Supplementary Material for "Saddlepoint approximations for spatial panel data models"

Chaonan Jiang, Davide La Vecchia, Elvezio Ronchetti, Olivier Scaillet

Appendix A includes proofs of Lemma 1-2 in A.1-2, Proposition 3-4 in A.3-4.

Appendix B provides the first, second and third derivatives of the log-likelihood in B.1-3 as well as some additional computations in B.4.

Appendix C contains the algorithms which explain how to implement our saddlepoint techniques.

Appendix D gives additional numerical results for the SAR(1) model in D.1 and D.3-6, and the analytical check of Assumption D(iv) in D.2.

A Proofs of Lemmas and Propositions

A.1 Proof of Lemma 1

Proof. Since the MLE is an M-estimator, we derive its second-order von Mises expansion using the results of von Mises (1947) (see also Filippova (1962), and Cabrera and Fernholz (1999)), and we get

$$\vartheta(P_{n,T}) - \vartheta(P_{\theta_0}) = \frac{1}{n}\sum_{i=1}^{n} IF_{i,T}(\psi, P_{\theta_0}) + \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\varphi_{i,j,T}(\psi, P_{\theta_0}) + O_P(m^{-3/2}), \quad \text{(A.1)}$$

where we make use of the fact that $O_P(n^{-3/2})$ is also $O_P(m^{-3/2})$, since $m = n(T-1)$. The expression of $IF_{i,T}(\psi, P_{\theta_0})$ and of $\varphi_{i,j,T}(\psi, P_{\theta_0})$ for general M-estimators are available in Gatto and Ronchetti (1996). Specifically, for the i-th observation and for the whole time span, $IF_{i,T}(\psi, P_{\theta_0})$ is the Influence Function (IF) of the MLE, having likelihood score $(T-1)^{-1}\sum_{t=1}^{T}\psi_{i,t}(\theta_0)$, which reads as: $IF_{i,T}(\psi, P_{\theta_0}) = M_{i,T}^{-1}(\psi, P_{\theta_0})(T-1)^{-1}\sum_{t=1}^{T}\psi_{i,t}(\theta_0)$, From Withers (1983), it follows that the second term of the von Mises expansion in (A.1) is given by (14) and (15).

To compute the second-order von Mises expansion, we need the matrices of partial derivatives $\frac{\partial^2\psi_{i,t,l}(\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\theta_0}$ for $l = 1, 2, \cdots, d$, whose expressions are provided in Appendix B. □

A.2 Proof of Lemma 2

Proof. For $m = n(T-1)$, the second-order von Mises expansion for $q[\vartheta(P_{n,T})]$ is:

$$\begin{aligned} q[\vartheta(P_{n,T})] - q[\vartheta(P_{\theta_0})] &= (\vartheta(P_{n,T}) - \vartheta(P_{\theta_0}))' \frac{\partial q(\vartheta)}{\partial\vartheta}\Big|_{\theta=\theta_0} \\ &+\frac{1}{2}(\vartheta(P_{n,T}) - \vartheta(P_{\theta_0}))' \frac{\partial^2 q(\vartheta)}{\partial\vartheta\partial\vartheta'}\Big|_{\theta=\theta_0}(\vartheta(P_{n,T}) - \vartheta(P_{\theta_0})) \\ &O_P(||\vartheta(P_{n,T}) - \vartheta(P_{\theta_0})||^3). \end{aligned} \tag{A.2}$$

Making use of (A.1), (13) and (14) into (A.2), we get

$$\begin{aligned} q[\vartheta(P_{n,T})] - q[\vartheta(P_{\theta_0})] &= \frac{1}{n}\sum_{i=1}^{n} IF'_{i,T}(\psi, P_{\theta_0})\frac{\partial q(\vartheta)}{\partial\vartheta}\Big|_{\theta=\theta_0} + \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Bigg[\varphi'_{i,j,T}(\psi, P_{\theta_0})\frac{\partial q(\vartheta)}{\partial\vartheta}\Big|_{\theta=\theta_0} \\ &+ IF'_{i,T}(\psi, P_{\theta_0})\frac{\partial^2 q(\vartheta)}{\partial\vartheta\partial\vartheta'}\Big|_{\theta=\theta_0} IF_{j,T}(\psi, P_{\theta_0})\Bigg] + O_P(m^{-3/2}). \end{aligned} \tag{A.3}$$

Similarly to Gatto and Ronchetti (1996), we delete the diagonal terms from (A.3), and we define the following U-statistic of order two (see, e.g., Serfling (2009) or Van der Vaart (1998), page 295) by making use of (16), (17) and (18):

$$
\begin{aligned}
U_{n,T}(\psi,\theta_0) &= \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} h_{i,j,T}\left(\psi, P_{\theta_0}\right) \\
&= \frac{2}{n} \sum_{i=1}^{n} g_{i,T}\left(\psi, P_{\theta_0}\right) + \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \gamma_{i,j,T}(\psi, P_{\theta_0}). \qquad \text{(A.4)}
\end{aligned}
$$

Then, we remark that $q[\vartheta(P_{n,T})] - q[\vartheta(P_{\theta_0})]$ in (A.3) is equivalent (up to $O_P(m^{-3/2})$) to $U_{n,T}(\psi,\theta_0)$, namely

$$
q[\vartheta(P_{n,T})] - q[\vartheta(P_{\theta_0})] = U_{n,T}(\psi,\theta_0) + O_P(m^{-3/2}), \qquad \text{(A.5)}
$$

which concludes the proof. □

A.3 Proof of Proposition 3

To derive the Edgeworth expansion for $\sigma_{n,T}^{-1}\{q[\vartheta(P_{n,T})] - q[\vartheta(P_{\theta_0})]\}$, we first introduce two lemmas.

Lemma A.1. *Under Assumptions A-D, for all i, j and $i \neq j$,*

$$
M_{i,T}^{-1}(\psi, P_{\theta_0}) M_{j,T}(\psi, P_{\theta_0}) = I_d + O(n^{-1}), \qquad \text{(A.6)}
$$

with I_d representing the $(d \times d)$ identity matrix.

Proof. From the definition of $M_{i,T}(\psi, P_{\theta_0})$ in (11), we have:

$$
\begin{aligned}
& M_{i,T}^{-1}(\psi, P_{\theta_0}) M_{j,T}(\psi, P_{\theta_0}) \\
= \; & \left\{M_{j,T}(\psi, P_{\theta_0}) - [M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})]\right\}^{-1} M_{j,T}(\psi, P_{\theta_0})
\end{aligned}
$$

$$
\begin{aligned}
&= \left(M_{j,T}(\psi, P_{\theta_0}) \left\{I_d - M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right]\right\}\right)^{-1} M_{j,T}(\psi, P_{\theta_0}) \\
&= \left\{I_d - M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right]\right\}^{-1} M_{j,T}^{-1}(\psi, P_{\theta_0}) M_{j,T}(\psi, P_{\theta_0}) \\
&= \left\{I_d - M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right]\right\}^{-1}. \qquad \text{(A.7)}
\end{aligned}
$$

By a Taylor expansion in (A.7), we get

$$
\begin{aligned}
& \left\{I_d - M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right]\right\}^{-1} \\
= \;& I_d + M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right] + \left(M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right]\right)^2 \\
+ \;& \left(M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right]\right)^3 + \cdots \\
= \;& I_d + M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right] \Big\{ I_d + M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right] \\
+ \;& \left(M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right]\right)^2 + \cdots \Big\} \\
= \;& I_d + M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right] \left\{I_d - M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right]\right\}^{-1}.
\end{aligned}
\tag{A.8}
$$

From Assumption A, we know that $M_{j,T}^{-1}(\psi, P_{\theta_0}) = O(1)$. Under Assumption D($iv$), we get

$$
||M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right]|| = O(n^{-1}). \tag{A.9}
$$

By (A.7), (A.8) and (A.9), we finally find

$$
\begin{aligned}
& ||M_{i,T}^{-1}(\psi, P_{\theta_0}) M_{j,T}(\psi, P_{\theta_0}) - I_d|| \\
= \;& ||M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right] \left\{I_d - M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right]\right\}^{-1} || \\
\leq \;& ||M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right]|| \; || \left\{I_d - M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right]\right\}^{-1} || \\
= \;& ||M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right]|| \; ||I_d + M_{j,T}^{-1}(\psi, P_{\theta_0}) \left[M_{j,T}(\psi, P_{\theta_0}) - M_{i,T}(\psi, P_{\theta_0})\right] + \cdots || \\
= \;& O(n^{-1}).
\end{aligned}
\tag{A.10}
$$

Thus,

$$M_{i,T}^{-1}(\psi, P_{\theta_0}) M_{j,T}(\psi, P_{\theta_0}) = I_d + O(n^{-1}). \tag{A.11}$$

$\square$

Lemma A.2. *Under Assumptions A-D,*

(i) $g_{i,T}(\psi, P_{\theta_0})$ *and* $\gamma_{i,j,T}(\psi, P_{\theta_0})$, $1 \le i < j \le n$ *are asymptotically pairwise uncorrelated. The mean, variance, the standardized third and fourth cumulant of* $U_{n,T}$ *(defined by (A.4)) are given by the following expressions.*

(ii) Mean:

$$\mu_{n,T} = \mathbb{E}\left[U_{n,T}(\psi, \theta_0)\right] = 0. \tag{A.12}$$

(iii) Variance:

$$\begin{aligned} \sigma_{n,T}^2 &= Var\left[U_{n,T}(\psi, \theta_0)\right] \\ &= \frac{4}{n}\sigma_g^2 + \frac{4}{n^2(n-1)^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}\left[\gamma_{i,j,T}^2(\psi, P_{\theta_0})\right] + O(n^{-3}), \end{aligned} \tag{A.13}$$

where

$$\sigma_g^2 = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[g_{i,T}^2(\psi, P_{\theta_0})\right]. \tag{A.14}$$

(iv) Third standardized cumulant is such that

$$\tilde{\kappa}_{n,T}^{(3)} = \mathbb{E}\left[U_{n,T}^3(\psi, \theta_0)/\sigma_{n,T}^3\right] = n^{-1/2}\kappa_{n,T}^{(3)} + O(n^{-3/2})$$

where:

$$\kappa_{n,T}^{(3)} = \sigma_g^{-3}(\overline{g^3} + 3\overline{g_1 g_2 \gamma_{12}}), \tag{A.15}$$

$$\overline{g^3} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[g_{i,T}^3(\psi, P_{\theta_0})\right], \tag{A.16}$$

$$\overline{g_1 g_2 \gamma_{12}} = \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathbb{E}\left[g_{i,T}\left(\psi, P_{\theta_0}\right) g_{j,T}\left(\psi, P_{\theta_0}\right) \gamma_{i,j,T}(\psi, P_{\theta_0})\right]. \tag{A.17}$$

(v) *The fourth standardized cumulant is such that:*

$$\tilde{\kappa}_{n,T}^{(4)} = \mathbb{E}\left[U_{n,T}^4(\psi, \theta_0)/\sigma_{n,T}^4\right] - 3 = n^{-1}\kappa_{n,T}^{(4)} + O(n^{-2}),$$

where

$$\kappa_{n,T}^{(4)} = \sigma_g^{-4}(\overline{g^4} + 12\overline{g_1 g_2 \gamma_{13}\gamma_{23}} + 12\overline{g_1^2 g_2 \gamma_{12}}) - 3, \tag{A.18}$$

$$\overline{g^4} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[g_{i,T}^4(\psi, P_{\theta_0})\right], \tag{A.19}$$

$$\overline{g_1 g_2 \gamma_{13}\gamma_{23}} = \frac{2}{n(n-1)(n-2)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{\substack{k=1\\k\neq i,j}}^{n} \mathbb{E}\left[g_{i,T}\left(\psi, P_{\theta_0}\right) g_{j,T}\left(\psi, P_{\theta_0}\right) \gamma_{i,k,T}(\psi, P_{\theta_0})\gamma_{j,k,T}(\psi, P_{\theta_0})\right], \tag{A.20}$$

$$\overline{g_1^2 g_2 \gamma_{12}} = \frac{1}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathbb{E}\left[\left(g_{i,T}\left(\psi, P_{\theta_0}\right) + g_{j,T}\left(\psi, P_{\theta_0}\right)\right) g_{i,T}\left(\psi, P_{\theta_0}\right) g_{j,T}\left(\psi, P_{\theta_0}\right) \gamma_{i,j,T}(\psi, P_{\theta_0})\right]. \tag{A.21}$$

Proof. (i) Under Lemma A.1 and the fact that $IF_{i,T}(\psi, P_{\theta_0}) = O_P(1)$, using the definitions of $M_{i,T}(\psi, P_{\theta_0})$ in (11), $\gamma_{i,j,T}(\psi, P_{\theta_0})$ in (18), and $g_{i,T}(\psi, P_{\theta_0})$ in (17), we get for the conditional expectation:

$$\begin{aligned}
& \mathbb{E}\left[\gamma_{i,j,T}(\psi, P_{\theta_0})\Big|\frac{1}{T-1}\sum_{t=1}^{T}\psi_{i,t}(\theta_0)\right] \\
= \ & \mathbb{E}\left[\frac{1}{2}\left(\varphi'_{i,j,T}(\psi, P_{\theta_0})\frac{\partial q(\vartheta)}{\partial \vartheta}\Big|_{\theta=\theta_0} + IF'_{i,T}(\psi, P_{\theta_0})\frac{\partial^2 q(\vartheta)}{\partial \vartheta \partial \vartheta'}\Big|_{\theta=\theta_0} IF_{j,T}(\psi, P_{\theta_0})\right)\Big|\frac{1}{T-1}\sum_{t=1}^{T}\psi_{i,t}(\theta_0)\right] \\
= \ & \frac{1}{2}IF'_{i,T}(\psi, P_{\theta_0})\frac{\partial q(\vartheta)}{\partial \vartheta}\Big|_{\theta=\theta_0}
\end{aligned}$$

$$
\begin{aligned}
&+ \quad \frac{1}{2}\left(M_{i,T}^{-1}(\psi, P_{\theta_0})\mathbb{E}\left[(T-1)^{-1}\sum_{t=1}^{T}\frac{\partial\psi_{j,t}(\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\right] IF_{i,T}(\psi, P_{\theta_0})\right)'\frac{\partial q(\vartheta)}{\partial\vartheta}\Big|_{\theta=\theta_0} \\
&= \quad \frac{1}{2}IF_{i,T}'(\psi, P_{\theta_0})\frac{\partial q(\vartheta)}{\partial\vartheta}\Big|_{\theta=\theta_0} - \frac{1}{2}\left(M_{i,T}^{-1}(\psi, P_{\theta_0})M_{j,T}(\psi, P_{\theta_0})IF_{i,T}(\psi, P_{\theta_0})\right)'\frac{\partial q(\vartheta)}{\partial\vartheta}\Big|_{\theta=\theta_0} \\
&= \quad \frac{1}{2}IF_{i,T}'(\psi, P_{\theta_0})\frac{\partial q(\vartheta)}{\partial\vartheta}\Big|_{\theta=\theta_0} - \frac{1}{2}IF_{i,T}'(\psi, P_{\theta_0})\frac{\partial q(\vartheta)}{\partial\vartheta}\Big|_{\theta=\theta_0} + O_P(n^{-1}) \\
&= \quad O_P(n^{-1}). \qquad\qquad (A.22)
\end{aligned}
$$

So we deduce that $g_{i,T}(\psi, P_{\theta_0})$ in (17) and $\gamma_{i,j,T}(\psi, P_{\theta_0})$ in (18), $1 \le i < j \le n$ are pairwise uncorrelated, up to an $O_P(n^{-1})$ term.

(ii) From the independence of the estimating function and using (A.4) and (16), we get for the mean $\mu_{n,T}$ of $U_{n,T}(\psi, \theta_0)$:

$$
\begin{aligned}
\mu_{n,T} \quad &= \quad \mathbb{E}\left[U_{n,T}(\psi, \theta_0)\right] \\
&= \quad \mathbb{E}\left[\frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} h_{i,j,T}(\psi, P_{\theta_0})\right] \\
&= \quad \frac{1}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathbb{E}\left[\left\{IF_{i,T}'(\psi, P_{\theta_0}) + IF_{j,T}'(\psi, P_{\theta_0}) + \varphi_{i,j,T}'(\psi, P_{\theta_0})\right\}\frac{\partial q(\vartheta)}{\partial\vartheta}\Big|_{\theta=\theta_0}\right. \\
&+ \quad \left. IF_{i,T}'(\psi, P_{\theta_0})\frac{\partial^2 q(\vartheta)}{\partial\vartheta\partial\vartheta'}\Big|_{\theta=\theta_0} IF_{j,T}(\psi, P_{\theta_0})\right] \\
&= \quad 0. \qquad\qquad (A.23)
\end{aligned}
$$

(iii) From the asymptotic pairwise uncorrelation of $g_{i,T}(\psi, P_{\theta_0})$ and $\gamma_{i,j,T}(\psi, P_{\theta_0})$, we know that

$$
\mathbb{E}\left[g_{i,T}(\psi, P_{\theta_0})\gamma_{i,j,T}(\psi, P_{\theta_0})\right] = O(n^{-1}).
$$

Then by using (A.4) and (A.14), we get for the variance $\sigma_{n,T}^2$ of $U_{n,T}(\psi, \theta_0)$:

$$
\sigma_{n,T}^2 \quad = \quad Var\left[U_{n,T}(\psi, \theta_0)\right]
$$

$$
\begin{aligned}
&= \frac{4}{n^2}\sum_{i=1}^{n}\mathbb{E}\left[g_{i,T}^2(\psi,P_{\theta_0})\right] + \frac{4}{n^2(n-1)^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathbb{E}\left[\gamma_{i,j,T}^2(\psi,P_{\theta_0})\right] \\
&+ \frac{8}{n^2(n-1)^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathbb{E}\left[g_{i,T}(\psi,P_{\theta_0})\gamma_{i,j,T}(\psi,P_{\theta_0})\right] \\
&= \frac{4}{n}\sigma_g^2 + \frac{4}{n^2(n-1)^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathbb{E}\left[\gamma_{i,j,T}^2(\psi,P_{\theta_0})\right] + O(n^{-3}) \qquad \text{(A.24)}
\end{aligned}
$$

(iv) Due to the asymptotic pairwise uncorrelation of $g_{i,T}(\psi, P_{\theta_0})$ and $\gamma_{i,j,T}(\psi, P_{\theta_0})$, several expectations in the calculation of cumulants are of order $O(n^{-1})$, for example $\mathbb{E}\left[g_{i,T}^2(\psi,P_{\theta_0})\gamma_{i,j,T}(\psi,P_{\theta_0})\right]$, $1 \le i < j \le n$. Making use of (A.4), (A.13), (A.16), (A.17), and (A.15), we get for the third cumulant of $\sigma_{n,T}^{-1}U_{n,T}(\psi,\theta_0)$:

$$
\begin{aligned}
\tilde{\kappa}_{n,T}^{(3)} &= \mathbb{E}\left[U_{n,T}^3(\psi,\theta_0)/\sigma_{n,T}^3\right] \\
&= \sigma_{n,T}^{-3}\mathbb{E}\left[\left(\frac{2}{n}\sum_{i=1}^{n}g_{i,T}(\psi,P_{\theta_0}) + \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\gamma_{i,j,T}(\psi,P_{\theta_0})\right)^3\right] \\
&= \sigma_{n,T}^{-3}\frac{8}{n^3(n-1)^3}\sum_{i=1}^{n}(n-1)^3\mathbb{E}\left[g_{i,T}^3(\psi,P_{\theta_0})\right] + \sigma_{n,T}^{-3}\frac{8}{n^3(n-1)^3}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathbb{E}\left[\gamma_{i,j,T}^3(\psi,P_{\theta_0})\right] \\
&+ \sigma_{n,T}^{-3}\frac{8}{n^3(n-1)^3}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}6(n-1)^2\mathbb{E}\left[g_{i,T}(\psi,P_{\theta_0})\,g_{j,T}(\psi,P_{\theta_0})\,\gamma_{i,j,T}(\psi,P_{\theta_0})\right] \\
&+ \sigma_{n,T}^{-3}\frac{8}{n^3(n-1)^3}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}3(n-1)\mathbb{E}\left[\left\{g_{i,T}(\psi,P_{\theta_0}) + g_{j,T}(\psi,P_{\theta_0})\right\}\gamma_{i,j,T}^2(\psi,P_{\theta_0})\right] \\
&= \sigma_{n,T}^{-3}\frac{8}{n^3(n-1)^3}\sum_{i=1}^{n}(n-1)^3\mathbb{E}\left[g_{i,T}^3(\psi,P_{\theta_0})\right] \\
&+ \sigma_{n,T}^{-3}\frac{8}{n^3(n-1)^3}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}6(n-1)^2\mathbb{E}\left[g_{i,T}(\psi,P_{\theta_0})\,g_{j,T}(\psi,P_{\theta_0})\,\gamma_{i,j,T}(\psi,P_{\theta_0})\right] + O(n^{-3/2}) \\
&= \sigma_g^{-3}n^{-1/2}(\overline{g^3} + 3\overline{g_1g_2\gamma_{12}}) + O(n^{-3/2}) \\
&= n^{-1/2}\kappa_{n,T}^{(3)} + O(n^{-3/2}). \qquad \text{(A.25)}
\end{aligned}
$$

(v) Similarly, making use of (A.4), (A.13), (A.19), (A.20), (A.21), and (A.18), we get for the fourth cumulant of $\sigma_{n,T}^{-1} U_{n,T}(\psi, \theta_0)$:

$$
\begin{aligned}
&\tilde{\kappa}_{n,T}^{(4)} \\
= \;& \mathbb{E}\left[U_{n,T}^4(\psi,\theta_0)/\sigma_{n,T}^4\right] - 3 \\
= \;& -3 + \sigma_{n,T}^{-4}\mathbb{E}\left[\left(\frac{2}{n}\sum_{i=1}^{n} g_{i,T}\left(\psi, P_{\theta_0}\right) + \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\gamma_{i,j,T}(\psi, P_{\theta_0})\right)^4\right] \\
= \;& -3 + \sigma_{n,T}^{-4}\frac{16}{n^4(n-1)^4}\sum_{i=1}^{n}(n-1)^4\mathbb{E}\left[g_{i,T}^4\left(\psi, P_{\theta_0}\right)\right] + \sigma_{n,T}^{-4}\frac{16}{n^4(n-1)^4}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathbb{E}\left[\gamma_{i,j,T}^4(\psi, P_{\theta_0})\right] \\
+ \;& \sigma_{n,T}^{-4}\frac{16}{n^4(n-1)^4}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 4(n-1)\mathbb{E}\left[\left(g_{i,T}\left(\psi, P_{\theta_0}\right) + g_{j,T}\left(\psi, P_{\theta_0}\right)\right)\gamma_{i,j,T}^3(\psi, P_{\theta_0})\right] \\
+ \;& \sigma_{n,T}^{-4}\frac{16}{n^4(n-1)^4}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 6(n-1)^2\mathbb{E}\left[\left(g_{i,T}^2\left(\psi, P_{\theta_0}\right) + g_{j,T}^2\left(\psi, P_{\theta_0}\right)\right)\gamma_{i,j,T}^2(\psi, P_{\theta_0})\right] \\
+ \;& \sigma_{n,T}^{-4}\frac{16}{n^4(n-1)^4}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 12(n-1)^3\mathbb{E}\left[\left(g_{i,T}^2\left(\psi, P_{\theta_0}\right) g_{j,T}\left(\psi, P_{\theta_0}\right) + g_{i,T}\left(\psi, P_{\theta_0}\right) g_{j,T}^2\left(\psi, P_{\theta_0}\right)\right)\gamma_{i,j,T}(\psi, P_{\theta_0})\right] \\
+ \;& \sigma_{n,T}^{-4}\frac{16}{n^4(n-1)^4}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 6(n-1)^2\mathbb{E}\left[g_{i,T}\left(\psi, P_{\theta_0}\right) g_{j,T}\left(\psi, P_{\theta_0}\right)\gamma_{i,j,T}^2(\psi, P_{\theta_0})\right] \\
+ \;& \sigma_{n,T}^{-4}\frac{16}{n^4(n-1)^4}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 6(n-1)^4\mathbb{E}\left[g_{i,T}^2\left(\psi, P_{\theta_0}\right) g_{j,T}^2\left(\psi, P_{\theta_0}\right)\right] \\
+ \;& \sigma_{n,T}^{-4}\frac{16}{n^4(n-1)^4}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{\substack{k=1 \\ k\neq i,j}}^{n} 24(n-1)^2\mathbb{E}\left[g_{i,T}\left(\psi, P_{\theta_0}\right) g_{j,T}\left(\psi, P_{\theta_0}\right)\gamma_{i,k,T}(\psi, P_{\theta_0})\gamma_{j,k,T}(\psi, P_{\theta_0})\right] \\
= \;& -3 + \sigma_{n,T}^{-4}\frac{16}{n^4(n-1)^4}\sum_{i=1}^{n}(n-1)^4\mathbb{E}\left[g_{i,T}^4\left(\psi, P_{\theta_0}\right)\right] \\
+ \;& \sigma_{n,T}^{-4}\frac{16}{n^4(n-1)^4}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 6(n-1)^4\mathbb{E}\left[g_{i,T}^2\left(\psi, P_{\theta_0}\right) g_{j,T}^2\left(\psi, P_{\theta_0}\right)\right] \\
+ \;& \sigma_{n,T}^{-4}\frac{16}{n^4(n-1)^4}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{\substack{k=1 \\ k\neq i,j}}^{n} 24(n-1)^2\mathbb{E}\left[g_{i,T}\left(\psi, P_{\theta_0}\right) g_{j,T}\left(\psi, P_{\theta_0}\right)\gamma_{i,k,T}(\psi, P_{\theta_0})\gamma_{j,k,T}(\psi, P_{\theta_0})\right] \\
+ \;& \sigma_{n,T}^{-4}\frac{16}{n^4(n-1)^4}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 12(n-1)^3\mathbb{E}\left[\left(g_{i,T}\left(\psi, P_{\theta_0}\right) + g_{j,T}\left(\psi, P_{\theta_0}\right)\right) g_{i,T}\left(\psi, P_{\theta_0}\right) g_{j,T}\left(\psi, P_{\theta_0}\right)\gamma_{i,j,T}(\psi, P_{\theta_0})\right] \\
+ \;& O(n^{-2})
\end{aligned}
$$

$$
\begin{aligned}
&= \sigma_g^{-4} n^{-1}(\overline{g^4} + 12\overline{g_1 g_2 \gamma_{13}\gamma_{23}} + 12\overline{g_1^2 g_2 \gamma_{12}}) - 3n^{-1} + O(n^{-2}) \\
&= n^{-1}\kappa_{n,T}^{(4)} + O(n^{-2}).
\end{aligned}
$$

□

Now we can prove Proposition 3. Let $\Psi_{n,T}$ be the characteristic function (c.f.) of $\sigma_{n,T}^{-1}\{q[\vartheta(P_{n,T})] - q[\vartheta(P_{\theta_0})]\}$,

$$
\Psi_{n,T}(z) = \mathbb{E}\left[\exp\left(\iota t \sigma_{n,T}^{-1}\{q[\vartheta(P_{n,T})] - q[\vartheta(P_{\theta_0})]\}\right)\right], \tag{A.26}
$$

where $\iota^2 = -1$. Making use of $\kappa_{n,T}^{(3)}$ in (A.15) and $\kappa_{n,T}^{(4)}$ in (A.18), we define

$$
\Psi_{n,T}^*(z) = \left\{1 + n^{-1/2}\kappa_{n,T}^{(3)}\frac{(\iota z)^3}{6} + n^{-1}\kappa_{n,T}^{(4)}\frac{(\iota z)^4}{24} + n^{-1}\left(\kappa_{n,T}^{(3)}\right)^2\frac{(\iota z)^6}{72}\right\} e^{-z^2/2}. \tag{A.27}
$$

as the approximate c.f.. To prove (19) in Proposition 3, we use Esseen smoothing lemma as in Feller (1971) and show that there exist sequences $\{Z_n\}$ and $\{\varepsilon'_n\}$ such that $n^{-1}Z_n \to \infty$, $\varepsilon'_n \to 0$, and

$$
\int_{-Z_n}^{Z_n}\left|\frac{\Psi_{n,T}(z) - \Psi_{n,T}^*(z)}{z}\right| dz \le \varepsilon'_n n^{-1}. \tag{A.28}
$$

We proceed along the same lines as in Bickel et al. (1986). We work on a compact subset of $\mathbb{R}$ and we consider the c.f. for small $|z|$. Then, we prove Lemma A.3, which essentially shows the validity of the Edgeworth by means of the Esseen lemma. With this regard, we flag that we need to prove the Esseen lemma within our setting (we are dealing with independent but not identically distributed random variables) and we cannot invoke directly the results in the the paper by Bickel et al. (1986). To this end, we prove a new result similar to the one in Lemma 2.1 of the last paper, adapting their proof to our context—see Lemma A.3 below. Finally, the application of the derived results concludes the proof.

Lemma A.3. *Under Assumptions A-D, there exists a sequence $\varepsilon_n'' \downarrow 0$ such that for*

$$z_n = n^{(r-1)/r}(\log n)^{-1}, \tag{A.29}$$

$$\int_{-z_n}^{z_n} \Big| \frac{\Psi_{n,T}(z) - \Psi^*_{n,T}(z)}{z} \Big| dz \le \varepsilon_n'' n^{-1}. \tag{A.30}$$

Proof. For the sake of readability, we split the proof in five steps. At the beginning of each step, we explain the goal of the derivation.

Step 1. We approximate the characteristic function (c.f.) of $\hat{\theta}_{n,T}$ via the c.f. of $U_{n,T}$, up to the suitable order. This yields (A.33).

Let us decompose the U-statistic $U_{n,T}$ in (A.4) as $U_{n,T} = U_{1,n,T} + U_{2,n,T}$ with $U_{1,n,T} = \frac{2}{n}\sum_{i=1}^{n} g_{i,T}(\psi, P_{\theta_0})$ and $U_{2,n,T} = \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \gamma_{i,j,T}(\psi, P_{\theta_0})$. Making use of (2.6), (2.7) as in Bickel et al. (1986), (A.5) and the U-statistic decomposition, we can write

$$\begin{aligned}\Psi_{n,T}(z) &= \mathbb{E}\left[\exp\left(\iota z\sigma_{n,T}^{-1}U_{1,n,T}\right)\left(1 + \iota z\sigma_{n,T}^{-1}U_{2,n,T} - \frac{1}{2}z^2\sigma_{n,T}^{-2}U_{2,n,T}^2\right)\right] \\ &\quad + O(E|z\sigma_{n,T}^{-1}U_{2,n,T}|^{2+\delta}) + O(n^{-3/2}|z\sigma_{n,T}^{-1}|),\end{aligned} \tag{A.31}$$

for $\delta \in (0,1]$. Let

$$\Psi_{g,i,T}(z) = \mathbb{E}\left[\exp\left(\iota z\sigma_{n,T}^{-1}\frac{2}{n}g_{i,T}(\psi, P_{\theta_0})\right)\right] \tag{A.32}$$

be the c.f. of $\sigma_{n,T}^{-1}\frac{2}{n}g_{i,T}(\psi, P_{\theta_0})$. In view of (A.13), and the fact that $\mathbb{E}\left[|U_{2,n,T}|^{2+\delta}\right] = O(n^{2+\delta})$ (see Callaert and Janssen (1978)), we rewrite (A.31) as

$$\begin{aligned}&\Psi_{n,T}(z) \\ =\ & \prod_{i=1}^{n}\Psi_{g,i,T}(z) \\ +\ & \sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\left\{\prod_{\substack{i=1\\ i\neq j,k}}^{n}\Psi_{g,i,T}(z)\right\}\left\{\iota z\sigma_{n,T}^{-1}\mathbb{E}\left[\exp\left(\iota z\sigma_{n,T}^{-1}\frac{2}{n}\left\{g_{j,T}(\psi, P_{\theta_0}) + g_{k,T}(\psi, P_{\theta_0})\right\}\right)\frac{2\gamma_{j,k,T}(\psi, P_{\theta_0})}{n(n-1)}\right]\right.\end{aligned}$$

$$
\begin{aligned}
- & \quad \frac{1}{2} z^2 \sigma_{n,T}^{-2} \mathbb{E}\left[\exp\left(\iota z \sigma_{n,T}^{-1} \frac{2}{n}\left\{g_{j,T}(\psi, P_{\theta_0}) + g_{k,T}(\psi, P_{\theta_0})\right\}\right) \frac{4}{n^2(n-1)^2} \gamma_{j,k,T}^2(\psi, P_{\theta_0})\right]\Bigg\} \\
- & \quad \sum_{j=1}^{n-2} \sum_{k=j+1}^{n-1} \sum_{m=k+1}^{n} \left\{\prod_{\substack{i=1 \\ i \neq j,k,m}}^{n} \Psi_{g,i,T}(z)\right\} z^2 \sigma_{n,T}^{-2} \frac{4}{n^2(n-1)^2} \\
\times & \quad \mathbb{E}\Bigg[\exp\left(\iota z \sigma_{n,T}^{-1} \frac{2}{n}\left\{g_{j,T}(\psi, P_{\theta_0}) + g_{k,T}(\psi, P_{\theta_0}) + g_{m,T}(\psi, P_{\theta_0})\right\}\right) \{\gamma_{j,k,T}(\psi, P_{\theta_0}) \gamma_{j,m,T}(\psi, P_{\theta_0}) \\
+ & \quad \gamma_{j,m,T}(\psi, P_{\theta_0}) \gamma_{k,m,T}(\psi, P_{\theta_0}) + \gamma_{j,k,T}(\psi, P_{\theta_0}) \gamma_{k,m,T}(\psi, P_{\theta_0})\}\Bigg] \\
- & \quad \sum_{j=1}^{n-3} \sum_{k=j+1}^{n-2} \sum_{m=k+1}^{n-1} \sum_{l=m+1}^{n} \left\{\prod_{\substack{i=1 \\ i \neq j,k,m,l}}^{n} \Psi_{g,i,T}(z)\right\} z^2 \sigma_{n,T}^{-2} \frac{4}{n^2(n-1)^2} \\
\times & \quad \mathbb{E}\Bigg[\exp\left(\frac{2\iota z}{n\sigma_{n,T}}\left\{g_{j,T}(\psi, P_{\theta_0}) + g_{k,T}(\psi, P_{\theta_0}) + g_{m,T}(\psi, P_{\theta_0}) + g_{l,T}(\psi, P_{\theta_0})\right\}\right) \{\gamma_{j,k,T}(\psi, P_{\theta_0}) \gamma_{m,l,T}(\psi, P_{\theta_0}) \\
+ & \quad \gamma_{j,m,T}(\psi, P_{\theta_0}) \gamma_{k,l,T}(\psi, P_{\theta_0}) + \gamma_{j,l,T}(\psi, P_{\theta_0}) \gamma_{k,m,T}(\psi, P_{\theta_0})\}\Bigg] \\
+ & \quad O(|n^{-1/2} z|^{2+\delta} + |n^{-1} z|).
\end{aligned} \tag{A.33}
$$

Step 2. To match the expression of $\Psi_{n,T}^*(z)$ as in (A.27), we need an expansion for each of the terms in (A.33). To this end, we work on the exponential terms in (A.33). Here, we focus on the first exponential term and get (A.34). We can repeat the computations for the other terms, and those tedious developments follow similar arguments. We expand the first exponential term in (A.33) by using (2.7) in Bickel et al. (1986), (A.13) and Assumption D. Thus, we obtain

$$
\begin{aligned}
& \quad \mathbb{E}\left[\exp\left(\iota z \sigma_{n,T}^{-1} \frac{2}{n}\left\{g_{j,T}(\psi, P_{\theta_0}) + g_{k,T}(\psi, P_{\theta_0})\right\}\right) \frac{2}{n(n-1)} \gamma_{j,k,T}(\psi, P_{\theta_0})\right] \\
= & \quad \mathbb{E}\Bigg[\left(\exp\left\{\iota z \sigma_{n,T}^{-1} \frac{2}{n} g_{j,T}(\psi, P_{\theta_0})\right\} - 1 - \iota z \sigma_{n,T}^{-1} \frac{2}{n} g_{j,T}(\psi, P_{\theta_0})\right) \\
\times & \quad \left(\exp\left\{\iota z \sigma_{n,T}^{-1} \frac{2}{n} g_{k,T}(\psi, P_{\theta_0})\right\} - 1 - \iota z \sigma_{n,T}^{-1} \frac{2}{n} g_{k,T}(\psi, P_{\theta_0})\right) \frac{2}{n(n-1)} \gamma_{j,k,T}(\psi, P_{\theta_0})\Bigg] \\
+ & \quad \mathbb{E}\left[\iota z \sigma_{n,T}^{-1} \frac{4 g_{k,T}(\psi, P_{\theta_0}) \gamma_{j,k,T}(\psi, P_{\theta_0})}{n^2(n-1)} \left(\exp\left\{\iota z \sigma_{n,T}^{-1} \frac{2}{n} g_{j,T}(\psi, P_{\theta_0})\right\} - \sum_{\nu=0}^{2} \frac{\left\{\iota z \sigma_{n,T}^{-1} \frac{2}{n} g_{j,T}(\psi, P_{\theta_0})\right\}^\nu}{\nu!}\right)\right] \\
+ & \quad \mathbb{E}\left[\iota z \sigma_{n,T}^{-1} \frac{4 g_{j,T}(\psi, P_{\theta_0}) \gamma_{j,k,T}(\psi, P_{\theta_0})}{n^2(n-1)} \left(\exp\left\{\iota z \sigma_{n,T}^{-1} \frac{2}{n} g_{k,T}(\psi, P_{\theta_0})\right\} - \sum_{\nu=0}^{2} \frac{\left\{\iota z \sigma_{n,T}^{-1} \frac{2}{n} g_{k,T}(\psi, P_{\theta_0})\right\}^\nu}{\nu!}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&- \mathbb{E}\left[z^2\sigma_{n,T}^{-2}\frac{4}{n^2}g_{j,T}(\psi,P_{\theta_0})g_{k,T}(\psi,P_{\theta_0})\frac{2}{n(n-1)}\gamma_{j,k,T}(\psi,P_{\theta_0})\right] \\
&- \mathbb{E}\left[\left(\iota z^3\sigma_{n,T}^{-3}\frac{4}{n^3}g_{j,T}^2(\psi,P_{\theta_0})g_{k,T}(\psi,P_{\theta_0})+\iota z^3\sigma_{n,T}^{-3}\frac{4}{n^3}g_{j,T}(\psi,P_{\theta_0})g_{k,T}^2(\psi,P_{\theta_0})\right)\frac{2}{n(n-1)}\gamma_{j,k,T}(\psi,P_{\theta_0})\right] \\
&= -z^2\sigma_{n,T}^{-2}\frac{8}{n^3(n-1)}\mathbb{E}\left[g_{j,T}(\psi,P_{\theta_0})g_{k,T}(\psi,P_{\theta_0})\gamma_{j,k,T}(\psi,P_{\theta_0})\right] \\
&- \iota z^3\sigma_{n,T}^{-3}\frac{8}{n^4(n-1)}\mathbb{E}\left[\left\{g_{j,T}^2(\psi,P_{\theta_0})g_{k,T}(\psi,P_{\theta_0})+g_{j,T}(\psi,P_{\theta_0})g_{k,T}^2(\psi,P_{\theta_0})\right\}\gamma_{j,k,T}(\psi,P_{\theta_0})\right] \\
&+ O(n^{-4}z^4+n^{-2}|n^{-1/2}z|^{3+\delta}), \qquad \text{(A.34)}
\end{aligned}
$$

with $\delta \in (0,1]$. Similarly, we expand all the other exponentials in (A.33) and after some algebraic simplifications, we get

$$
\begin{aligned}
&\Psi_{n,T}(z) \\
&= \left\{\prod_{i=1}^{n}\Psi_{g,i,T}(z)\right\}+\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\left\{\prod_{\substack{i=1\\ i\neq j,k}}^{n}\Psi_{g,i,T}(z)\right\}\left(\frac{-8\iota z^3\mathbb{E}\left[g_{j,T}(\psi,P_{\theta_0})g_{k,T}(\psi,P_{\theta_0})\gamma_{j,k,T}(\psi,P_{\theta_0})\right]}{n^3(n-1)\sigma_{n,T}^3}\right. \\
&+ z^4\sigma_{n,T}^{-4}\frac{8\mathbb{E}\left[\left\{g_{j,T}^2(\psi,P_{\theta_0})g_{k,T}(\psi,P_{\theta_0})+g_{j,T}(\psi,P_{\theta_0})g_{k,T}^2(\psi,P_{\theta_0})\right\}\gamma_{j,k,T}(\psi,P_{\theta_0})\right]}{n^4(n-1)} \\
&- \left.\frac{2}{n^2(n-1)^2}z^2\sigma_{n,T}^{-2}\mathbb{E}\left[\gamma_{j,k,T}^2(\psi,P_{\theta_0})\right]\right) \\
&+ \sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\sum_{\substack{m=1\\ m\neq j,k}}^{n}\left\{\prod_{\substack{i=1\\ i\neq j,k,m}}^{n}\Psi_{g,i,T}(z)\right\}\frac{16z^4\mathbb{E}\left[g_{j,T}(\psi,P_{\theta_0})g_{k,T}(\psi,P_{\theta_0})\gamma_{j,m,T}(\psi,P_{\theta_0})\gamma_{k,m,T}(\psi,P_{\theta_0})\right]}{n^4(n-1)^2\sigma_{n,T}^4} \\
&- \sum_{j=1}^{n-3}\sum_{k=j+1}^{n-2}\sum_{m=k+1}^{n-1}\sum_{l=m+1}^{n}\left\{\prod_{\substack{i=1\\ i\neq j,k,m,l}}^{n}\Psi_{g,i,T}(z)\right\}z^6\sigma_{n,T}^{-6}\frac{64}{n^6(n-1)^2} \\
&\times \left\{\mathbb{E}\left[g_{j,T}(\psi,P_{\theta_0})g_{k,T}(\psi,P_{\theta_0})\gamma_{j,k,T}(\psi,P_{\theta_0})\right]\mathbb{E}\left[g_{m,T}(\psi,P_{\theta_0})g_{l,T}(\psi,P_{\theta_0})\gamma_{m,l,T}(\psi,P_{\theta_0})\right]\right. \\
&+ \mathbb{E}\left[g_{j,T}(\psi,P_{\theta_0})g_{m,T}(\psi,P_{\theta_0})\gamma_{j,m,T}(\psi,P_{\theta_0})\right]\mathbb{E}\left[g_{k,T}(\psi,P_{\theta_0})g_{l,T}(\psi,P_{\theta_0})\gamma_{k,l,T}(\psi,P_{\theta_0})\right] \\
&+ \left.\mathbb{E}\left[g_{j,T}(\psi,P_{\theta_0})g_{l,T}(\psi,P_{\theta_0})\gamma_{j,l,T}(\psi,P_{\theta_0})\right]\mathbb{E}\left[g_{k,T}(\psi,P_{\theta_0})g_{m,T}(\psi,P_{\theta_0})\gamma_{k,m,T}(\psi,P_{\theta_0})\right]\right\} \\
&+ O\Big(\prod_{i=1}^{n}\Psi_{g,i,T}(z)|z|\mathcal{P}(|z|)n^{-1-\delta/2}+|n^{-1/2}z|^{2+\delta}+|n^{-1}z|\Big), \qquad \text{(A.35)}
\end{aligned}
$$

where $\mathcal{P}$ is a fixed polynomial.

Step 3. We need to derive the expansions (up to a suitable order) of the products of $\Psi_{g,i,T}(z)$

represented as the four curly brackets in (A.35), to have similar expressions as the terms in $\Psi^*_{n,T}(z)$; see (A.27). To achieve it, we introduce the approximate c.f. of $\sigma_g^{-1}\sum_{i=1}^n g_{i,T}(\psi, P_{\theta_0})$ in (A.37) and find a connection to the c.f. of $\sigma_{n,T}^{-1}\sum_{i=1}^n g_{i,T}(\psi, P_{\theta_0})$ so that we can get the expressions of the four curly brackets in (A.35).

Let

$$\Psi_{i,T}(z) = \mathbb{E}\left[\exp\left(\iota z \sigma_g^{-1} g_{i,T}(\psi, P_{\theta_0})\right)\right] \tag{A.36}$$

denote the c.f. of $\sigma_g^{-1} g_{i,T}(\psi, P_{\theta_0})$, where σ_g^2 is defined in (A.14). For sufficient small $\varepsilon' > 0$ and for $|z| \le \varepsilon' n^{1/2}$, we get for the c.f. of $\sigma_g^{-1}\sum_{i=1}^n g_{i,T}(\psi, P_{\theta_0})$

$$\prod_{i=1}^n \Psi_{i,T}(n^{-1/2}z) = e^{-z^2/2}\left[1 - \frac{\iota\tilde{\kappa}_3}{6} n^{-1/2} z^3 + \frac{\tilde{\kappa}_4}{24} n^{-1} z^4 - \frac{\tilde{\kappa}_3^2}{72} n^{-1} z^6\right] + o(n^{-1}|z|e^{-z^2/4}), \tag{A.37}$$

where $\tilde{\kappa}_3 = n^{-1}\sigma_g^{-3}\sum_{i=1}^n \mathbb{E}\left[g_{i,T}^3(\psi, P_{\theta_0})\right]$ and $\tilde{\kappa}_4 = n^{-1}\sigma_g^{-4}\sum_{i=1}^n \mathbb{E}\left[g_{i,T}^4(\psi, P_{\theta_0})\right] - 3$. Since $\Psi_{g,i,T}(z) = \Psi_{i,T}(\sigma_g \sigma_{n,T}^{-1} \frac{2}{n} z)$, we can investigate the behaviour of the four curly brackets in (A.35), namely

$$\begin{aligned} \prod_{i=1}^n \Psi_{g,i,T}(z) &= \prod_{i=1}^n \Psi_{i,T}(n^{-1/2}z) + e^{-z^2/2}\left[\frac{1}{n(n-1)^2}\sigma_g^{-2}\sum_{u=1}^{n-1}\sum_{v=u+1}^n \mathbb{E}\left[\gamma_{u,v,T}^2(\psi, P_{\theta_0})\right]\right] z^2 \\ &+ o(n^{-1}|z|e^{-z^2/4}), \end{aligned} \tag{A.38}$$

$$\begin{aligned} \prod_{\substack{i=1 \\ i\neq j,k}}^n \Psi_{g,i,T}(z) &= \prod_{i=1}^n \Psi_{i,T}(n^{-1/2}z) \\ &+ e^{-z^2/2}\left[\sum_{u=1}^{n-1}\sum_{v=u+1}^n \frac{\mathbb{E}\left[\gamma_{u,v,T}^2(\psi, P_{\theta_0})\right]}{n(n-1)^2\sigma_g^2} + \frac{\mathbb{E}\left[g_{j,T}^2(\psi, P_{\theta_0})\right] + \mathbb{E}\left[g_{k,T}^2(\psi, P_{\theta_0})\right]}{n\sigma_g^2}\right] z^2 \\ &+ o(n^{-1}|z|e^{-z^2/4}), \end{aligned} \tag{A.39}$$

$$\prod_{\substack{i=1 \\ i\neq j,k,m}}^n \Psi_{g,i,T}(z)$$

$$
\begin{aligned}
&= \prod_{i=1}^{n} \Psi_{i,T}(n^{-1/2}z) \\
&+ e^{-z^2/2}\left[\sum_{u=1}^{n-1}\sum_{v=u+1}^{n} \frac{\mathbb{E}\left[\gamma^2_{u,v,T}(\psi, P_{\theta_0})\right]}{n(n-1)^2\sigma_g^2} + \frac{\mathbb{E}\left[g^2_{j,T}(\psi, P_{\theta_0})\right] + \mathbb{E}\left[g^2_{k,T}(\psi, P_{\theta_0})\right] + \mathbb{E}\left[g^2_{m,T}(\psi, P_{\theta_0})\right]}{n\sigma_g^2}\right] z^2 \\
&+ o(n^{-1}|z|e^{-z^2/4}), \qquad \text{(A.40)}
\end{aligned}
$$

$$
\begin{aligned}
\prod_{\substack{i=1 \\ i\neq j,k,m,l}}^{n} \Psi_{g,i,T}(z) &= \prod_{i=1}^{n} \Psi_{i,T}(n^{-1/2}z) \\
&+ e^{-z^2/2}\left[\sum_{u=1}^{n-1}\sum_{v=u+1}^{n} \frac{\mathbb{E}\left[\gamma^2_{u,v,T}(\psi, P_{\theta_0})\right]}{n(n-1)^2\sigma_g^2}\right. \\
&+ \left.\frac{\mathbb{E}\left[g^2_{j,T}(\psi, P_{\theta_0})\right] + \mathbb{E}\left[g^2_{k,T}(\psi, P_{\theta_0})\right] + \mathbb{E}\left[g^2_{m,T}(\psi, P_{\theta_0})\right] + \mathbb{E}\left[g^2_{l,T}(\psi, P_{\theta_0})\right]}{n\sigma_g^2}\right] z^2 \\
&+ o(n^{-1}|z|e^{-z^2/4}), \qquad \text{(A.41)}
\end{aligned}
$$

for $|z| \leq \varepsilon' n^{1/2}$.

Step 4. We combine the remainders and derive an expression for $\Psi_{n,T}(z)$ such that $\Psi^*_{n,T}(z)$ is the leading term and we characterize the order of the remainder. This yields (A.42).

Substitution of (A.37), (A.38), (A.39), (A.40), (A.41), (A.13), and (A.27) into (A.35) shows that for $|z| \leq \varepsilon' n^{1/2}$,

$$
\Psi_{n,T}(z) = \Psi^*_{n,T}(z) + o(n^{-1}|z|\mathcal{P}(|z|)e^{-z^2/4}) + O(|n^{-1/2}z|^{2+\delta}), \qquad \text{(A.42)}
$$

the same as (2.13) in Bickel et al. (1986).

Step 5. Moving along the lines of (2.13) in Bickel et al. (1986), we prove (A.30).

□

A.4 Proof of Proposition 4

Proof. We derive (20) by the tilted-Edgeworth technique; see e.g. Barndorff-Nielsen and Cox (1989) for a book-length presentation. Our proof follows from standard arguments, like e.g. those in Field (1982), Easton and Ronchetti (1986), and Gatto and Ronchetti (1996). The main difference between the available proofs and ours is related to the fact that we need to use our approximate c.g.f., as obtained via the (approximate) cumulants in (A.12), (A.13), (A.15) and (A.18). To this end, we set

$$\tilde{\mathcal{K}}_{n,T}(\nu) = \mu_{n,T}\nu + \frac{1}{2}n\sigma^2_{n,T}\nu^2 + \frac{1}{6}n^2\varkappa^{(3)}_{n,T}\sigma^3_{n,T}\nu^3 + \frac{1}{24}n^3\varkappa^{(4)}_{n,T}\sigma^4_{n,T}\nu^4, \tag{A.43}$$

where we use the cumulants $\varkappa^{(3)}_{n,T} = n^{-1/2}\kappa^{(3)}_{n,T}$, $\varkappa^{(4)}_{n,T} = n^{-1}\kappa^{(4)}_{n,T}$, with $\varkappa^{(3)}_{n,T}$ and $\varkappa^{(4)}_{n,T}$ being of order $O(m^{-1})$, as derived in Lemma A.2. Then, following the argument of Remark 2 in Easton and Ronchetti (1986), we obtain the required result $f_{n,T}(z) = p_{n,T}(z)\left[1 + O\left(m^{-1}\right)\right]$; see also Field (1982), p. 677. Finally, a straightforward application of Lugannani-Rice formula yields (22); see Lugannani and Rice (1980) and Gatto and Ronchetti (1996).

□

B The first and second derivatives of the log-likelihood

Lee and Yu (2010) have already provided a few calculations for the first-order asymptotics. To go further, our online materials give additional and more explicit mathematical expressions for the higher-order terms needed for the saddlepoint approximation.

We recall the following notations, which are frequently used:

$$\begin{aligned} S_n(\lambda) &= I_n - \lambda W_n \\ R_n(\rho) &= I_n - \rho M_n \end{aligned}$$

$$
\begin{aligned}
G_n(\lambda) &= W_n S_n^{-1} \\
H_n(\rho) &= M_n R_n^{-1} \\
\ddot{W}_n &= R_n W_n R_n^{-1}, \\
\ddot{G}_n(\lambda_0) &= \ddot{W}_n (I_n - \lambda_0 \ddot{W}_n)^{-1} = R_n G_n R_n^{-1}, \\
\ddot{X}_{nt} &= R_n \tilde{X}_{nt} \\
H_n^s &= H_n^{'} + H_n, \\
G_n^s &= G_n^{'} + G_n \\
\hbar_{nt}(\zeta) &= \frac{1}{m} \sum_{t=1}^{T} \left(\ddot{X}_{nt}, \ddot{G}_n(\lambda_0) \ddot{X}_{nt} \beta_0 \right)' \left(\ddot{X}_{nt}, \ddot{G}_n(\lambda_0) \ddot{X}_{nt} \beta_0 \right)
\end{aligned}
$$

B.1 The first derivative of the log-likelihood

B.1.1 Common terms

First, consider the following elements which are common to many partial derivatives that we are going to compute. To this end, we set $\xi = (\beta', \lambda, \rho)'$ and we compute:

- the matrix

$$\partial_\lambda S_n(\lambda) = -W_n \tag{B.1}$$

$$\partial_\rho R_n(\rho) = -M_n \tag{B.2}$$

- the vector

$$\partial_\xi \tilde{V}_{nt}(\xi) = (\partial_{\beta'} \tilde{V}_{nt}(\xi), \partial_\lambda \tilde{V}_{nt}(\xi), \partial_\rho \tilde{V}_{nt}(\xi)),$$

where

$$\partial_{\beta'}\tilde{V}_{nt}(\xi) \;=\; \partial_{\beta'}\left\{R_n(\rho)[S_n(\lambda)\tilde{Y}_{nt}-\tilde{X}_{nt}\beta]\right\} = -R_n(\rho)\tilde{X}_{nt} \tag{B.3}$$

and

$$\partial_{\lambda}\tilde{V}_{nt}(\xi) \;=\; \partial_{\lambda}\left\{R_n(\rho)S_n(\lambda)\tilde{Y}_{nt}\right\} = -R_n(\rho)W_n\tilde{Y}_{nt} \tag{B.4}$$

and making use of (B.2), we have

$$\begin{aligned}
\partial_{\rho}\tilde{V}_{nt}(\xi) &= \partial_{\rho}\left\{R_n(\rho)[S_n(\lambda)\tilde{Y}_{nt}-\tilde{X}_{nt}\beta]\right\} \\
&= -M_n[S_n(\lambda)\tilde{Y}_{nt}-\tilde{X}_{nt}\beta] \\
&= -M_nR_n^{-1}(\rho)R_n(\rho)[S_n(\lambda)\tilde{Y}_{nt}-\tilde{X}_{nt}\beta] \\
&= -H_n(\rho)\tilde{V}_{nt}(\xi),
\end{aligned} \tag{B.5}$$

- the vector

$$\partial_{\xi}G_n(\lambda) = (\partial_{\beta'}G_n(\lambda), \partial_{\lambda}G_n(\lambda), \partial_{\rho}G_n(\lambda)),$$

where

$$\partial_{\beta'}G_n(\lambda) = 0 \tag{B.6}$$

$$\partial_{\rho}G_n(\lambda) = 0 \tag{B.7}$$

and

$$\partial_{\lambda}G_n(\lambda) \;=\; \partial_{\lambda}(W_nS_n^{-1}) = W_n\partial_{\lambda}S_n^{-1}$$

$$
\begin{aligned}
&= W_n(-S_n^{-1}\underbrace{\partial_\lambda(S_n)}_{B.1}S_n^{-1}) \\
&= (W_nS_n^{-1})^2 = G_n^2
\end{aligned} \tag{B.8}
$$

- the vector

$$\partial_\xi H_n(\rho) = (\partial_{\beta'}H_n(\rho), \partial_\lambda H_n(\rho), \partial_\rho H_n(\rho)),$$

where

$$\partial_{\beta'}H_n(\rho) = 0 \tag{B.9}$$

$$\partial_\lambda H_n(\rho) = 0 \tag{B.10}$$

and

$$
\begin{aligned}
\partial_\rho H_n(\rho) &= \partial_\rho(M_nR_n^{-1}) = M_n\partial_\rho R_n^{-1} \\
&= M_n(-R_n^{-1}\underbrace{\partial_\rho(R_n)}_{B.2}R_n^{-1}) \\
&= (M_nR_n^{-1}(\rho))^2 = H_n^2,
\end{aligned} \tag{B.11}
$$

- the vector

$$\partial_\xi G_n^2(\lambda) = (\partial_{\beta'}G_n^2(\lambda), \partial_\lambda G_n^2(\lambda), \partial_\rho G_n^2(\lambda)),$$

where

$$\partial_\lambda G_n^2(\lambda) = \partial_\lambda\{G_n(\lambda)G_n(\lambda)\} = \underbrace{\partial_\lambda G_n(\lambda)}_{B.8}G_n(\lambda) + G_n(\lambda)\underbrace{\partial_\lambda G_n(\lambda)}_{B.8}$$

$$
\begin{aligned}
&= G_n^2(\lambda)G_n(\lambda) + G_n(\lambda)G_n^2(\lambda) \\
&= 2G_n^3(\lambda) \qquad \text{(B.12)}
\end{aligned}
$$

$\partial_{\beta'} G_n^2(\lambda) = 0$ and $\partial_\rho G_n^2(\lambda) = 0$

- $$\partial_\xi H_n^2(\rho) = (\partial_{\beta'} H_n^2(\rho), \partial_\lambda H_n^2(\rho), \partial_\rho H_n^2(\rho)),$$

  where

  $\partial_{\beta'} H_n^2(\rho) = 0, \quad \partial_\lambda H_n^2(\rho) = 0,$

$$
\begin{aligned}
\partial_\rho H_n^2(\rho) &= \partial_\rho\{H_n(\rho)H_n(\rho)\} = \underbrace{\partial_\rho H_n(\rho)}_{B.11} H_n(\rho) + H_n(\rho)\underbrace{\partial_\rho H_n(\rho)}_{B.11} \\
&= H_n(\rho)^2 H_n(\rho) + H_n(\rho)H_n(\rho)^2 \\
&= 2H_n^3(\rho) \qquad \text{(B.13)}
\end{aligned}
$$

B.1.2 Component-wise calculation of the log-likelihood

$$\frac{\partial \ell_{n,T}(\theta)}{\partial \theta} = \{\partial_{\beta'}\ell_{n,T}(\theta), \partial_\lambda \ell_{n,T}(\theta), \partial_\rho \ell_{n,T}(\theta), \partial_{\sigma^2}\ell_{n,T}(\theta)\}$$

- $$
\begin{aligned}
\partial_{\beta'}\ell_{n,T}(\theta) &= \partial_{\beta'}\{-\frac{1}{2\sigma^2}\sum_{t=1}^T \tilde{V}_{nt}'(\xi)\tilde{V}_{nt}(\xi)\} \\
&= -\frac{1}{2\sigma^2}\sum_{t=1}^T (\underbrace{\partial_{\beta'}\tilde{V}_{nt}(\xi)}_{B.3})'\tilde{V}_{nt}(\xi) - \frac{1}{2\sigma^2}\sum_{t=1}^T \left(\tilde{V}_{nt}'(\xi)\underbrace{\partial_{\beta'}\tilde{V}_{nt}(\xi)}_{B.3}\right)' \\
&= \frac{1}{2\sigma^2}\sum_{t=1}^T (R_n(\rho)\tilde{X}_{nt})'\tilde{V}_{nt}(\xi) + \frac{1}{2\sigma^2}\sum_{t=1}^T \left(\tilde{V}_{nt}'(\xi)R_n(\rho)\tilde{X}_{nt}\right)'
\end{aligned}
$$

$$= \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'\tilde{V}_{nt}(\xi) \quad \text{(B.14)}$$

•

$$\begin{aligned}
\partial_\lambda \ell_{n,T}(\theta) &= \partial_\lambda\{(T-1)\ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}_{nt}'(\xi)\tilde{V}_{nt}(\xi)\} \\
&= (T-1)\mathrm{tr}(S_n^{-1}(\lambda)\underbrace{\partial_\lambda S_n(\lambda)}_{B.1}) \\
&- \frac{1}{2\sigma^2}\sum_{t=1}^{T}\{(\underbrace{\partial_\lambda \tilde{V}_{nt}(\xi)}_{B.4})'\tilde{V}_{nt}(\xi) + \tilde{V}_{nt}'(\xi)\underbrace{\partial_\lambda \tilde{V}_{nt}(\xi)}_{B.4}\} \\
&= -(T-1)\mathrm{tr}(S_n^{-1}(\lambda)W_n) \\
&+ \frac{1}{2\sigma^2}\sum_{t=1}^{T}\{(R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi) + \tilde{V}_{nt}'(\xi)R_n(\rho)W_n\tilde{Y}_{nt}\} \\
&= -(T-1)\mathrm{tr}(G_n(\lambda)) + \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi) \quad \text{(B.15)}
\end{aligned}$$

•

$$\begin{aligned}
\partial_\rho \ell_{n,T}(\theta) &= \partial_\rho\{(T-1)\ln|R_n(\rho)| - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}_{nt}'(\xi)\tilde{V}_{nt}(\xi)\} \\
&= (T-1)\mathrm{tr}(R_n^{-1}(\rho)\underbrace{\partial_\rho R_n(\rho)}_{B.2}) \\
&- \frac{1}{2\sigma^2}\sum_{t=1}^{T}\{(\underbrace{\partial_\rho \tilde{V}_{nt}(\xi)}_{B.5})'\tilde{V}_{nt}(\xi) + \tilde{V}_{nt}'(\xi)\underbrace{\partial_\rho \tilde{V}_{nt}(\xi)}_{B.5}\} \\
&= -(T-1)\mathrm{tr}(R_n^{-1}(\rho)M_n) \\
&+ \frac{1}{2\sigma^2}\sum_{t=1}^{T}\{(H_n(\rho)\tilde{V}_{nt}(\xi))'\tilde{V}_{nt}(\xi) + \tilde{V}_{nt}'(\xi)H_n(\rho)\tilde{V}_{nt}(\xi)\} \\
&= -(T-1)\mathrm{tr}(H_n(\rho)) + \frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\xi))'\tilde{V}_{nt}(\xi) \quad \text{(B.16)}
\end{aligned}$$

•

$$
\begin{aligned}
\partial_{\sigma^2}\ell_{n,T}(\theta) &= \partial_{\sigma^2}\{-\frac{n(T-1)}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}'_{nt}(\xi)\tilde{V}_{nt}(\xi)\} \\
&= -\frac{n(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}\tilde{V}'_{nt}(\xi)\tilde{V}_{nt}(\xi) \qquad \text{(B.17)}
\end{aligned}
$$

$$
\frac{\partial \ell_{n,T}(\theta)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'\tilde{V}_{nt}(\xi) \\ -(T-1)\mathrm{tr}(G_n(\lambda)) + \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi) \\ -(T-1)\mathrm{tr}(H_n(\rho)) + \frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\xi))'\tilde{V}_{nt}(\xi) \\ -\frac{n(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}\tilde{V}'_{nt}(\xi)\tilde{V}_{nt}(\xi) \end{pmatrix} \qquad \text{(B.18)}
$$

$$
\frac{\partial \ell_{n,T}(\theta)}{\partial\theta} = \frac{1}{(T-1)}\sum_{t=1}^{T}\psi((Y_{nt}, X_{nt}), \theta) = 0.
$$

where $\psi((Y_{nt}, X_{nt}), \theta_{n,T})$ represents the likelihood score function and its expression is

$$
\psi(Y_{nt}, X_{nt}, \theta) = \begin{pmatrix} \frac{(T-1)}{\sigma^2}(R_n(\rho)\tilde{X}_{nt})'\tilde{V}_{nt}(\zeta) \\ \frac{(T-1)}{\sigma^2}(R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\zeta) - \frac{(T-1)^2}{T}\mathrm{tr}(G_n(\lambda)) \\ \frac{(T-1)}{\sigma^2}(H_n(\rho)\tilde{V}_{nt}(\zeta))'\tilde{V}_{nt}(\zeta) - \frac{(T-1)^2}{T}\mathrm{tr}(H_n(\rho)) \\ \frac{(T-1)}{2\sigma^4}\left(\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta) - \frac{n(T-1)}{T}\sigma^2\right) \end{pmatrix} \qquad \text{(B.19)}
$$

B.2 The second derivative the log-likelihood

- The first row of $\frac{\partial^2 \ell_{n,T}(\theta_0)}{\partial\theta\partial\theta'}$ is $\frac{\partial\{\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'\tilde{V}_{nt}(\xi)\}}{\partial\theta'}$

$$
\begin{aligned}
\partial_{\beta'}(\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'\tilde{V}_{nt}(\xi)) &= \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'\underbrace{\partial'_{\beta}\tilde{V}_{nt}(\xi)}_{B.3} \\
&= -\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'R_n(\rho)\tilde{X}_{nt} \qquad \text{(B.20)}
\end{aligned}
$$

$$\partial_\lambda\Big(\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'\tilde{V}_{nt}(\xi)\Big) = \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'\underbrace{\partial_\lambda\tilde{V}_{nt}(\xi)}_{B.4}$$
$$= -\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'R_n(\rho)W_n\tilde{Y}_{nt} \quad \text{(B.21)}$$

$$\partial_\rho\Big(\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'\tilde{V}_{nt}(\xi)\Big) = \frac{1}{\sigma^2}\sum_{t=1}^{T}\{(\underbrace{\partial_\rho R_n(\rho)}_{B.2}\tilde{X}_{nt})'\tilde{V}_{nt}(\xi) + (R_n(\rho)\tilde{X}_{nt})'\underbrace{\partial_\rho\tilde{V}_{nt}(\xi)}_{B.5}\}$$
$$= -\frac{1}{\sigma^2}\sum_{t=1}^{T}\{(M_n\tilde{X}_{nt})'\tilde{V}_{nt}(\xi) + (R_n(\rho)\tilde{X}_{nt})'H_n(\rho)\tilde{V}_{nt}(\xi)\} \quad \text{(B.22)}$$

$$\partial_{\sigma^2}\Big(\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'\tilde{V}_{nt}(\xi)\Big) = -\frac{1}{\sigma^4}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'\tilde{V}_{nt}(\xi) \quad \text{(B.23)}$$

- The second row of $\dfrac{\partial^2\ell_{n,T}(\theta_0)}{\partial\theta\partial\theta'}$ is $\dfrac{\partial\{-(T-1)\mathrm{tr}(G_n(\lambda)) + \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi)\}}{\partial\theta'}$. The matrix $\dfrac{\partial^2\ell_{n,T}(\theta_0)}{\partial\theta\partial\theta'}$ is symmetric. The first element is the transpose of the second one in the first row. So

$$\partial_{\beta'}\Big(-(T-1)\mathrm{tr}(G_n(\lambda)) + \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi)\Big) = -\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'R_n(\rho)\tilde{X}_{nt}$$

$$\partial_\lambda\big(-(T-1)\mathrm{tr}(G_n(\lambda))\quad + \quad \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi)\big)$$

$$= -(T-1)\mathrm{tr}(\underbrace{\partial_\lambda G_n(\lambda)}_{B.8}) + \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'\underbrace{\partial_\lambda \tilde{V}_{nt}(\xi)}_{B.4}$$

$$= -(T-1)\mathrm{tr}(G_n^2(\lambda)) - \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'R_n(\rho)W_n\tilde{Y}_{nt} \tag{B.24}$$

$$\begin{aligned}\partial_\rho(-(T-1)\mathrm{tr}(G_n(\lambda)) &+ \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi))\\ &= \frac{1}{\sigma^2}\sum_{t=1}^{T}\{(\underbrace{\partial_\rho R_n(\rho)}_{B.2}W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi) + (R_n(\rho)W_n\tilde{Y}_{nt})'\underbrace{\partial_\rho \tilde{V}_{nt}(\xi)}_{B.5}\}\\ &= -\frac{1}{\sigma^2}\sum_{t=1}^{T}\{(M_nW_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi) + (R_n(\rho)W_n\tilde{Y}_{nt})'H_n(\rho)\tilde{V}_{nt}(\xi)\}\end{aligned} \tag{B.25}$$

$$\begin{aligned}\partial_{\sigma^2}(-(T-1)\mathrm{tr}(G_n(\lambda)) &+ \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi))\\ &= -\frac{1}{\sigma^4}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi)\end{aligned} \tag{B.26}$$

- The third row of $\dfrac{\partial^2 \ell_{n,T}(\theta_0)}{\partial\theta\partial\theta'}$ is

$$\frac{\partial\{-(T-1)\mathrm{tr}(H_n(\rho)) + \frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\xi))'\tilde{V}_{nt}(\xi)\}}{\partial\theta'}.$$

We could get the first two elements from the transpose of the third ones in the first two rows. So we only need to calculate the following two derivatives.

$$\begin{aligned}
&\partial_\rho(-(T-1)\mathrm{tr}(H_n(\rho)) + \frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\xi))'\tilde{V}_{nt}(\xi))\\
&= -(T-1)\mathrm{tr}(\underbrace{\partial_\rho H_n(\rho)}_{B.11}) + \frac{1}{\sigma^2}\sum_{t=1}^{T}\{(\underbrace{\partial_\rho H_n(\rho)}_{B.11}\tilde{V}_{nt}(\xi))'\tilde{V}_{nt}(\xi)\\
&+(H_n(\rho)\underbrace{\partial_\rho \tilde{V}_{nt}(\xi)}_{B.5})'\tilde{V}_{nt}(\xi) + (H_n(\rho)\tilde{V}_{nt}(\xi))'\underbrace{\partial_\rho\tilde{V}_{nt}(\xi)}_{B.5}\}\\
&= -(T-1)\mathrm{tr}(H_n^2(\rho)) + \frac{1}{\sigma^2}\sum_{t=1}^{T}\{(H_n^2(\rho)\tilde{V}_{nt}(\xi))'\tilde{V}_{nt}(\xi)\\
&-(H_n^2(\rho)\tilde{V}_{nt}(\xi))'\tilde{V}_{nt}(\xi) - (H_n(\rho)\tilde{V}_{nt}(\xi))'H_n(\rho)\tilde{V}_{nt}(\xi)\}\\
&= -(T-1)\mathrm{tr}(H_n^2(\rho)) - \frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\xi))'H_n(\rho)\tilde{V}_{nt}(\xi) \qquad \text{(B.27)}
\end{aligned}$$

$$\begin{aligned}
\partial_{\sigma^2}(-(T-1)\mathrm{tr}(H_n(\rho)) \quad &+ \quad \frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\xi))'\tilde{V}_{nt}(\xi))\\
&= \quad -\frac{1}{\sigma^4}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\xi))'\tilde{V}_{nt}(\xi) \qquad \text{(B.28)}
\end{aligned}$$

- The fourth row of $\dfrac{\partial^2 \ell_{n,T}(\theta_0)}{\partial\theta\partial\theta'}$ is $\dfrac{\partial\{-\frac{n(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}\tilde{V}_{nt}'(\xi)\tilde{V}_{nt}(\xi)\}}{\partial\theta'}$. We only need to calculate the derivative in respect with σ^2.

$$\partial_{\sigma^2}(-\frac{n(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}\tilde{V}_{nt}'(\xi)\tilde{V}_{nt}(\xi)) \quad = \quad \frac{n(T-1)}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{t=1}^{T}\tilde{V}_{nt}'(\xi)\tilde{V}_{nt}(\xi) \qquad \text{(B.29)}$$

We report the matrix

$$-\frac{\partial^2 \ell_{n,T}(\theta)}{\partial\theta\partial\theta'} =$$

$$\begin{bmatrix} \frac{1}{\sigma^2}\sum_t (R_n(\rho)\tilde{X}_{nt})'(R_n(\rho)\tilde{X}_{nt}) & * & * & * \\ \frac{1}{\sigma^2}\sum_t (R_n(\rho)W_n\tilde{Y}_{nt})'(R_n(\rho)\tilde{X}_{nt}) & \frac{1}{\sigma^2}\sum_t (R_n(\rho)W_n\tilde{Y}_{nt})'(R_n(\rho)W_n\tilde{Y}_{nt}) + (T-1)\mathrm{tr}(G_n^2(\lambda)) & * & * \\ \frac{1}{\sigma^2}\sum_t \{(H_n(\rho)\tilde{V}_{nt}(\xi))'(R_n(\rho)\tilde{X}_{nt}) + \tilde{V}_{nt}(\xi)'M_n\tilde{X}_{nt}\} & \frac{1}{\sigma^2}\sum_t \{(R_n(\rho)W_n\tilde{Y}_{nt})'(H_n(\rho)\tilde{V}_{nt}(\xi)) + (M_nW_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi)\} & 0 & 0 \\ \frac{1}{\sigma^4}\sum_t \tilde{V}_{nt}(\xi)'R_n(\rho)\tilde{X}_{nt} & \frac{1}{\sigma^4}\sum_t (R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi) & 0 & 0 \end{bmatrix}$$

$$+\begin{bmatrix} 0_{k\times k} & 0_{k\times 1} & 0_{k\times 1} & 0_{k\times 1} \\ 0_{1\times k} & 0 & 0 & 0 \\ 0_{1\times k} & 0 & \sigma^{-2}\sum_t (H_n(\rho)\tilde{V}_{nt}(\xi))'H_n(\rho)\tilde{V}_{nt}(\xi) + (T-1)\mathrm{tr}\{H_n(\rho)^2\} & * \\ 0_{1\times k} & 0 & \sigma^{-4}\sum_t (H_n(\rho)\tilde{V}_{nt}(\xi))'\tilde{V}_{nt}(\xi) & -\frac{m}{2\sigma^4} + \sigma^{-6}\sum_t \tilde{V}_{nt}'(\xi)\tilde{V}_{nt}(\xi) \end{bmatrix}. \tag{B.30}$$

B.3 The third derivative of the log-likelihood

Assuming, for the third derivative w.r.t . θ of $\ell_{n,T}(\theta)$, that derivation and integration can be exchanged (namely the dominated convergence theorem holds, component-wise for in θ, for the third derivative of the log-likelihood), we derive $\frac{\partial^2 \ell_{n,T}(\theta)}{\partial\theta\partial\theta'}$ to compute the term $\Gamma(i,j,T,\theta_0)$, for each i and j. To this end, we have:

-
$$\frac{\partial\left\{\sigma^{-2}\sum_t (R_n(\rho)\tilde{X}_{nt})'(R_n(\rho)\tilde{X}_{nt})\right\}}{\partial\theta} = \begin{bmatrix} A^{(1,1,\beta')}(\rho,\sigma^2) & A^{(1,1,\lambda)}(\rho,\sigma^2) & A^{(1,1,\rho)}(\rho,\sigma^2) & A^{(1,1,\sigma^2)}(\rho,\sigma^2) \end{bmatrix}$$

  where
$$A^{(1,1,\beta')}(\rho,\sigma^2) = 0_{k\times k}, \quad A^{(1,1,\lambda)}(\rho,\sigma^2) = 0, \tag{B.31}$$

$$A^{(1,1,\rho)}(\rho,\sigma^2) = -2\sigma^{-2}\sum_{t=1}^{T}\left\{\tilde{X}'_{nt}M'_n\tilde{X}_{nt} - \rho\tilde{X}'_{nt}M'_nM_n\tilde{X}_{nt}\right\} = -2\sigma^{-2}\sum_{t=1}^{T}(M_n\tilde{X}_{nt})'R_n(\rho)\tilde{X}_{nt}, \tag{B.32}$$

and

$$A^{(1,1,\sigma^2)}(\rho,\sigma^2) = -\sigma^{-4}\sum_{t=1}^{T}\left\{(R_n(\rho)\tilde{X}_{nt})'(R_n(\rho)\tilde{X}_{nt})\right\}. \tag{B.33}$$

- $$\frac{\partial\left\{\sigma^{-2}\sum_t(R_n(\rho)W_n\tilde{Y}_{nt})'(R_n(\rho)\tilde{X}_{nt})\right\}}{\partial\theta} = \left[\begin{array}{cccc} A^{(2,1,\beta')}(\rho,\sigma^2) & A^{(2,1,\lambda)}(\rho,\sigma^2) & A^{(2,1,\rho)}(\rho,\sigma^2) & A^{(2,1,\sigma^2)}(\rho,\sigma^2) \end{array}\right]$$

  where

$$A^{(2,1,\beta')}(\rho,\sigma^2) = 0_{k\times k}, \quad A^{(2,1,\lambda)}(\rho,\sigma^2) = 0, \tag{B.34}$$

$$A^{(2,1,\rho)}(\rho,\sigma^2) = \sigma^{-2}\sum_{t=1}^{T}\left\{(-M_nW_n\tilde{Y}_{nt})'R_n(\rho)\tilde{X}_{nt} - (R_n(\rho)W_n\tilde{Y}_{nt})'M_n\tilde{X}_{nt}\right\}, \tag{B.35}$$

  and

$$A^{(2,1,\sigma^2)}(\rho,\sigma^2) = -\sigma^{-4}\sum_{t=1}^{T}\left\{(R_n(\rho)W_n\tilde{Y}_{nt})'(R_n(\rho)\tilde{X}_{nt})\right\}. \tag{B.36}$$

- $$\frac{\partial\left\{\sigma^{-2}\sum_t(R_n(\rho)W_n\tilde{Y}_{nt})'(R_n(\rho)W_n\tilde{Y}_{nt}) + (T-1)\mathrm{tr}(G_n^2(\lambda))\right\}}{\partial\theta} = \left[\begin{array}{cccc} A^{(2,2,\beta')}(\lambda,\rho,\sigma^2) & A^{(2,2,\lambda)}(\lambda,\rho,\sigma^2) & A^{(2,2,\rho)}(\lambda,\rho,\sigma^2) & A^{(2,2,\sigma^2)}(\lambda,\rho,\sigma^2) \end{array}\right]$$

  where

$$A^{(2,2,\beta')}(\lambda,\rho,\sigma^2) = 0_{k\times k} \tag{B.37}$$

$$A^{(2,2,\lambda)}(\lambda,\rho,\sigma^2) = (T-1)\mathrm{tr}\left\{\partial_\lambda(G_n^2(\lambda))\right\} = (T-1)\mathrm{tr}\left\{2G_n^3(\lambda)\right\}, \tag{B.38}$$

where we make use of (B.12), moreover, using (B.2)

$$
\begin{aligned}
A^{(2,2,\rho)}(\lambda,\rho,\sigma^2) &= \frac{1}{\sigma^2}\sum_{t=1}^{T}\{(\partial_\rho R_n(\rho)W_n\tilde{Y}_{nt})'(R_n(\rho)W_n\tilde{Y}_{nt}) \\
&+ (R_n(\rho)W_n\tilde{Y}_{nt})'(\partial_\rho R_n(\rho)W_n\tilde{Y}_{nt})\} \\
&= -\frac{1}{\sigma^2}\sum_{t=1}^{T}\{(M_nW_n\tilde{Y}_{nt})'(R_n(\rho)W_n\tilde{Y}_{nt}) \\
&+ (R_n(\rho)W_n\tilde{Y}_{nt})'(M_nW_n\tilde{Y}_{nt})\} \\
&= -\frac{2}{\sigma^2}\sum_{t=1}^{T}(M_nW_n\tilde{Y}_{nt})'(R_n(\rho)W_n\tilde{Y}_{nt}) \qquad \text{(B.39)}
\end{aligned}
$$

and

$$
A^{(2,2,\sigma^2)}(\lambda,\rho,\sigma^2) = -\frac{1}{\sigma^4}\sum_{t=1}^{T}(R_n(\rho)M_n\tilde{Y}_{nt})'R_n(\rho)W_n\tilde{Y}_{nt}, \qquad \text{(B.40)}
$$

- $$
\frac{\partial\left\{\sigma^{-2}\sum_{t=1}^{T}\left((H_n(\rho)\tilde{V}_{nt}(\xi))'(R_n(\rho)\tilde{X}_{nt})+\sigma^{-2}\sum_{t=1}^{T}\tilde{V}'_{nt}(\xi)M_n\tilde{X}_{nt}\right)\right\}}{\partial\theta} =
$$
$$
\begin{bmatrix} A^{(3,1,\beta')}(\lambda,\rho,\sigma^2) & A^{(3,1,\lambda)}(\lambda,\rho,\sigma^2) & A^{(3,1,\rho)}(\beta',\lambda,\rho,\sigma^2) & A^{(3,1,\sigma^2)}(\beta',\lambda,\rho,\sigma^2) \end{bmatrix}
$$

where

$$
\begin{aligned}
A^{(3,1,\beta')}(\lambda,\rho,\sigma^2) &= \sigma^{-2}\sum_{t}\left\{(H_n(\rho)\underbrace{\partial_{\beta'}\tilde{V}_{nt}(\xi)}_{\text{see }(B.3)})'(R_n(\rho)\tilde{X}_{nt})+\underbrace{\partial_{\beta'}\tilde{V}'_{nt}(\xi)}_{\text{see }(B.3)}M_n\tilde{X}_{nt}\right\} \\
&= -\frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{(H_n(\rho)R_n(\rho)\tilde{X}_{nt})'R_n(\rho)\tilde{X}_{nt}+(R_n(\rho)\tilde{X}_{nt})'M_n\tilde{X}_{nt}\right\}
\end{aligned}
$$
(B.41)

$$A^{(3,1,\lambda)}(\lambda,\rho,\sigma^2) = \frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{\left(H_n(\rho)\underbrace{\partial_\lambda \tilde{V}_{nt}(\xi)}_{\text{see}\quad(B.4)}\right)' R_n(\rho)\tilde{X}_{nt} + \left(\underbrace{\partial_\lambda \tilde{V}_{nt}(\xi)}_{\text{see}\quad(B.4)}\right)' M_n\tilde{X}_{nt}\right\}$$
$$= -\frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{(H_n(\rho)R_n(\rho)W_n\tilde{Y}_{nt})'R_n(\rho)\tilde{X}_{nt} + (R_n(\rho)W_n\tilde{Y}_{nt})'M_n\tilde{X}_{nt}\right\} \tag{B.42}$$

where $\partial_\lambda \tilde{V}_{nt}(\xi) = -R_n(\rho)W_n\tilde{Y}_{nt}$ as in (B.4), and

$$A^{(3,1,\rho)}(\beta',\lambda,\rho,\sigma^2)$$
$$= \frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{\left(\underbrace{\partial_\rho H_n(\rho)}_{\text{see}\quad(B.11)}\tilde{V}_{nt}(\xi)\right)' R_n(\rho)\tilde{X}_{nt} + \left(H_n(\rho)\underbrace{\partial_\rho \tilde{V}_{nt}(\xi)}_{\text{see}\quad(B.5)}\right)' R_n(\rho)\tilde{X}_{nt}\right\}$$
$$+\frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{\left(H_n(\rho)\tilde{V}_{nt}(\xi)\right)'\underbrace{\partial_\rho R_n(\rho)}_{\text{see}\quad(B.2)}\tilde{X}_{nt} + \underbrace{\partial_\rho \tilde{V}'_{nt}(\xi)}_{\text{see}\quad(B.5)}M_n\tilde{X}_{nt}\right\}$$
$$= \frac{1}{\sigma^2}\sum_{t=1}^{T}\left(H_n(\rho)^2\tilde{V}_{nt}(\xi)\right)' R_n(\rho)\tilde{X}_{nt} - \frac{1}{\sigma^2}\sum_{t=1}^{T}\left(H_n(\rho)^2\tilde{V}_{nt}(\xi)\right)' R_n(\rho)\tilde{X}_{nt}$$
$$-\frac{1}{\sigma^2}\sum_{t=1}^{T}\left(H_n(\rho)\tilde{V}_{nt}(\xi)\right)' M_n\tilde{X}_{nt} - \frac{1}{\sigma^2}\sum_{t=1}^{T}\left(H_n(\rho)\tilde{V}_{nt}(\xi)\right)' M_n\tilde{X}_{nt}$$
$$= -\frac{2}{\sigma^2}\sum_{t=1}^{T}\left(H_n(\rho)\tilde{V}_{nt}(\xi)\right)' M_n\tilde{X}_{nt} \tag{B.43}$$

$$A^{(3,1,\sigma^2)}(\beta',\lambda,\rho,\sigma^2) = -\frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{(H_n(\rho)\tilde{V}_{nt}(\xi))'(R_n(\rho)\tilde{X}_{nt}) + \tilde{V}'_{nt}(\xi)M_n\tilde{X}_{nt}\right\} \tag{B.44}$$

-

$$\frac{\partial \left\{\sigma^{-2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'(H_n(\rho)\tilde{V}_{nt}(\xi)) + \sigma^{-2}\sum_{t=1}^{T}(M_nW_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi)\right\}}{\partial\theta} =$$

$$\left[\begin{array}{cccc} A^{(3,2,\beta')}(\lambda,\rho,\sigma^2) & A^{(3,2,\lambda)}(\lambda,\rho,\sigma^2) & A^{(3,2,\rho)}(\beta',\lambda,\rho,\sigma^2) & A^{(3,2,\sigma^2)}(\beta',\lambda,\rho,\sigma^2) \end{array}\right]$$

where

$$\begin{aligned} A^{(3,2,\beta')}(\lambda,\rho,\sigma^2) &= \frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{\left((R_n(\rho)W_n\tilde{Y}_{nt})'(H_n(\rho)\underbrace{\partial_{\beta'}\tilde{V}_{nt}(\xi)}_{\text{see (B.3)}})\right)' + \left((M_nW_n\tilde{Y}_{nt})'\underbrace{\partial_{\beta'}\tilde{V}_{nt}(\xi)}_{\text{see }(B.3)}\right)'\right\} \\ &= -\frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{(H_n(\rho)R_n(\rho)\tilde{X}_{nt})'R_n(\rho)W_n\tilde{Y}_{nt} + (R_n(\rho)\tilde{X}_{nt})'M_nW_n\tilde{Y}_{nt}\right\} \end{aligned} \tag{B.45}$$

$$\begin{aligned} A^{(3,2,\lambda)}(\lambda,\rho,\sigma^2) &= \frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{\left(R_n(\rho)W_m\tilde{Y}_{nt}\right)'H_n(\rho)\underbrace{\partial_{\lambda}\tilde{V}_{nt}(\xi)}_{\text{see }(B.4)} + \left(M_nW_n\tilde{Y}_{nt}\right)'\underbrace{\partial_{\lambda}\tilde{V}_{nt}(\xi)}_{\text{see }(B.4)}\right\} \\ &= -\frac{1}{\sigma^2}\sum_{t=1}^{T}\left(R_n(\rho)W_m\tilde{Y}_{nt}\right)'H_n(\rho)R_n(\rho)W_n\tilde{Y}_{nt} \\ &- \frac{1}{\sigma^2}\sum_{t=1}^{T}\left(M_nW_n\tilde{Y}_{nt}\right)'R_n(\rho)W_n\tilde{Y}_{nt} \\ &= -\frac{2}{\sigma^2}\sum_{t=1}^{T}(M_nW_n\tilde{Y}_{nt})'(R_n(\rho)W_n\tilde{Y}_{nt}) \end{aligned} \tag{B.46}$$

and

$$A^{(3,2,\rho)}(\beta',\lambda,\rho,\sigma^2)$$

$$
\begin{aligned}
&= \frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{\left(\underbrace{\partial_\rho R_n(\rho)}_{\text{see } (B.2)} W_n\tilde{Y}_{nt}\right)' H_n(\rho)\tilde{V}_{nt}(\xi) + \left(R_n(\rho)W_n\tilde{Y}_{nt}\right)' \underbrace{\partial_\rho H_n(\rho)}_{\text{see } (B.11)} \tilde{V}_{nt}(\xi)\right\} \\
&+\frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{\left(R_n(\rho)W_n\tilde{Y}_{nt}\right)' H_n(\rho) \underbrace{\partial_\rho\tilde{V}_{nt}(\xi)}_{\text{see } (B.5)} + \left(M_nW_n\tilde{Y}_{nt}\right)' \underbrace{\partial_\rho\tilde{V}_{nt}(\xi)}_{\text{see } (B.5)}\right\} \\
&= -\frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{\left(M_nW_n\tilde{Y}_{nt}\right)' H_n(\rho)\tilde{V}_{nt}(\xi) - \left(R_n(\rho)W_n\tilde{Y}_{nt}\right)' H_n(\rho)^2\tilde{V}_{nt}(\xi)\right\} \\
&-\frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{\left(R_n(\rho)W_n\tilde{Y}_{nt}\right)' H_n(\rho)^2\tilde{V}_{nt}(\xi) + \left(M_nW_n\tilde{Y}_{nt}\right)' H_n(\rho)\tilde{V}_{nt}(\xi)\right\} \\
&= -\frac{2}{\sigma^2}\sum_{t=1}^{T}\left(M_nW_n\tilde{Y}_{nt}\right)' H_n(\rho)\tilde{V}_{nt}(\xi) \qquad \text{(B.47)}
\end{aligned}
$$

$$
A^{(3,2,\sigma^2)}(\beta',\lambda,\rho,\sigma^2) = -\frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{(R_n(\rho)W_n\tilde{Y}_{nt})'(H_n(\rho)\tilde{V}_{nt}(\xi)) + (M_nW_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi)\right\}. \qquad \text{(B.48)}
$$

- $$
\frac{\partial\left\{\sigma^{-4}\sum_t \tilde{V}_{nt}(\xi)'R_n(\rho)\tilde{X}_{nt}\right\}}{\partial\theta} =
$$
$$
\left[\begin{array}{cccc} A^{(4,1,\beta')}(\lambda,\rho,\sigma^2) & A^{(4,1,\lambda)}(\lambda,\rho,\sigma^2) & A^{(4,1,\rho)}(\beta',\lambda,\rho,\sigma^2) & A^{(4,1,\sigma^2)}(\beta',\lambda,\rho,\sigma^2)\end{array}\right],
$$

where

$$
A^{(4,1,\beta')}(\lambda,\rho,\sigma^2) = \frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{\underbrace{(\partial_{\beta'}\tilde{V}_{nt}(\xi))}_{\text{see } (B.3)}{}'R_n(\rho)\tilde{X}_{nt}\right\}
$$

$$= -\frac{1}{\sigma^4}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})' R_n(\rho)\tilde{X}_{nt} \tag{B.49}$$

$$\begin{aligned} A^{(4,1,\lambda)}(\lambda,\rho,\sigma^2) &= \frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{(\underbrace{\partial_\lambda \tilde{V}_{nt}(\xi)}_{\text{see } (B.4)})' R_n(\rho)\tilde{X}_{nt}\right\} \\ &= -\frac{1}{\sigma^4}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})' R_n(\rho)\tilde{X}_{nt} \end{aligned} \tag{B.50}$$

and

$$\begin{aligned} A^{(4,1,\rho)}(\beta',\lambda,\rho,\sigma^2) &= \frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{(\underbrace{\partial_\rho \tilde{V}_{nt}(\xi)}_{\text{see } (B.5)})' R_n(\rho)\tilde{X}_{nt} + \tilde{V}'_{nt}(\xi)\underbrace{\partial_\rho R_n(\rho)}_{\text{see } (B.2)}\tilde{X}_{nt}\right\} \\ &= -\frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{\left(H_n(\rho)\tilde{V}_{nt}(\xi)\right)' R_n(\rho)\tilde{X}_{nt} + \tilde{V}'_{nt}(\xi)M_n\tilde{X}_{nt}\right\} \end{aligned} \tag{B.51}$$

$$A^{(4,1,\sigma^2)}(\beta',\lambda,\rho,\sigma^2) = -2\sigma^{-6}\sum_{t=1}^{T}\left\{\tilde{V}'_{nt}(\xi)R_n(\rho)\tilde{X}_{nt}\right\}, \tag{B.52}$$

- $$\frac{\partial\left\{\sigma^{-4}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi)\right\}}{\partial\theta} =$$
  $$\left[\begin{array}{cccc} A^{(4,2,\beta')}(\lambda,\rho,\sigma^2) & A^{(4,2,\lambda)}(\lambda,\rho,\sigma^2) & A^{(4,2,\rho)}(\beta',\lambda,\rho,\sigma^2) & A^{(4,2,\sigma^2)}(\beta',\lambda,\rho,\sigma^2) \end{array}\right],$$

where

$$A^{(4,2,\beta')}(\lambda,\rho,\sigma^2) = \frac{1}{\sigma^4}\sum_{t=1}^{T}\left((R_n(\rho)W_n\tilde{Y}_{nt})'\underbrace{\partial_{\beta'}\tilde{V}_{nt}(\xi)}_{\text{see}\quad(B.3)}\right)'$$

$$= -\frac{1}{\sigma^4}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'R_n(\rho)W_n\tilde{Y}_{nt} \tag{B.53}$$

$$A^{(4,2,\lambda)}(\lambda,\rho,\sigma^2) = \frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{\Big(R_n(\rho)W_n\tilde{Y}_{nt}\Big)'\underbrace{\partial_{\lambda}\tilde{V}_{nt}(\xi)}_{\text{see}\quad(B.4)}\right\}$$

$$= -\frac{1}{\sigma^4}\sum_{t=1}^{T}\Big(R_n(\rho)W_n\tilde{Y}_{nt}\Big)'R_n(\rho)W_n\tilde{Y}_{nt} \tag{B.54}$$

and

$$A^{(4,2,\rho)}(\beta',\lambda,\rho,\sigma^2) = \frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{\left(\underbrace{\partial_{\rho}R_n(\rho)}_{\text{see}\quad(B.2)}W_n\tilde{Y}_{nt}\right)'\tilde{V}_{nt}(\xi)+\Big(R_n(\rho)W_n\tilde{Y}_{nt}\Big)'\underbrace{\partial_{\rho}\tilde{V}_{nt}(\xi)}_{\text{see}\quad(B.5)}\right\}$$

$$= -\frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{\Big(M_nW_n\tilde{Y}_{nt}\Big)'\tilde{V}_{nt}(\xi)+\Big(R_n(\rho)W_n\tilde{Y}_{nt}\Big)'H_n(\rho)\tilde{V}_{nt}(\xi)\right\} \tag{B.55}$$

$$A^{(4,2,\sigma^2)}(\beta',\lambda,\rho,\sigma^2) = -2\sigma^{-6}\sum_{t=1}^{T}\left\{(R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi)\right\}. \tag{B.56}$$

-

$$\frac{\partial \left\{\sigma^{-2} \sum_t^T (H_n(\rho)\tilde{V}_{nt}(\xi))' H_n(\rho)\tilde{V}_{nt}(\xi) + (T-1)\mathrm{tr}\{H_n^2(\rho)\}\right\}}{\partial\theta} =$$
$$\left[\begin{array}{cccc} B^{(3,3,\beta')}(\beta',\lambda,\rho,\sigma^2) & B^{(3,3,\lambda)}(\beta',\lambda,\rho,\sigma^2) & B^{(3,3,\rho)}(\beta',\lambda,\rho,\sigma^2) & B^{(3,3,\sigma^2)}(\beta',\lambda,\rho,\sigma^2) \end{array}\right],$$

where, making use of $\partial_{\beta'} H_n(\rho) = 0$,

$$\begin{aligned} B^{(3,3,\beta')}(\beta',\lambda,\rho,\sigma^2) &= \frac{1}{\sigma^2}\sum_{t=1}^T \left\{ \left(H_n(\rho) \underbrace{\partial_{\beta'} \tilde{V}_{nt}(\xi)}_{\text{see} \quad (B.3)} \right)' H_n(\rho)\tilde{V}_{nt}(\xi) \right\} \\ &+ \frac{1}{\sigma^2}\sum_{t=1}^T \left\{ \left(H_n(\rho)\tilde{V}_{nt}(\xi)\right)' H_n(\rho) \underbrace{\partial_{\beta'} \tilde{V}_{nt}(\xi)}_{\text{see} \quad (B.3)}{}' \right\} \\ &= -\frac{1}{\sigma^2}\sum_{t=1}^T \left(H_n(\rho)R_n(\rho)\tilde{X}_{nt}\right)' H_n(\rho)\tilde{V}_{nt}(\xi) \\ &- \frac{1}{\sigma^2}\sum_{t=1}^T \left(H_n(\rho)R_n(\rho)\tilde{X}_{nt}\right)' H_n(\rho)\tilde{V}_{nt}(\xi) \\ &= -\frac{2}{\sigma^2}\sum_{t=1}^T \left(M_n\tilde{X}_{nt}\right)' H_n(\rho)\tilde{V}_{nt}(\xi) \qquad \text{(B.57)} \end{aligned}$$

$$\begin{aligned} B^{(3,3,\lambda)}(\beta',\lambda,\rho,\sigma^2) &= \frac{1}{\sigma^2}\sum_{t=1}^T \left\{ \left(H_n(\rho) \underbrace{\partial_{\lambda} \tilde{V}_{nt}(\xi)}_{\text{see} \quad (B.4)} \right)' H_n(\rho)\tilde{V}_{nt}(\xi) \right\} \\ &+ \frac{1}{\sigma^2}\sum_{t=1}^T \left\{ \left(H_n(\rho)\tilde{V}_{nt}(\xi)\right)' H_n(\rho) \underbrace{\partial_{\lambda} \tilde{V}_{nt}(\xi)}_{\text{see} \quad (B.4)} \right\} \\ &= -\frac{1}{\sigma^2}\sum_{t=1}^T \left\{ \left(H_n(\rho)R_n(\rho)W_n\tilde{Y}_{nt}\right)' H_n(\rho)\tilde{V}_{nt}(\xi) \right\} \end{aligned}$$

$$- \quad \frac{1}{\sigma^2}\sum_{t=1}^{T}\left\{\left(H_n(\rho)\tilde{V}_{nt}(\xi)\right)' H_n(\rho)R_n(\rho)W_n\tilde{Y}_{nt}\right\} \quad \text{(B.58)}$$

$$\begin{aligned}
B^{(3,3,\rho)}(\beta',\lambda,\rho,\sigma^2) &= \sigma^{-2}\sum_{t}^{T}(\underbrace{\partial_\rho H_n(\rho)}_{\text{see}\quad (B.13)}\tilde{V}_{nt}(\xi))' H_n(\rho)\tilde{V}_{nt}(\xi) \\
&+ \sigma^{-2}\sum_{t}^{T}(H_n(\rho)\underbrace{\partial_\rho \tilde{V}_{nt}(\xi)}_{\text{see}\quad (B.5)})' H_n(\rho)\tilde{V}_{nt}(\xi) \\
&+ \sigma^{-2}\sum_{t}^{T}(H_n(\rho)\tilde{V}_{nt}(\xi))' \underbrace{\partial_\rho H_n(\rho)}_{\text{see}\quad (B.13)}\tilde{V}_{nt}(\xi) \\
&+ \sigma^{-2}\sum_{t}^{T}(H_n(\rho)\tilde{V}_{nt}(\xi))' H_n(\rho)\underbrace{\partial_\rho \tilde{V}_{nt}(\xi)}_{\text{see}\quad (B.5)} \\
&+ (T-1)\mathrm{tr}\{\underbrace{\partial_\rho H_n^2(\rho)}_{\text{see}\quad (B.13)}\} \\
&= \frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n^2(\rho)\tilde{V}_{nt}(\xi))' H_n(\rho)\tilde{V}_{nt}(\xi) \\
&- \frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n^2(\rho)\tilde{V}_{nt}(\xi))' H_n(\rho)\tilde{V}_{nt}(\xi) \\
&+ \frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\xi))' H_n^2(\rho)\tilde{V}_{nt}(\xi) \\
&- \frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\xi))' H_n^2(\rho)\tilde{V}_{nt}(\xi) \\
&+ (T-1)\mathrm{tr}\{2H_n(\rho)^3\} \\
&= (T-1)\mathrm{tr}\{2H_n(\rho)^3\} \qquad \text{(B.59)}
\end{aligned}$$

$$B^{(3,3,\sigma^2)}(\beta', \lambda, \rho, \sigma^2) = \;=\; -\sigma^{-4}\sum_{t=1}^{T}\left\{(H_n(\rho)\tilde{V}_{nt}(\xi))' H_n(\rho)\tilde{V}_{nt}(\xi)\right\}. \quad \text{(B.60)}$$

- $$\frac{\partial\left\{\sigma^{-4}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\xi))'\tilde{V}_{nt}(\xi)\right\}}{\partial\theta} =$$
$$\left[\begin{array}{cccc} B^{(4,3,\beta')}(\beta', \lambda, \rho, \sigma^2) & B^{(4,3,\lambda)}(\beta', \lambda, \rho, \sigma^2) & B^{(4,3,\rho)}(\beta', \lambda, \rho, \sigma^2) & B^{(4,3,\sigma^2)}(\beta', \lambda, \rho, \sigma^2) \end{array}\right],$$

  where

$$\begin{aligned} B^{(4,3,\beta')}(\beta', \lambda, \rho, \sigma^2) &= \frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{\left(H_n(\rho)\underbrace{\partial_{\beta'}\tilde{V}_{nt}(\xi)}_{\text{see } (B.3)}\right)'\tilde{V}_{nt}(\xi) + \left(\left(H_n(\rho)\tilde{V}_{nt}(\xi)\right)'\underbrace{\partial_{\beta'}\tilde{V}_{nt}(\xi)}_{\text{see } (B.3)}\right)'\right\} \\ &= -\frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{\left(H_n(\rho)R_n(\rho)\tilde{X}_{nt}\right)'\tilde{V}_{nt}(\xi) + \left(R_n(\rho)\tilde{X}_{nt}\right)' H_n(\rho)\tilde{V}_{nt}(\xi)\right\} \end{aligned} \quad \text{(B.61)}$$

$$\begin{aligned} B^{(4,3,\lambda)}(\beta', \lambda, \rho, \sigma^2) &= \frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{\left(H_n(\rho)\underbrace{\partial_{\lambda}\tilde{V}_{nt}(\xi)}_{\text{see } (B.4)}\right)'\tilde{V}_{nt}(\xi) + \left(H_n(\rho)\tilde{V}_{nt}(\xi)\right)'\underbrace{\partial_{\lambda}\tilde{V}_{nt}(\xi)}_{\text{see } (B.4)}\right\} \\ &= \frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{\left(H_n(\rho)R_n(\rho)W_n\tilde{Y}_{nt}\right)'\tilde{V}_{nt}(\xi)\right\} \\ &+ \frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{\left(H_n(\rho)\tilde{V}_{nt}(\xi)\right)' R_n(\rho)W_n\tilde{Y}_{nt}\right\} \end{aligned} \quad \text{(B.62)}$$

$$\begin{aligned}
B^{(4,3,\rho)}(\beta',\lambda,\rho,\sigma^2) &= \frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{\left(\underbrace{\partial_\rho H_n(\rho)}_{\text{see}\quad(B.2)}\tilde{V}_{nt}(\xi)+H_n(\rho)\underbrace{\partial_\rho \tilde{V}_{nt}(\xi)}_{\text{see}\quad(B.5)}\right)'\tilde{V}_{nt}(\xi)\right\} \\
&+ \frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{\left(H_n(\rho)\tilde{V}_{nt}(\xi)\right)'\underbrace{\partial_\rho \tilde{V}_{nt}(\xi)}_{\text{see}\quad(B.5)}\right\} \\
&= \frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{\left(H_n^2(\rho)\tilde{V}_{nt}(\xi)-H_n^2(\rho)\tilde{V}_{nt}(\xi)\right)'\tilde{V}_{nt}(\xi)\right\} \\
&- \frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{\left(H_n(\rho)\tilde{V}_{nt}(\xi)\right)'H_n(\rho)\tilde{V}_{nt}(\xi)\right\} \\
&= -\frac{1}{\sigma^4}\sum_{t=1}^{T}\left\{\left(H_n(\rho)\tilde{V}_{nt}(\xi)\right)'H_n(\rho)\tilde{V}_{nt}(\xi)\right\} \qquad \text{(B.63)}
\end{aligned}$$

$$B^{(4,3,\sigma^2)}(\beta',\lambda,\rho,\sigma^2) = -2\sigma^{-6}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\xi))'\tilde{V}_{nt}(\xi). \qquad \text{(B.64)}$$

•

$$\frac{\partial\left\{-\frac{m}{2\sigma^4}+\sigma^{-6}\sum_t \tilde{V}'_{nt}(\xi)\tilde{V}_{nt}(\xi)\right\}}{\partial\theta} =$$
$$\left[\begin{array}{cccc} B^{(4,4,\beta')}(\beta',\lambda,\rho,\sigma^2) & B^{(4,4,\lambda)}(\beta',\lambda,\rho,\sigma^2) & B^{(4,4,\rho)}(\beta',\lambda,\rho,\sigma^2) & B^{(4,4,\sigma^2)}(\beta',\lambda,\rho,\sigma^2)\end{array}\right],$$

where

$$\begin{aligned}
B^{(4,4,\beta')}(\beta',\lambda,\rho,\sigma^2) &= \frac{1}{\sigma^6}\sum_{t=1}^{T}\left\{\left(\underbrace{\partial_{\beta'}\tilde{V}_{nt}(\xi)}_{\text{see}\quad(B.3)}\right)'\tilde{V}_{nt}(\xi)+\left(\tilde{V}'_{nt}(\xi)\left(\underbrace{\partial_{\beta'}\tilde{V}_{nt}(\xi)}_{\text{see}\quad(B.3)}\right)\right)'\right\} \\
&= -\frac{1}{\sigma^6}\sum_{t=1}^{T}\left\{\left(R_n(\rho)\tilde{X}_{nt}\right)'\tilde{V}_{nt}(\xi)+\left(R_n(\rho)\tilde{X}_{nt}\right)'\tilde{V}_{nt}(\xi)\right\}
\end{aligned}$$

$$= -\frac{2}{\sigma^6}\sum_{t=1}^{T}\left(R_n(\rho)\tilde{X}_{nt}\right)'\tilde{V}_{nt}(\xi) \tag{B.65}$$

$$\begin{aligned} B^{(4,4,\lambda)}(\beta',\lambda,\rho,\sigma^2) &= \frac{1}{\sigma^6}\sum_{t=1}^{T}\left\{\left(\underbrace{\partial_\lambda \tilde{V}_{nt}(\xi)}_{\text{see } (B.4)}\right)'\tilde{V}_{nt}(\xi)+\tilde{V}'_{nt}(\xi)\left(\underbrace{\partial_\lambda \tilde{V}_{nt}(\xi)}_{\text{see } (B.4)}\right)\right\} \\ &= -\frac{1}{\sigma^6}\sum_{t=1}^{T}\left\{\left(R_n(\rho)W_n\tilde{Y}_{nt}\right)'\tilde{V}_{nt}(\xi)+\tilde{V}'_{nt}(\xi)R_n(\rho)W_n\tilde{Y}_{nt}\right\} \end{aligned} \tag{B.66}$$

$$\begin{aligned} B^{(4,4,\rho)}(\beta',\lambda,\rho,\sigma^2) &= \frac{1}{\sigma^6}\sum_{t=1}^{T}\left\{\left(\underbrace{\partial_\rho \tilde{V}_{nt}(\xi)}_{\text{see } (B.5)}\right)'\tilde{V}_{nt}(\xi)+\tilde{V}'_{nt}(\xi)\left(\underbrace{\partial_\rho \tilde{V}_{nt}(\xi)}_{\text{see } (B.5)}\right)\right\} \\ &= -\frac{1}{\sigma^6}\sum_{t=1}^{T}\left\{\left(H_n(\rho)\tilde{V}_{nt}(\xi)\right)'\tilde{V}_{nt}(\xi)+\tilde{V}'_{nt}(\xi)H_n(\rho)\tilde{V}_{nt}(\xi)\right\} \end{aligned} \tag{B.67}$$

$$B^{(4,4,\sigma^2)}(\beta',\lambda,\rho,\sigma^2) = m\sigma^{-6}-3\sigma^{-8}\sum_{t=1}^{T}\tilde{V}'_{nt}(\xi)\tilde{V}_{nt}(\xi) \tag{B.68}$$

B.4 Component-wise calculation of $-\mathbb{E}\left(\frac{1}{m}\frac{\partial^2\ell_{n,T}(\theta_0)}{\partial\theta\partial\theta'}\right)$

From the model setting, we have $V_{nt} = (V_{1t}, V_{2t}, \cdots V_{nt})$ and V_{it} is i.i.d. across i and t with zero mean and variance σ_0^2. So $\mathbb{E}(V_{nt}) = 0_{n\times 1}$, $Var(V_{nt}) = \sigma_0^2 I_n$. Knowing that $\tilde{V}_{nt}(\xi) = V_{nt} - \sum_{t=1}^{T} V_{nt}/T$,

$\mathbb{E}(\tilde{V}_{nt}(\xi)) = 0_{n\times 1}$, $Var(\tilde{V}_{nt}(\xi)) = Var((1-1/T)V_{nt} + 1/T\sum_{j=1,j\neq t}^{T} V_{nj}) = \frac{T-1}{T}\sigma_0^2 I_n$.

$\tilde{V}_{nt}(\xi) = R_n(\rho)[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\beta]$, so $\mathbb{E}(\tilde{Y}_{nt}) = S_n^{-1}(\lambda)\tilde{X}_{nt}\beta$.

Some other notations: $\ddot{X}_{nt} = R_n \tilde{X}_{nt}, H_n = M_n R_n^{-1}, G_n = W_n S_n^{-1}, \ddot{G}_n = R_n G_n R_n^{-1}$.

- $$\mathbb{E}\left[\frac{1}{m\sigma_0^2}\sum_t (R_n(\rho_0)\tilde{X}_{nt})'(R_n(\rho_0)\tilde{X}_{nt})\right] = \frac{1}{m\sigma_0^2}\sum_{t=1}^{T}\ddot{X}_{nt}'\ddot{X}_{nt} \tag{B.69}$$

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{m\sigma_0^2}\sum_t (R_n(\rho_0)W_n\tilde{Y}_{nt})'(R_n(\rho_0)\tilde{X}_{nt})\right] &= \frac{1}{m\sigma_0^2}\sum_t \left(R_n(\rho_0)W_n E(\tilde{Y}_{nt})\right)'(R_n(\rho_0)\tilde{X}_{nt}) \\
&= \frac{1}{m\sigma_0^2}\sum_t (R_n(\rho_0)\underbrace{W_n S_n^{-1}}_{G_n}\tilde{X}_{nt}\beta_0)'\ddot{X}_{nt} \\
&= \frac{1}{m\sigma_0^2}\sum_t (R_n(\rho_0)G_n R_n^{-1}R_n\tilde{X}_{nt}\beta_0)'\ddot{X}_{nt} \\
&= \frac{1}{m\sigma_0^2}\sum_t (\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0)'\ddot{X}_{nt} \qquad \text{(B.70)}
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{m\sigma_0^2}\sum_t \left\{(H_n(\rho_0)\tilde{V}_{nt}(\xi_0))'(R_n(\lambda_0)\tilde{X}_{nt}) + \tilde{V}_{nt}'(\xi_0)M_n\tilde{X}_{nt}\right\}\right] \\
&= \frac{1}{m\sigma_0^2}\sum_t \left\{(H_n(\rho_0)\mathbb{E}\left[\tilde{V}_{nt}(\xi_0)\right])'(R_n(\lambda_0)\tilde{X}_{nt}) + \mathbb{E}\left[\tilde{V}_{nt}'(\xi_0)\right]M_n\tilde{X}_{nt}\right\} = 0_{1\times k}
\end{aligned} \tag{B.71}$$

$$\mathbb{E}\left[\frac{1}{m\sigma_0^4}\sum_t \tilde{V}_{nt}'(\xi_0)R_n(\lambda_0)\tilde{X}_{nt}\right] = \frac{1}{m\sigma_0^4}\sum_t \mathbb{E}\left[\tilde{V}_{nt}'(\xi_0)\right]R_n(\lambda_0)\tilde{X}_{nt} = 0_{1\times k} \tag{B.72}$$

So we prove the first column of $\Sigma_{0,n,T}$, that is:

$$\frac{1}{m\sigma_0^2}\begin{pmatrix} \sum_t \ddot{X}_{nt}'\ddot{X}_{nt} \\ \sum_t (\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0)'\ddot{X}_{nt} \\ 0_{1\times k} \\ 0_{1\times k} \end{pmatrix}$$

-

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{m\sigma_0^2}\sum_t \left(R_n(\rho_0)W_n\tilde{Y}_{nt})'(R_n(\rho_0)W_n\tilde{Y}_{nt}\right) + \frac{T-1}{m}\mathrm{tr}(G_n^2(\lambda_0))\right] \\
&= \mathbb{E}\left[\frac{1}{m\sigma_0^2}\sum_t \left(R_nW_nS_n^{-1}(R_n^{-1}\tilde{V}_{nt} + \tilde{X}_{nt}\beta_0)\right)'\left(R_nW_nS_n^{-1}(R_n^{-1}\tilde{V}_{nt} + \tilde{X}_{nt}\beta_0)\right)\right] \\
&+\frac{1}{n}\mathrm{tr}(G_n^2(\lambda_0)) \\
&= \mathbb{E}\left[\left(\frac{1}{m\sigma_0^2}\sum_t \left(\ddot{G}_n\tilde{V}_{nt}(\xi_0) + \ddot{G}_n\ddot{X}_{nt}\beta_0\right)'\left(\ddot{G}_n\tilde{V}_{nt}(\xi_0) + \ddot{G}_n\ddot{X}_{nt}\beta_0\right)\right)\right] + \frac{1}{n}\mathrm{tr}(G_n^2(\lambda_0)) \\
&= \mathbb{E}\left[\frac{1}{m\sigma_0^2}\sum_t \left(\ddot{G}_n(\lambda_0)\tilde{V}_{nt}(\xi_0)\right)'\left(\ddot{G}_n(\lambda_0)\tilde{V}_{nt}(\xi_0)\right)\right] \\
&+\frac{1}{m\sigma_0^2}\sum_t \left(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0\right)'\left(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0\right) \\
&+\frac{1}{m\sigma_0^2}\sum_t \left(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0\right)'\ddot{G}_n(\lambda_0)\underbrace{\mathbb{E}\left[\tilde{V}_{nt}(\xi_0)\right]}_{0_{n\times 1}} + \frac{1}{n}\mathrm{tr}(G_n^2(\lambda_0)) \\
&= \frac{1}{n}\mathrm{tr}(\ddot{G}_n'(\lambda_0)\ddot{G}_n(\lambda_0)) + \frac{1}{n}\mathrm{tr}(R_n^{-1}(\rho_0)R_n(\rho_0)G_n^2(\lambda_0)) \\
&+\frac{1}{m\sigma_0^2}\sum_t \left(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0\right)'\left(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0\right) \\
&= \frac{1}{n}\mathrm{tr}\left(\ddot{G}_n'(\lambda_0)\ddot{G}_n(\lambda_0) + R_n(\rho_0)G_n(\lambda_0)R_n^{-1}(\rho_0)R_n(\rho_0)G_n(\lambda_0)R_n^{-1}\right)
\end{aligned}$$

$$+\frac{1}{m\sigma_0^2}\sum_t \left(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0\right)'\left(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0\right)$$
$$=\frac{1}{n}\mathrm{tr}\left(\ddot{G}_n'(\lambda_0)\ddot{G}_n(\lambda_0)+\ddot{G}_n(\lambda_0)\ddot{G}_n(\lambda_0)\right)+\frac{1}{m\sigma_0^2}\sum_t \left(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0\right)'\left(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0\right)$$
$$=\frac{1}{n}\mathrm{tr}\left(\ddot{G}_n^S(\lambda_0)\ddot{G}_n(\lambda_0)\right)+\frac{1}{m\sigma_0^2}\sum_t \left(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0\right)'\left(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0\right) \tag{B.73}$$

$$\mathbb{E}\left[\frac{1}{m\sigma_0^2}\sum_t (R_n(\rho_0)W_n\tilde{Y}_{nt})'(H_n(\lambda_0)\tilde{V}_{nt}(\xi_0))+\frac{1}{m\sigma^2}\sum_t (M_nW_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\xi_0)\right]$$
$$=\mathbb{E}\left[\frac{1}{m\sigma_0^2}\sum_t \left(R_n(\rho_0)W_nS_n^{-1}(\lambda_0)(R_n^{-1}(\rho_0)\tilde{V}_{nt}(\xi_0)+\tilde{X}_{nt}\beta_0)\right)'(H_n(\lambda_0)\tilde{V}_{nt}(\xi_0))\right]$$
$$+\mathbb{E}\left[\frac{1}{m\sigma_0^2}\sum_t \left(M_nW_nS_n^{-1}(\lambda_0)(R_n^{-1}(\rho_0)\tilde{V}_{nt}(\xi_0)+\tilde{X}_{nt}\beta_0)\right)'\tilde{V}_{nt}(\xi_0)\right]$$
$$=\mathbb{E}\left[\frac{1}{m\sigma_0^2}\sum_t \left(\left(\ddot{G}_n(\lambda_0)\tilde{V}_{nt}(\xi_0)\right)'(H_n(\lambda_0)\tilde{V}_{nt}(\xi_0))\right)\right]$$
$$+\mathbb{E}\left[\frac{1}{m\sigma_0^2}\sum_t \left(\left(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0)\right)'(H_n(\lambda_0)\tilde{V}_{nt}(\xi_0))\right)\right]$$
$$+\mathbb{E}\left[\frac{1}{m\sigma_0^2}\sum_t \left(\left(H_n(\rho_0)\ddot{G}_n(\lambda_0)\tilde{V}_{nt}(\xi_0)\right)'\tilde{V}_{nt}(\xi_0)+\left(M_nG_n(\lambda_0)\tilde{X}_{nt}\beta_0\right)'\tilde{V}_{nt}(\xi_0)\right)\right]$$
$$=\frac{1}{n}\mathrm{tr}\left(H_n'(\rho_0)\ddot{G}_n(\lambda_0)\right)+\frac{1}{n}\mathrm{tr}\left(H_n(\rho_0)\ddot{G}_n(\lambda_0)\right)=\frac{1}{n}\mathrm{tr}\left(H_n^S(\rho_0)\ddot{G}_n(\lambda_0)\right) \tag{B.74}$$

$$\mathbb{E}\left[\frac{1}{m\sigma_0^4}\sum_t \left(R_n(\rho_0)W_n\tilde{Y}_{nt}\right)'\tilde{V}_{nt}(\xi_0)\right]$$
$$=\mathbb{E}\left[\frac{1}{m\sigma_0^4}\sum_t \left(R_n(\rho_0)W_nS_n^{-1}(\lambda_0)(R_n^{-1}(\rho_0)\tilde{V}_{nt}(\xi_0)+\tilde{X}_{nt}\beta_0)\right)'\tilde{V}_{nt}(\xi_0)\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[\frac{1}{m\sigma_0^4}\sum_t \left(\underbrace{R_n(\rho_0)W_nS_n^{-1}(\lambda_0)R_n^{-1}(\rho_0)}_{\ddot{G}_n(\lambda_0)}\tilde{V}_{nt}(\xi_0)\right)'\tilde{V}_{nt}(\xi_0)\right] \\
&+\mathbb{E}\left[\frac{1}{m\sigma_0^4}\sum_t \left(R_n(\rho_0)W_n\tilde{X}_{nt}\beta_0\right)'\tilde{V}_{nt}(\xi_0)\right] \\
&= \frac{1}{n\sigma_0^2}\mathrm{tr}(\ddot{G}_n(\lambda_0)) \qquad \text{(B.75)}
\end{aligned}
$$

The second column of $\Sigma_{0,n,T}$ is:

$$
\begin{pmatrix}
\frac{1}{m\sigma_0^2}\sum_t \ddot{X}_{nt}'(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0) \\
\frac{1}{n}\mathrm{tr}\left(\ddot{G}_n^S(\lambda_0)\ddot{G}_n(\lambda_0)\right) + \frac{1}{m\sigma_0^2}\sum_t\left(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0\right)'\left(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0\right) \\
\frac{1}{n}\mathrm{tr}\left(H_n^S(\rho_0)\ddot{G}_n(\lambda_0)\right) \\
\frac{1}{n\sigma_0^2}\mathrm{tr}(\ddot{G}_n(\lambda_0))
\end{pmatrix}
$$

- $$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{m\sigma_0^2}\sum_t \left(H_n(\rho_0)\tilde{V}_{nt}(\xi_0)\right)' H_n(\rho_0)\tilde{V}_{nt}(\xi_0) + \frac{T-1}{m}\mathrm{tr}\left(H_n^2(\rho_0)\right)\right] \\
&= \frac{1}{n}\mathrm{tr}\left(H_n^{'}(\rho_0)H_n(\rho_0) + H_n^2(\rho_0)\right) = \frac{1}{n}\mathrm{tr}\left(H_n^S(\rho_0)H_n(\rho_0)\right) \qquad \text{(B.76)}
\end{aligned}
$$

$$
\mathbb{E}\left[\frac{1}{m\sigma_0^4}\sum_t\left(H_n(\rho_0)\tilde{V}_{nt}(\xi_0)\right)'\tilde{V}_{nt}(\xi_0)\right] = \frac{1}{n\sigma_0^2}\mathrm{tr}(H_n(\rho_0)) \qquad \text{(B.77)}
$$

The third column of $\Sigma_{0,n,T}$ is:

$$\begin{pmatrix} 0_{k\times 1} \\ \frac{1}{n}\mathrm{tr}\left(H_n^S(\rho_0)\ddot{G}_n(\lambda_0)\right) \\ \frac{1}{n}\mathrm{tr}\left(H_n^S(\rho_0)H_n(\rho_0)\right) \\ \frac{1}{n\sigma_0^2}\mathrm{tr}(H_n(\rho_0)) \end{pmatrix}$$

- $$\mathbb{E}\left[-\frac{m}{2m\sigma_0^4}+\frac{1}{m\sigma_0^6}\sum_t \tilde{V}_{nt}'(\xi_0)\tilde{V}_{nt}(\xi_0)\right] = -\frac{1}{2\sigma_0^4}+\frac{1}{m\sigma_0^6}T\frac{T-1}{T}\sigma_0^2\cdot n = \frac{1}{2\sigma_0^4} \tag{B.78}$$

The fourth column of $\Sigma_{0,n,T}$ is:

$$\begin{pmatrix} 0_{k\times 1} \\ \frac{1}{n\sigma_0^2}\mathrm{tr}(\ddot{G}_n(\lambda_0)) \\ \frac{1}{n\sigma_0^2}\mathrm{tr}(H_n(\rho_0)) \\ \frac{1}{2\sigma_0^4} \end{pmatrix}.$$

Thus, we have:

$$\Sigma_{0,T} = \begin{pmatrix} \frac{1}{m\sigma_0^2}\sum_t \ddot{X}_{nt}'\ddot{X}_{nt} & \frac{1}{m\sigma_0^2}\sum_t \ddot{X}_{nt}'(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0) & 0_{k\times 1} & 0_{k\times 1} \\ \frac{1}{m\sigma_0^2}\sum_t (\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0)'\ddot{X}_{nt} & \frac{1}{m\sigma_0^2}\sum_t \left(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0\right)'\left(\ddot{G}_n(\lambda_0)\ddot{X}_{nt}\beta_0\right) & 0 & 0 \\ 0_{1\times k} & 0 & 0 & 0 \\ 0_{1\times k} & 0 & 0 & 0 \end{pmatrix}$$

$$
+\begin{pmatrix}
0_{k\times k} & 0_{k\times 1} & 0_{k\times 1} & 0_{k\times 1} \\
0_{1\times k} & \frac{1}{n}\mathrm{tr}\left(\ddot{G}_n^S(\lambda_0)\ddot{G}_n(\lambda_0)\right) & \frac{1}{n}\mathrm{tr}\left(H_n^S(\rho_0)\ddot{G}_n(\lambda_0)\right) & \frac{1}{n\sigma_0^2}\mathrm{tr}(\ddot{G}_n(\lambda_0)) \\
0_{1\times k} & \frac{1}{n}\mathrm{tr}\left(H_n^S(\rho_0)\ddot{G}_n(\lambda_0)\right) & \frac{1}{n}\mathrm{tr}\left(H_n^S(\rho_0)H_n(\rho_0)\right) & \frac{1}{n\sigma_0^2}\mathrm{tr}(H_n(\rho_0)) \\
0_{1\times k} & \frac{1}{n\sigma_0^2}\mathrm{tr}(\ddot{G}_n(\lambda_0)) & \frac{1}{n\sigma_0^2}\mathrm{tr}(H_n(\rho_0)) & \frac{1}{2\sigma_0^4}
\end{pmatrix}
$$

C Algorithms

C.1 Algorithm 1

Most of the quantities related to the saddlepoint density approximation $p_{n,T}$ and the tail area in (22) are available in closed-form. Here, we provide an algorithm (see Algorithm 1) in which we itemize the main computational steps needed to implement the saddlepoint tail area approximation, for a given transformation q and for a given reference parameter θ_0—it is, e.g., the parameter characterizing the null hypothesis in a simple hypothesis testing, where the tail area probability is an approximate p-value.

Algorithm: Density and tail-area saddlepoint approximation

Input: a sample of Y_{nt} and X_{nt}; the reference parameter θ_0.
Output: density and tail area saddlepoint approximation.

1: Given a sample of Y_{nt} and X_{nt}, first compute the transformed values $\tilde{Y}_{nt}$ and $\tilde{X}_{nt}$, as in (5)
2: Compute $\mu_{n,T}$, σ_g^2, $\varkappa_{n,T}^{(3)}$ and $\varkappa_{n,T}^{(4)}$, using the formulae available in Appendix A.3 and A.4
3: Combine the expressions of the approximate cumulants into the analytical expression of the approximate c.g.f. $\tilde{\mathcal{K}}_{n,T}(\nu)$ given by (A.42) in Appendix A.4
4: Define a grid $\mathcal{Z}$ of points $\{z_j, j = 1, ..., j_{\max}\}$. Select the min (i.e. z_1) and max (i.e. $z_{j_{\max}}$) value of this grid according to the min and max values taken by the parameter of interest
5: **for** $j \leftarrow 1$ ***to*** $j_{\max}$ **do**
| Compute the saddlepoint density approximation $p_{n,T}(z_j)$ as in (20)
end
6: Compute the tail area using formula (22) or by numerical integration of the density $p_{n,T}$

Some additional comments are in order. Step 2 requires the approximation of some expected values, like, e.g., $\mathbb{E}\left[g_{i,T}^2(\psi, P_{\theta_0})\right]$, for $g_{i,T}(\psi, P_{\theta_0})$ as in (17), which numerical methods can provide. For instance, we can rely on numerical integration with respect to the underlying Gaussian distribution P_{θ_0}, or on the Laplace method, or on the approximation of the integrals by Riemann sums, using simulated data. In our experience, the latter approximation represents a good compromise between accuracy, simplicity to code, and computational burden. Moreover for Step 5, one needs to solve Eq. (21) for each grid point. Some well-known methods are available. For instance, Kolassa (2006) p. 84 suggests the use of Newton-Raphson derivative-based methods. Specifically, for a given starting value ν_0 (which is an approximate solution to the saddlepoint equation), a first-order Taylor expansion of the saddlepoint equation yields $\tilde{\mathcal{K}}'_{n,T}(\nu_0)+\tilde{\mathcal{K}}''_{n,T}(\nu_0)(\nu-\nu_0) \approx z$, whose solution is $\nu = \nu_0 + (z - \tilde{\mathcal{K}}'_{n,T}(\nu_0))/\tilde{\mathcal{K}}''_{n,T}(\nu_0)$. We apply this solution to update the approximate solution ν_0, yielding a new approximation ν to the saddlepoint. We iterate the procedure until the approximate solution is accurate enough—e.g., we can set a tolerance value (say, *tol*) and iterate the procedure till $|z - \tilde{\mathcal{K}}'_{n,T}(\nu)| < tol$. An alternative option is the secant method; see Kolassa (2006) p. 86. As noticed by Gatto and Ronchetti (1996), due to the approximate nature of the c.g.f. in (A.42), the saddlepoint equation can admit multiple solutions in some areas of the density. To solve this problem, one may use the modified c.g.f. proposed by Wang (1992).

C.2 Algorithm 2

To implement the test (24) for the problem (23), we propose the following algorithm:

Algorithm: Saddlepoint test in the presence of nuisance parameters

Input: a sample of Y_{nt} and X_{nt}; the testing problem in (23).

Output: saddlepoint test based on the test statistic $SAD_n(\hat{\lambda})$

1: Given a sample of Y_{nt} and X_{nt}, first compute the transformed values $\tilde{Y}_{nt}$ and $\tilde{X}_{nt}$, as in (5)

2: Compute the maximum likelihood estimate $(\hat{\lambda}, \hat{\theta}_2)$

3: **for** $i \leftarrow 1$ ***to*** n **do**

Plug-in $\hat{\lambda}$ in $\psi_i^{(T)}(\lambda, \theta_2)$ and compute $\mathcal{K}_{\psi_i}^{(T)}(\nu, \hat{\lambda}, \theta_2)$

end

4: Compute $\mathcal{K}_{\psi}(\nu, \hat{\lambda}, \theta_2)$ as in (25)

5: Obtain the value of $SAD_n(\hat{\lambda})$ and compare it to the quantile of χ_1^2— a chi-square with one degree-of-freedom.

As far as Step 3 is concerned, we suggest to apply a standard MC integration to approximate numerically the c.g.f. $\mathcal{K}_{\psi}$—this is in line with our suggestion for Algorithm . Specifically, for each location i, we may approximate $\mathcal{K}_{\psi_i}^{(T)}(\nu, \lambda, \theta_2) = \ln E_{P_{(\lambda_0, \theta_2)}}[\exp\{\nu^T \psi_i^{(T)}(\lambda, \theta_2)\}]$ by simulating from the corresponding SARAR model under the null. To this end, for a user-specified MC size (MC.size), in each j-th MC run, we simulate the variables $\{Y_{nt}^{(j)}\}_{t=1,\cdots,T}$, under the probability $P_{(\lambda_0, \theta_2)}$. We remark that $P_{(\lambda_0, \theta_2)}$ is given by the composite null hypothesis, where the nuisance parameters are not specified. Thus, θ_2 is not fixed in the MC runs: the numerical optimization needed to compute the infimum over θ_2 takes care of that aspect. Then, we set

$$\mathcal{K}_{\psi_i^{(T)}}(\nu, \hat{\lambda}, \theta_2) \approx \ln\left(\frac{1}{\text{MC.size}} \sum_{j=1}^{\text{MC.size}} \exp\{\nu^T \psi_{i,j}^{(T)}(\hat{\lambda}, \theta_2)\}\right),$$

where $\psi_{i,j}^{(T)}$ is the estimating function at location i as evaluated in the j-th MC run. Finally, we build the test using the MC approximated c.g.f. $\mathcal{K}_{\psi}(\nu, \hat{\lambda}, \theta_2) = -n^{-1} \sum_{i=1}^{n} \mathcal{K}_{\psi_i^{(T)}}(\nu, \hat{\lambda}, \theta_2)$.

D Additional results for the SAR(1) model

D.1 First-order asymptotics

As it is customary in the statistical/econometric software, we consider three different spatial weight matrices: Rook matrix, Queen matrix, and Queen matrix with torus. In Figure 1, we display the geometry of Y_{nt} as implied by each considered spatial matrix for $n = 100$: the plots highlight that different matrices imply different spatial relations. For instance, we see that the Rook matrix implies fewer links than the Queen matrix. Indeed, the Rook criterion defines neighbours by the existence of a common edge between two spatial units, whilst the Queen criterion is less rigid and defines neighbours as spatial units sharing an edge or a vertex. Besides, we may interpret $\{Y_{nt}\}$ as a n-dimensional random field on the network graph which describes the known underlying spatial structure. Then, W_n represents the weighted adjacency matrix (in the spatial econometrics literature, W_n is called contiguity matrix). In Figure 1, we display the geometry of a random field on a regular lattice (undirected graph).

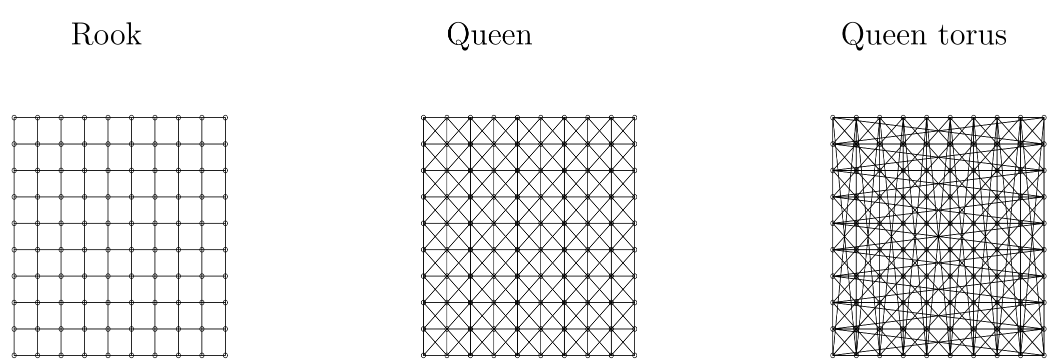

Figure 1: Different types of neighboring structure for Y_{nt}, as implied by different types of W_n matrix, for $n = 100$.

We complement the motivating example of our research illustrating the low accuracy of the routinely applied first-order asymptotics in the setting of the spatial autoregressive process of order one, henceforth SAR(1):

$$Y_{nt} = \lambda_0 W_n Y_{nt} + c_{n0} + V_{nt}, \quad \text{for} \quad t = 1, 2, \tag{D.1}$$

where $V_{nt} = (v_{1t}, v_{2t}, .., v_{nt})'$ are $n \times 1$ vectors, and $v_{it} \sim \mathcal{N}(0, 1)$, i.i.d. across i and t. The model is a special case of the general model in (1), the spatial autoregressive process with spatial autoregressive error (SARAR) of Lee and Yu (2010). Since c_{n0} creates an incidental parameter issue, we eliminate it by the standard differentiation procedure. Given that we have only two periods, the transformed (differentiated) model is formally equivalent to the cross-sectional SAR(1) model, in which $c_{n0} \equiv 0$, a priori; see Robinson and Rossi (2015) for a related discussion.

In the MC exercise, we set $\lambda_0 = 0.2$, and we estimate it through Gaussian likelihood maximisation. The resulting M-estimator (the maximum likelihood estimator) is consistent and asymptotically normal; see §4.1. We consider the same types of W_n as in the motivating example of §2. In Figure 2, we display the MC results. Via QQ-plot, we compare the distribution of $\hat{\lambda}$ to the Gaussian asymptotic distribution (as implied by the first-order asymptotic theory).

The plots show that, for both sample sizes $n = 24$ and $= 100$, the Gaussian approximation can be either too thin or too thick in the tails with respect to the "exact" distribution (as obtained via simulation). The more complex is the geometry of W_n (e.g., W_n is Queen), the more pronounced are the departures from the Gaussian approximation.

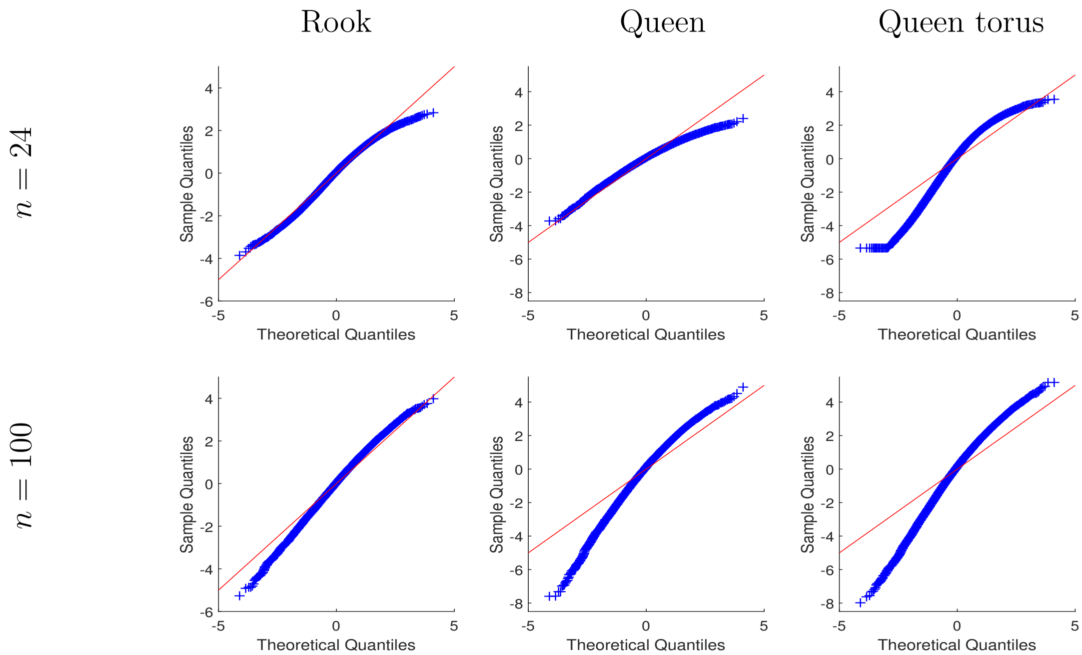

Figure 2: SAR(1) model: QQ-plot vs normal of the MLE $\hat{\lambda}$, for different sample sizes ($n = 24$ and $n = 100$), $\lambda_0 = 0.2$ and different types of W_n matrix.

D.2 Assumption D(iv) and its relation to the spatial weights

We can obtain $M_{i,T}(\psi, P_{\theta_0})$ from (11) on a SAR(1) model as in (D.1),

$$M_{i,T}(\psi, P_{\theta_0}) = (T-1)(\tilde{g}_{ii} + g_{ii}),$$

where $\tilde{g}_{ii}$ and g_{ii} are the i_{th} element of the diagonals of $G_n G_n'$ and $G_n^2(\lambda_0)$, respectively. Note that $S_n(\lambda_0) = I_n - \lambda_0 W_n$ and $G_n(\lambda_0) = W_n S_n^{-1}(\lambda_0)$.

To check Assumption D(iv): $||M_{i,T}(\psi, P_{\theta_0}) - M_{j,T}(\psi, P_{\theta_0})|| = O(n^{-1})$, first let us find the expression for the difference between $M_{i,T}(\psi, P_{\theta_0})$ and $M_{j,T}(\psi, P_{\theta_0})$,

$$M_{i,T}(\psi, P_{\theta_0}) - M_{j,T}(\psi, P_{\theta_0}) \quad = \quad (T-1)\left[(\tilde{g}_{ii} - \tilde{g}_{jj}) + (g_{ii} - g_{jj})\right], \qquad \text{(D.2)}$$

where $\tilde{g}_{ii}$ and $\tilde{g}_{ii}$ are i_{th} and j_{th} elements of the diagonal of $G_n G_n'$. g_{ii} snd g_{jj} the i_{th} and j_{th} elements of the diagonal of $G_n^2(\lambda_0)$. Then, we rewrite the expression of $G_n^2(\lambda_0)$ to check its diagonal:

$$\begin{aligned} G_n^2(\lambda_0) &= W_n^2 S_n^{-2}(\lambda_0) = W_n^2 \left(I_n - \lambda_0 W_n\right)^{-2} \\ &= W_n^2 \left(I_n + \lambda_0 W_n + \lambda_0^2 W_n^2 + \lambda_0^3 W_n^3 + \cdots\right)^2 \\ &= \sum_{k=2}^{\infty} a_k W_n^k, \end{aligned} \tag{D.3}$$

where a_k is the coefficient of W_n^k. By comparing the entries of diagonals of W_n^k, we can make an assumption on the differences between any two entries of the diagonal of W_n^k that should be at most of order n^{-1}, for all integers $k \geq 2$, such that $||g_{ii} - g_{jj}|| = O(n^{-1})$, for any $1 \leq i$, $j \leq n$ and $i \neq j$.

We do the same for $G_n G_n'$:

$$\begin{aligned} G_n G_n' &= W_n S_n^{-1}(\lambda_0) \left(W_n S_n^{-1}(\lambda_0)\right)' = W_n \left(I_n - \lambda_0 W_n\right)^{-1} \left(W_n \left(I_n - \lambda_0 W_n\right)^{-1}\right)' \\ &= W_n \left(I_n + \lambda_0 W_n + \lambda_0^2 W_n^2 + \lambda_0^3 W_n^3 + \cdots\right) \left(I_n + \lambda_0 W_n + \lambda_0^2 W_n^2 + \lambda_0^3 W_n^3 + \cdots\right)' W_n' \\ &= \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} a_{k_1 k_2} W_n^{k_1} \left(W_n^{k_2}\right)', \end{aligned} \tag{D.4}$$

where $a_{k_1 k_2}$ is the coefficient of $W_n^{k_1} \left(W_n^{k_2}\right)'$. Then, we can make a second assumption on the differences between any two entries of the diagonal of $W_n^{k_1} \left(W_n^{k_2}\right)'$ that should be at most of order n^{-1}, for all positive integers k_1 and k_2, such that $||\tilde{g}_{ii} - \tilde{g}_{jj}|| = O(n^{-1})$, for any for any $1 \leq i$, $j \leq n$ and $i \neq j$. From those two assumptions on the behaviour of diagonal entries, we notice that W_n plays a leading role.

To exemplify those two conditions, we consider numerical examples for different choices of W_n. The sample size n is 24 with the benchmark $1/n = 1/24 = 0.041$. When W_n is

Rook, the elements of W_n are $\{1/2, 1/3, 1/4, 0\}$. On top of that, we can find the differences between the entries of the diagonals of W_n^k and $W_n^{k_1}\left(W_n^{k_2}\right)'$ are small. For instance, the entries of the diagonal of W_n^2 are $\{0.27, 0.29, 0.31, 0.33, 0.36\}$. The elements of diagonal of W_n^3 are all 0. For $W_n^2(W_n^2)'$, the entries of its diagonal are {0.17,0.21,0.22,0.37}. We can obtain a similar result for Queen weight matrix. For example, the largest difference between diagonal entries of W_n^2 is 0.044. When W_n is Queen with torus, W_n is a symmetric matrix and its elements are 0 or 1/8. The entries of the diagonals of W_n^k and $W_n^{k_1}\left(W_n^{k_2}\right)'$ are always identical to each other. Thus, the differences are always 0. Since the entries of W_n are all positive and less than 1, when the exponent becomes larger, all the values of the powers of W_n and W_n' will go close to 0.00. Thus, large powers of W_n definitely satisfy the two assumptions.

D.3 First-order asymptotics vs saddlepoint approximation

For the SAR(1) model, we analyse the behaviour of the MLE of λ_0, whose PP-plots are available in Figure 3. For each type of W_n, for $n = 100$, the plots show that the saddlepoint approximation is closer to the "exact" probability than the first-order asymptotics approximation. For W_n Rook, the saddlepoint approximation improves on the routinely-applied first-order asymptotics. In Figure 3, the accuracy gains are evident also for W_n Queen and Queen with torus, where the first-order asymptotic theory displays large errors essentially over the whole support (specially in the tails). On the contrary, the saddlepoint approximation is close to the 45-degree line.

Density plots show the same information as PP-plots displayed in Figure 3 (of the paper) and Figure 3 (in this Appendix), but provide a better visualization of the behavior of the considered approximations. Thus, we compute the Gaussian density implied by the asymptotic theory, and we compare it to our saddlepoint density approximation. In Figure 4, we plot the histogram of the "exact" estimator density (as obtained using 25,000

Rook Queen Queen torus

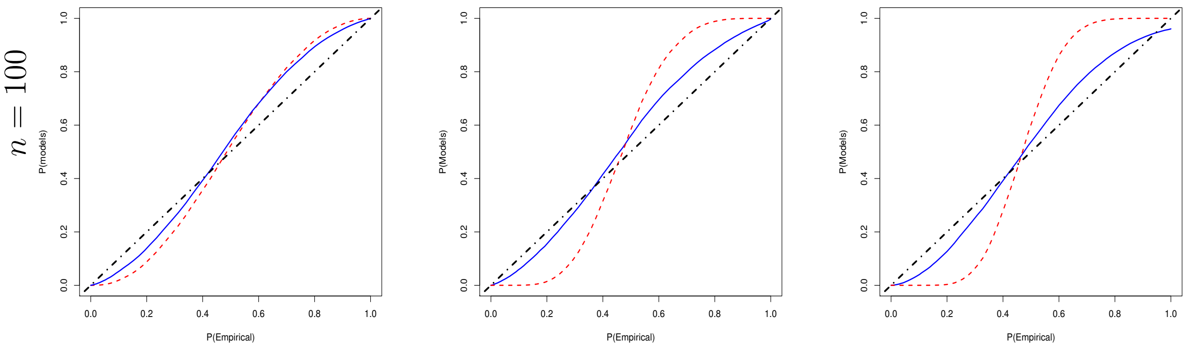

Figure 3: SAR(1) model: PP-plots for saddlepoint (continuous line) vs asymptotic normal (dotted line) probability approximation, for the MLE $\hat{\lambda}$, for $n = 100$, $\lambda_0 = 0.2$, and different W_n.

Monte Carlo runs) to which we superpose both the Gaussian and the saddlepoint density approximation. W_n are Rook and Queen. The plots illustrate that the saddlepoint technique provides an approximation to the true density which is more accurate than the one obtained using the first-order asymptotics.

D.4 Saddlepoint approximation vs Edgeworth expansion

The Edgeworth expansion derived in Proposition 3 represents the natural alternative to the saddlepoint approximation since it is fully analytic. Thus, we compare the performance of the two approximations, looking at their relative error for the approximation of the tail area probability. We keep the same Monte Carlo design as in Section 5.1, namely $n = 24$, and we consider different values of z, as in (22). Figure 5 displays the absolute value of the relative error, i.e., |approximation/exact − 1|, when W_n is Rook, Queen, and Queen torus. The plots illustrate that the relative error yielded by the saddlepoint approximation is smaller (down to ten times smaller in the case of Rook and Queen Torus) than the

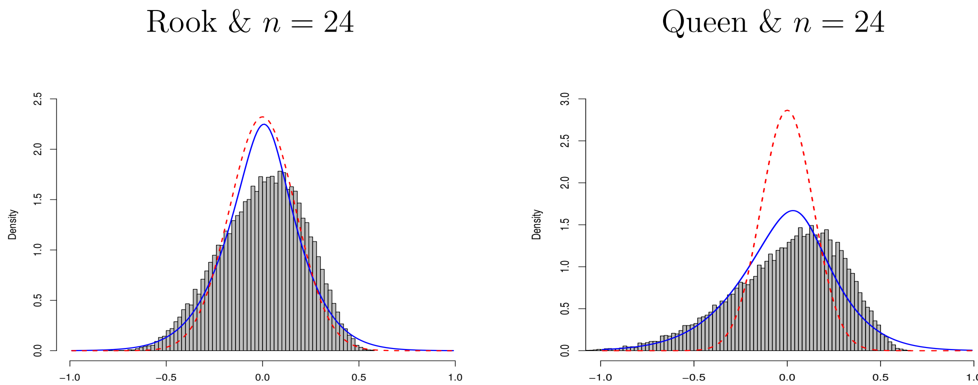

Figure 4: SAR(1) model: Density plots for saddlepoint (continuous line) vs asymptotic normal (dotted line) probability approximation to the exact density (as expressed by the histogram and obtained using MC with size 25000), for the MLE $\hat{\lambda}$ and W_n is Rook (left panel) and Queen (right panel). Sample size is $n = 24$, while $\lambda_0 = 0.2$.

relative error entailed by the first-order asymptotic approximation (which is always about 100%). The Edgeworth expansion entails a relative error which is typically higher than the one entailed by the saddlepoint approximation—the expansion can even become negative in some parts of the support, with relative error above 100%.

D.5 Saddlepoint vs parametric bootstrap

The parametric bootstrap represents a (computer-based) competitor, commonly applied in statistics and econometrics. To compare our saddlepoint approximation to the one obtained by bootstrap, we consider different numbers of bootstrap repetitions, labeled as B: we use $B = 499$ and $B = 999$. For space constraints, in Figure 6, we display the results for $B = 499$ (similar plots are available for $B = 999$) showing the functional boxplots (as obtained iterating the procedure 100 times) of the bootstrap approximated density, for sample size $n = 24$ and for W_n is Queen. To visualize the variability entailed by the bootstrap, we display the first and third quartile curves (two-dash lines) and the median

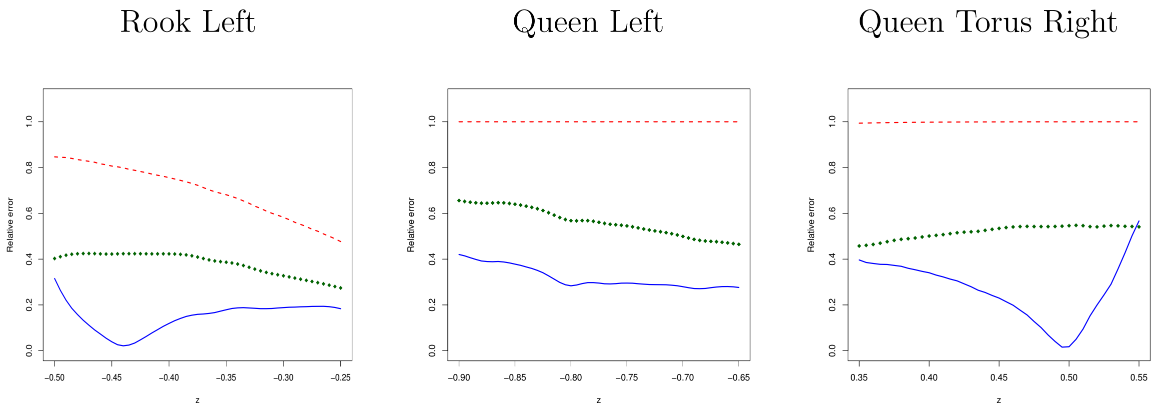

Figure 5: SAR(1) model: Relative error (in absolute value) for the approximate right and left tail probability, as obtained using the Gaussian asymptotic theory (dotted line), the Edgeworth approximation (dotted line with diamonds) and saddlepoint approximation (continuous line), for the MLE $\hat{\lambda}$. In each plot, on the x-axes we display different values of z. The sample size is $n = 24$, $\lambda_0 = 0.2$, and W_n is Rook (left tail), Queen (left tail), and Queen Torus (right tail).

functional curve (dotted line with crosses); for details about functional boxplots, we refer to Sun and Genton (2011) and to R routine `fbplot`. We notice that, while the bootstrap median functional curve (representing a typical bootstrap density approximation) is close to the actual density (as represented by the histogram), the range between the quartile curves illustrates that the bootstrap approximation has a variability. Clearly, the variability depends on B: the larger is B, the smaller is the variability. However, larger values of B entail bigger computational costs: when $B = 499$, the bootstrap is almost as fast as the saddlepoint density approximation (computation time about 7 minutes, on a 2.3 GHz Intel Core i5 processor), but for $B = 999$, it is three times slower. We refer to Jeganathan et al. (2015) for comments on the computational burden of the parametric bootstrap and on the possibility to use saddlepoint approximations to speed up the computation. For $B = 499$ and zooming on the tails, we notice that in the right tail and in the center of the density, the bootstrap yields an approximation (slightly) more accurate than the saddlepoint method,

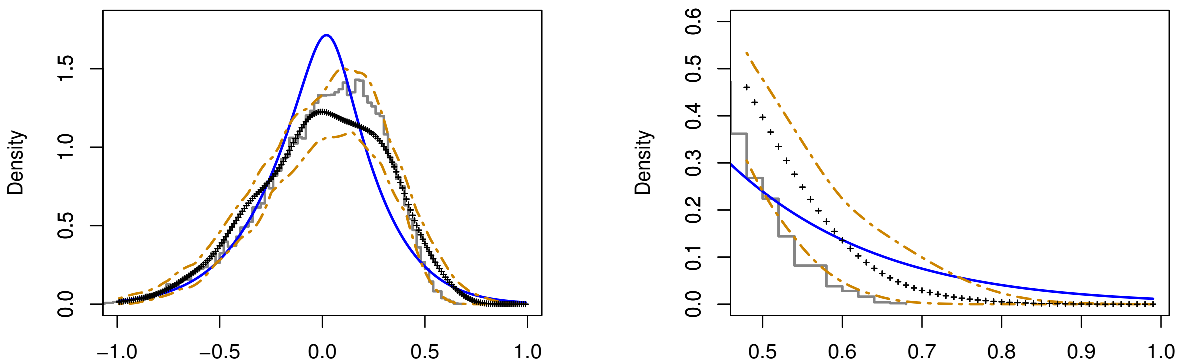

Figure 6: SAR(1) model. Left panel: Density plots for saddlepoint (continuous line) vs the functional boxplot of the parametric bootstrap probability approximation to the exact density (as expressed by the histogram and obtained using MC with size 25000), for the MLE $\hat{\lambda}$ and W_n is Queen. Sample size is $n = 24$, while $\lambda_0 = 0.2$. Right panel: zoom on the right tail. In each plot, we display the functional central curve (dotted line with crosses), the 1_{st} and 3_{rd} functional quartile (two-dash lines).

but the saddlepoint approximation is either inside or extremely close to the bootstrap quartile curves (see right panel of Figure 6). In the left tail, the saddlepoint density approximation is closer to the true density than the bootstrap typical functional curve or $\lambda \leq -0.85$. Thus, overall, we cannot conclude that the bootstrap dominates uniformly (in terms of accuracy improvements over the whole domain) the saddlepoint approximation. Even if we are ready to accept a larger computational cost, the accuracy gains yielded by the bootstrap are yet not fully clear: also for $B = 999$, the bootstrap does not dominate uniformly the saddlepoint approximation. Finally, for $B = 49$, the bootstrap is about eight times faster than the saddlepoint approximation, but this gain in speed comes with a large cost in terms of accuracy. As an illustration, for $\hat{\lambda} - \lambda_0 = -0.8$ (left tail), the true density is 0.074, the saddlepoint density approximation is 0.061, while the bootstrap median value is 0.040, with a wide spread between the first and the third quartile, being 0.009 and 0.108 respectively.

D.6 Saddlepoint test for composite hypotheses: plug-in approach

Our saddlepoint density and/or tail approximations are helpful for testing simple hypotheses about θ_0; see §5.1 of the paper. Another interesting case suggested by the Associate Editor and an anonymous referee that has a strong practical relevance is related to testing a composite null hypothesis. It is a problem which is different from the one considered so far in the paper, because it raises the issue of dealing with nuisance parameters.

To tackle this problem, several possibilities are available. For instance, we may fix the nuisance parameters at the MLE estimates. Alternatively, we may consider to use the (re-centered) profile estimators, as suggested, e.g., in Hillier and Martellosio (2018) and Martellosio and Hillier (2020). Combined with the saddlepoint density in (20), these techniques yield a ready solution to the nuisance parameter problem. In our numerical experience, these solutions may preserve reasonable accuracy in some cases.

To illustrate this aspect, we consider a numerical exercise about SAR(1) model, in which we set the nuisance parameter equal to the MLE estimate. Specifically, we consider a SAR(1) with parameter $\theta = (\lambda, \sigma^2)'$ and our goal is to test $\mathcal{H}_0 : \lambda = \lambda_0 = 0$ versus $\mathcal{H}_1 : \lambda > 0$, while leaving σ^2 unspecified. To study the performance of the saddlepoint density approximation for the sampling distribution of $\hat{\lambda}$, we perform a MC study, in which we set $n = 24$, $T = 2$, and $\sigma_0^2 = 1$. For each simulated (panel data) sample, we estimate the model parameter via MLE and get $\hat{\theta} = (\hat{\lambda}, \hat{\sigma}^2)'$. Then, we compute the saddlepoint density in (20) using $\hat{\sigma}^2$ in lieu of σ_0^2: for each sample, we have a saddlepoint density approximation. We repeat this procedure 100 times, for W_n Rook and Queen, so we obtain 100 saddlepoint density approximations for each W_n type. In Figure 7, we display functional boxplots of the resulting saddlepoint density approximation for the MLE $\hat{\lambda}$. To have a graphical idea of the saddlepoint approximation accuracy, on the same plot we superpose the histogram of $\hat{\lambda}$, which represents the sampling distribution of the estimated parameter. Even a quick inspection of the right and left plots suggests that the resulting saddlepoint density approximation preserves (typically) a good performance. Indeed, we find that the median functional curve (dotted line with crosses, which we consider as the typical saddlepoint density approximation) is close to the histogram and it gives a good performance in the tails, for both W_n Rook and Queen. The range between the first and third quartile curves (two-dash lines) illustrates the variability of the saddlepoint approximation. When W_n is Queen, even though there is a departure from the median curve from the histogram over the x-axis intervals $(-0.75, 0)$ and $(0.25, 0.5)$, the histogram is inside the functional confidence interval expressed by the first and third curves. Thus, we conclude that computing the p-value for testing $\mathcal{H}_0$ using $\hat{\sigma}^2$ in the expression of the saddlepoint density approximation seems to preserve accuracy in this SAR(1) model, even for such a small sample.

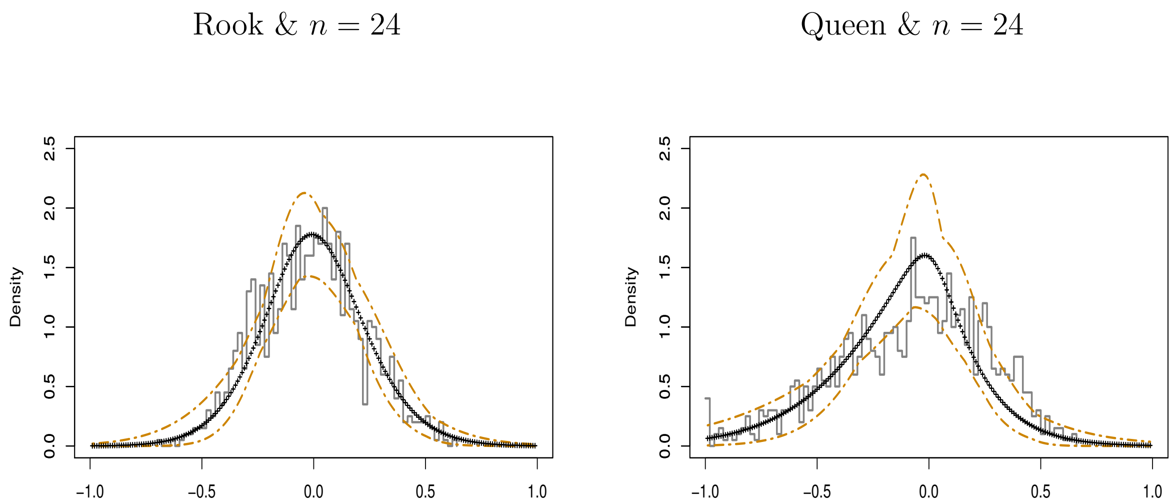

Figure 7: SAR(1) model: Functional boxplots of saddlepoint density approximation to the exact density (as expressed by the histogram), for the MLE $\hat{\lambda}$ and W_n is Rook (left panel) and Queen (right panel). Sample size is $n = 24$, while $\lambda_0 = 0$.

References

Barndorff-Nielsen, O. and Cox, D. (1989), *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, London.

Bickel, P., Götze, F., and Van Zwet, W. (1986), The Edgeworth Expansion for U-statistics of Degree Two, *The Annals of Statistics*, 14, 1463–1484.

Cabrera, J. and Fernholz, L. T. (1999), Target Estimation for Bias and Mean Square Error Reduction, *The Annals of Statistics*, 27, 1080–1104.

Callaert, H. and Janssen, P. (1978), The Berry-Esseen Theorem for U-Statistics, *The Annals of Statistics*, 6, 417–421.

Easton, G. S. and Ronchetti, E. (1986), General Saddlepoint Approximations with Applications to L-statistics, *Journal of the American Statistical Association*, 81, 420–430.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, vl. 2, Wiley, New York.

Field, C. (1982), Small Sample Asymptotic Expansions for Multivariate M-estimates, *The Annals of Statistics*, 10, 672–689.

Filippova, A. (1962), Mises' Theorem on the Asymptotic Behavior of Functionals of Empirical Distribution Functions and Its Statistical Applications, *Theory of Probability & Its Applications*, 7, 24–57.

Gatto, R. and Ronchetti, E. (1996), General Saddlepoint Approximations of Marginal Densities and Tail Probabilities, *Journal of the American Statistical Association*, 91, 666–673.

Hillier, G. and Martellosio, F. (2018), Exact and Higher-order Properties of the MLE in Spatial Autoregressive Models, With Applications to Inference, *Journal of Econometrics*, 205, 402–422.

Jeganathan, P., Paige, R. L., and Trindade, A. A. (2015), Saddlepoint-based Bootstrap Inference for the Spatial Dependence Parameter in the Lattice Process, *Spatial Statistics*, 12, 1–14.

Kolassa, J. (2006), *Series Approximation Methods in Statistics*, vl. 88, Springer, New York.

Lee, L. and Yu, J. (2010), Estimation of Spatial Autoregressive Panel Data Models With Fixed Effects, *Journal of Econometrics*, 154, 165–185.

Lugannani, R. and Rice, S. (1980), Saddlepoint Approximation for the Distribution of the Sum of Independent Random Variables, *Advances in applied probability*, 12, 475–490.

Martellosio, F. and Hillier, G. (2020), Adjusted QMLE for the Spatial Autoregressive Parameter, *Journal of Econometrics*, 219, 488–506.

Robinson, P. M. and Rossi, F. (2015), Refinements in Maximum Likelihood Inference on Spatial Autocorrelation in Panel Data, *Journal of Econometrics*, 189, 447–456.

Serfling, R. J. (2009), *Approximation Theorems of Mathematical Statistics*, vl. 162, Wiley, New York.

Sun, Y. and Genton, M. G. (2011), Functional Boxplots, *Journal of Computational and Graphical Statistics*, 20, 316–334.

Van der Vaart, A. W. (1998), *Asymptotic Statistics*, vl. 3, Cambridge University Press.

von Mises, R. (1947), On the Asymptotic Distribution of Differentiable Statistical Functions, *Annals of Mathematical Statistics*, 18, 309–348.

Wang, S. (1992), General Saddlepoint Approximations in the Bootstrap, *Statistics & Probability Letters*, 13, 61–66.

Withers, C. (1983), Expansions for the Distribution and Quantiles of a Regular Functional of the Empirical Distribution With Applications to Nonparametric Confidence Intervals, *The Annals of Statistics*, 11, 577–587.