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

Koby Todros
Ben-Gurion University of the Negev

***Abstract*—In this paper, we extend the measure-transformed Gaussian quasi score test (MT-GQST) for the case where nuisance parameters are present. The proposed extension is based on a zero-expectation property of a partial Gaussian quasi score function under the transformed null distribution. The nuisance parameters are estimated under the null hypothesis via the measure-transformed Gaussian quasi MLE. In the paper, we analyze the effect of the probability measure-transformation on the asymptotic detection performance of the extended MT-GQST. This leads to a data-driven procedure for selection of the generating function of the considered transform, called MT-function, which, in practice, weights the data points. Furthermore, we provide conditions on the MT-function to ensure stability of the asymptotic false-alarm-rate in the presence of noisy outliers. The extended MT-GQST is applied for testing a vector parameter of interest comprising a noisy multivariate linear data model in the presence of nuisance parameters. Simulation study illustrates its advantages over other robust detectors.**

I. Introduction

The Gaussian quasi score test (GQST) [1]-[6] is a popular alternative to Rao's score test [7]-[9] when the likelihood function is unknown. Similarly to the score test, it has the advantage that is does not involve parameter estimation under the alternative hypothesis. The GQST operates under a hypothesized Gaussian probability model, and thus, utilizes only the first and second-order statistical moments, leading to implementation simplicity and tractable performance analysis. However, when the underlying probability distribution of the data largely deviates from the assumed Gaussian probability model, the GQST may perform poorly, especially in the presence of heavy-tailed noise that produces outliers.

Recently, we developed in [10], [11] a robust generalization of the GQST, called measure-transformed GQST (MT-GQST), that assumes a Gaussian probability model after applying a transform to the probability measure (distribution) of the data. The considered measure-transformation is structured by a non-negative scalar function, called the MT-function, that weights the observation space. By proper selection of the MT-function, we have shown that the MT-GQST can gain resilience against outliers along with attractive implementation simplicity that arises from the convenient Gaussian model. The MT-GQST was developed for a parametric composite binary hypothesis testing problem that *does not* involve nuisance parameters.

In this paper, we extend the MT-GQST for the case where nuisance parameters are present. The proposed extension is based on the property that the partial measure-transformed Gaussian quasi score function (MT-GQSF), evaluated at the null vector parameter, has a zero expectation under the transformed null probability distribution. The partial MT-GQSF is defined as the partial gradient, w.r.t. the vector parameter of interest, of a Gaussian log-likelihood function that is characterized by the parametric measure-transformed mean vector and covariance matrix. Unlike the MT-GQST [10], [11], the extended MT-GQST involves an empirical estimate of the nuisance vector parameter under the null hypothesis. This empirical estimate is obtained using the measure-transformed Gaussian quasi MLE (MT-GQMLE) [12].

Under some regularity assumptions, we show that the extended version of the MT-GQST is an asymptotically constant-false-alarm-rate (CFAR) detector, w.r.t. the MT-function and the null vector parameter. Similarly to [10], [11], the asymptotic distribution of the test-statistic is shown to be central chi-squared under the null hypothesis and non-central chi-squared under a sequence of contiguous local alternatives. To analyze robustness against outliers, we obtain an expansion of the asymptotic test-size (false-alarm-rate) under a contaminated distribution in the neighbourhood of the null hypothesis. Using this expansion, sufficient conditions on the MT-function to ensure stability of the asymptotic test-size are derived. Similarly to [11], we show that the asymptotic local power is monotonically increasing with the inverse error-covariance matrix of the MT-GQMLE for estimating the vector parameter of interest under the null hypothesis. This property leads to a data driven procedure for selection of the MT-function that minimizes the spectral norm of an empirical estimate of this error-covariance.

The extended MT-GQST is applied for testing a vector parameter of interest of a noisy multivariate linear data model in the presence of nuisance parameters. The MT-function is selected via the procedure discussed above out of a class of zero-centered pseudo-Gaussian functions parameterized by a width parameter. We show that the extended MT-GQST is easy to implement and outperforms other robust detectors in the presence of heavy-tailed noise.

II. Probability measure transform: review

In this section, we briefly review the parametric probability measure transformation [12]. Based on this transformation, we redefine the measure-transformed mean vector and covariance matrix. These, will be used in the following section to derive the proposed extension of the MT-GQST [10], [11].

This work was supported in part by the Israel Science Foundation under Grant 1568/17.

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

A. Probability measure transform

We define the measure space $(\mathcal{X}, \mathcal{S}_{\mathcal{X}}, P_{\mathbf{X};\boldsymbol{\theta}})$, where $\mathcal{X} \subseteq \mathbb{C}^p$ is the observation space of a random vector $\mathbf{X}$, $\mathcal{S}_{\mathcal{X}}$ is a σ-algebra over $\mathcal{X}$ and $P_{\mathbf{X};\boldsymbol{\theta}}$ is a probability measure on $\mathcal{S}_{\mathcal{X}}$ parameterized by a vector parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^m$.

Definition 1 (Definition of the transform)**.** *Given a non-negative function $u : \mathbb{C}^p \to \mathbb{R}_+$ satisfying*

$$0 < \mathrm{E}\left[u\left(\mathbf{X}\right); P_{\mathbf{X};\boldsymbol{\theta}}\right] < \infty, \tag{1}$$

where $\mathrm{E}\left[u\left(\mathbf{X}\right); P_{\mathbf{X};\boldsymbol{\theta}}\right] \triangleq \int_{\mathcal{X}} u\left(\mathbf{x}\right) dP_{\mathbf{X};\boldsymbol{\theta}}\left(\mathbf{x}\right)$ and $\mathbf{x} \in \mathcal{X}$, a transform on $P_{\mathbf{X};\boldsymbol{\theta}}$ is defined via the relation:

$$Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}\left(A\right) \triangleq \mathrm{T}_u\left[P_{\mathbf{X};\boldsymbol{\theta}}\right]\left(A\right) = \int_A \varphi_u\left(\mathbf{x};\boldsymbol{\theta}\right) dP_{\mathbf{X};\boldsymbol{\theta}}\left(\mathbf{x}\right), \tag{2}$$

where $A \in \mathcal{S}_{\mathcal{X}}$ and $\varphi_u\left(\mathbf{x};\boldsymbol{\theta}\right) \triangleq u\left(\mathbf{x}\right)/\mathrm{E}\left[u\left(\mathbf{X}\right); P_{\mathbf{X};\boldsymbol{\theta}}\right]$. The function $u\left(\cdot\right)$ is called the MT-function.

By Definition 1, $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ (2) is a probability measure on $\mathcal{S}_{\mathcal{X}}$ that is absolutely continuous w.r.t. $P_{\mathbf{X};\boldsymbol{\theta}}$, with Radon-Nikodym derivative [14]:

$$dQ^{(u)}_{\mathbf{X};\boldsymbol{\theta}}\left(\mathbf{x}\right)/dP_{\mathbf{X};\boldsymbol{\theta}}\left(\mathbf{x}\right) = \varphi_u\left(\mathbf{x};\boldsymbol{\theta}\right). \tag{3}$$

B. The measure-transformed mean and covariance

By relation (3), it follows that the mean vector and the covariance matrix of $\mathbf{X}$ under $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$, called the MT-mean and MT-covariance, respectively, can be expressed as weighted mean and covariance under $P_{\mathbf{X};\boldsymbol{\theta}}$, i.e.,

$$\boldsymbol{\mu}_u(\boldsymbol{\theta}) \triangleq \mathrm{E}[\mathbf{X}; Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}] = \mathrm{E}[\mathbf{X}\varphi_u\left(\mathbf{X};\boldsymbol{\theta}\right); P_{\mathbf{X};\boldsymbol{\theta}}] \tag{4}$$

and

$$\begin{aligned} \boldsymbol{\Sigma}_u(\boldsymbol{\theta}) &\triangleq \mathrm{cov}[\mathbf{X}; Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}] \\ &= \mathrm{E}[\mathbf{X}\mathbf{X}^H\varphi_u\left(\mathbf{X};\boldsymbol{\theta}\right); P_{\mathbf{X};\boldsymbol{\theta}}] - \boldsymbol{\mu}_u(\boldsymbol{\theta})\boldsymbol{\mu}_u^H(\boldsymbol{\theta}). \end{aligned} \tag{5}$$

The weighting function $\varphi_u\left(\mathbf{x};\boldsymbol{\theta}\right)$ is defined below (1). Notice that when $u\left(\mathbf{x}\right)$ is non-zero and constant valued $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}} = P_{\mathbf{X};\boldsymbol{\theta}}$ and then (4) and (5) coincide with the standard mean vector and covariance matrix, respectively. Unlike the standard mean and covariance, which do not exist for certain types of heavy-tailed distributions, such as Cauchy's, the MT-mean vector (4) and the MT-covariance matrix (5) universally take finite values when $u(\mathbf{x})$ and $u(\mathbf{x})\|\mathbf{x}\|^2$ are bounded over $\mathbb{C}^p$.

III. Measure-transformed Gaussian quasi score test: the case of nuisance parameters

In this section, we extend the MT-GQST [10], [11] to the case where nuisance parameters are present.

A. Problem statement

We partition the vector parameter $\boldsymbol{\theta}$ as $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_r^T, \boldsymbol{\theta}_s^T\right]^T$, where $\boldsymbol{\theta}_r \in \boldsymbol{\Theta}_r \subseteq \mathbb{R}^{m_r}$ is a vector parameter of interest and $\boldsymbol{\theta}_s \in \boldsymbol{\Theta}_s \subseteq \mathbb{R}^{m_s}$ is a nuisance vector parameter, such that $m_r + m_s = m$. Given a sequence of samples $\mathbf{X}_1, \ldots, \mathbf{X}_N$ from $P_{\mathbf{X};\boldsymbol{\theta}}$, we consider the following composite binary hypothesis testing problem:

$$\begin{aligned} H_0: &\quad \boldsymbol{\theta}_r = \boldsymbol{\theta}_{r_0}, \quad \boldsymbol{\theta}_s = \boldsymbol{\theta}_{s_0} \\ H_1: &\quad \boldsymbol{\theta}_r \neq \boldsymbol{\theta}_{r_0}, \quad \boldsymbol{\theta}_s = \boldsymbol{\theta}_{s_0} \end{aligned} \tag{6}$$

where $\boldsymbol{\theta}_{r_0}$ is *known* and $\boldsymbol{\theta}_{s_0}$ is *unknown*. Similarly to [10], [11], we consider the case where underlying parametric class of distributions $\{P_{\mathbf{X};\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ is unknown. Partial statistical information is available through the MT-mean vector $\boldsymbol{\mu}_u(\boldsymbol{\theta})$ and the MT-covariance matrix $\boldsymbol{\Sigma}_u(\boldsymbol{\theta})$, that are assumed to be known parameterized functions up to some redundant factors.

B. Derivation of the test

We define the partial MT-GQSF corresponding to the vector parameter of interest $\boldsymbol{\theta}_r$ as the partial gradient:

$$\boldsymbol{\psi}_{u,r}(\mathbf{X};\boldsymbol{\theta}) \triangleq \nabla_{\boldsymbol{\theta}_r}\Lambda_u(\mathbf{X};\boldsymbol{\theta}), \tag{7}$$

where $\Lambda_u(\mathbf{X};\boldsymbol{\theta}) \triangleq \log\phi(\mathbf{X};\boldsymbol{\mu}_u(\boldsymbol{\theta}), \boldsymbol{\Sigma}_u(\boldsymbol{\theta}))$ and $\phi(\cdot;\cdot,\cdot)$ denotes a proper complex Gaussian probability density function. Let $\boldsymbol{\theta}_0 \triangleq \left[\boldsymbol{\theta}_{r_0}^T, \boldsymbol{\theta}_{s_0}^T\right]^T$ denote the value of $\boldsymbol{\theta}$ under H_0. By [11, Eqs. (66), (67)] it follows that the partial MT-GQSF satisfies:

$$\mathrm{E}[\boldsymbol{\psi}_{u,r}(\mathbf{X};\boldsymbol{\theta}_0); Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}] = \mathbf{0} \ \text{ for } \ \boldsymbol{\theta} = \boldsymbol{\theta}_0. \tag{8}$$

Therefore, by (1) and (3) we conclude that

$$\mathrm{E}[u(\mathbf{X})\boldsymbol{\psi}_{u,r}\left(\mathbf{X};\boldsymbol{\theta}_0\right); P_{\mathbf{X};\boldsymbol{\theta}}] = \mathbf{0} \ \text{ for } \ \boldsymbol{\theta} = \boldsymbol{\theta}_0. \tag{9}$$

This property suggests that a test for the decision problem (6) can be obtained through a weighted Euclidean norm of an empirical estimate of the expectation in (9).

Hence, the extended MT-GQST is defined as:

$$T_u \triangleq N \cdot \hat{\boldsymbol{\eta}}_{u,r}^T(\tilde{\boldsymbol{\theta}}_{u,0})\hat{\mathbf{B}}_{u,r}^{-1}(\tilde{\boldsymbol{\theta}}_{u,0})\hat{\boldsymbol{\eta}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0}) \underset{H_0}{\overset{H_1}{\gtrless}} t, \tag{10}$$

where $t \in \mathbb{R}_+$ denotes a threshold,

$$\hat{\boldsymbol{\eta}}_{u,r}(\boldsymbol{\theta}) \triangleq \frac{1}{N}\sum\nolimits_{n=1}^{N} u(\mathbf{X}_n)\boldsymbol{\psi}_{u,r}(\mathbf{X}_n;\boldsymbol{\theta}), \tag{11}$$

$\tilde{\boldsymbol{\theta}}_{u,0} \triangleq [\boldsymbol{\theta}_{r_0}^T, \hat{\boldsymbol{\theta}}_{u,s_0}^T]^T$ and $\hat{\boldsymbol{\theta}}_{u,s_0}$ is the MT-GQMLE [12] of the nuisance vector parameter $\boldsymbol{\theta}_{s_0}$ under the constraint $\boldsymbol{\theta}_r = \boldsymbol{\theta}_{r_0}$. Note that $\hat{\boldsymbol{\eta}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0})$ is an empirical estimate of the expectation in (9). The matrix $\hat{\mathbf{B}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0})$ is an empirical estimate of the asymptotic covariance matrix of $\sqrt{N}\hat{\boldsymbol{\eta}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0})$ under H_0. The normalization in (10) by its inverse will result in an asymptotically CFAR detector w.r.t. $\boldsymbol{\theta}_{r_0}$, $\boldsymbol{\theta}_{s_0}$ and the MT-function $u(\cdot)$. Explicitly, $\hat{\mathbf{B}}_{u,r}(\boldsymbol{\theta})$ is defined as:

$$\hat{\mathbf{B}}_{u,r}(\boldsymbol{\theta}) \triangleq \hat{\mathbf{H}}_{u,r}(\boldsymbol{\theta})\hat{\mathbf{R}}_{u,r}(\boldsymbol{\theta})\hat{\mathbf{H}}_{u,r}(\boldsymbol{\theta}), \tag{12}$$

where $\hat{\mathbf{R}}_{u,r}(\boldsymbol{\theta}) \in \mathbb{R}^{m_r \times m_r}$ is formed by the intersection of the first m_r rows and columns of the empirical asymptotic error-covariance of the MT-GQMLE [12, Eq. (36)], given by:

$$\hat{\mathbf{R}}_u(\boldsymbol{\theta}) \triangleq \hat{\mathbf{F}}_u^{-1}(\boldsymbol{\theta})\hat{\mathbf{G}}_u(\boldsymbol{\theta})\hat{\mathbf{F}}_u^{-1}(\boldsymbol{\theta}), \tag{13}$$

with

$$\hat{\mathbf{G}}_u(\boldsymbol{\theta}) \triangleq \frac{1}{N}\sum\nolimits_{n=1}^{N} u^2\left(\mathbf{X}_n\right)\boldsymbol{\psi}_u\left(\mathbf{X}_n;\boldsymbol{\theta}\right)\boldsymbol{\psi}_u^T\left(\mathbf{X}_n;\boldsymbol{\theta}\right), \tag{14}$$

$$\hat{\mathbf{F}}_u(\boldsymbol{\theta}) \triangleq -\frac{1}{N}\sum_{n=1}^{N} u(\mathbf{X}_n)\,\boldsymbol{\Gamma}_u(\mathbf{X}_n;\boldsymbol{\theta}), \quad (15)$$

$\boldsymbol{\psi}_u(\mathbf{X};\boldsymbol{\theta}) \triangleq \nabla_{\boldsymbol{\theta}}\Lambda_u(\mathbf{X};\boldsymbol{\theta})$ and $\boldsymbol{\Gamma}_u(\mathbf{X};\boldsymbol{\theta}) \triangleq \nabla^2_{\boldsymbol{\theta}}\Lambda_u(\mathbf{X};\boldsymbol{\theta})$. The matrix $\hat{\mathbf{H}}_{u,r}(\boldsymbol{\theta})$ in (12) is defined as:

$$\hat{\mathbf{H}}_{u,r}(\boldsymbol{\theta}) \triangleq \hat{\mathbf{F}}_{u,r}(\boldsymbol{\theta}) - \hat{\mathbf{F}}_{u,rs}(\boldsymbol{\theta})\,\hat{\mathbf{F}}^{-1}_{u,s}(\boldsymbol{\theta})\,\hat{\mathbf{F}}^{T}_{u,rs}(\boldsymbol{\theta}), \quad (16)$$

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

$$\hat{\mathbf{F}}_u(\boldsymbol{\theta}) = \begin{bmatrix} \hat{\mathbf{F}}_{u,r}(\boldsymbol{\theta}) & \hat{\mathbf{F}}_{u,rs}(\boldsymbol{\theta}) \\ \hat{\mathbf{F}}^{T}_{u,rs}(\boldsymbol{\theta}) & \hat{\mathbf{F}}_{u,s}(\boldsymbol{\theta}) \end{bmatrix}. \quad (17)$$

Notice that when no nuisance parameters are present, i.e., $m_r = m$ and $m_s = 0$, we obtain that $\boldsymbol{\theta}_r = \boldsymbol{\theta}$, $\boldsymbol{\psi}_{u,r}(\mathbf{X};\boldsymbol{\theta}) = \boldsymbol{\psi}_u(\mathbf{X};\boldsymbol{\theta})$, $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_{r_0} = \tilde{\boldsymbol{\theta}}_{u,0}$, $\hat{\mathbf{R}}_{u,r}(\boldsymbol{\theta}) = \hat{\mathbf{R}}_u(\boldsymbol{\theta})$ and $\hat{\mathbf{H}}_{u,r}(\boldsymbol{\theta}) = \hat{\mathbf{F}}_u(\boldsymbol{\theta})$. Therefore, in this case, it can be shown that the test (10) coincides with the one in [11, Eq. (19)]. The effect of the MT-function $u(\cdot)$ on the detection performance and the resilience against outliers will be discussed in the sequel.

C. Asymptotic performance analysis

In this subsection we analyze the asymptotic performance of the extended MT-GQST (10). Throughout the analysis we shall assume that the test-statistic is implemented using a sequence of i.i.d. samples from $P_{\mathbf{x};\boldsymbol{\theta}}$, and that the following regularity conditions are satisfied:

(A-1) The parameter space $\boldsymbol{\Theta}$ is compact.
(A-2) $\boldsymbol{\theta}_0 \triangleq [\boldsymbol{\theta}^T_{r_0}, \boldsymbol{\theta}^T_{s_0}]^T$ lies in the interior of $\boldsymbol{\Theta}$.
(A-3) $\boldsymbol{\mu}_u(\boldsymbol{\theta}) \neq \boldsymbol{\mu}_u(\boldsymbol{\theta}_0)$ or $\boldsymbol{\Sigma}_u(\boldsymbol{\theta}) \neq \boldsymbol{\Sigma}_u(\boldsymbol{\theta}_0)$ for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.
(A-4) $\boldsymbol{\mu}_u(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_u(\boldsymbol{\theta})$ are twice continuously differentiable.
(A-5) $\boldsymbol{\Sigma}_u(\boldsymbol{\theta})$ is non-singular over $\boldsymbol{\Theta}$.
(A-6) $\mathrm{E}[u^4(\mathbf{X}); P_{\mathbf{x};\boldsymbol{\theta}}]$ and $\mathrm{E}[\|\mathbf{X}\|^8 u^4(\mathbf{X}); P_{\mathbf{x};\boldsymbol{\theta}}]$ are bounded.
(A-7) The matrices $\mathbf{G}_u(\boldsymbol{\theta})$ (23) and $\mathbf{F}_u(\boldsymbol{\theta})$ (24) and $\mathbf{F}_{u,s}(\boldsymbol{\theta})$ [16, Eq. (S-4)] are non-singular at $\boldsymbol{\theta}_0$.
(A-8) The probability measure $P_{\mathbf{x};\boldsymbol{\theta}}$ has a density function $f(\mathbf{x};\boldsymbol{\theta})$ that is differentiable in $\boldsymbol{\Theta}$ for a.e. $\mathbf{x} \in \mathcal{X}$.
(A-9) The Fisher information under $P_{\mathbf{x};\boldsymbol{\theta}}$ is continuous at $\boldsymbol{\theta}_0$.

The following theorem states the asymptotic distribution of the test statistic under the null hypothesis H_0.

Theorem 1 (Asymptotic distribution under the H_0)**.** *Assume that conditions (A-1)-(A-7) are satisfied. Then,*

$$T_u \xrightarrow[N\to\infty]{D} \chi^2_{m_r} \;\; \text{under } H_0, \quad (18)$$

where $\chi^2_{m_r}$ denotes a central chi-squared distribution with m_r-degrees of freedom. [A proof is given in [16, Sec. I-A]]

Hence, the extended MT-GQST (10) is an asymptotically CFAR detector w.r.t. $\boldsymbol{\theta}_0$ and the MT-function $u(\cdot)$. In the following, we derive the asymptotic distribution of the test-statistic under a sequence of local alternatives. This will enable to obtain the asymptotic power in a neighbourhood of $\boldsymbol{\theta}_{r_0}$.

Theorem 2 (Asymptotic distribution under local alternatives)**.** *Assume that conditions (A-1)–(A-9) are satisfied. Furthermore, consider the following sequence of local alternatives that converges to $\boldsymbol{\theta}_{r_0}$ at a rate of $1/\sqrt{N}$:*

$$H_1: \boldsymbol{\theta}_r = \boldsymbol{\theta}_{r_0} + \mathbf{r}/\sqrt{N}, \;\; \boldsymbol{\theta}_s = \boldsymbol{\theta}_{s_0}, \quad (19)$$

where $\mathbf{r} \in \mathbb{R}^{m_r}$ is a non-zero locality parameter. Then,

$$T_u \xrightarrow[N\to\infty]{D} \chi^2_{m_r}(\lambda_u(\mathbf{r})) \;\; \text{under } H_1, \quad (20)$$

where $\chi^2_{m_r}(\lambda_u(\mathbf{r}))$ is a non-central chi-squared distribution with m_r-degrees of freedom and non-centrality parameter

$$\lambda_u(\mathbf{r}) \triangleq \mathbf{r}^T\mathbf{R}^{-1}_{u,r}(\boldsymbol{\theta}_0)\mathbf{r}. \quad (21)$$

The matrix $\mathbf{R}_{u,r}(\boldsymbol{\theta})$ is formed by intersection of the first m_r rows and column of the matrix

$$\mathbf{R}_u(\boldsymbol{\theta}) \triangleq \mathbf{F}^{-1}_u(\boldsymbol{\theta})\,\mathbf{G}_u(\boldsymbol{\theta})\,\mathbf{F}^{-1}_u(\boldsymbol{\theta}), \quad (22)$$

where

$$\mathbf{G}_u(\boldsymbol{\theta}) \triangleq \mathrm{E}[u^2(\mathbf{X})\,\boldsymbol{\psi}_u(\mathbf{X};\boldsymbol{\theta})\,\boldsymbol{\psi}^T_u(\mathbf{X};\boldsymbol{\theta}); P_{\mathbf{x};\boldsymbol{\theta}_0}], \quad (23)$$

$$\mathbf{F}_u(\boldsymbol{\theta}) \triangleq -\mathrm{E}[u(\mathbf{X})\,\boldsymbol{\Gamma}_u(\mathbf{X};\boldsymbol{\theta}); P_{\mathbf{x};\boldsymbol{\theta}_0}], \quad (24)$$

and $\boldsymbol{\psi}_u(\mathbf{x};\boldsymbol{\theta})$, $\boldsymbol{\Gamma}_u(\mathbf{x};\boldsymbol{\theta})$ are defined below (15). [A proof is given in [16, Sec. I-B]]

The matrix $\mathbf{R}_{u,r}(\boldsymbol{\theta}_0)$ is the asymptotic error covariance of the MT-GQMLE [12] for estimating the vector parameter of interest at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. By theorems 1 and 2, we conclude that under the local alternatives (19), the asymptotic power of the proposed test at a fixed asymptotic size α is given by:

$$\beta^{(\alpha)}_u(\mathbf{r}) = H_{m_r}(\lambda_u(\mathbf{r}), \alpha), \quad (25)$$

where $H_{m_r}(\lambda,\alpha) \triangleq Q_{\chi^2_{m_r}(\lambda)}\left(Q^{-1}_{\chi^2_{m_r}}(\alpha)\right)$, and the functions $Q_{\chi^2_{m_r}}(\cdot)$ and $Q_{\chi^2_{m_r}(\cdot)}(\cdot)$ denote the tail probabilities of a central and non-central chi-squared distributions, respectively. The relation in (25) will be used in Subsection III-E for selection of the MT-function $u(\cdot)$.

D. Robustness analysis

In this subsection, we analyze the robustness of the asymptotic test-size to outliers. To do so, we quantify the deviation from the nominal asymptotic test-size that is caused by a small contamination. We consider the following ϵ-contaminated probability measure, also applied in [13, Ch. 13.2]:

$$P_{\epsilon,N} \triangleq \left(1 - \epsilon/\sqrt{N}\right)P_{\mathbf{x};\boldsymbol{\theta}_0} + (\epsilon/\sqrt{N})\delta_{\mathbf{y}}, \quad (26)$$

where $0 \leq \epsilon \leq 1$, $\mathbf{y} \in \mathbb{C}^p$, and $\delta_{\mathbf{y}}$ is a Dirac probability measure at a point $\mathbf{y}$ that represents an outlier. 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. The following Theorem states a relation between α_ϵ and α.

Theorem 3. *Under (A-1)–(A-7) and the ϵ-contaminated measure (26)*

$$\alpha_\epsilon = \alpha + \epsilon^2 I_u(\mathbf{y};\boldsymbol{\theta}_0) + o(\epsilon^2), \quad (27)$$

where

$$I_u(\mathbf{y};\boldsymbol{\theta}_0) \triangleq c\left\|u(\mathbf{y})\boldsymbol{\psi}_{u,r}(\mathbf{y};\boldsymbol{\theta}_0)\right\|^2_{\mathbf{R}^{-1}_{u,r}(\boldsymbol{\theta}_0)}, \quad (28)$$

the constant $c \triangleq 2\left.\frac{\partial H_{m_r}(\lambda,\alpha)}{\partial\lambda}\right|_{\lambda=0}$, and $H_{m_r}(\lambda,\alpha)$ is defined below (25). [A proof is given in [16, Sec. I-C]]

Therefore, in order to obtain a stable test-size in the presence of an infinitesimal contamination at $\mathbf{y}$, the function $I_u(\mathbf{y};\boldsymbol{\theta}_0)$ must be bounded over $\mathbb{C}^p$. It can be shown that this condition is satisfied whenever, the MT-function $u(\mathbf{y})$ and the function $u(\mathbf{y})\|\mathbf{y}\|^2$ are bounded over $\mathbb{C}^p$.

E. Selection of the MT-function

Based on (25), we define the worst-case asymptotic local power $\bar{\beta}_u^{(\alpha)}(c) \triangleq \min_{\mathbf{r}:\|\mathbf{r}\|\geq c} \beta_u^{(\alpha)}(\mathbf{r})$, where $c>0$ denotes a constant that lower bounds the Euclidean norm of the locality parameter $\mathbf{r}$. Similarly to [11, Corollary 1], it can be shown that

$$\bar{\beta}_u^{(\alpha)}(c) = H_{m_r}(\gamma_u(c),\alpha), \quad (29)$$

where, by the definition below (25), $H_{m_r}(\gamma_u(c),\alpha)$ is a tail probability of a non-central chi-squared distribution with non-centrality parameter $\gamma_u(c) \triangleq c^2\|\mathbf{R}_{u,r}(\boldsymbol{\theta}_0)\|_S^{-1}$, $\mathbf{R}_{u,r}(\boldsymbol{\theta})$ is defined below (21) and $\|\cdot\|_S$ denotes the spectral norm. By (29), one sees that the MT-function $u(\cdot)$ controls the worst-case asymptotic local power through the non-centrality parameter $\gamma_u(c)$. Therefore, by the property that the tail probability of the non-central chi-squared distribution is monotonically increasing in the non-centrality parameter [17], it follows that for any fixed values of the constant c and the asymptotic test-size α, minimization of the spectral norm $\|\mathbf{R}_{u,r}(\boldsymbol{\theta}_0)\|_S$ amounts to maximization of the worst-case local power in (29).

Hence, we propose to choose the MT-function $u(\cdot)$ that minimizes $\|\hat{\mathbf{R}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0})\|_S$, where $\hat{\mathbf{R}}_{u,r}(\boldsymbol{\theta})$ is the empirical version of $\mathbf{R}_{u,r}(\boldsymbol{\theta})$, defined below (12), and $\tilde{\boldsymbol{\theta}}_{u,0}$ is the restricted estimator of $\boldsymbol{\theta}_0$ defined below (11). Similarly to [11, Prop. 4], it can be shown that under conditions (A-1)−(A-9) $\hat{\mathbf{R}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0})$ is a consistent estimator of $\mathbf{R}_{u,r}(\boldsymbol{\theta}_0)$ under the local alternatives (19). We emphasize that $\hat{\mathbf{R}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0})$ is obtained using the same data samples $\{\mathbf{X}_n\}_{n=1}^N$ comprising the extended MT-GQST (10). Similarly to [10], [11] we restrict the class of MT-functions to some family of parametric functions $\{u(\mathbf{X};\boldsymbol{\omega}),\boldsymbol{\omega}\in\mathbb{R}^k\}$ that satisfy the conditions in Definition 1 and conditions (A-3)−(A-7). The optimal MT-function parameter $\boldsymbol{\omega}_{\rm opt}$ is the minimizer of $\|\hat{\mathbf{R}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0}(\boldsymbol{\omega});\boldsymbol{\omega})\|_S$, i.e.,

$$\boldsymbol{\omega}_{\rm opt} = \arg\min_{\boldsymbol{\omega}} \|\hat{\mathbf{R}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0}(\boldsymbol{\omega});\boldsymbol{\omega})\|_S. \quad (30)$$

IV. Example

In this section, the extended MT-GQST (10) is applied for testing a vector parameter of interest that comprises the following linear multivariate data model:

$$\mathbf{X}_n = \mathbf{A}_r\boldsymbol{\vartheta}_r + \mathbf{A}_s\boldsymbol{\vartheta}_s + \mathbf{W}_n, \;\; n=1,\ldots,N, \quad (31)$$

where $\{\mathbf{X}_n\in\mathbb{C}^p\}$ is an observation process, $\mathbf{A}_r\in\mathbb{C}^{p\times q_r}$ and $\mathbf{A}_s\in\mathbb{C}^{p\times q_s}$, $q_r+q_s<p$, are known full column-rank matrices with linearly independent range spaces, and $\{\mathbf{W}_n\in\mathbb{C}^p\}$ is an i.i.d. noise process with centered complex spherical distribution [15]. The real vector parameter of interest $\boldsymbol{\theta}_r$ and the nuisance vector parameter $\boldsymbol{\theta}_s$ in (6) are related to the complex vector parameters $\boldsymbol{\vartheta}_r$ and $\boldsymbol{\vartheta}_s$ in (31) through the complex-to-real mappings:

$$\boldsymbol{\theta}_r = \left[\Re(\boldsymbol{\vartheta}_r)^T,\Im(\boldsymbol{\vartheta}_r)^T\right]^T \text{ and } \boldsymbol{\theta}_s = \left[\Re(\boldsymbol{\vartheta}_s)^T,\Im(\boldsymbol{\vartheta}_s)^T\right]^T, \quad (32)$$

respectively. Under the settings above, the exact probability distribution of the noise process is unknown. Therefore, the extended version of the standard score test for the case of nuisance parameters [7, p. 513], [8, p. 190], that requires complete knowledge of the likelihood function, is not implementable.

In the following, we show that, unlike the standard score test, the extended MT-GQST (10), that requires only partial statistical information, can be implemented when the MT-function $u(\cdot)$ belongs to the following wide class of functions:

$$\{u(\mathbf{x}) = v(\|\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{x}\|), \;\; v:\mathbb{R}_+\to\mathbb{R}_+\}, \quad (33)$$

where $\mathbf{A}\triangleq[\mathbf{A}_r,\mathbf{A}_s]$ and $\mathbf{P}_{\mathbf{A}}^{\perp}$ is a projection matrix onto the null-space of $\mathbf{A}^H$. Using (31)-(33), it can be shown that the MT-mean (4) and the MT-covariance (5) are given by:

$$\boldsymbol{\mu}_u(\boldsymbol{\theta}) = \mathbf{A}_r\boldsymbol{\vartheta}_r+\mathbf{A}_s\boldsymbol{\vartheta}_s \;\text{ and }\; \boldsymbol{\Sigma}_u(\boldsymbol{\theta}) = c_v\mathbf{P}_{\mathbf{A}}+r_v\mathbf{P}_{\mathbf{A}}^{\perp}, \quad (34)$$

respectively, where c_v and r_v are some strictly positive constants, and $\mathbf{P}_{\mathbf{A}}$ is the projection matrix onto the range-space of $\mathbf{A}$. Hence, maximization of the objective function in [12, Eq. (23)] w.r.t. $\boldsymbol{\theta}_s$ under the constraint $\boldsymbol{\theta}_r=\boldsymbol{\theta}_{r_0}$ leads to the following MT-GQMLE of the nuisance vector parameter under H_0:

$$\hat{\boldsymbol{\theta}}_{u,s_0} = (\tilde{\mathbf{A}}_s^T\tilde{\mathbf{A}}_s)^{-1}\tilde{\mathbf{A}}_s^T(\tilde{\boldsymbol{\mu}}_u-\tilde{\mathbf{A}}_r\boldsymbol{\theta}_{r_0}), \quad (35)$$

where

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

$\tilde{\boldsymbol{\mu}}_u \triangleq \sum_{n=1}^N\hat{\varphi}_u(\mathbf{X}_n)\tilde{\mathbf{X}}_n$, $\hat{\varphi}_u(\mathbf{X}_n)\triangleq u(\mathbf{X}_n)/\sum_{k=1}^N u(\mathbf{X}_k)$ and $\tilde{\mathbf{X}}\triangleq[\Re^T(\mathbf{X}),\Im^T(\mathbf{X})]^T$. Note that by [12, Eq. (16)] it follows that $\tilde{\boldsymbol{\mu}}_u$ is the empirical MT-mean of the real random vector $\tilde{\mathbf{X}}$. Therefore, using (11)-(17), (34) and (35) it can be shown that under the observation model (31), the test-statistic (10) of the extended MT-GQST, for the composite hypothesis testing problem (6), takes the simple form:

$$T_u = \frac{(\tilde{\boldsymbol{\mu}}_u-\tilde{\mathbf{A}}_r\boldsymbol{\theta}_{r_0})^T\mathbf{P}_{\tilde{\mathbf{A}}_s}^{\perp}\tilde{\mathbf{A}}_r\tilde{\mathbf{M}}_{u^2}^{-1}\tilde{\mathbf{A}}_r^T\mathbf{P}_{\tilde{\mathbf{A}}_s}^{\perp}(\tilde{\boldsymbol{\mu}}_u-\tilde{\mathbf{A}}_r\boldsymbol{\theta}_{r_0})}{\sum_{n=1}^N\hat{\varphi}_u^2(\mathbf{X}_n)}, \quad (36)$$

where $\tilde{\mathbf{M}}_u \triangleq \tilde{\mathbf{A}}_r^T\mathbf{P}_{\tilde{\mathbf{A}}_s}^{\perp}\tilde{\mathbf{C}}_u\mathbf{P}_{\tilde{\mathbf{A}}_s}^{\perp}\tilde{\mathbf{A}}_r$, $\tilde{\mathbf{C}}_u \triangleq \tilde{\boldsymbol{\Sigma}}_u+(\tilde{\boldsymbol{\mu}}_u-\tilde{\mathbf{A}}_r\boldsymbol{\theta}_{r_0})(\tilde{\boldsymbol{\mu}}_u-\tilde{\mathbf{A}}_r\boldsymbol{\theta}_{r_0})^T$ and $\tilde{\boldsymbol{\Sigma}}_u\triangleq\sum_{n=1}^N\hat{\varphi}_u(\mathbf{X}_n)\tilde{\mathbf{X}}_n\tilde{\mathbf{X}}_n^T-\tilde{\boldsymbol{\mu}}_u\tilde{\boldsymbol{\mu}}_u^T$. Notice that by [12, Eq. (17)] it follows that $\tilde{\boldsymbol{\Sigma}}_u$ is the empirical MT-covariance of the real random vector $\tilde{\mathbf{X}}$. Furthermore, it can be shown that the empirical error-covariance of the MT-GQMLE associated with $\boldsymbol{\theta}_{r_0}$, that is required for the optimization in (30), is given by:

$$\hat{\mathbf{R}}_{u,r}(\tilde{\boldsymbol{\theta}}_{u,0}) = \frac{(\tilde{\mathbf{A}}_r^T\mathbf{P}_{\tilde{\mathbf{A}}_s}^{\perp}\tilde{\mathbf{A}}_r)^{-1}\tilde{\mathbf{M}}_{u^2}(\tilde{\mathbf{A}}_r^T\mathbf{P}_{\tilde{\mathbf{A}}_s}^{\perp}\tilde{\mathbf{A}}_r)^{-1}}{(N\sum_{n=1}^N\hat{\varphi}_u^2(\mathbf{X}_n))^{-1}}, \quad (37)$$

where $\tilde{\boldsymbol{\theta}}_{u,0}$ is defined below (11).

To gain robustness against heavy-tailed noise outliers, we specify the MT-function in a subset of (33), that is comprised of zero-centered pseudo-Gaussian MT-functions parameterized by a width parameter:

$$\{u(\mathbf{x};\omega) = \exp(-\|\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{x}\|^2/\omega), \;\; \omega\in\mathbb{R}_{++}\}. \quad (38)$$

Similarly to [11, Sec. 5.1] it can be shown that $u_{\rm PG}(\mathbf{x};\omega)$ satisfies the robustness conditions stated below Eq. (28) with high probability.

In the following simulation study, we compare the detection performance of the extended MT-GQST, to the extended versions (for the case of nuisance parameters) of the GQST, the omniscient score test, the quasi GLRT [6], and the quasi score test [6]. The latter two tests, called here, GGD-QGLRT and GGD-QST, respectively, were obtained under the assumption that the noise process obeys a generalized Gaussian distribution (GGD) [15]. The test-statistic of the GQST was obtained form (36) by setting $u(\mathbf{x}) = 1$. The omniscient score test was implemented according to [7, p. 513], [8, p. 190]. Exact implementation details of the GGD-QGLRT and GGD-QST appear in [16, Sec. II]. One can verify that implementation of the MT-GQST (36) is significantly easier as compared to the GGD-QGLRT and GGD-QST.

Throughout the simulation study, the vector parameter of interest at the null hypothesis H_0 was set to $\boldsymbol{\theta}_{r_0} = \mathbf{0}_{2q_r}$, $q_r = 2$, where $\mathbf{0}_q$ denotes a vector comprised of q zeros ($2q$ follows from the real-imaginary decomposition in (32)). We considered a specific local alternative $\boldsymbol{\theta}_{r_1} = \boldsymbol{\theta}_{r_0} + 0.05 \times \mathbf{1}_{2q_r}$, corresponding to $\mathbf{r} = \sqrt{N}(\boldsymbol{\theta}_{r_1} - \boldsymbol{\theta}_{r_0})$ in (19), where $\mathbf{1}_q$ denotes a vector with q unit entries. The nuisance vector parameter was set to $\boldsymbol{\theta}_{s_0} = \mathbf{1}_{2q_s}$, $q_s = 2$. The observations dimensionality was set to $p = 16$. The matrices $\mathbf{A}_r$ and $\mathbf{A}_s$ in (31) were set to $\mathbf{A}_r = \frac{1}{\sqrt{4}}[\mathbf{a}_0, \mathbf{a}_1]$ and $\mathbf{A}_s = \frac{1}{\sqrt{4}}[\mathbf{a}_3, \mathbf{a}_4]$, where $\mathbf{a}_k \triangleq \frac{1}{\sqrt{p}}[1, \exp[(i\phi_k), \ldots, \exp[(i(p-1)\phi_k)]^T$, $\phi_k = \pi/(3+k)$, $k = 0, 1, 2, 3$. Two types of spherical noise distributions with scatter matrix $\sigma^2_{\mathbf{W}}\mathbf{I}_p$ were examined: 1) Gaussian and 2) Cauchy [15].

For each type of noise, we examined the empirical power of the extended MT-GQST as compared to the empirical powers of the detectors specified above versus the signal-to-noise-ratio defined here as SNR $\triangleq {\rm tr}[\mathbf{A}^H\mathbf{A}]/\sigma^2_{\mathbf{W}}$, where ${\rm tr}[\cdot]$ denotes the trace operator. The optimum theoretical asymptotic local power of the extended MT-GQST is also reported. The sample-size and test-size were set to $N = 1000$ and $\alpha = 10^{-3}$, respectively. For all compared tests, the threshold parameter and the empirical power curves were obtained via 10^5 Monte-Carlo trials. The optimal scale parameter $\omega_{\rm opt}$ of the pseudo-Gaussian MT-function (38) was obtained according to (30) by minimizing the spectral norm of (37) over 100 equally spaced grid points of the closed interval $\Omega = [0.1, 5]$.

Observing Fig. 1, one sees that for the Gaussian noise the compared tests perform similarly. However, for the heavy-tailed Cauchy noise, the MT-GQST outperforms the compared tests and attains detection performance that are significantly closer to those obtained by the omniscient score test that assumes complete knowledge of the likelihood function.

V. Conclusion

In this paper, the MT-GQST [10], [11] was extended to the case where nuisance parameters are present. The extended MT-GQST was applied for testing a vector parameter of interest of a linear data model in the presence of Gaussian and heavy-tailed Cauchy noise. Simulation study demonstrates implementation and performance advantage of the extended MT-GQST over other robust detectors.

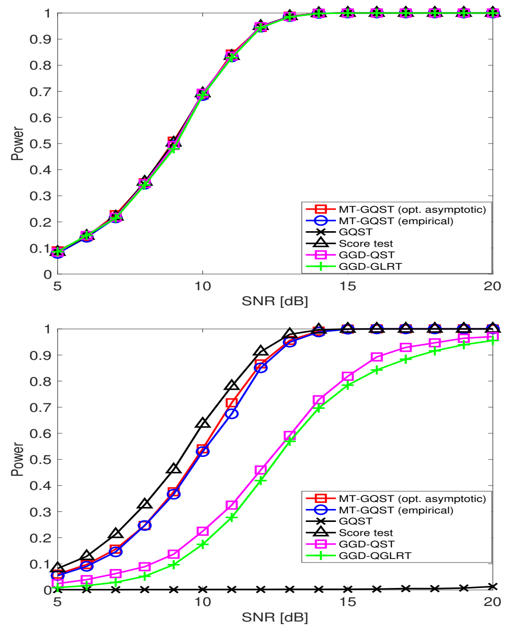

Fig. 1. Testing a vector parameter of interest (in the presence of nuisance parameter) of a linear model in Gaussian noise (**top**) and heavy-tailed Cauchy noise (**bottom**).

References

[1] H. White, "Maximum likelihood estimation of misspecified models," *Econometrica: Journal of the Econometric Society,* pp. 1-25, 1982.
[2] H. White, *Estimation, inference and specification analysis,* Cambridge university press, 1996.
[3] R. F. Engle, "Wald, likelihood ratio, and Lagrange multiplier tests in econometrics," *Handbook of econometrics,* vol. 2, pp. 775-826, 1984.
[4] G. Fiorentini and E. Sentana, "Tests for serial dependence in static, non-Gaussian factor models," *Unobserved components and time series econometrics,* pp. 118-163, Oxford University Press, 2015.
[5] T. Bollerslev and J. M. Wooldridge, "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances," *Econometric reviews,* vol. 11, no. 2, pp. 143-172, 1992.
[6] J. T. Kent, "Robust properties of likelihood ratio tests," *Biometrika,* vol. 69, no. 1, pp. 19-27, 1982.
[7] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses.* Springer Texts in Statistics, 2005.
[8] S. M. Kay, *Fundamentals of statistical signal processing: detection theory,* Prentice-Hall, 1993.
[9] C. R. Rao, "Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation," *Mathematical Proceedings of the Cambridge Philosophical Society,* vol. 44, no. 1, pp. 50-57, 1948.
[10] K. Todros, "Measure-transformed Gaussian quasi score test," *Proc. of EUSIPCO 2017,* pp. 2115-2119.
[11] K. Todros, Robust composite binary hypothesis testing via measure transformed quasi score test," *Signal Processing*, vol. 155, pp. 202-217, Feb. 2019.
[12] 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.
[13] P. J. Huber, *Robust statistics,* Springer, 2011.
[14] K. B. Athreya and S. N. Lahiri, *Measure theory and probability theory,* Springer-Verlag, 2006.
[15] 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.
[16] K. Todros, "Measure-transformed Gaussian quasi score test in the presence of nuisance parameters: Supplementary material," *Technical report,* Mar. 2019. Online version: *http://www.ee.bgu.ac.il/~todros/Report.pdf*
[17] A. Aubel and W. Gawronski, "Analytic properties of non-central distributions," *Applied Math. and Comp.,* vol. 141, pp. 3-12, 2003.