Measure-Transformed Gaussian Quasi Score Test in the Presence of Nuisance Parameters: Supplementary Material

Koby Todros
Ben-Gurion University of the Negev

In this supplementary material document, we provide proofs for the theorems stated throughout the paper. Furthermore, we provide implementation details of the quasi GLRT and the quasi score test under the assumption that the observations obey a generalized Gaussian distribution (GGD) [s1].

I. Proofs for the Theorems

A. Proof of Theorem 1

Under assumptions (A-1)−(A-7) we show in Lemma 1 that

$$\sqrt{N}\hat{\boldsymbol{\eta}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0}) \xrightarrow[N\to\infty]{D} \mathcal{N}\left(\mathbf{0},\mathbf{B}_{u,r}\left(\boldsymbol{\theta}_0\right)\right) \text{ under } H_0, \quad \text{(S-1)}$$

where

$$\mathbf{B}_{u,r}(\boldsymbol{\theta}) \triangleq \mathbf{H}_{u,r}(\boldsymbol{\theta})\mathbf{R}_{u,r}(\boldsymbol{\theta})\mathbf{H}_{u,r}(\boldsymbol{\theta}), \quad \text{(S-2)}$$

and $\mathbf{R}_{u,r}(\boldsymbol{\theta})$ is a matrix formed by intersection of the first m_r rows and column of the matrix $\mathbf{R}_u(\boldsymbol{\theta})$ defined in (22). The marix

$$\mathbf{H}_{u,r}(\boldsymbol{\theta}) \triangleq \mathbf{F}_{u,r}\left(\boldsymbol{\theta}\right) - \mathbf{F}_{u,rs}\left(\boldsymbol{\theta}\right)\mathbf{F}^{-1}_{u,s}\left(\boldsymbol{\theta}\right)\mathbf{F}^{T}_{u,rs}\left(\boldsymbol{\theta}\right), \quad \text{(S-3)}$$

where $\mathbf{F}_{u,r}\left(\boldsymbol{\theta}\right) \in \mathbb{R}^{m_r\times m_r}$, $\mathbf{F}_{u,rs}\left(\boldsymbol{\theta}\right) \in \mathbb{R}^{m_r\times m_s}$ and $\mathbf{F}_{u,s}\left(\boldsymbol{\theta}\right) \in \mathbb{R}^{m_s\times m_s}$ are the subblocks of the matrix $\mathbf{F}_u\left(\boldsymbol{\theta}\right)$ (24) that are obtained from the partition:

$$\mathbf{F}_u\left(\boldsymbol{\theta}\right) = \begin{bmatrix} \mathbf{F}_{u,r}\left(\boldsymbol{\theta}\right) & \mathbf{F}_{u,rs}\left(\boldsymbol{\theta}\right) \\ \mathbf{F}^{T}_{u,rs}\left(\boldsymbol{\theta}\right) & \mathbf{F}_{u,s}\left(\boldsymbol{\theta}\right)) \end{bmatrix}. \quad \text{(S-4)}$$

Furthermore, by Assumptions (A-1) and (A-3)−(A-6), it follows from [s2, Th. 1] that $\hat{\boldsymbol{\theta}}_{s_0}$ is a strongly consistent estimator of the nuisance vector parameter $\boldsymbol{\theta}_{s_0}$. Therefore, similarly to the proof of Theorem 3 in [s2], it can be shown that when conditions (A-4) and (A-6) are satisfied, then

$$\hat{\mathbf{B}}_{u,r}(\tilde{\boldsymbol{\theta}}_0) \xrightarrow[N\to\infty]{P} \mathbf{B}_{u,r}(\boldsymbol{\theta}_0) \text{ under } H_0. \quad \text{(S-5)}$$

By Lemma 2, stated below, the covariance matrix $\mathbf{B}_{u,r}(\boldsymbol{\theta}_0)$ is non-singular under condition (A-7).

Hence, relation (18) follows from (10), (S-1), (S-5), Slutskey's Theorem [s3], Mann-Wald's Theorem [s4], and the properties of quadratic forms of Gaussian random variables [s5].

Lemma 1. *The convergence in (S-1) holds under assumptions (A-1)−(A-7).*

Proof. We first recall that under H_0, $\mathbf{X}_1,\ldots,\mathbf{X}_N$ is a sequence of i.i.d. samples from $P_{\mathbf{x};\boldsymbol{\theta}_0}$, where $\boldsymbol{\theta}_0 \triangleq [\boldsymbol{\theta}^T_{r_0},\boldsymbol{\theta}^T_{s_0}]^T$. By (11) and the definition of $\tilde{\boldsymbol{\theta}}_{u,0}$

$$\sqrt{N}\hat{\boldsymbol{\eta}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0}) = \frac{1}{\sqrt{N}}\sum_{n=1}^{N} u(\mathbf{X}_n)\boldsymbol{\psi}_{u,r}(\mathbf{X}_n;\boldsymbol{\theta}_{r_0},\hat{\boldsymbol{\theta}}_{s_0}). \quad \text{(S-6)}$$

As stated below Eq. (S-4), $\hat{\boldsymbol{\theta}}_{s_0}$ is a strongly consistent estimator of the nuisance vector parameter $\boldsymbol{\theta}_{s_0}$, which by Assumption (A-2) lies in the interior of $\boldsymbol{\Theta}_s$. Therefore, we conclude that $\hat{\boldsymbol{\theta}}_{s_0}$ lies in an open neighbourhood $\mathcal{U}_s \subset \boldsymbol{\Theta}_s$ of $\boldsymbol{\theta}_{s_0}$ with sufficiently high probability as N gets large. Therefore, since by Assumption (A-4) $\boldsymbol{\psi}_{u,r}(\mathbf{X}_n;\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}_s)$ is continuous over $\boldsymbol{\Theta}_s$ w.p.1, then by the mean-value Theorem [s6, Th. 3.4]

$$\begin{aligned}\sqrt{N}\hat{\boldsymbol{\eta}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0}) &= \frac{1}{\sqrt{N}}\sum_{n=1}^{N} u(\mathbf{X}_n)\boldsymbol{\psi}_{u,r}(\mathbf{X}_n;\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}_{s_0}) \\ &- \hat{\mathbf{F}}_{u,rs}(\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}^{*}_{s_0})\sqrt{N}(\hat{\boldsymbol{\theta}}_{s_0}-\boldsymbol{\theta}_{s_0}), \quad \text{(S-7)}\end{aligned}$$

where $\hat{\mathbf{F}}_{u,rs}(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s) = \hat{\mathbf{F}}_{u,rs}(\boldsymbol{\theta})$ is given in (17) and $\boldsymbol{\theta}^{*}_{s_0}$ lies in the line segment connecting $\hat{\boldsymbol{\theta}}_{s_0}$ and $\boldsymbol{\theta}_{s_0}$.

By [s2, Eq. (72)], that follows from (A-1)−(A-7), we have

$$\begin{aligned}\sqrt{N}(\hat{\boldsymbol{\theta}}_{s_0}-\boldsymbol{\theta}_{s_0}) &= \hat{\mathbf{F}}^{-1}_{u,s}(\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}^{**}_{s_0}) \quad \text{(S-8)} \\ &\times \frac{1}{\sqrt{N}}\sum_{n=1}^{N} u(\mathbf{X}_n)\boldsymbol{\psi}_{u,s}(\mathbf{X}_n;\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}_{s_0}),\end{aligned}$$

with sufficiently high probability as N gets large, where $\hat{\mathbf{F}}_{u,s}(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s) = \hat{\mathbf{F}}_{u,s}(\boldsymbol{\theta})$ is given in (17), $\boldsymbol{\theta}^{**}_{s_0}$ lies in the line segment connecting $\hat{\boldsymbol{\theta}}_{s_0}$ and $\boldsymbol{\theta}_{s_0}$, and $\boldsymbol{\psi}_{u,s}(\mathbf{X}_n;\boldsymbol{\theta}_r,\boldsymbol{\theta}_s) = \boldsymbol{\psi}_{u,s}(\mathbf{X};\boldsymbol{\theta}) \triangleq \nabla_{\boldsymbol{\theta}_s}\Lambda_u(\mathbf{X};\boldsymbol{\theta})$. Therefore, by (S-7), (S-8) and the definition of $\boldsymbol{\psi}_u(\mathbf{X};\boldsymbol{\theta})$, below (15), we conclude that

$$\begin{aligned}\sqrt{N}\hat{\boldsymbol{\eta}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0}) &= [\mathbf{I}_{m_r}, -\hat{\mathbf{F}}_{u,rs}(\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}^{*}_{s_0})\hat{\mathbf{F}}^{-1}_{u,s}(\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}^{**}_{s_0})] \\ &\times \frac{1}{\sqrt{N}}\sum_{n=1}^{N} u(\mathbf{X}_n)\boldsymbol{\psi}_{u}(\mathbf{X}_n;\boldsymbol{\theta}_0), \quad \text{(S-9)}\end{aligned}$$

where $\mathbf{I}_p$ denotes a $p\times p$ identity matrix. By the strong consistency of $\hat{\boldsymbol{\theta}}_{s_0}$, assumptions (A-4), (A-6), (A-7), [s2, Lemma 3] and the Mann-Wald Theorem [s4], we conclude that

$$\hat{\mathbf{F}}_{u,rs}(\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}^{*}_{s_0})\hat{\mathbf{F}}^{-1}_{u,s}(\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}^{**}_{s_0}) \xrightarrow[N\to\infty]{P} \mathbf{F}_{u,rs}(\boldsymbol{\theta}_0)\mathbf{F}^{-1}_{u,s}(\boldsymbol{\theta}_0), \quad \text{(S-10)}$$

where "P" denotes convergence in probability [s3]. Furthermore, by Assumptions (A-4), (A-6) and [s2, Lemma 5] it follows that

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}u(\mathbf{X}_n)\boldsymbol{\psi}_u(\mathbf{X}_n;\boldsymbol{\theta}_0)\xrightarrow[N\to\infty]{D}\mathcal{N}\left(\mathbf{0},\mathbf{G}_u\left(\boldsymbol{\theta}_0\right)\right), \tag{S-11}$$

where $\mathbf{G}_u\left(\boldsymbol{\theta}\right)$ is defined in (23). Hence, by (S-9)-(S-11) and Slutskey's Theorem [s3], we conclude that

$$\sqrt{N}\hat{\boldsymbol{\eta}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0})\xrightarrow[N\to\infty]{D}\mathcal{N}\left(\mathbf{0},\mathbf{A}_{u,r}\left(\boldsymbol{\theta}_0\right)\mathbf{G}_u\left(\boldsymbol{\theta}_0\right)\mathbf{A}_{u,r}\left(\boldsymbol{\theta}_0\right)\right), \tag{S-12}$$

where $\mathbf{A}_{u,r}\left(\boldsymbol{\theta}_0\right)\triangleq[\mathbf{I}_{m_r},\mathbf{F}_{u,rs}(\boldsymbol{\theta}_0)\mathbf{F}_{u,s}^{-1}(\boldsymbol{\theta}_0)]$. Since by (22) $\mathbf{G}_u\left(\boldsymbol{\theta}\right)=\mathbf{F}_u(\boldsymbol{\theta})\mathbf{R}_u(\boldsymbol{\theta})\mathbf{F}_u(\boldsymbol{\theta})$, it follows from (S-2)-(S-4) that $\mathbf{A}_{u,r}\left(\boldsymbol{\theta}_0\right)\mathbf{G}_u\left(\boldsymbol{\theta}_0\right)\mathbf{A}_{u,r}\left(\boldsymbol{\theta}_0\right)=\mathbf{B}_{u,r}\left(\boldsymbol{\theta}_0\right)$, which completes the proof. □

Lemma 2. *By Assumption (A-7), the matrix $\mathbf{B}_{u,r}(\boldsymbol{\theta})$ is non-singular at $\boldsymbol{\theta}_0$.*

Proof. To prove the Lemma, we need to show that $\mathbf{H}_{u,r}(\boldsymbol{\theta}_0)$ and $\mathbf{R}_{u,r}(\boldsymbol{\theta}_0)$ are non-singular matrices. Since by Assumption (A-7) $\mathbf{F}_u\left(\boldsymbol{\theta}_0\right)$ and $\mathbf{F}_{u,s}\left(\boldsymbol{\theta}_0\right)$ are non-singular, then by [s7, Th. A.3.3], it follows that $\mathbf{H}_{u,r}(\boldsymbol{\theta}_0)$ is a non-singular matrix. Furthermore, since by Assumption (A-7) $\mathbf{G}_u\left(\boldsymbol{\theta}_0\right)$ is non-singular, it follows from (22) and (23) that $\mathbf{R}_u(\boldsymbol{\theta}_0)$ is a positive-definite matrix. Thus, by [s8, Th. 7.7.7], $\mathbf{R}_{u,r}(\boldsymbol{\theta}_0)$ must be non-singular. □

B. Proof of Theorem 2

In Proposition 1 and Lemma 3 stated below, we show that under the sequence of local alternatives (19)

$$\sqrt{N}\hat{\boldsymbol{\eta}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0})\xrightarrow[N\to\infty]{D}\mathcal{N}\left(\mathbf{H}_{u,r}\left(\boldsymbol{\theta}_0\right)\mathbf{r},\mathbf{B}_{u,r}\left(\boldsymbol{\theta}_0\right)\right)$$

and

$$\hat{\mathbf{B}}_{u,r}(\tilde{\boldsymbol{\theta}}_0)\xrightarrow[N\to\infty]{P}\mathbf{B}_{u,r}(\boldsymbol{\theta}_0).$$

Hence, by (10) the convergence in (20) follows from Slutskey's Theorem [s3], Mann-Wald's Theorem [s4], and the properties of quadratic forms of Gaussian random variables [s5].

Remark 1 (Triangular array). *The proofs of Proposition 1 and Lemma 3 are based on the fact that since by (19) the vector parameter of interest $\boldsymbol{\theta}_r$ varies with the sample size N, the observations form a triangular array [s9, p. 31] (rather than a sequence) of random vectors:*

$$\mathbf{X}_{N,k},\quad k=1,\ldots,N,\quad N\geq 1, \tag{S-13}$$

where $\mathbf{X}_{N,k}$, $k=1,\ldots,N$ are i.i.d. samples from the probability distribution $P_{\mathbf{x};\boldsymbol{\theta}_N}$, $\boldsymbol{\theta}_N\triangleq[(\boldsymbol{\theta}_{r_0}+\frac{\mathbf{r}}{\sqrt{N}})^T,\boldsymbol{\theta}_{s_0}^T]^T$.

Proposition 1. *Assume that conditions (A-1)–(A-9) are satisfied. Under the local alternatives (19)*

$$\sqrt{N}\hat{\boldsymbol{\eta}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0})\xrightarrow[N\to\infty]{D}\mathcal{N}\left(\mathbf{H}_{u,r}\left(\boldsymbol{\theta}_0\right)\mathbf{r},\mathbf{B}_{u,r}\left(\boldsymbol{\theta}_0\right)\right) \tag{S-14}$$

Proof. By Assumption (A-2) $\boldsymbol{\theta}_{r_0}$ lies in the interior of $\boldsymbol{\Theta}_r$. Thus, for large enough N, we can conclude from (19) that $\boldsymbol{\theta}_r=\boldsymbol{\theta}_{r_0}+\frac{\mathbf{r}}{\sqrt{N}}$ lies in an open neighbourhood $\mathcal{U}_r\subset\boldsymbol{\Theta}_r$ of $\boldsymbol{\theta}_{r_0}$. Furthermore, by Assumption (A-4) $\boldsymbol{\psi}_{u,r}(\mathbf{X}_n;\boldsymbol{\theta}_r,\boldsymbol{\theta}_s)$ is continuous over $\boldsymbol{\Theta}_r$ w.p.1. Therefore, by the mean-value Theorem [s6, Th. 3.4] and (S-6) we have:

$$\begin{aligned}\sqrt{N}\hat{\boldsymbol{\eta}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0}) &= \frac{1}{\sqrt{N}}\sum_{n=1}^{N}u(\mathbf{X}_n)\boldsymbol{\psi}_{u,r}(\mathbf{X}_n;\boldsymbol{\theta}_r,\hat{\boldsymbol{\theta}}_{s_0})\\ &+ \hat{\mathbf{F}}_{u,r}(\boldsymbol{\theta}_r^*,\hat{\boldsymbol{\theta}}_{s_0})\mathbf{r},\end{aligned} \tag{S-15}$$

where $\hat{\mathbf{F}}_{u,r}(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s)=\hat{\mathbf{F}}_{u,r}(\boldsymbol{\theta})$ is given in (S-4) and $\boldsymbol{\theta}_r^*$ lies in the line segment connecting $\boldsymbol{\theta}_r=\boldsymbol{\theta}_{r_0}+\frac{\mathbf{r}}{\sqrt{N}}$ and $\boldsymbol{\theta}_{r_0}$. In Lemma 4, stated below, we show that under the local alternatives (19) $\hat{\boldsymbol{\theta}}_{s_0}$ is a consistent estimator of $\boldsymbol{\theta}_{s_0}$, which by Assumption (A-2) lies in the interior of $\boldsymbol{\Theta}_s$. Therefore, we conclude that $\hat{\boldsymbol{\theta}}_{s_0}$ lies in an open neighbourhood $\mathcal{U}_s\subset\boldsymbol{\Theta}_s$ of $\boldsymbol{\theta}_{s_0}$ with sufficiently large probability as N gets large. Therefore, since by Assumption (A-4) $\boldsymbol{\psi}_{u,r}(\mathbf{X}_n;\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}_s)$ is continuous over $\boldsymbol{\Theta}_s$ w.p.1, then by the mean-value theorem [s6, Th. 3.4]

$$\begin{aligned}&\frac{1}{\sqrt{N}}\sum_{n=1}^{N}u(\mathbf{X}_n)\boldsymbol{\psi}_{u,r}(\mathbf{X}_n;\boldsymbol{\theta}_r,\hat{\boldsymbol{\theta}}_{s_0})\\ &= \frac{1}{\sqrt{N}}\sum_{n=1}^{N}u(\mathbf{X}_n)\boldsymbol{\psi}_{u,r}(\mathbf{X}_n;\boldsymbol{\theta}_r,\boldsymbol{\theta}_{s_0})\\ &- \hat{\mathbf{F}}_{u,rs}(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s^*)\sqrt{N}(\hat{\boldsymbol{\theta}}_{s_0}-\boldsymbol{\theta}_{s_0}),\end{aligned} \tag{S-16}$$

where $\boldsymbol{\theta}_{s_0}^*$ lies in the line segment connecting $\hat{\boldsymbol{\theta}}_{s_0}$ and $\boldsymbol{\theta}_{s_0}$. Similarly to Lemma 1, it can be shown that when conditions (A-1)–(A-7) are satisfied, then

$$\begin{aligned}\sqrt{N}(\hat{\boldsymbol{\theta}}_{s_0}-\boldsymbol{\theta}_{s_0}) &= \hat{\mathbf{F}}_{u,s}^{-1}(\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}_{s_0}^{**})\\ &\times \frac{1}{\sqrt{N}}\sum_{n=1}^{N}u(\mathbf{X}_n)\boldsymbol{\psi}_{u,s}(\mathbf{X}_n;\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}_{s_0}),\end{aligned} \tag{S-17}$$

where $\boldsymbol{\theta}_{s_0}^{**}$ lies in the line segment connecting $\hat{\boldsymbol{\theta}}_{s_0}$ and $\boldsymbol{\theta}_{s_0}$, and $\boldsymbol{\psi}_{u,s}(\mathbf{X}_n;\boldsymbol{\theta}_r,\boldsymbol{\theta}_s)=\boldsymbol{\psi}_{u,s}(\mathbf{X};\boldsymbol{\theta})\triangleq\nabla_{\boldsymbol{\theta}_s}\Lambda_u(\mathbf{X};\boldsymbol{\theta})$. By arguments similar to those stated above Eq. (S-15) it can be shown using the mean-value theorem that:

$$\begin{aligned}&\frac{1}{\sqrt{N}}\sum_{n=1}^{N}u(\mathbf{X}_n)\boldsymbol{\psi}_{u,s}(\mathbf{X}_n;\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}_{s_0})\\ &= \frac{1}{\sqrt{N}}\sum_{n=1}^{N}u(\mathbf{X}_n)\boldsymbol{\psi}_{u,s}(\mathbf{X}_n;\boldsymbol{\theta}_r,\boldsymbol{\theta}_{s_0})+\hat{\mathbf{F}}_{u,sr}(\boldsymbol{\theta}_r^{**},\boldsymbol{\theta}_{s_0})\mathbf{r},\end{aligned} \tag{S-18}$$

where $\boldsymbol{\theta}_r^{**}$ lies in the line segment connecting $\boldsymbol{\theta}_r=\boldsymbol{\theta}_{r_0}+\frac{\mathbf{r}}{\sqrt{N}}$ and $\boldsymbol{\theta}_{r_0}$. Hence, by (S-15)-(S-18) it follows that

$$\begin{aligned}&\sqrt{N}\hat{\boldsymbol{\eta}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0})=[\mathbf{I}_{m_r},-\hat{\mathbf{F}}_{u,rs}(\boldsymbol{\theta}_r,\boldsymbol{\theta}_{s_0}^*)\hat{\mathbf{F}}_{u,s}^{-1}(\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}_{s_0}^{**})]\\ &\times \frac{1}{\sqrt{N}}\sum_{n=1}^{N}u(\mathbf{X}_n)\boldsymbol{\psi}_u(\mathbf{X}_n;\boldsymbol{\theta}_r,\boldsymbol{\theta}_{s_0})+\Big(\hat{\mathbf{F}}_{u,r}(\boldsymbol{\theta}_r^*,\hat{\boldsymbol{\theta}}_{s_0})\\ &- \hat{\mathbf{F}}_{u,rs}(\boldsymbol{\theta}_r,\boldsymbol{\theta}_{s_0}^*)\hat{\mathbf{F}}_{u,s}^{-1}(\boldsymbol{\theta}_{r_0},\boldsymbol{\theta}_{s_0}^{**})\hat{\mathbf{F}}_{u,sr}(\boldsymbol{\theta}_r^{**},\boldsymbol{\theta}_{s_0})\Big)\mathbf{r}.\end{aligned} \tag{S-19}$$

Therefore, by (S-2)-(S-4), (S-19), Lemma 4, [s10, Lemma 2], Lemma 7, the definitions of $\boldsymbol{\theta}_{s_0}^*$, $\boldsymbol{\theta}_{s_0}^{**}$, $\boldsymbol{\theta}_{r_0}^*$, $\boldsymbol{\theta}_{r_0}^{**}$ and Slutskey's Theorem [s3], the relation in (S-14) is verified. □

Lemma 3. *Assume that conditions (A-1), (A-3)–(A-6), (A-8) and (A-9) are satisfied. Then, under the local alternatives (19)*

$$\hat{\mathbf{B}}_{u,r}(\tilde{\boldsymbol{\theta}}_0) \xrightarrow[N\to\infty]{P} \mathbf{B}_{u,r}(\boldsymbol{\theta}_0). \tag{S-20}$$

Proof. By Lemma 4, $\tilde{\boldsymbol{\theta}}_0$ is a consistent estimator of $\boldsymbol{\theta}_0$. Thus, by (12) and (S-2) the relation in (S-20) follows directly from Lemma 7. □

Lemma 4. *Assume that conditions (A-1) and (A-3)–(A-6) are satisfied. Then, under the local alternatives (19)*

$$\hat{\boldsymbol{\theta}}_{s_0} \xrightarrow[N\to\infty]{P} \boldsymbol{\theta}_{s_0}. \tag{S-21}$$

Proof. Notice that by [s2, Eq. (23)],

$$\hat{\boldsymbol{\theta}}_{s_0} = \arg\max_{\boldsymbol{\theta}_s\in\boldsymbol{\Theta}_s} J_u(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s), \tag{S-22}$$

where the random objective function

$$\begin{aligned} J_u(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) \triangleq & -D_{\text{LD}}[\hat{\boldsymbol{\Sigma}}_u||\boldsymbol{\Sigma}_u(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s)] \\ & -\|\hat{\boldsymbol{\mu}}_u - \boldsymbol{\mu}_u(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s)\|^2_{\boldsymbol{\Sigma}_u^{-1}(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s)}. \end{aligned} \tag{S-23}$$

The operator $D_{\text{LD}}[\cdot||\cdot]$ denotes the log-determinant divergence [s11], defined below [s2, Eq. (8)]. The random quantities $\hat{\boldsymbol{\mu}}_u$ and $\hat{\boldsymbol{\Sigma}}_u$ are the empirical MT-mean [s2, Eq. (16)] and MT-covariance [s2, Eq. (17)], respectively.

Define the deterministic objective function:

$$\begin{aligned} \bar{J}_u(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) \triangleq & -D_{\text{LD}}[\boldsymbol{\Sigma}_u(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_{s_0})||\boldsymbol{\Sigma}_u(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s)] \\ & -\|\boldsymbol{\mu}_u(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_{s_0}) - \boldsymbol{\mu}_u(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s)\|^2_{\boldsymbol{\Sigma}_u^{-1}(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s)}. \end{aligned} \tag{S-24}$$

Similarly to [s2, Lemma 1], under Assumptions (A-1), (A-2), (A-5) and (A-6), it can be shown that $\bar{J}_u(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s)$ is uniquely maximized at $\boldsymbol{\theta}_s = \boldsymbol{\theta}_{s_0}$. Furthermore, in Lemma 5 below, we show that under the local alternatives (19) $\sup_{\boldsymbol{\theta}_s\in\boldsymbol{\Theta}_s} |J_u(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s) - \bar{J}_u(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)| \xrightarrow[N\to\infty]{P} 0$. Furthermore, by Assumptions (A-4) and (A-5) it follows that $J_u(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)$ is continuous over the compact space $\boldsymbol{\Theta}_s$ w.p. 1. Therefore, the relation in (S-21) follows directly from Theorem [s12, Th. 3.4]. □

Lemma 5. *Assume that conditions (A-1) and (A-4)–(A-6) are satisfied. Then, under the local alternatives (19)*

$$\sup_{\boldsymbol{\theta}_s\in\boldsymbol{\Theta}_s} |J_u(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s) - \bar{J}_u(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)| \xrightarrow[N\to\infty]{P} 0. \tag{S-25}$$

Proof. According to [s2, Eq. (68)]

$$\begin{aligned} & \sup_{\boldsymbol{\theta}_s\in\boldsymbol{\Theta}_s} |J_u(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s) - \bar{J}_u(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)| \\ \leq & \Big(\|\boldsymbol{\Sigma}_u(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_{s_0}) - \hat{\boldsymbol{\Sigma}}_u\|_{\text{Fro}} \\ + & \|\boldsymbol{\mu}_u(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_{s_0})\boldsymbol{\mu}_u^H(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_{s_0}) - \hat{\boldsymbol{\mu}}_u\hat{\boldsymbol{\mu}}_u^H\|_{\text{Fro}}\Big) \\ \times & \sup_{\boldsymbol{\theta}_s\in\boldsymbol{\Theta}_s} \|\boldsymbol{\Sigma}_u^{-1}(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)\|_{\text{Fro}} \\ + & |\log\det[\hat{\boldsymbol{\Sigma}}_u\boldsymbol{\Sigma}_u^{-1}(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_{s_0})]| \\ + & 2\|\boldsymbol{\mu}_u(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_{s_0}) - \hat{\boldsymbol{\mu}}_u\| \sup_{\boldsymbol{\theta}_s\in\boldsymbol{\Theta}_s} \left\|\boldsymbol{\Sigma}_u^{-1}(\boldsymbol{\theta})\boldsymbol{\mu}_u(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)\right\|. \end{aligned} \tag{S-26}$$

Notice that by Assumptions (A-1), (A-4) and (A-5), the supremum terms $\sup_{\boldsymbol{\theta}_s\in\boldsymbol{\Theta}_s}\|\boldsymbol{\Sigma}_u^{-1}(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)\|_{\text{Fro}}$ and $\sup_{\boldsymbol{\theta}_s\in\boldsymbol{\Theta}_s}\left\|\boldsymbol{\Sigma}_u^{-1}(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)\boldsymbol{\mu}_u(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_s)\right\|$ in (S-26) are finite. Also notice that by Lemma 6, under the local alternatives (19), we have $\hat{\boldsymbol{\mu}}_u \xrightarrow[N\to\infty]{P} \boldsymbol{\mu}_u(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_{s_0})$ and $\hat{\boldsymbol{\Sigma}}_u \xrightarrow[N\to\infty]{P} \boldsymbol{\Sigma}_u(\boldsymbol{\theta}_{r_0}, \boldsymbol{\theta}_{s_0})$. Furthermore, note that the operators $\|\cdot\|_{\text{Fro}}$, $\|\cdot\|$ and $\log\det[\cdot]$ define real continuous mappings of $\hat{\boldsymbol{\mu}}_u$ and $\hat{\boldsymbol{\Sigma}}_u$. Hence, by Mann-Wald's Theorem [s4] the upper bound in (S-26) converges to zero in probability as $N\to\infty$. Therefore, the relation in (S-25) must hold. □

Lemma 6. *Assume that conditions (A-6), (A-8) and (A-9) are satisfied. Then, under the local alternatives (19)*

$$\hat{\boldsymbol{\mu}}_u \xrightarrow[N\to\infty]{P} \boldsymbol{\mu}_u(\boldsymbol{\theta}_0) \tag{S-27}$$

and

$$\hat{\boldsymbol{\Sigma}}_u \xrightarrow[N\to\infty]{P} \boldsymbol{\Sigma}_u(\boldsymbol{\theta}_0). \tag{S-28}$$

Proof. We shall prove only the convergence in (S-27). The convergence in (S-28) can be carried out using similar argumentations, and therefore omitted.

By [s2, Eq. (16)] the l-th enrty of $\hat{\boldsymbol{\mu}}_u$ can be written as:

$$[\hat{\boldsymbol{\mu}}_u]_l = \frac{\frac{1}{N}\sum_{k=1}^N v_{1,l}(\mathbf{X}_{N,k})}{\frac{1}{N}\sum_{k=1}^N v_{0,l}(\mathbf{X}_{N,k})}, \tag{S-29}$$

where

$$v_{m,l}(\mathbf{x}) \triangleq u(\mathbf{x})([\mathbf{x}]_l)^m, \quad m = 0, 1. \tag{S-30}$$

and $\{\mathbf{X}_{N,k}\}$ is the row wise i.i.d triangular array defined in (S-13). Hence, if

$$\frac{1}{N}\sum_{k=1}^N v_{m,l}(\mathbf{X}_{N,k}) \xrightarrow[N\to\infty]{P} \text{E}[v_{m,l}(\mathbf{X}); P_{\mathbf{x};\boldsymbol{\theta}_0}] \tag{S-31}$$

for any $m = 0, 1$ and $l = 1, \ldots, p$, then by (4) and (S-29), it follows from Mann-Wald's theorem [s4] that (S-27) holds.

In the following we prove (S-31). By the triangle inequality

$$\left|\frac{1}{N}\sum_{k=1}^N v_{m,l}(\mathbf{X}_{N,k}) - \text{E}[v_{m,l}(\mathbf{X}); P_{\mathbf{x};\boldsymbol{\theta}_0}]\right| \leq A_{m,l} + B_{m,l}, \tag{S-32}$$

where

$$A_{m,l} \triangleq \left|\frac{1}{N}\sum_{k=1}^N v_{m,l}(\mathbf{X}_{N,k}) - \text{E}[v_{m,l}(\mathbf{X}); P_{\mathbf{x};\boldsymbol{\theta}_N}]\right| \tag{S-33}$$

and

$$B_{m,l} \triangleq |\text{E}[v_{m,l}(\mathbf{X}); P_{\mathbf{x};\boldsymbol{\theta}_0}] - \text{E}[v_{m,l}(\mathbf{X}); P_{\mathbf{x};\boldsymbol{\theta}_N}]| \tag{S-34}$$

By Assumption (A-6) and Hölder's inequality [s3], one can verify that $\sup_{N\in\mathbb{N}} \text{E}[v_{m,l}^2(\mathbf{X}); P_{\mathbf{x};\boldsymbol{\theta}_N}] < \infty$ for any $m = 0, 1$ and $l = 1, \ldots, p$. Therefore, by Lemma 8, stated in Section I-D, we conclude that

$$A_{m,l} \xrightarrow[N\to\infty]{P} 0 \;\; \forall m = 0, 1 \text{ and } l = 1, \ldots, p. \tag{S-35}$$

Under Assumption (A-8), it can be shown using the mean-value-theorem that

$$\mathrm{E}[v_{m,l}(\mathbf{X}); P_{\mathbf{x};\boldsymbol{\theta}_N}] = \mathrm{E}[v_{m,l}(\mathbf{X}); P_{\mathbf{x};\boldsymbol{\theta}_0}] + \frac{\mathrm{E}[v_{m,l}(\mathbf{X})\mathbf{r}^T\boldsymbol{\xi}_r(\mathbf{X};\boldsymbol{\theta}'); P_{\mathbf{x};\boldsymbol{\theta}'}]}{\sqrt{N}}, \quad \text{(S-36)}$$

where $\boldsymbol{\theta}' \triangleq [\boldsymbol{\theta}_r^{*T}, \boldsymbol{\theta}_{s_0}^T]^T$, $\boldsymbol{\theta}_r^*$ it a vector that lies on the line segment connecting $\boldsymbol{\theta}_r = \boldsymbol{\theta}_{r_0} + \frac{\mathbf{r}}{\sqrt{N}}$ and $\boldsymbol{\theta}_{r_0}$, and $\boldsymbol{\xi}_r(\mathbf{x};\boldsymbol{\theta}) \triangleq \nabla_{\boldsymbol{\theta}_r} \log f(\mathbf{x};\boldsymbol{\theta})$. Hence, for any $m = 0, 1$ and $l = 1, \ldots, p$

$$\begin{aligned} B_{m,l} &= \frac{1}{\sqrt{N}} |\mathrm{E}[v_{m,l}(\mathbf{X})\mathbf{r}^T\boldsymbol{\xi}_r(\mathbf{X};\boldsymbol{\theta}'); P_{\mathbf{x};\boldsymbol{\theta}'}]| \quad \text{(S-37)} \\ &\leq \frac{1}{\sqrt{N}} \sqrt{\mathrm{E}[\|v_{m,l}(\mathbf{X})\|^2; P_{\mathbf{x};\boldsymbol{\theta}'}]} \|\mathbf{r}\|_{\mathbf{I}_{\mathrm{FIM_r}}(\boldsymbol{\theta}')} \\ &\leq \frac{c}{\sqrt{N}} \|\mathbf{r}\|_{\mathbf{I}_{\mathrm{FIM_r}}(\boldsymbol{\theta}')}, \end{aligned}$$

where $\mathbf{I}_{\mathrm{FIM_r}}(\boldsymbol{\theta})$ is the Fisher information matrix (FIM) [s13] for estimating $\boldsymbol{\theta}_r$ (obtained via intersection of the first m_r rows and columns of the FIM for estimation of $\boldsymbol{\theta} = [\boldsymbol{\theta}_r^T, \boldsymbol{\theta}_s^T]^T$), and c denotes a positive constant. The first inequality in (S-37) follows from Hölder's inequality [s3] and the definition of the FIM. The second inequality is a consequence of Assumption (A-6). Furthermore, since $\boldsymbol{\theta}' \xrightarrow[N\to\infty]{} \boldsymbol{\theta}_0$, then by Assumption (A-9), the weighted Euclidean norm $\|\mathbf{r}\|_{\mathbf{I}_{\mathrm{FIM_r}}(\boldsymbol{\theta}')} \xrightarrow[N\to\infty]{} \|\mathbf{r}\|_{\mathbf{I}_{\mathrm{FIM_r}}(\boldsymbol{\theta}_0)} < \infty$. Hence, we conclude that

$$B_{m,l} \xrightarrow[N\to\infty]{P} 0 \;\; \forall m = 0, 1 \;\text{ and }\; l = 1, \ldots, p. \quad \text{(S-38)}$$

Thus, (S-31) follows directly from (S-32), (S-35) and (S-38). □

Lemma 7. *Let $\bar{\boldsymbol{\theta}}$ be a consistent estimator of $\boldsymbol{\theta}_0$ satisfying $\bar{\boldsymbol{\theta}} \xrightarrow[N\to\infty]{P} \boldsymbol{\theta}_0$. Assume that conditions (A-4)–(A-6), (A-8) and (A-9) are satisfied. Then, under (19)*

$$\hat{\mathbf{F}}_u(\bar{\boldsymbol{\theta}}) \xrightarrow[N\to\infty]{P} \mathbf{F}_u(\boldsymbol{\theta}_0) \quad \text{(S-39)}$$

and

$$\hat{\mathbf{G}}_u(\bar{\boldsymbol{\theta}}) \xrightarrow[N\to\infty]{P} \mathbf{G}_u(\boldsymbol{\theta}_0). \quad \text{(S-40)}$$

Proof. We shall prove the convergence in (S-39). Proof of the convergence in (S-40) follows similar argumentations and therefore omitted.

By the triangle inequality

$$\|\hat{\mathbf{F}}_u(\bar{\boldsymbol{\theta}}) - \mathbf{F}_u(\boldsymbol{\theta}_0)\| \leq \|\hat{\mathbf{F}}_u(\bar{\boldsymbol{\theta}}) - \mathbf{F}_u(\bar{\boldsymbol{\theta}})\| + \|\mathbf{F}_u(\bar{\boldsymbol{\theta}}) - \mathbf{F}_u(\boldsymbol{\theta}_0)\|.$$

Under conditions (A-4) and (A-6) it follows from [s2, Lemma 4] that $\mathbf{F}_u(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\Theta}$. Therefore, since $\bar{\boldsymbol{\theta}} \xrightarrow[N\to\infty]{P} \boldsymbol{\theta}_0$, by Mann-Wald's Theorem we conclude that $\|\mathbf{F}_u(\bar{\boldsymbol{\theta}}) - \mathbf{F}_u(\boldsymbol{\theta}_0)\| \xrightarrow[N\to\infty]{P} 0$. Hence, to complete the proof we need to show that

$$\|\hat{\mathbf{F}}_u(\bar{\boldsymbol{\theta}}) - \mathbf{F}_u(\bar{\boldsymbol{\theta}})\| \xrightarrow[N\to\infty]{P} 0. \quad \text{(S-41)}$$

To do so, we shall prove that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\mathbf{F}}_u(\boldsymbol{\theta}) - \mathbf{F}_u(\boldsymbol{\theta})\| \xrightarrow[N\to\infty]{P} 0. \quad \text{(S-42)}$$

By (15), (24) and the triangle inequality, one can verify that:

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\mathbf{F}}_u(\boldsymbol{\theta}) - \mathbf{F}_u(\boldsymbol{\theta})\| \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{A}(\boldsymbol{\theta})\| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{B}(\boldsymbol{\theta})\|, \quad \text{(S-43)}$$

where

$$\mathbf{A}(\boldsymbol{\theta}) \triangleq \frac{1}{N}\sum_{k=1}^{N} u(\mathbf{X}_{N,k})\boldsymbol{\Gamma}_u(\mathbf{X}_{N,k};\boldsymbol{\theta}) - \mathrm{E}[u(\mathbf{X})\boldsymbol{\Gamma}_u(\mathbf{X};\boldsymbol{\theta}); P_{\mathbf{x};\boldsymbol{\theta}_N}], \quad \text{(S-44)}$$

$$\mathbf{B}(\boldsymbol{\theta}) \triangleq -\mathrm{E}[u(\mathbf{X})\boldsymbol{\Gamma}_u(\mathbf{X};\boldsymbol{\theta}); P_{\mathbf{x};\boldsymbol{\theta}_N}]) - \mathbf{F}_u(\boldsymbol{\theta}), \quad \text{(S-45)}$$

and $\{\mathbf{X}_{N,k}\}$ and $P_{\mathbf{x};\boldsymbol{\theta}_N}$ are the triangular array and the associated probability measure defined in Remark 1.

We begin with convergence analysis of the first summand in the r.h.s. of (S-43). Define

$$g_{l,m}(\mathbf{x};\boldsymbol{\theta}) \triangleq u(\mathbf{x})[\boldsymbol{\Gamma}_u(\mathbf{x};\boldsymbol{\theta})]_{l,m}. \quad \text{(S-46)}$$

Notice that

$$[\mathbf{A}(\boldsymbol{\theta})]_{l,m} = \frac{1}{N}\sum_{k=1}^{N} g_{l,m}(\mathbf{X}_{N,k};\boldsymbol{\theta}) - \mathrm{E}[g_{l,m}(\mathbf{X};\boldsymbol{\theta}); P_{\mathbf{x};\boldsymbol{\theta}_N}]. \quad \text{(S-47)}$$

By Assumption (A-4), $g(\mathbf{x};\boldsymbol{\theta})$ is continuous in $\boldsymbol{\Theta}$ for a.e. $\mathbf{x} \in \mathcal{X}$. Furthermore, by [s10, Eq. (102)] there exists a positive constant b, such that $h(\mathbf{x}) \triangleq b u^2(\mathbf{x}) \sum_{r=0}^{4} \|\mathbf{x}\|^r \geq g^2(\mathbf{x};\boldsymbol{\theta})$. Note that by (S-13) and Assumption (A-6)

$$\sup_{N\in\mathbb{N}} \mathrm{E}[h(\mathbf{X}_{N,k}); P_{\mathbf{x}_{N,k}}] \leq \sum_{r=0}^{4} \sup_{N\in\mathbb{N}} \mathrm{E}[u^2(\mathbf{X})\|\mathbf{X}\|^r; P_{\mathbf{x};\boldsymbol{\theta}_N}] < \infty. \quad \text{(S-48)}$$

Therefore, by (S-44), (S-47) and Lemma 9, stated in Section I-D, we conclude that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{A}(\boldsymbol{\theta})\| \xrightarrow[N\to\infty]{P} 0. \quad \text{(S-49)}$$

Next, we analyze convergence of the second summand in the r.h.s. of (S-43). Notice that by (24) and (S-45)

$$[\mathbf{B}(\boldsymbol{\theta})]_{l,m} = -\mathrm{E}[g_{l,m}(\mathbf{X};\boldsymbol{\theta}); P_{\mathbf{x};\boldsymbol{\theta}_N}] + \mathrm{E}[g_{l,m}(\mathbf{X};\boldsymbol{\theta}); P_{\mathbf{x};\boldsymbol{\theta}_0}]. \quad \text{(S-50)}$$

Under Assumption (A-8), it can be shown using the mean-value-theorem that

$$\mathrm{E}[g_{l,m}(\mathbf{X};\boldsymbol{\theta}); P_{\mathbf{x};\boldsymbol{\theta}_N}] = \mathrm{E}[g_{l,m}(\mathbf{X};\boldsymbol{\theta}); P_{\mathbf{x};\boldsymbol{\theta}_0}] + \frac{\mathrm{E}[g_{l,m}(\mathbf{X};\boldsymbol{\theta})\mathbf{r}^T\boldsymbol{\xi}_r(\mathbf{X};\boldsymbol{\theta}'); P_{\mathbf{x};\boldsymbol{\theta}'}]}{\sqrt{N}}, \quad \text{(S-51)}$$

where $\boldsymbol{\theta}'$ and $\boldsymbol{\xi}_r(\mathbf{x};\boldsymbol{\theta})$ are defined below Eq. (S-36). Hence,

$$\begin{aligned} |[\mathbf{B}(\boldsymbol{\theta})]_{l,m}| &= \frac{1}{\sqrt{N}} |\mathrm{E}[g_{l,m}(\mathbf{X};\boldsymbol{\theta})\mathbf{r}^T\boldsymbol{\xi}_r(\mathbf{X};\boldsymbol{\theta}'); P_{\mathbf{x};\boldsymbol{\theta}'}]| \\ &\leq \frac{1}{\sqrt{N}} \sqrt{\mathrm{E}[g_{l,m}^2(\mathbf{X};\boldsymbol{\theta}); P_{\mathbf{x};\boldsymbol{\theta}'}]} \|\mathbf{r}\|_{\mathbf{I}_{\mathrm{FIM_r}}(\boldsymbol{\theta}')} \\ &\leq \frac{c}{\sqrt{N}} \|\mathbf{r}\|_{\mathbf{I}_{\mathrm{FIM_r}}(\boldsymbol{\theta}')}, \quad \text{(S-52)} \end{aligned}$$

where $\mathbf{I}_{\mathrm{FIM_r}}(\boldsymbol{\theta})$ is the FIM defined below (S-37), and c denotes a positive constant. The first inequality in (S-52) follows from Hölder's inequality [s3]. The second inequality is a consequence of Assumptions (A-6), [s10, Eq. (102)], and the definition of the FIM [s13]. Furthermore, since $\boldsymbol{\theta}' \xrightarrow[N\to\infty]{} \boldsymbol{\theta}_0$, the by Assumption (A-9), the weighted Euclidean norm $\|\mathbf{r}\|_{\mathbf{I}_{\mathrm{FIM_r}}(\boldsymbol{\theta}')} \xrightarrow[N\to\infty]{} \|\mathbf{r}\|_{\mathbf{I}_{\mathrm{FIM_r}}(\boldsymbol{\theta}_0)} < \infty$. Hence,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{B}(\boldsymbol{\theta})\| \xrightarrow[N\to\infty]{} 0. \tag{S-53}$$

Therefore, by (S-42), (S-43), (S-49) and (S-53) the relation in (S-41) must hold. □

C. Proof of Theorem 3

Similarly to the proof of Theorem 2, it can be shown that when conditions (A-1)−(A-7) are satisfied, then under the contaminated probability measure $P_{\epsilon,N}$, defined in (26), the asymptotic distribution of the test-statistic is non-central chi-squared, i.e.,

$$T_u \xrightarrow[N\to\infty]{D} \chi^2_{m_r}\left(\lambda_u(\epsilon)\right), \tag{S-54}$$

where the non-centrality parameter

$$\lambda_u(\epsilon) = \epsilon^2 \left\|u(\mathbf{y})\boldsymbol{\psi}_{u,r}(\mathbf{y};\boldsymbol{\theta}_0)\right\|^2_{\mathbf{R}^{-1}_{u,r}(\boldsymbol{\theta}_0)}. \tag{S-55}$$

Let α and α_ϵ denote the asymptotic test-sizes under the uncontaminated and contaminated probability distributions $P_{\mathbf{x};\boldsymbol{\theta}_0}$ and $P_{\epsilon,N}$, respectively, for a fixed threshold t. By (18), it follows that $t = Q^{-1}_{\chi^2_{m_r}}(\alpha)$, where $Q_{\chi^2_{m_r}}(\cdot)$ denotes the tail probability of a central chi-squared distribution. Therefore, by (S-54) we conclude that

$$\alpha_\epsilon = H_{m_r}(\lambda_u(\epsilon),\alpha) \triangleq Q_{\chi^2_{m_r}(\lambda_u(\epsilon))}\left(Q^{-1}_{\chi^2_{m_r}}(\alpha)\right), \tag{S-56}$$

where $Q_{\chi^2_{m_r}(\cdot)}(\cdot)$ denotes the tail probability of a non-central chi-squared distribution. Second-order Taylor series expansion of $H_{m_r}(\lambda_u(\epsilon),\alpha)$ about $\epsilon = 0$ yields:

$$\begin{aligned} H_{m_r}(\lambda_u(\epsilon),\alpha) &= H_{m_r}(\lambda_u(0),\alpha) + \left.\frac{dH_{m_r}(\lambda_u(\epsilon),\alpha)}{d\epsilon}\right|_{\epsilon=0} \\ &+ \left.\frac{d^2H_{m_r}(\lambda_u(\epsilon),\alpha)}{d\epsilon^2}\right|_{\epsilon=0}\epsilon^2 + o(\epsilon^2). \end{aligned} \tag{S-57}$$

By (S-55), we have

$$\lambda_u(0) = 0, \tag{S-58}$$

$$\left.\frac{d\lambda_u(\epsilon)}{d\epsilon}\right|_{\epsilon=0} = 0, \tag{S-59}$$

and

$$\left.\frac{d^2\lambda_u(\epsilon)}{d\epsilon^2}\right|_{\epsilon=0} = 2\left\|u(\mathbf{y})\boldsymbol{\psi}_{u,r}(\mathbf{y};\boldsymbol{\theta}_0)\right\|^2_{\mathbf{R}^{-1}_{u,r}(\boldsymbol{\theta}_0)} \tag{S-60}$$

Therefore, by (S-56) and (S-58)-(S-60) we obtain the following equalities:

$$H_{m_r}(\lambda_u(0),\alpha) = \alpha, \tag{S-61}$$

$$\left.\frac{dH_{m_r}(\lambda_u(\epsilon),\alpha)}{d\epsilon}\right|_{\epsilon=0} = \left.\frac{dH_{m_r}(\lambda,\alpha)}{d\lambda}\right|_{\lambda=0}\left.\frac{d\lambda_u(\epsilon)}{d\epsilon}\right|_{\epsilon=0} = 0, \tag{S-62}$$

and

$$\begin{aligned} \left.\frac{d^2H_{m_r}(\lambda_u(\epsilon),\alpha)}{d\epsilon^2}\right|_{\epsilon=0} &= \left.\frac{d^2H_{m_r}(\lambda,\alpha)}{d\lambda^2}\right|_{\lambda=0}\left(\left.\frac{d\lambda_u(\epsilon)}{d\epsilon}\right|_{\epsilon=0}\right)^2 \\ &+ \left.\frac{dH_{m_r}(\lambda,\alpha)}{d\lambda}\right|_{\lambda=0}\left.\frac{d^2\lambda_u(\epsilon)}{d\epsilon^2}\right|_{\epsilon=0} \\ &= c\left\|u(\mathbf{y})\boldsymbol{\psi}_{u,r}(\mathbf{y};\boldsymbol{\theta}_0)\right\|^2_{\mathbf{R}^{-1}_{u,r}(\boldsymbol{\theta}_0)}, \end{aligned} \tag{S-63}$$

where the constant $c \triangleq 2\left.\frac{dH_{m_r}(\lambda,\alpha)}{d\lambda}\right|_{\lambda=0}$. The relation in (27) follows now directly from (S-56), (S-57) and (S-61)-(S-63).

D. Auxiliary lemmas:

Lemma 8 (Law of large numbers for triangular arrays)**.** *Let $\{Y_{N,k}\}$, be a row wise i.i.d. triangular array of random variables. Assume that*

$$\sup_{N\in\mathbb{N}} \mathrm{E}[|Y_{N,k}|^2; P_{Y_{N,k}}] < \infty \tag{S-64}$$

or

$$\frac{\mathrm{E}[|Y_{N,k}|^2; P_{Y_{N,k}}]}{N} \xrightarrow[N\to\infty]{} 0. \tag{S-65}$$

Then,

$$\frac{1}{N}\sum_{k=1}^{N}(Y_{N,k} - \mathrm{E}[Y_{N,k}; P_{Y_{N,k}}]) \xrightarrow[N\to\infty]{P} 0. \tag{S-66}$$

Proof. Define the zero-mean random variable

$$Z \triangleq \frac{1}{N}\sum_{k=1}^{N}(Y_{N,k} - \mathrm{E}[Y_{N,k}; P_{Y_{N,k}}]).$$

By Chebychev's inequality [s3], $\Pr[|Z| > \epsilon] \leq \frac{\mathrm{E}[|Z|^2;P_Z]}{\epsilon^2}$. Since $\{Y_{N,k}\}$, be a row wise i.i.d. triangular array, it follows that $\mathrm{E}[|Z|^2; P_Z] = \mathrm{var}[Y_{N,k}; P_{Y_{N,k}}]/N$. Hence, we conclude that

$$\begin{aligned} &\Pr\left[\left|\frac{1}{N}\sum_{k=1}^{N}(Y_{N,k} - \mathrm{E}[Y_{N,k}; P_{Y_{N,k}}])\right| > \epsilon\right] \\ &\leq \frac{\mathrm{var}[Y_{N,k}; P_{Y_{N,k}}]}{N\epsilon^2} \leq \frac{\mathrm{E}[|Y_{N,k}|^2; P_{Y_{N,k}}]}{N\epsilon^2}. \end{aligned} \tag{S-67}$$

Hence, the relation (S-66) holds if either condition (S-64) or (S-65) are satisfied. □

Lemma 9 (Uniform law of large numbers for triangular arrays)**.** *Let $g(\mathbf{x};\boldsymbol{\theta})$ be a real scalar function that is continuous over $\boldsymbol{\Theta}$ for a.e. $\mathbf{x} \in \mathcal{X}$. Assume that there exists a real scalar function $h(\mathbf{x})$, such that $g^2(\mathbf{x};\boldsymbol{\theta}) \leq h(\mathbf{x})$ and $\sup_{N\in\mathbb{N}} \mathrm{E}[h(\mathbf{X}_{N,k}); P_{\mathbf{x}_{N,k}}] < \infty$, where $\{\mathbf{X}_{N,k}\}$ is a row wise i.i.d. triangular array of random vectors, then*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{N}\sum_{k=1}^{N}(g(\mathbf{X}_{N,k};\boldsymbol{\theta}) - \mathrm{E}[g(\mathbf{X}_{N,k};\boldsymbol{\theta}); P_{\mathbf{x}_{N,k}}])\right| \xrightarrow[N\to\infty]{P} 0. \tag{S-68}$$

The proof is similar to the one of the standard uniform law of large numbers [s14] and therefore omitted.

II. Quasi GLRT and quasi score test under the assumption of generalized Gaussian distribution

In this section, we provide implementation details of the GGD based quasi GLRT and the quasi score test, called here GGD-QGLRT and GGD-QST, respectively, for the detection problem considered in Section IV. Both detectors are based on the assumption that under the linear model (31), the observations obey a GGD having location parameter $\mathbf{A}_r\boldsymbol{\vartheta}_r + \mathbf{A}_s\boldsymbol{\vartheta}_s$ and a scaled-identity scatter matrix, with probability density function [s1]:

$$\tilde{f}(\mathbf{x};\boldsymbol{\xi}) = \frac{s\Gamma(p)\exp\left(-\|\mathbf{x}-\mathbf{A}_r\boldsymbol{\vartheta}_r-\mathbf{A}_s\boldsymbol{\vartheta}_s\|^{2h}/c^h\right)}{\pi^p\Gamma(p/h)c^p}, \tag{S-69}$$

where $\boldsymbol{\xi} \triangleq [\boldsymbol{\theta}_r^T, \boldsymbol{\theta}_s^T, s, c]^T$, $\boldsymbol{\theta}_r = [\Re(\boldsymbol{\vartheta}_r)^T, \Im(\boldsymbol{\vartheta}_r)^T]^T$, $\boldsymbol{\theta}_s = [\Re(\boldsymbol{\vartheta}_s)^T, \Im(\boldsymbol{\vartheta}_s)^T]^T$, h is the shape parameter of the GGD, c is a scale parameter, and $\Gamma(\cdot)$ denotes the Gamma function. The function $\tilde{f}(\mathbf{x};\boldsymbol{\xi})$ will be called the quasi-likelihood function. Note that in addition to $\boldsymbol{\theta}_s$, the parameters h and c are also nuisance.

A. Quasi GLRT

Given a sequence of i.i.d. samples $\mathbf{X}_1,\ldots,\mathbf{X}_N$ from $P_{\mathbf{x};\boldsymbol{\theta}}$, the test-statistic of the GGD-QGLRT is given by:

$$T_{\text{GGD-QGLRT}} = J(\hat{\boldsymbol{\xi}}_1) - J(\tilde{\boldsymbol{\xi}}_0), \tag{S-70}$$

where $J(\boldsymbol{\xi}) \triangleq \sum_{n=1}^N \log\tilde{f}(\mathbf{X}_n;\boldsymbol{\xi})$ is the assumed (quasi) joint log-likelihood, $\hat{\boldsymbol{\xi}}_1 \triangleq [\hat{\boldsymbol{\theta}}_{r_1}^T, \hat{\boldsymbol{\theta}}_{s_1}^T, \hat{h}_1, \hat{c}_1]^T$ is the vector of quasi maximum-likelihood (QML) estimators of $(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s, h, c)$ under H_1 and $\tilde{\boldsymbol{\xi}}_0 \triangleq [\boldsymbol{\theta}_{r_0}^T, \hat{\boldsymbol{\theta}}_{s_0}^T, \hat{h}_0, \hat{c}_0]^T$ is the vector of QML estimators of $(\boldsymbol{\theta}_s, h, c)$ under H_0.

1) Derivation of $\hat{\boldsymbol{\theta}}_{r_1}$, $\hat{\boldsymbol{\theta}}_{s_1}$, $\hat{h}_1$ and $\hat{c}_1$: By equating the partial gradients of $J(\boldsymbol{\xi})$ to zero, one can verify that $\hat{\boldsymbol{\theta}}_{r_1}$, $\hat{\boldsymbol{\theta}}_{s_1}$, $\hat{h}_1$ and $\hat{c}_1$ are the solutions of the equations:

$$\begin{aligned}\boldsymbol{\theta}_r &= (\tilde{\mathbf{A}}_r^T\tilde{\mathbf{A}}_r)^{-1}\tilde{\mathbf{A}}_r^T \\ &\times \frac{\sum_{n=1}^N \|\tilde{\mathbf{Y}}_n(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s)\|^{2(h-1)}(\tilde{\mathbf{X}}_n - \tilde{\mathbf{A}}_s\boldsymbol{\theta}_s)}{\sum_{n=1}^N \|\tilde{\mathbf{Y}}_n(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s)\|^{2(h-1)}},\end{aligned} \tag{S-71}$$

$$\begin{aligned}\boldsymbol{\theta}_s &= (\tilde{\mathbf{A}}_s^T\tilde{\mathbf{A}}_s)^{-1}\tilde{\mathbf{A}}_s^T \\ &\times \frac{\sum_{n=1}^N \|\tilde{\mathbf{Y}}_n(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s)\|^{2(h-1)}(\tilde{\mathbf{X}}_n - \tilde{\mathbf{A}}_r\boldsymbol{\theta}_r)}{\sum_{n=1}^N \|\tilde{\mathbf{Y}}_n(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s)\|^{2(h-1)}},\end{aligned} \tag{S-72}$$

$$c = \left(\frac{h}{pN}\sum_{n=1}^N \|\tilde{\mathbf{Y}}_n(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s)\|^{2h}\right)^{1/h}, \tag{S-73}$$

$$h = \frac{p\psi_0(p/h)}{\frac{h}{N}\sum_{n=1}^N \frac{\|\tilde{\mathbf{Y}}_n(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s)\|^{2h}}{c^h}\log\frac{\|\tilde{\mathbf{Y}}_n(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s)\|^2}{c}}, \tag{S-74}$$

where the real random vector $\tilde{\mathbf{X}} \triangleq [\text{Re}^T\{\mathbf{X}\}, \text{Im}^T\{\mathbf{X}\}]^T$, $\tilde{\mathbf{Y}}(\boldsymbol{\theta}_r,\boldsymbol{\theta}_s) \triangleq \tilde{\mathbf{X}} - \tilde{\mathbf{A}}_r\boldsymbol{\theta}_r - \tilde{\mathbf{A}}_s\boldsymbol{\theta}_s$, and the real regression matrices $\tilde{\mathbf{A}}_r$ and $\tilde{\mathbf{A}}_s$ are defined as:

$$\tilde{\mathbf{A}}_r \triangleq \begin{bmatrix} \Re(\mathbf{A}_r) & -\Im(\mathbf{A}_r) \\ \Im(\mathbf{A}_r) & \Re(\mathbf{A}_r)\end{bmatrix}, \; \tilde{\mathbf{A}}_s \triangleq \begin{bmatrix} \Re(\mathbf{A}_s) & -\Im(\mathbf{A}_s) \\ \Im(\mathbf{A}_s) & \Re(\mathbf{A}_s)\end{bmatrix}.$$

$\psi_0(\cdot)$ denotes the zero-order polygamma function. The solution of (S-71)-(S-74) was obtained numerically via fixed-point iteration. The maximum number of iterations and the stopping criterion were set to 100 and $\frac{|\varphi_1(l)-\varphi_1(l-1)|}{\varphi_1(l-1)} < 10^{-6}$, where $\varphi_1(l) \triangleq J(\hat{\boldsymbol{\xi}}_1^{(l)})$ and $\hat{\boldsymbol{\xi}}_1^{(l)}$ is the vector comprised of $\hat{\boldsymbol{\theta}}_{r_1}^{(l)}$, $\hat{\boldsymbol{\theta}}_{s_1}^{(l)}$, $\hat{h}_1^{(l)}$ and $\hat{c}_1^{(l)}$ denoting the estimates at iteration index $l \geq 1$. The initial conditions $\hat{\boldsymbol{\theta}}_{r_1}^{(0)}$, $\hat{\boldsymbol{\theta}}_{s_1}^{(0)}$, $\hat{h}_1^{(0)}$,and $\hat{c}_1^{(0)}$ were set ot $\mathbf{0}$, $\mathbf{0}$, 1/2 and 1, respectively.

2) Derivation of $\hat{\boldsymbol{\theta}}_{s_0}$, $\hat{h}_0$ and $\hat{c}_0$: Similarly, by setting $\boldsymbol{\theta}_r = \boldsymbol{\theta}_{r_0}$ and equating the partial gradients of $J(\boldsymbol{\xi})$ w.r.t. $\boldsymbol{\theta}_s$, h and c to zero, one can verify that $\hat{\boldsymbol{\theta}}_{s_0}$, $\hat{h}_0$ and $\hat{c}_0$ are the solutions of (S-72)-(S-74). Also here, the solution was obtained numerically via fixed-point iteration. The maximum number of iterations and the stopping criterion were set to 100 and $\frac{|\varphi_0(l)-\varphi_0(l-1)|}{\varphi_0(l-1)} < 10^{-6}$, where $\varphi_0(l) \triangleq J(\tilde{\boldsymbol{\xi}}_0^{(l)})$ and $\tilde{\boldsymbol{\xi}}_0^{(l)}$ is the vector comprised of $\boldsymbol{\theta}_{r_0}, \hat{\boldsymbol{\theta}}_{s_0}^{(l)}, \hat{h}_0^{(l)}, \hat{c}_0^{(l)}$ with the latter three denoting the estimates at iteration index $l \geq 1$. The initial conditions $\hat{\boldsymbol{\theta}}_{s_0}^{(0)}$, $\hat{h}_0^{(0)}$,and $\hat{c}_0^{(0)}$ were set to $\mathbf{0}$, 1/2 and 1, respectively.

B. Quasi score test

Implementation of the GGD-QST, based on a sequence of i.i.d. samples $\mathbf{X}_1,\ldots,\mathbf{X}_N$ from $P_{\mathbf{x};\boldsymbol{\theta}}$, is comprised of two stages. First, the QML estimators $\hat{\boldsymbol{\theta}}_{s_0}$, $\hat{h}_0$ and $\hat{c}_0$ are obtained exactly as described in Subsection II-A2 to construct the vector $\tilde{\boldsymbol{\xi}}_0$ defined below (S-70). Second, the test statistic is constructed according to:

$$T_{\text{GGD-QST}} = N\hat{\boldsymbol{\eta}}_r^T(\tilde{\boldsymbol{\xi}}_0)\hat{\mathbf{B}}_r^{-1}(\tilde{\boldsymbol{\xi}}_0)\hat{\boldsymbol{\eta}}_r(\tilde{\boldsymbol{\xi}}_0), \tag{S-75}$$

where

$$\hat{\boldsymbol{\eta}}_r(\boldsymbol{\xi}) \triangleq \frac{1}{N}\sum_{n=1}^N \boldsymbol{\psi}_r(\mathbf{X}_n;\boldsymbol{\xi}), \tag{S-76}$$

and $\boldsymbol{\psi}_r(\mathbf{x};\boldsymbol{\xi}) \triangleq \nabla_{\boldsymbol{\theta}_r}\log\tilde{f}(\mathbf{x};\boldsymbol{\xi})$. The matrix $\hat{\mathbf{B}}_r(\boldsymbol{\xi})$ is defined as:

$$\hat{\mathbf{B}}_r(\boldsymbol{\xi}) \triangleq \hat{\mathbf{H}}_r(\boldsymbol{\xi})\hat{\mathbf{R}}_r(\boldsymbol{\xi})\hat{\mathbf{H}}_r(\boldsymbol{\xi}), \tag{S-77}$$

where $\hat{\mathbf{R}}_r(\boldsymbol{\xi}) \in \mathbb{R}^{m_r\times m_r}$ is formed by the intersection of the first m_r rows and columns of the matrix:

$$\hat{\mathbf{R}}(\boldsymbol{\xi}) \triangleq \hat{\mathbf{F}}^{-1}(\boldsymbol{\xi})\hat{\mathbf{G}}(\boldsymbol{\xi})\hat{\mathbf{F}}^{-1}(\boldsymbol{\xi}), \tag{S-78}$$

with

$$\hat{\mathbf{G}}(\boldsymbol{\xi}) \triangleq \frac{1}{N}\sum_{n=1}^N \boldsymbol{\psi}\left(\mathbf{X}_n;\boldsymbol{\xi}\right)\boldsymbol{\psi}^T\left(\mathbf{X}_n;\boldsymbol{\xi}\right), \tag{S-79}$$

$$\hat{\mathbf{F}}\left(\boldsymbol{\xi}\right) \triangleq -\frac{1}{N}\sum_{n=1}^N \boldsymbol{\Gamma}\left(\mathbf{X}_n;\boldsymbol{\xi}\right), \tag{S-80}$$

$\boldsymbol{\psi}(\mathbf{X};\boldsymbol{\xi}) \triangleq \nabla_{\boldsymbol{\xi}}\log\tilde{f}(\mathbf{X};\boldsymbol{\xi})$ and $\boldsymbol{\Gamma}(\mathbf{X};\boldsymbol{\xi}) \triangleq \nabla^2_{\boldsymbol{\xi}}\log\tilde{f}(\mathbf{X};\boldsymbol{\xi})$. The matrix $\hat{\mathbf{H}}_{u,r}(\boldsymbol{\xi})$ in (S-77) is defined as:

$$\hat{\mathbf{H}}_r(\boldsymbol{\theta}) \triangleq \hat{\mathbf{F}}_r\left(\boldsymbol{\xi}\right) - \hat{\mathbf{F}}_{rs}\left(\boldsymbol{\xi}\right)\hat{\mathbf{F}}_s^{-1}\left(\boldsymbol{\xi}\right)\hat{\mathbf{F}}_{rs}^T\left(\boldsymbol{\xi}\right), \tag{S-81}$$

where $\hat{\mathbf{F}}_r\left(\boldsymbol{\theta}\right) \in \mathbb{R}^{m_r\times m_r}$, $\hat{\mathbf{F}}_{rs}\left(\boldsymbol{\xi}\right) \in \mathbb{R}^{m_r\times(m_s+2)}$ and $\hat{\mathbf{F}}_{u,s}\left(\boldsymbol{\xi}\right) \in \mathbb{R}^{(m_s+2)\times(m_s+2)}$ are obtained from the partition:

$$\hat{\mathbf{F}}_u\left(\boldsymbol{\xi}\right) = \begin{bmatrix} \hat{\mathbf{F}}_r\left(\boldsymbol{\xi}\right) & \hat{\mathbf{F}}_{rs}\left(\boldsymbol{\xi}\right) \\ \hat{\mathbf{F}}_{rs}^T\left(\boldsymbol{\xi}\right) & \hat{\mathbf{F}}_s\left(\boldsymbol{\xi}\right)\end{bmatrix}. \tag{S-82}$$

References

[s1] E. Ollila, D. E. Tyler, V. Koivunen and H. V. Poor, "Complex elliptically symmetric distributions: survey, new results and applications," *IEEE Transactions on Signal Processing,* vol. 60, no. 1, pp. 5597-5625, 2012.

[s2] K. Todros and A. O. Hero, "Measure-transformed quasi maximum likelihood estimation," *IEEE Transactions on Signal Processing,* vol. 60, no. 9, pp. 4570-4585, 2016.

[s3] K. B. Athreya and S. N. Lahiri, *Measure theory and probability theory,* Springer-Verlag, 2006.

[s4] H. B. Mann and A. Wald, "On stochastic limit and order relationships," *Ann. Math. Stat.,* vol. 14, pp. 217-226, 1943.

[s5] S. M. Kay, *Fundamentals of statistical signal processing: detection theory,* Prentice-Hall, 1993.

[s6] C. H. Edwards, *Advanced calculus of several variables,* Courier Corporation, 2012.

[s7] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis,* third edition, John Wiley & Sons, 2003.

[s8] A. R. Horn and C. R. Johnson, *Matrix Analysis,* Cambridge University Press, 2012.

[s9] R. J. Serfling, *Approximation theorems of mathematical statistics.* John Wiley & Sons, 1980.

[s10] K. Todros, Robust composite binary hypothesis testing via measure transformed quasi score test," *Signal Processing*, vol. 155, pp. 202-217, Feb. 2019.

[s11] I. S. Dhillon and J. A. Tropp, "Matrix nearness problems with Bregman divergences," *SIAM Journal on Matrix Analysis and Applications,* vol. 29, no. 4, pp. 1120-1146, 2007.

[s12] H. White, *Estimation, inference and specification analysis,* Cambridge university press, 1996.

[s13] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory,* Prentice-Hall, 1993.

[s14] W. K. Newey and D. McFadden, "Large sample estimation and hypothesis testing," *Handbook of econometrics,* vol. 4, pp. 2111-2245, 1994.