_{j+t,j+t+1},...,\hat{\pi}_{j,j+t},\hat{\pi}_{j+1,j+t+1}\}$,
and $\hat{P}_{j,j+t+1\backslash j+t+1}=\{\hat{\rho}_{j,j+1},...,\hat{\rho}_{j+t,j+t+1},...,\hat{\rho}_{j,j+t},\hat{\rho}_{j+1,j+t+1}\}$.

We have

$$
\begin{aligned}
&\widetilde{\rho}^{j,j+t+1} \\
&\quad = (\widetilde{\rho}_{j,j+1}, \widetilde{\rho}_{j+1,j+2}, ..., \widetilde{\rho}_{j+t,j+t+1}, ..., \widetilde{\rho}_{j,j+t}, \widetilde{\rho}_{j+1,j+t+1}, \widetilde{\rho}_{j,j+t+1}) \\
&\quad = (\hat{\rho}_{j,j+1}, \hat{\rho}_{j+1,j+2}, ..., \hat{\rho}_{j+t,j+t+1}, ..., \hat{\rho}_{j,j+t}, \hat{\rho}_{j+1,j+t+1}, \widetilde{\rho}_{j,j+t+1}) \\
&\quad = \{\hat{P}_{j,j+t+1\backslash j+t+1}, \widetilde{\rho}_{j,j+t+1}\}
\end{aligned}
$$

and

$$
\begin{aligned}
&\hat{\rho}^{j,j+t+1} \\
&\quad = (\hat{\rho}_{j,j+1}, \hat{\rho}_{j+1,j+2}, ..., \hat{\rho}_{j+t,j+t+1}, ..., \hat{\rho}_{j,j+t}, \hat{\rho}_{j+1,j+t+1}, \hat{\rho}_{j,j+t+1}) \\
&\quad = \{\hat{P}_{j,j+t+1\backslash j+t+1}, \hat{\rho}_{j,j+t+1}\}
\end{aligned}
$$

which is the maximizer of $L(\Pi_{j,j+t+1})$,

Moreover,

$$
\begin{aligned}
&\widetilde{\rho}_{jj+t+1} \\
&\quad = \text{maxarg}_{\rho}\{G(\pi_{j,j+t+1})\} = \text{maxarg}_{\rho}\{L(\Pi_{jj+t+1}|\hat{\Pi}_{j,j+t+1\backslash j+t+1})\} \\
&\quad = \hat{\rho}_{j,j+t+1}.
\end{aligned}
$$

Therefore,

$$
\widetilde{\rho}^{j,j+t+1} = \hat{\rho}^{j,j+t+i}.
$$

Correspondingly,

$$
\begin{aligned}
&\widetilde{\pi}_{jj+t+1} \\
&\quad = \text{maxarg}_{\pi}\{G(\pi_{j,j+t+1})\} = \text{maxarg}_{\pi}\{L(\Pi_{jj+t+1}|\hat{\Pi}_{j,j+t+1\backslash j+t+1})\} \\
&\quad = \hat{\pi}_{j,j+t+1}.
\end{aligned}
$$

Hence, by induction, $\widetilde{\pi}^{1p}$ is the mle of the multivariate normal likelihood function.

Proof of Theorem 2: The proof follows directly from the following Lemma.

Lemma: Suppose $Y_1, Y_2, ..., Y_n$ $(n > p)$ are iid $N(\underline{0}, DRD)$, where D is a diagonal matrix of marginal standard deviations. Let $\hat{\Pi} = (\hat{\pi}_{j,t})_{p\times p}$ be the mle of the partial autocorrelation matrix. For a band k_0 partial autocorrelation matrix, the mles of the partial autocorrelations with lags greater than k_0 are independent with marginal distributions given by

$$
f(\hat{\pi}_{j,j+k}) \propto (1 - \hat{\pi}_{j,j+k})^{\alpha}(1 + \hat{\pi}_{j,j+k})^{\beta}
$$

where

$$
\alpha = \beta = \begin{cases} \frac{n-k-2}{2} & \text{for } k_0 < k \leq p-2 \\ \frac{n-p-1}{2} & \text{for } k_0 < k = p-1 \end{cases}.
$$

Proof: Since $Y_1, Y_2, ..., Y_n$ are iid multivariate normal random vectors $N(0, DRD)$, the sample covariance S follows a Wishart distribution $W_n(\Sigma)$ for $n \geq p+1$ with pdf,

$$p(s) \propto |S|^{\frac{n-p-1}{2}} exp[-\frac{1}{2} trace(D^{-1}R^{-1}D^{-1}S)].$$

Let $S = \{\hat{\sigma}_{j,l} : j, l = 1, \ldots, p\}$, $A = (\hat{\sigma}_{1,2}, \hat{\sigma}_{2,3}, \hat{\sigma}_{1,3}, ..., \hat{\sigma}_{1,p}, \hat{\sigma}_{1,1}, ..., \hat{\sigma}_{p,p})$,
$B = (\hat{\rho}_{1,2}, \hat{\rho}_{2,3}, \hat{\rho}_{1,3}, ..., \hat{\rho}_{1,p}, \hat{\sigma}_{1,1}, ..., \hat{\sigma}_{p,p})$, and $\hat{\Pi} = (\hat{\pi}_{1,2}, \hat{\pi}_{2,3}, \hat{\pi}_{1,3}, ..., \hat{\pi}_{1,p}, \hat{\sigma}_{1,1}, ..., \hat{\sigma}_{p,p}) = (\hat{\pi}, \hat{\sigma}_0)$.
where $\hat{\pi} = (\hat{\pi}_{1,2}, \hat{\pi}_{2,3}, \hat{\pi}_{1,3}, ..., \hat{\pi}_{1,p})$, and $\hat{\sigma}_0 = (\hat{\sigma}_{1,1}, ..\hat{\sigma}_{p,p})$.
The Jacobian from A to B is
$\hat{J} = \begin{pmatrix} \hat{J}_{11} & \hat{J}_{12} \\ 0 & I_{p\times p} \end{pmatrix}$, where $\hat{J}_{11} = diag(vech(\hat{\sigma}_0^{1/2} \otimes \hat{\sigma}_0^{1/2}))$.
According to Joe (2006) , the determinant of the Jacobian for $\hat{J}_{B\to\pi}$ is

$$|\hat{J}_{B\to\pi}| = \prod_{j=1}^{p-1} (1 - \hat{\pi}_{j,j+1}^2)^{\frac{p-2}{2}} \cdot \prod_{k=2}^{p-2} \prod_{j=1}^{p-k} (1 - \hat{\pi}_{j,j+k}^2)^{\frac{p-1-k}{2}}.$$

Also, recall $|\hat{R}| = \prod_{k=1}^{p-1} \prod_{j=1}^{p-k} (1 - \hat{\pi}_{j,j+k}^2)$ (Result S3), and $\hat{S} = \hat{D}\hat{R}\hat{D}$ where $\hat{D}$ is a diagonal matrix with sample standard deviations on its main diagonal.
Therefore,

$$\begin{aligned} p(B) &\propto |\hat{J}||S|^{\frac{n-p-1}{2}} exp[-\frac{1}{2} trace(D^{-1}R^{-1}D^{-1}S)] \text{ and} \\ p(\hat{\pi}, \hat{\sigma}_0) &\propto |\hat{J}_{11}||\hat{J}_{B\to\pi}||\hat{D}\hat{R}\hat{D}|^{\frac{n-p-1}{2}} exp[-\frac{1}{2} trace(D^{-1}R^{-1}D^{-1}S)] \\ &= |\hat{J}_{11}||\hat{J}_{B\to\pi}||\hat{D}\hat{R}\hat{D}|^{\frac{n-p-1}{2}} h_{(\hat{\Pi}_0, \hat{\sigma}_0)}, \end{aligned}$$

where $h_{(\hat{\Pi}_0,\hat{\sigma}_0)} = exp[-\frac{1}{2} trace(D^{-1}R^{-1}D^{-1}S)]$. We can simplify this as follows,

$$\begin{aligned} p(\hat{\pi}, \hat{\sigma}_0) &\propto |\hat{J}_{11}||\hat{D}|^{n-p-1} [\prod_{j=1}^{p-1} (1 - \hat{\pi}_{j,j+1}^2)^{\frac{p-2}{2}} \cdot \prod_{k=2}^{p-2} \prod_{j=1}^{p-k} (1 - \hat{\pi}_{j,j+k}^2)^{\frac{p-1-k}{2}}] \cdot \\ & [\prod_{k=1}^{p-1} \prod_{j=1}^{p-k} (1 - \hat{\pi}_{j,j+k}^2)]^{\frac{n-p-1}{2}} h(\hat{\Pi}_{k_0}, \hat{\sigma}_0) \\ &= (1 - \hat{\pi}_{1,p}^2)^{\frac{n-p-1}{2}} \prod_{k=k_0+1}^{p-2} \prod_{j=k}^{p-k} (1 - \hat{\pi}_{j,j+k}^2)^{\frac{n-k-2}{2}} \\ & \cdot \prod_{j=1}^{p-1} (1 - \hat{\pi}_{j,j+1}^2)^{\frac{p-2}{2}} \cdot [\prod_{k=2}^{k_0} \prod_{j=1}^{p-k} (1 - \hat{\pi}_{j,j+k}^2)]^{\frac{n-k-2}{2}} |\hat{J}_{11}||\hat{D}|^{n-p-1} h(\hat{\Pi}_{k_0}, \hat{\sigma}_0) \\ &= (1 - \hat{\pi}_{1,p}^2)^{\frac{n-p-1}{2}} \prod_{k=k_0+1}^{p-2} \prod_{j=1}^{p-k} (1 - \hat{\pi}_{j,j+k}^2)^{\frac{n-k-2}{2}} \cdot h^\star(\hat{\Pi}_{k_0}, \hat{\sigma}_0), \end{aligned}$$

where
$h^\star(\hat{\Pi}_{k_0}, \hat{\sigma}_0) = [\prod_{j=1}^{p-1}(1-\pi_{j,j+1}^2)^{\frac{p-2}{2}}] \cdot [\prod_{k=2}^{k_0}\prod_{j=1}^{p-k}(1-\hat{\pi}_{j,j+k}^2)]^{\frac{n-k-2}{2}} |\hat{J}_{11}||\hat{D}|^{n-p-1} h(\hat{\Pi}_{k_0}, \hat{\sigma}_0)$.
Hence, all sample partial autocorrelations with lags not less than k_0 are independent with marginal distributions given by

$$f(\hat{\pi}_{jj+k}) \propto \begin{cases} (1-\hat{\pi}_{j,j+k}^2)^{\frac{n-k-2}{2}} & \text{for } k \in \{k_0+1, \ldots, p-2\} \\ (1-\hat{\pi}_{1,p}^2)^{\frac{n-p-1}{2}} & \text{for } k_0 < k = p-1 \end{cases} .$$

3 Form of estimating equations for $\pi_{j,k}$

Before deriving the equations for $\pi_{j,k}$, we introduce some notation. Let $\{Y_i : i = 1, ..., n\}$ be $p \times 1$ vectors of independent, normally distributed random variables with mean 0 and covariance matrix Σ. The mle of Σ is $S_\Sigma = \frac{1}{n}\sum_{i=1}^n Y_i Y_i'$. After applying the transformation $X_i = TY_i$, where $T = \text{diag}(s_{jj,\Sigma}^{-\frac{1}{2}})$, we define

$$S = \frac{1}{n}\sum_{i=1}^n X_i X_i' = (s_{(j,k)}).$$

The likelihood for $\pi_{j,k}$, $G(\pi_{j,k})$ is proportional to

$$G(\pi_{j,k}) \propto (1-\pi_{j,k}^2)^{-\frac{n}{2}} \exp\{-\frac{n}{2} tr(R^{-1}[j:k]S[j:k])\}.$$

The corresponding log likelihood can be re-written as

$$\log G(\pi_{j,k}) = g(\pi_{j,k}) \propto -\frac{n}{2}\log(1-\pi_{j,k}^2) - \frac{n}{2\det(R[j:k])} tr(A[j:k]S[j:k]),$$

where $A[j:k]$ is the adjoint matrix of $R[j:k]$ and $R^{-1}[j:k] = \frac{1}{\det(R[j:k])}A[j:k]$. So $A[j:k]$ is a quadratic function of $\rho_{j,k}$, i.e. $A[j:k] = A_0 + A_1\rho_{j,k} + A_2\rho_{j,k}^2$, where A_0, A_1, A_2 are $(k-j+1)\times(k-j+1)$ matrices.
The first derivative of the log likelihood for $\pi_{j,k}$ is

$$\begin{aligned} \frac{\partial g_{(\pi_{j,k})}}{\partial \pi_{j,k}} \\ &= -\frac{n}{2}\frac{-2\pi_{j,k}}{1-\pi_{j,k}^2} - \frac{n}{2a(1-\pi_{j,k}^2)^2}\{D_{jk}(1-\pi_{j,k}^2)[tr(A_1S[j:k]) + 2\rho_{j,k}tr(A_2S[j,k])] \\ &\quad +2\pi_{j,k}tr(A[j:k]S[j:k])\} \\ &= \frac{-n}{2a(1-\pi_{j,k}^2)^2}\{-2a\pi_{j,k}(1-\pi_{j,k}^2) + D_{jk}(1-\pi_{j,k}^2)[tr(A_1S[j:k]) + 2\rho_{j,k}tr(A_2S[j:k])] \\ &\qquad +2\pi_{j,k}[tr(A_0S[j:k]) + \rho_{j,k}tr(A_1S[j:k]) + \rho_{j,k}^2 tr(A_2S[j:k])]\}. \end{aligned}$$

where $a = det(R[j:k])/(1-\pi_{j,k}^2)$, which is not a function of $\pi_{j,k}$ (cf: Result S3).

By letting $\frac{\partial g_{(\pi_{j,k})}}{\partial \pi_{j,k}} = 0$, we have

$$-2\pi_{j,k}(1-\pi_{j,k}^2) + \frac{j}{a}\{D_{jk}(1-\pi_{j,k}^2)[tr(A_1 S_{jk}) + 2\rho_{j,k} tr(A_2 S_{jk})]$$

$$+2\pi_{j,k}\det(R_{jk}) \cdot tr(A_{[j:k]}S_{jk})\} = 0$$

where $S_{jk} = S[j:k, j:k]$, $R_{jk} = R[j:k, j:k]$, and $A_{[j:k]}$ is the adjoint matrix of $R_{[j:k]}$. Therefore, by solving (2), we obtain $\widetilde{\pi}_{j,k}$.

4 ADDITIONAL SIMULATION RESULTS

Scenario 3: Four band matrices

- a: Four band *correlation* matrices with correlation function $\rho_{j,j+l} = 0.4I(|l|=1) + 0.2I(2 \leq |l| \leq 3) + 0.1I(|l|=4)$.
- b: Four band *partial autocorrelation* matrices with partial autocorrelation $\pi_{j,j+l} = 0.4I(|l|=1) + 0.2I(2 \leq |l| \leq 3) + 0.1I(|l|=4)$.

p=60	$\hat{R}$		AIC		Chol-Testing		PAC-Testing
Size(n)	Risk (Var)	EB	Risk (Var)	EB	Risk (Var)	EB	Risk (Var)
30	10.95 (0.06)	1.00	2.67 (0.04)	1.12	2.76 (0.13)	1.09	2.74 (0.11)
60	7.67 (0.03)	1.00	2.24 (0.02)	1.14	2.28 (0.03)	1.12	2.27 (0.03)
100	5.96 (0.02)	1.02	2.06 (0.01)	1.21	2.06 (0.01)	1.20	2.06 (0.01)
300	3.43 (0.01)	1.48	1.76 (0.02)	1.81	1.70 (0.03)	1.81	1.70 (0.03)
500	2.65 (0.00)	2.26	1.53 (0.05)	2.28	1.54 (0.07)	2.24	1.54 (0.07)
1000	1.87 (0.00)	3.76	1.06 (0.05)	3.26	1.19 (0.10)	3.24	1.20 (0.10)
n=100	$\hat{R}$		AIC		Chol-Testing		PAC-Testing
Dim(p)	Risk (Var)	EB	Risk (Var)	EB	Risk (Var)	EB	Risk (Var)
30	2.92 (0.02)	1.04	1.42 (0.01)	1.21	1.43 (0.01)	1.21	1.43 (0.01)
100	9.98 (0.02)	1.00	2.67 (0.01)	1.28	2.71 (0.02)	1.23	2.70 (0.02)
200	20.02 (0.02)	1.00	3.79 (0.01)	1.18	3.82 (0.02)	1.16	3.82 (0.02)
500	50.18 (0.02)	1.00	6.02 (0.01)	1.31	6.08 (0.05)	1.25	6.06 (0.04)

Table 1: Results for scenario 3a
Note: Frobenius matrix norms (Risk) and Monte Carlo variance (Var) over 100 replicates for the banded partial autocorrelation estimator by multiple hypotheses testing (PAC-Testing), banded Choleski factors of inverse covariance matrix by multiple hypotheses testing (Chol-Testing) or (AIC) for scenario 3a. EB is the estimated number of bands of the partial autocorrelation matrix or Choleski factors.

p=60	$\hat{R}$		AIC		Chol-Testing		PAC-Testing
Size	Risk (Var)	EB	Risk (Var)	EB	Risk (Var)	EB	Risk (Var)
30	10.85 (0.11)	2.60	4.69 (1.09)	1.63	5.67 (1.09)	1.57	5.72 (1.00)
60	7.57 (0.07)	3.13	2.99 (0.09)	2.88	3.60 (1.32)	2.76	3.71 (1.44)
100	5.88 (0.05)	3.54	2.40 (0.04)	3.40	2.48 (0.35)	3.39	2.48 (0.35)
300	3.37 (0.01)	4.00	1.31 (0.02)	3.99	1.36 (0.04)	3.99	1.36 (0.04)
500	2.60 (0.01)	4.00	1.00 (0.01)	4.01	1.00 (0.01)	4.01	1.00 (0.01)
1000	1.84 (0.00)	4.00	0.71 (0.00)	4.04	0.72 (0.00)	4.04	0.72 (0.00)
n=100	$\hat{R}$		AIC		Chol-Testing		PAC-Testing
Dim(p)	Risk (Var)	EB	Risk (Var)	EB	Risk (Var)	EB	Risk (Var)
30	2.85 (0.04)	3.50	1.61 (0.05)	3.26	1.68 (0.14)	3.23	1.70 (0.15)
100	9.82 (0.04)	3.50	3.10 (0.07)	3.43	3.14 (0.09)	3.40	3.14 (0.09)
200	19.95 (0.05)	3.61	4.49 (0.11)	3.61	4.52 (0.11)	3.61	4.52 (0.11)
500	50.07 (0.05)	3.64	7.11 (0.13)	3.68	7.17 (0.14)	3.65	7.20 (0.13)

Table 2: Results for scenario 3b

Note: Frobenius matrix norms (Risk) and Monte Carlo variance (Var) over 100 replicates for the banded partial autocorrelation estimator by multiple hypotheses testing (PAC-Testing), banded Choleski factors of inverse covariance matrix by multiple hypotheses testing (Chol-Testing) or (AIC) for scenario 3b. EB is the estimated number of bands of the partial autocorrelation matrix or Choleski factors.