UNIVERSITY OF SURREY

Discussion Papers in Economics

RELATIVE ERROR ACCURATE STATISTIC BASED ON NONPARAMETRIC LIKELIHOOD

By

Lorenzo Camponovo
(University of Surrey),

&

Taisuke Otsu
(London School of Economics).

DP 05/18

School of Economics
University of Surrey
Guildford
Surrey GU2 7XH, UK
Telephone +44 (0)1483 689380
Facsimile +44 (0)1483 689548
Web www.econ.surrey.ac.uk
ISSN: 1749-5075

RELATIVE ERROR ACCURATE STATISTIC BASED ON NONPARAMETRIC LIKELIHOOD

LORENZO CAMPONOVO AND TAISUKE OTSU

Abstract. This paper develops a new test statistic for parameters defined by moment conditions that exhibits desirable relative error properties for the approximation of tail area probabilities. Our statistic, called the tilted exponential tilting (TET) statistic, is constructed by estimating certain cumulant generating function under exponential tilting weights. We show that the asymptotic p-value of the TET statistic can provide an accurate approximation to the p-value of an infeasible saddlepoint statistic, which is asymptotically chi-squared distributed with a relative error of order n^{-1} both in normal and large deviation regions. Numerical results illustrate the accuracy of the proposed TET statistic. Our results cover both just- and over-identified moment condition models.

1. Introduction

This paper develops a new test statistic for parameters defined by moment conditions that exhibits desirable relative error properties for the approximation of tail area probabilities. For this problem, there are various test statistics available in the literature, such as the Wald, empirical likelihood (Owen, 1988), exponential tilting (Efron, 1981, 1982, Kitamura and Stutzer, 1997, and Imbens, Spady and Johnson, 1998), power divergence (Baggerly, 1998), and saddlepoint statistics (Robinson, Ronchetti and Young, 2003, and Ma and Ronchetti, 2011), among others. In particular, it is known that the empirical likelihood statistic admits the Bartlett correction, a higher-order refinement for the *absolute* error of the type I error probability (DiCiccio, Hall and Romano, 1991). This refinement in the absolute error is typically not achieved by other statistics, such as exponential tilting (Jing and Wood, 1996, and Baggerly, 1998).

For statistical inference, researchers are commonly interested in the accuracy of approximations for tail area probabilities or p-values of test statistics. For this purpose, the *relative* error rather than the absolute one would be a more relevant measure of accuracy, and various procedures typically based on saddlepoint approximations are developed (Tingley and Field, 1990, Daniels and Young, 1991, Jing and Robinson, 1994, Robinson, Ronchetti and Young, 2003, and Kolassa and Robinson, 2011, among others). In particular, Robinson, Ronchetti and Young (2003) considered the situation where the cumulant generating function is known to the researcher and developed a novel saddlepoint statistic that is asymptotically chi-squared distributed with a relative error of order n^{-1} even in the large deviation region. Although this statistic is generally infeasible due to the requirement on knowledge of the cumulant generating function, Robinson,

Date: April 2016.

The authors are grateful to helpful discussions with Elvezio Ronchetti and comments by seminar participants at University of Geneva. Financial support from the ERC Consolidator Grant (SNP 615882) is gratefully acknowledged (Otsu).

Ronchetti and Young (2003) and Ma and Ronchetti (2011) proposed some feasible versions of the saddlepoint statistic by using the exponential tilting weights (Efron, 1981, 1982).

In this paper, we propose a new test statistic that achieves desirable relative error properties for the approximation of tail area probabilities. More precisely, our statistic is asymptotically chi-squared distributed, and the asymptotic p-value approximations using the chi-squared distribution are very accurate even in the tails. The basic idea of our statistic is to note that the conventional exponential tilting statistic is constructed from estimating the cumulant generating function by the sample average, and to modify the cumulant estimation by using the exponential tilting weights instead of the uniform weights n^{-1}. In other words, we tilt the exponential tilting statistic. Thus, the new statistic is called the tilted exponential tilting (TET) statistic.

We show that the TET statistic is asymptotically chi-squared distributed, and demonstrate that its asymptotic p-value provides an accurate approximation to the p-value of some ideal (but infeasible) saddlepoint statistic, which is also asymptotically chi-squared distributed with a relative error of order n^{-1} both in normal and large deviation regions. We note that both the TET and ideal statistics are new in the literature, and different from the saddlepoint statistics discussed above. Furthermore, our results on the TET statistic cover both just- and over-identified moment condition models.

Finally, we study through Monte Carlo simulations the accuracy of the proposed TET statistic. We consider both just- and over-identified instrumental variable regression models. The numerical results highlight a desirable accuracy of our test statistic. In particular, the empirical quantiles of the TET statistic are extremely close to those of the limiting distribution even for very small sample sizes.

2. Benchmark case

In this section, we present the basic idea of the new test statistic and its theoretical and numerical properties under a benchmark setup. Section 2.1 introduces our TET statistic. Section 2.2 illustrates its finite sample accuracy through Monte Carlo simulation. In Section 2.3, we show that the TET statistic has desirable relative error properties for the approximation of tail area probabilities.

2.1. **Tilted exponential tilting statistic.** Suppose we observe an i.i.d. sample $\{X_i\}_{i=1}^n$ of X. In this section, we focus on hypothesis testing for the null $H_0 : \theta_0 = 0$ against the two-sided alternative $H_1 : \theta_0 \neq 0$, where the p-dimensional vector of parameters θ_0 is defined by p-dimensional moment conditions

$$E[g(X, \theta_0)] = 0.$$

Since we do not specify the parametric distribution form of X, testing methods based on parametric likelihood theory, such as the likelihood ratio and score tests, are not applicable. However, there are several ways to test H_0 in this setting. For example, we can implement the Wald test based on some estimator of θ_0. Also based on some nonparametric likelihood, we can conduct likelihood ratio or score type tests (see, e.g., Owen, 2001).

We propose a new test statistic for H_0, which exhibits a desirable finite sample accuracy. Our test statistic is constructed by evaluating the exponential tilting statistic (Efron, 1981, 1982, Kitamura and Stutzer, 1997, and Imbens, Spady and Johnson, 1998) under the exponential tilting weights based on the restriction $E[g(X,0)] = 0$. Let $\hat{\lambda}$ be the solution of $\sum_{i=1}^{n} e^{\hat{\lambda}' g(X_i,0)} g(X_i, 0) = 0$. The conventional exponential tilting statistic is written as

$$T_n^{\mathrm{et}} = -2 \log \left(\frac{1}{n} \sum_{i=1}^{n} e^{\hat{\lambda}' g(X_i,0)} \right).$$

It is known that nT_n^{et} converges in distribution to the chi-squared distribution with p degrees of freedom under H_0. This statistic is obtained by minimization of the empirical relative entropy

$$\min_{\pi_1,\dots,\pi_n} \sum_{i=1}^{n} n\pi_i \log(n\pi_i), \quad \text{s.t.} \quad \sum_{i=1}^{n} \pi_i g(X_i, 0) = 0, \quad \sum_{i=1}^{n} \pi_i = 1.$$

By applying the Lagrange multiplier method, the solution is written as

$$\hat{\pi}_i = \frac{e^{\hat{\lambda}' g(X_i,0)}}{\sum_{j=1}^{n} e^{\hat{\lambda}' g(X_j,0)}},$$

for $i = 1, \dots, n$. Note that by construction these optimal weights are positive and satisfy the moment condition $\sum_{i=1}^{n} \hat{\pi}_i g(X_i, 0) = 0$. Indeed, the empirical distribution using the weights $\{\hat{\pi}_i\}_{i=1}^{n}$ is an asymptotically efficient estimator of the distribution function of X under the restriction $E[g(X,0)] = 0$ (Brown and Newey, 1998).

Intuitively, the exponential tilting statistic T_n^{et} is constructed by taking expectation of $e^{\hat{\lambda}' g(X_i,0)}$ under the empirical distribution with weights $1/n$. Our proposal is to replace the uniform weights by the optimal ones under H_0, and to take expectation of $e^{\hat{\lambda}' g(X_i,0)}$ under the *tilted* empirical distribution with weights $\{\hat{\pi}_i\}_{i=1}^{n}$, that is

$$T_n^{\mathrm{tet}} = 2 \log \left(\sum_{i=1}^{n} \hat{\pi}_i e^{\hat{\lambda}' g(X_i,0)} \right) = 2 \left[\log \left(\sum_{i=1}^{n} e^{2\hat{\lambda}' g(X_i,0)} \right) - \log \left(\sum_{i=1}^{n} e^{\hat{\lambda}' g(X_i,0)} \right) \right].$$

We call this statistic the tilted exponential tilting (TET) statistic. As shown in the proof of Theorem 1, the reason for the positive sign of T_n^{tet} can be seen from a second-order expansion around $\hat{\lambda}' g(X_i, 0) = 0$,

$$nT_n^{\mathrm{tet}} = n\hat{\lambda}' \left[\sum_{i=1}^{n} \hat{\pi}_i g(X_i, 0) g(X_i, 0)' \right] \hat{\lambda} + o_p(1),$$

under H_0, where we used $\sum_{i=1}^{n} \hat{\pi}_i g(X_i, 0) = 0$. It will be shown that the right hand side converges in distribution to the chi-squared distribution with p degrees of freedom under H_0. To make the argument rigorous, we impose the following assumption.

Assumption 1. *$\{X_i\}_{i=1}^{n}$ is i.i.d., $E[|g(X, \theta_0)|^{\zeta}] < \infty$ for some $\zeta > 2$, and $E[g(X, \theta_0) g(X, \theta_0)']$ is nonsingular.*

All conditions are standard. Based on these conditions, the limiting null distribution of the TET statistic is obtained as follows.

Theorem 1. *Under Assumption 1 and $H_0 : \theta_0 = 0$, the TET statistic nT_n^{tet} converges in distribution to the chi-squared distribution with p degrees of freedom.*

Therefore, under H_0, the TET statistic nT_n^{tet} is asymptotically equivalent to the exponential tilting statistic nT_n^{et}. We can also show that they have the same local power function under local alternatives.

2.2. **Simulation for benchmark case.** To illustrate finite sample accuracy of the TET statistic, we provide some simulation results. We generate random samples $\{X_i\}_{i=1}^n = \{Y_i, W_i, Z_i\}_{i=1}^n$ of sizes $n = 20$, 40, 60, and 80 according to

$$\begin{aligned} Y_i &= W_i\theta_0 + U_i, \\ W_i &= Z_i\pi_0 + V_i, \end{aligned}$$

where $Z_i \sim N(0,1)$, $\begin{pmatrix} U_i \\ V_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}\right)$, $\theta_0 = 0$, and $\pi_0 = 0.8$. The parameter θ_0 is defined by the moment condition $E[g(X,\theta_0)] = E[Z(Y - W\theta_0)] = 0$. We are interested in testing the null hypothesis $H_0 : \theta_0 = 0$ against the alternative $H_1 : \theta_0 \neq 0$, and compare the exponential tilting statistic nT_n^{et}, and the TET statistic nT_n^{tet}. Both statistics converge in distribution to the χ_1^2 distribution under H_0. Figure 2.1 reports the q-q plots of the empirical quantiles of these test statistics against those of the χ_1^2 distribution. The number of Monte Carlo replications is $20,000$.

The empirical quantiles of the TET statistic are extremely close to those of the limiting χ_1^2 distribution. The accuracy of the exponential tilting statistic increases as the sample size increases. However, the TET always outperforms the exponential tilting. In the next subsection, we provide some theoretical arguments which clarify the desirable accuracy of the TET statistic.

2.3. **Relative error properties.** In the definition of the TET statistic, we propose to take expectation of $e^{\hat{\lambda}' g(X_i,0)}$ under the tilted weights $\{\hat{\pi}_i\}_{i=1}^n$ satisfying $\sum_{i=1}^n \hat{\pi}_i g(X_i, 0) = 0$. To see the rationale of our approach, consider the ideal (but infeasible) statistic

$$T_n = 2K(\hat{\lambda}),$$

where $K(\lambda) = \log E[e^{\lambda' g(X,0)}]$ is the cumulant generating function of $g(X,0)$, and $\hat{\lambda}$ solves $\sum_{i=1}^n e^{\hat{\lambda}' g(X_i,0)} g(X_i, 0) = 0$. Observe that T_n is infeasible because it involves expectation to evaluate the cumulant. Furthermore, T_n is different from the saddlepoint statistic proposed in Robinson, Ronchetti and Young (2003) because it does not involve any estimators of θ_0.

Let F_p be the cumulative distribution function of the chi-squared distribution with p degrees of freedom. The relative error property for the approximation of the tail area probability of T_n is established as follows.

Theorem 2. *Suppose that Assumptions 1 and 2 in the Appendix hold true. Then under H_0 : $\theta_0 = 0$,*

$$\Pr\{nT_n \geq nt : F\} = \{1 - F_p(n\xi(t))\}(1 + O(n^{-1})),$$

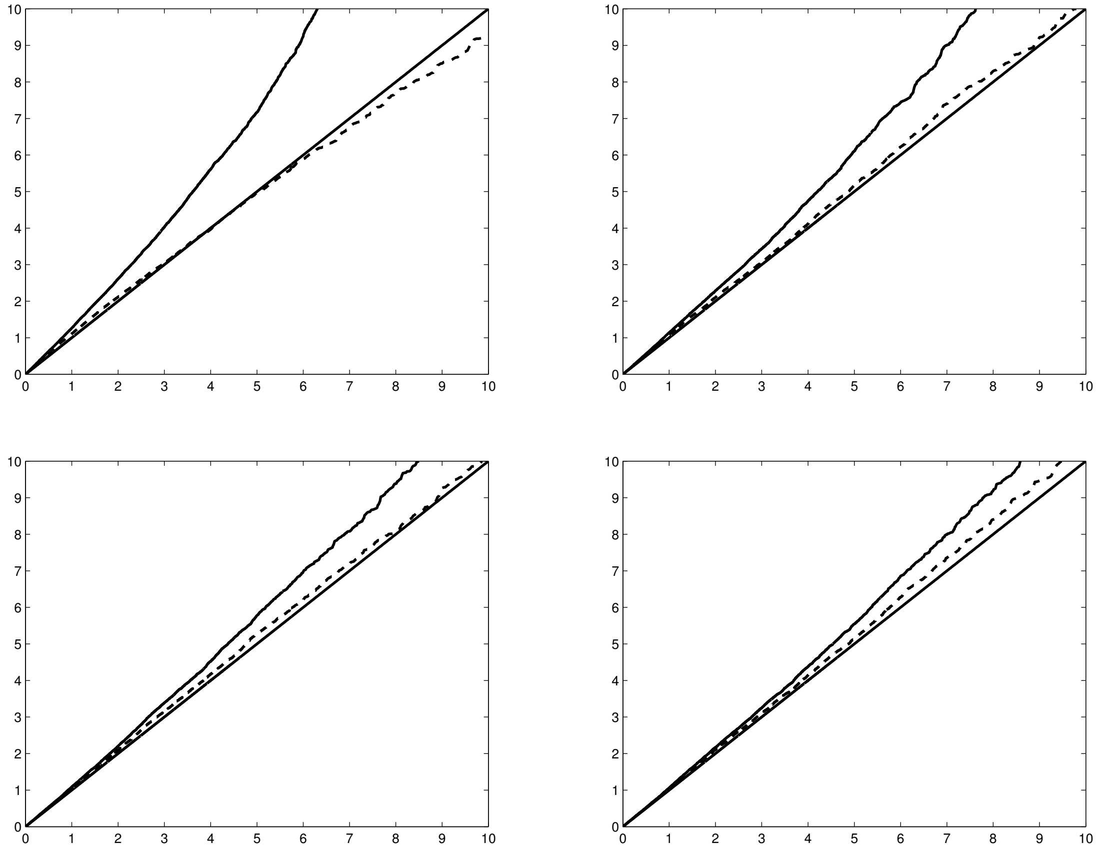

FIGURE 2.1. Empirical quantiles of the exponential tilting statistic (solid line) and TET statistic (dashed line) against quantiles of their asymptotic distribution. In the top panels, from the left to right, the sample sizes are $n = 20$ and 40. In the bottom panels, from the left to right, the sample sizes are $n = 60$ and 80.

uniformly over $t \in (0, \varepsilon)$ *for some* $\varepsilon > 0$*, where* $\xi(t) = \left(\sqrt{t} - \frac{\log G(\sqrt{t})}{n\sqrt{t}}\right)^2$ *and* $G(\cdot)$ *is defined in (B.3) in the Appendix.*

Assumption 2 is on the saddlepoint approximation of the density of $E[g(X, 0)g(X, 0)]^{1/2}\hat{\lambda}$. This assumption is satisfied under mild conditions; see, e.g., Field (1982), Skovgaard (1990), Jensen and Wood (1998), and Almudevar, Field and Robinson (2000). We note that the saddlepoint approximation error in (A.1) is of relative order $O(n^{-1})$.

Theorem 2 provides an accurate approximation formula for the tail area probability of T_n. This approximation holds not only for the normal region (i.e., $\sqrt{n}t$ is bounded) but also for the large deviation region (i.e., t is bounded). This theorem shows that the ideal statistic T_n admits relative error of order $O(n^{-1})$ up to the large deviation region. Note that the relative error would provide more meaningful measure for quality of tail area approximation compared to the absolute one. Robinson, Ronchetti and Young (2003) established an analogous desirable relative error property for their saddlepoint statistic, which is also infeasible in the present setup.

It does not seem to be possible to obtain such accurate approximation by other statistics. For example, one may consider a quadratic form of the sample average $\frac{1}{n}\sum_{i=1}^{n} g(X_i, 0)$ as a test statistic. However, as shown in Osipov (1981) and argued in Kolassa and Robinson (2011, p. 3358), the relative error of the quadratic form statistic is at best of order $n^{-1/4}$ in the large deviation region. Therefore, we treat T_n as the ideal statistic and focus on approximating its p-value by using the TET statistic.

The adjustment by the transform $\xi(t)$ to achieve relative error refinement is analogous to the one in Kolassa and Robinson (2011) for the (parametric) likelihood ratio statistic. In general, the function $G(\cdot)$ requires numerical integration over a sphere of dimension p, but a simple Monte Carlo approximation to any degree of accuracy required can be readily obtained; see Kolassa and Robinson (2011) for a detail.

Motivated by the desirable relative error property of T_n, we now argue that the TET statistic T_n^{tet} provides an accurate approximation to the tail area probabilities of the ideal statistic T_n. To this end, we introduce the function

$$K^{\text{tet}}(\lambda) = \log\left(\sum_{i=1}^{n} \hat{\pi}_i^{\text{o}} e^{\lambda' g(x_i^{\text{o}},0)}\right),$$

where $\{x_i^{\text{o}}\}_{i=1}^n$ and $\{\hat{\pi}_i^{\text{o}}\}_{i=1}^n$ are the observed values of $\{X_i\}_{i=1}^n$ and $\{\hat{\pi}_i\}_{i=1}^n$, respectively. We can see that the observed values of T_n and T_n^{tet} are given by $t_n = 2K(\hat{\lambda}^{\text{o}})$ and $t_n^{\text{tet}} = 2K^{\text{tet}}(\hat{\lambda}^{\text{o}})$, respectively, where $\hat{\lambda}^{\text{o}}$ is the observed value of $\hat{\lambda}$. Taylor expansions of $K(\lambda)$ and $K^{\text{tet}}(\lambda)$ around $\lambda = 0$ yield

$$\begin{aligned} K(\lambda) &= \frac{1}{2}\lambda' E[g(X,0)^2]\lambda + O(|\lambda|^3), \\ K^{\text{tet}}(\lambda) &= \frac{1}{2}\lambda'\left(\sum_{i=1}^{n} \hat{\pi}_i^{\text{o}} g(x_i^{\text{o}},0)^2\right)\lambda + O(|\lambda|^3). \end{aligned} \tag{2.1}$$

These expansions highlight some interesting analogies between T_n and T_n^{tet}. By the argument in the proof of Theorem 1, the sample counterpart of the difference $K(\lambda) - K^{\text{tet}}(\lambda)$ is of order $O_p(n^{-1/2}|\lambda|^2)$. On the other hand, if we consider an analogous function $K^{\text{et}}(\lambda) = -\log\left(\frac{1}{n}\sum_{i=1}^n e^{\lambda' g(x_i^{\text{o}},0)}\right)$ for the exponential tilting statistic so that $t_n^{\text{et}} = 2K^{\text{et}}(\hat{\lambda}^{\text{o}})$, then an expansion yields

$$K^{\text{et}}(\lambda) = -\lambda'\bar{g}^{\text{o}} - \frac{1}{2}\lambda'\hat{V}^{\text{o}}\lambda + O(|\lambda|^3), \tag{2.2}$$

where $\bar{g}^{\text{o}} = \frac{1}{n}\sum_{i=1}^n g(x_i^{\text{o}},0)$ and $\hat{V}^{\text{o}} = \frac{1}{n}\sum_{i=1}^n (g(x_i^{\text{o}},0) - \bar{g}^{\text{o}})(g(x_i^{\text{o}},0) - \bar{g}^{\text{o}})'$. In this case, the sample counterpart of the difference $K(\lambda) - K^{\text{et}}(\lambda)$ is of order $O_p(\max\{n^{-1/2}|\lambda|, |\lambda|^2\})$.

The following theorem shows that $K^{\text{tet}}(\lambda)$ can provide an accurate approximation to the tail area probabilities of T_n.

Theorem 3. *Suppose that Assumptions 1 and 2 in the Appendix hold true. Then under $H_0 : \theta_0 = 0$,*

$$1 - F_p(n\xi(K(\lambda))) = \{1 - F_p(nK^{\text{tet}}(\lambda))\}(1 + r_n),$$

where the sample counterpart of r_n is of order $O_p(n^{1/2}|\lambda|^2)$.

Theorem 3 shows that the TET statistic can provide an accurate approximation to the tail area probability formula $1 - F_p(n\xi(K(\lambda)))$ for the ideal statistic T_n. The error of this approximation is relative and of order $n^{1/2}|\lambda|^2$. Therefore, in the normal region for $\lambda = O(n^{-1/2})$, the relative error is of order $O(n^{-1/2})$. Beyond the normal region, e.g., $\lambda = O(n^{-1/3})$, the relative error approximation is of order $O(n^{-1/6})$. On the other hand, it is clear from (2.2) that the function $K^{\text{et}}(\lambda)$ for exponential tilting does not have such a relative error property.

3. General case

In this section, we generalize the theoretical results obtained in the last section. Sections 3.1 and 3.2 consider testing for composite hypotheses and overidentifying restrictions, respectively. Section 3.3 provides some simulation evidence.

3.1. **Composite hypothesis test for just-identified model.** In this subsection, we extend the results for the benchmark case to composite hypothesis testing for just-identified moment conditions. Let $\theta_0 = (\theta_{10}', \theta_{20}')'$. Suppose we wish to test the null hypothesis $H_0 : \theta_{20} = 0$ against the two-sided alternative $H_1 : \theta_{20} \neq 0$. In this case, the conventional exponential tilting statistic may be written as

$$T_{n,c}^{\text{et}} = -2 \max_{\theta_1 \in \Theta_1} \log \left(\frac{1}{n} \sum_{i=1}^{n} e^{\hat{\lambda}(\theta_1)' g(X_i, \theta_1, 0)} \right),$$

where $\hat{\lambda}(\theta_1)$ solves $\sum_{i=1}^n e^{\hat{\lambda}(\theta_1)' g(X_i,\theta_1,0)} g(X_i, \theta_1, 0) = 0$ for each θ_1. It is known that $nT_{n,c}^{\text{et}}$ converges in distribution to the chi-squared distribution with q degrees of freedom under $H_0 : \theta_{20} = 0$, where q is the dimension of θ_{20}. Let $\tilde{\theta}_1$ be the solution of the above constrained maximization for θ_1 and $\tilde{\theta} = (\tilde{\theta}_1', 0')'$. The TET statistic for the composite hypothesis is constructed as

$$T_{n,c}^{\text{tet}} = 2 \log \left(\sum_{i=1}^{n} \tilde{\pi}_i e^{\tilde{\lambda}' g(X_i, \tilde{\theta})} \right) = 2 \left[\log \left(\sum_{i=1}^{n} e^{2\tilde{\lambda}' g(X_i, \tilde{\theta})} \right) - \log \left(\sum_{i=1}^{n} e^{\tilde{\lambda}' g(X_i, \tilde{\theta})} \right) \right],$$

where $\tilde{\pi}_i = \frac{e^{\tilde{\lambda}' g(X_i, \tilde{\theta})}}{\sum_{j=1}^n e^{\tilde{\lambda}' g(X_j, \tilde{\theta})}}$ and $\tilde{\lambda}$ solves $\sum_{i=1}^n e^{\tilde{\lambda}' g(X_i, \tilde{\theta})} g(X_i, \tilde{\theta}) = 0$. Similar to the last section, we consider the ideal but infeasible statistic

$$T_{n,c} = 2K(\tilde{\lambda}, \tilde{\theta}_1(\tilde{\lambda})),$$

where $K(\lambda, \theta_1) = \log E[e^{\lambda' g(X, \theta_1, 0)}]$ and $\tilde{\theta}_1(\lambda)$ solves $E\left[e^{\lambda' g(X, \tilde{\theta}_1(\lambda), 0)} \lambda' \left(\frac{\partial g(X, \tilde{\theta}_1(\lambda), 0)}{\partial \theta_1'} \right) \right] = 0$ for each λ. To analyze the relation between the $T_{n,c}^{\text{tet}}$ and $T_{n,c}$ consider the function $K^{\text{tet}}(\lambda, \theta_1) = \log \left(\sum_{i=1}^n \tilde{\pi}_i^{\text{o}} e^{\lambda' g(x_i^{\text{o}}, \theta_1, 0)} \right)$, where $\{\tilde{\pi}_i^{\text{o}}\}_{i=1}^n$ are the observed values of $\{\tilde{\pi}_i\}_{i=1}^n$. Note that the observed value of $T_{n,c}^{\text{tet}}$ is given by $t_{n,c}^{\text{tet}} = 2K^{\text{tet}}(\tilde{\lambda}^{\text{o}}, \tilde{\theta}_1^{\text{tet}}(\tilde{\lambda}^{\text{o}}))$, where $\tilde{\lambda}^{\text{o}}$ is the observed value of $\tilde{\lambda}$ and $\tilde{\theta}_1^{\text{tet}}(\lambda)$ solves $\frac{1}{n} \sum_{i=1}^n e^{\lambda' g(x_i^{\text{o}}, \tilde{\theta}_1^{\text{tet}}(\lambda), 0)} \lambda' \left(\frac{\partial g(x_i^{\text{o}}, \tilde{\theta}_1^{\text{tet}}(\lambda), 0)}{\partial \theta_1'} \right) = 0$ for each λ. To analyze the properties of the TET statistic $T_{n,c}^{\text{tet}}$, we modify Assumption 1 as follows.

Assumption 1' 1. *$\{X_i\}_{i=1}^n$ is i.i.d., $\theta_{10} \in \text{int}\Theta_1$ is the unique solution of $E[g(X, \theta_{10}, \theta_{20})] = 0$, Θ_1 is compact, $g(x, \theta_1, \theta_{20})$ is continuous at each $\theta_1 \in \Theta_1$ and is continuously differentiable in a neighborhood $\mathcal{N}$ of θ_{10} for almost every x, $E[\sup_{\theta_1 \in \Theta_1} |g(X, \theta_1, \theta_{20})|^\zeta] < \infty$ for some $\zeta > 2$, $E[\sup_{\theta_1 \in \mathcal{N}} |\partial g(X, \theta_1, \theta_{20})/\partial \theta_1'|] < \infty$, $E[\partial g(X, \theta_{10}, \theta_{20})/\partial \theta_1']$ is full column rank, and $E[g(X, \theta_0) g(X, \theta_0)']$ is nonsingular.*

The relative error properties of $T_{n,c}$ and $T_{n,c}^{\text{tet}}$ are presented as follows.

Theorem 4. *Suppose that Assumption 1' holds true and that Assumption 2 is satisfied with $\tilde{g}(X,\theta_0)$ in (B.6) instead of $g(X,0)$. Then under $H_0 : \theta_{20} = 0$,*

(i): *the ideal statistic $T_{n,c}$ satisfies*

$$\Pr\{nT_{n,c} \geq nt : F\} = \{1 - F_q(n\xi_c(t))\}(1 + O(n^{-1})),$$

uniformly over $t \in (0,\varepsilon)$ for some $\varepsilon > 0$, where $\xi_c(t)$ is defined as in $\xi(t)$ (by replacing $g(X,0)$ with $\tilde{g}(X,\theta_0)$),

(ii): *$K^{\text{tet}}(\lambda,\theta_1)$ satisfies*

$$1 - F_q(n\xi_c(K(\lambda,\tilde{\theta}_1(\lambda)))) = \{1 - F_q(nK^{\text{tet}}(\lambda,\tilde{\theta}_1^{\text{tet}}(\lambda)))\}(1 + r_{n,c}),$$

where the sample counterpart of $r_{n,c}$ is of order $O_p(n^{1/2}|\lambda|^2)$.

Theorem 4 (i) highlights the desirable relative error property of the ideal statistic $T_{n,c}$. Theorem 4 (ii) shows that the TET statistic can provide very accurate approximations of the tail area probabilities of the ideal statistic $T_{n,c}$.

We close this section by a comparison with the saddlepoint statistic introduced in Ma and Ronchetti (2011). In this case, their statistic is written as

$$2\left[\log\left(\sum_{i=1}^{n} e^{\tilde{\lambda}' g(X_i,\tilde{\theta})}\right) - \log\left(\sum_{i=1}^{n} e^{\tilde{\lambda}' g(X_i,\tilde{\theta}) + \hat{\mu}' g(X_i,\bar{\theta}_1,\hat{\theta}_2)}\right)\right],$$

where $\hat{\theta} = (\hat{\theta}_1', \hat{\theta}_2')'$ solves $\sum_{i=1}^{n} g(X_i,\hat{\theta}) = 0$, $\hat{\mu}$ and $\bar{\theta}_1$ solve $\sum_{i=1}^{n} \tilde{\pi}_i e^{\hat{\mu}' g(X_i,\bar{\theta}_1,0)} g(X_i,\bar{\theta}_1,\hat{\theta}_2) = 0$ and $\hat{\mu}' \sum_{i=1}^{n} \tilde{\pi}_i e^{\hat{\mu}' g(X_i,\hat{\theta})} \partial g(X_i,\bar{\theta}_1,\hat{\theta}_2)/\partial\theta_1' = 0$. Note that this saddlepoint statistic requires to solve several equations to obtain $\hat{\theta}$, $\tilde{\theta}$, $\bar{\theta}_1$, $\tilde{\lambda}$, and $\hat{\mu}$. In contrast, the TET statistic $nT_{n,c}^{\text{tet}}$ only requires to solve for $\tilde{\theta}$ and $\tilde{\lambda}$.

3.2. **Overidentifying restriction test.** In this subsection, we consider the case of overidentifying moment restrictions $E[g(X,\theta_0)] = 0$, where the dimension d of the moment functions g is larger than the dimension p of the unknown parameters θ_0. In particular, we focus on testing overidentifying restrictions, i.e., $H_0 : E[g(X,\theta)] = 0$ for some θ against $H_1 : E[g(X,\theta)] \neq 0$ for any θ. This is a specification testing problem for the model specified by moment restrictions. In this case, the conventional exponential tilting statistic may be written as

$$T_{n,v}^{\text{et}} = -2\max_{\theta} \log\left(\frac{1}{n}\sum_{i=1}^{n} e^{\hat{\lambda}(\theta)' g(X_i,\theta)}\right),$$

where $\hat{\lambda}(\theta)$ is defined in the last subsection. Based on Newey and Smith (2004), we can show that $nT_{n,v}^{\text{et}}$ converges in distribution to the chi-squared distribution with $d-p$ degrees of freedom under the null hypothesis. Let $\bar{\theta}$ be the maximizer of the above optimization problem. The TET statistic for the overidentifying restriction test is constructed as

$$T_{n,v}^{\text{tet}} = 2\log\left(\sum_{i=1}^{n} \bar{\pi}_i e^{\bar{\lambda}' g(X_i,\bar{\theta})}\right) = 2\left[\log\left(\sum_{i=1}^{n} e^{2\bar{\lambda}' g(X_i,\bar{\theta})}\right) - \log\left(\sum_{i=1}^{n} e^{\bar{\lambda}' g(X_i,\bar{\theta})}\right)\right],$$

where $\bar{\pi}_i = \frac{e^{\bar{\lambda}' g(X_i,\bar{\theta})}}{\sum_{j=1}^n e^{\bar{\lambda}' g(X_j,\bar{\theta})}}$ and $\bar{\lambda}$ solves $\sum_{i=1}^n e^{\bar{\lambda}' g(X_i,\bar{\theta})} g(X_i,\bar{\theta}) = 0$. In this case, the ideal but infeasible statistic is defined as

$$T_{n,v} = 2K(\bar{\lambda}, \bar{\theta}(\lambda)).$$

where $\bar{\theta}(\lambda)$ solves $E\left[e^{\lambda' g(X,\bar{\theta}(\lambda))} \lambda' \left(\frac{\partial g(X,\bar{\theta}(\lambda))}{\partial \theta'}\right)\right] = 0$. Next, consider the function $K^{\text{tet}}(\lambda,\theta) = \log\left(\sum_{i=1}^n \bar{\pi}_i^{\text{o}} e^{\lambda' g(x_i^{\text{o}},\theta)}\right)$, where $\{\bar{\pi}_i^{\text{o}}\}_{i=1}^n$ are the observed values of $\{\bar{\pi}_i\}_{i=1}^n$. Note that the observed value of $T_{n,v}^{\text{tet}}$ is given by $t_{n,v}^{\text{tet}} = 2K^{\text{tet}}(\bar{\lambda}^{\text{o}}, \bar{\theta}^{\text{tet}}(\bar{\lambda}^{\text{o}}))$, where $\bar{\lambda}^{\text{o}}$ is the observed value of $\bar{\lambda}$ and $\bar{\theta}^{\text{tet}}(\lambda)$ solves $\frac{1}{n}\sum_{i=1}^n e^{\lambda' g(x_i^{\text{o}},\bar{\theta}^{\text{tet}}(\lambda))} \lambda' \left(\frac{\partial g(x_i^{\text{o}},\bar{\theta}^{\text{tet}}(\lambda))}{\partial \theta'}\right) = 0$ for each λ. To analyze the properties of the TET statistic $T_{n,v}^{\text{tet}}$, we modify Assumption 1 as follows.

Assumption 1" 1. *$\{X_i\}_{i=1}^n$ is i.i.d., $\theta_0 \in \text{int}\Theta$ is the unique solution of $E[g(X,\theta_0)] = 0$, Θ is compact, $g(x,\theta)$ is continuous at each $\theta \in \Theta$ and is continuously differentiable in a neighborhood $\mathcal{N}$ of θ_0 for almost every x, $E[\sup_{\theta\in\Theta} |g(X,\theta)|^{\zeta}] < \infty$ for some $\zeta > 2$, $E[\sup_{\theta\in\mathcal{N}} |\partial g(X,\theta)/\partial\theta'|] < \infty$, $E[\partial g(X,\theta_0)/\partial\theta']$ is full column rank, and $E[g(X,\theta_0)g(X,\theta_0)']$ is nonsingular.*

The relative error properties of $T_{n,v}$ and $T_{n,v}^{\text{tet}}$ are presented as follows.

Theorem 5. *Suppose that Assumption 1" holds true and the adapted version of Assumption 2 is satisfied. Then under $H_0 : E[g(X,\theta_0)] = 0$,*

(i): *the ideal statistic $T_{n,v}$ satisfies*

$$\Pr\{nT_{n,v} \geq nt : F\} = \{1 - F_{d-p}(n\xi_v(t))\}(1 + O(n^{-1})),$$

uniformly over $t \in (0,\varepsilon)$ for some $\varepsilon > 0$, where $\xi_v(t)$ is defined as in $\xi(t)$,

(ii): *$K^{\text{tet}}(\lambda,\theta)$ satisfies*

$$1 - F_{d-p}(n\xi_v(K(\lambda,\bar{\theta}(\lambda)))) = \{1 - F_{d-p}(nK^{\text{tet}}(\lambda,\bar{\theta}^{\text{tet}}(\lambda)))\}(1 + r_{n,v}),$$

where the sample counterpart of $r_{n,v}$ is of order $O_p(n^{1/2}|\lambda|^2)$.

Theorem 5 shows that the desirable relative error properties of the TET statistic also hold true in overidentified moment condition models.

3.3. **Simulation for general case.** In this subsection, we evaluate the finite sample performance of the TET statistic for the overidentifying restrictions proposed in Section 3.2. We generate random samples $\{W_i\}_{i=1}^n = \{Y_i, X_i, Z_i'\}_{i=1}^n$ of sizes $n = 30$, 60, 90, and 120 according to

$$\begin{aligned} Y_i &= X_i\theta_0 + U_i, \\ X_i &= Z_i'\pi_0 + V_i, \end{aligned}$$

where $Z_i = \begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$, $\begin{pmatrix} U_i \\ V_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}\right)$, and $\pi_0 = (0.8, 0.6)'$. The true parameter value is set as $\theta_0 = 1$. We are interested in testing the overidentifying restrictions $H_0 : E[Z(Y - X\theta)] = 0$ for some θ against $H_1 : E[Z(Y - X\theta)] \neq 0$ for any θ. To this end, we consider the exponential tilting statistic $T_{n,v}^{\text{et}}$ and TET statistic $T_{n,v}^{\text{tet}}$. Both statistics converge in distribution to the χ_1^2 distribution under H_0. Figure 3.1 reports the q-q plots of the empirical quantiles of these test statistics against those of the χ_1^2 distribution. The number of Monte Carlo replications is $20,000$.

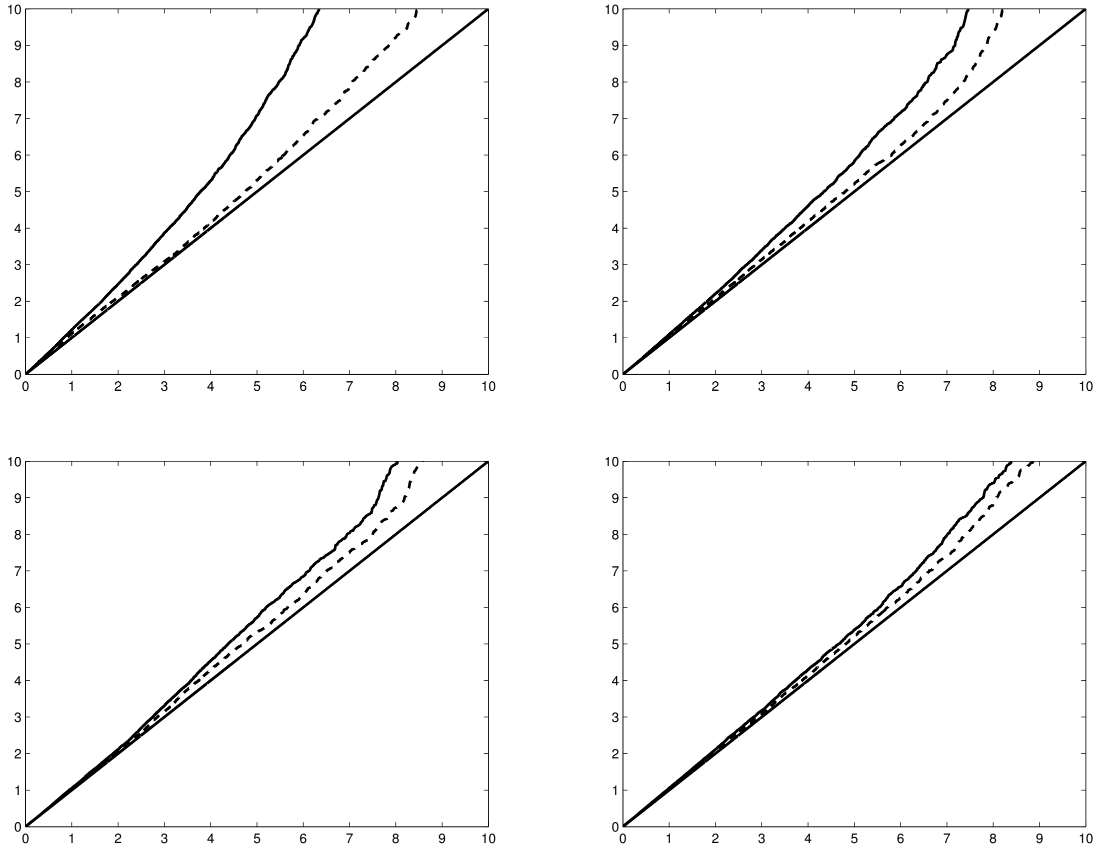

FIGURE 3.1. Empirical quantiles of the exponential tilting statistic (solid line) and TET statistic (dashed line) against quantiles of their asymptotic distribution. In the top panels, from the left to right, the sample sizes are $n = 30$ and 60. In the bottom panels, from the left to right, the sample sizes are $n = 90$ and 120.

The empirical quantiles of the TET statistic are extremely close to those of the limiting distribution. The accuracy of the exponential tilting statistic increases as the sample size increases. However, the TET always outperforms the exponential tilting.

Appendix A. Notation and assumptions

Let $\breve{g}(X) = E[g(X,0)g(X,0)']^{-1/2}g(X,0)$ and $\breve{\lambda} = E[g(X,0)g(X,0)']^{1/2}\hat{\lambda}$. To define the saddlepoint approximation for the density of $\breve{\lambda}$, we introduce the following notation:

$$\begin{aligned}
\psi(x,y) &= -e^{y'\breve{g}(x)}\breve{g}(x), \\
K(t,y) &= \log E[e^{t'\psi(X,y)}], \\
t(y) &= \text{solution of } \frac{\partial K(t(y),y)}{\partial t} = 0, \\
h(y) &= K(t(y),y), \\
B(y) &= e^{nK(t(y),y)} E\left[e^{t(y)'\psi(X,y)}\frac{\partial \psi(X,y)}{\partial y}\right], \\
\Sigma(y) &= e^{nK(t(y),y)} E[e^{t(y)'\psi(X,y)}\psi(X,y)\psi(X,y)'].
\end{aligned}$$

Let $\det A$ be the determinant of a matrix A. For Theorem 2, we impose the following assumption.

Assumption 2. *The density $f_{\breve{\lambda}}$ of $\breve{\lambda}$ exists and has the saddlepoint approximation*

$$f_{\breve{\lambda}}(y) = \left(\frac{n}{2\pi}\right)^{p/2} e^{-nh(y)} \frac{\det B(y)}{\sqrt{\det \Sigma(y)}}(1+O(n^{-1})). \tag{A.1}$$

Appendix B. Proofs

B.1. **Proof of Theorem 1.** Let $g_i = g(X_i,0)$. Using $\sum_{i=1}^n \hat{\pi}_i = 1$ and $\sum_{i=1}^n \hat{\pi}_i g_i = 0$, an expansion around $\hat{\lambda} = 0$ implies

$$nT_n^{\text{tet}} = n\hat{\lambda}'\left[\sum_{i=1}^n \hat{\pi}_i e^{\bar{\lambda}'g_i} g_i g_i'\right]\hat{\lambda},$$

where $\bar{\lambda}$ is a point on the line joining $\hat{\lambda}$ and 0. Let $\hat{M} = -\frac{1}{n}\sum_{j=1}^n e^{\hat{\lambda}'g_j}$. An expansion around $\hat{M} = -1$ implies

$$\begin{aligned}
nT_n^{\text{tet}} &= -\hat{M}^{-1}n\hat{\lambda}'\left[\frac{1}{n}\sum_{i=1}^n e^{(\hat{\lambda}+\bar{\lambda})'g_i}g_ig_i'\right]\hat{\lambda} \\
&= n\hat{\lambda}'\left[\frac{1}{n}\sum_{i=1}^n e^{(\hat{\lambda}+\bar{\lambda})'g_i}g_ig_i'\right]\hat{\lambda} + \bar{M}^{-2}n\hat{\lambda}'\left[\frac{1}{n}\sum_{i=1}^n e^{(\hat{\lambda}+\bar{\lambda})'g_i}g_ig_i'\right]\hat{\lambda}(\hat{M}+1) \\
&= T_1 + T_2,
\end{aligned}$$

where $\bar{M}$ is a point on the line joining $\hat{M}$ and -1. By applying the argument in Newey and Smith (2004, pp. 239-240), we can show $\max_{1\le i\le n}|-e^{\hat{\lambda}'g_i}+1| \xrightarrow{p} 0$ and $\max_{1\le i\le n}|e^{\bar{\lambda}'g_i}+1| \xrightarrow{p} 0$. An expansion of $\sum_{i=1}^n e^{\hat{\lambda}'g_i}g_i = 0$ around $\hat{\lambda} = 0$ implies

$$\hat{\lambda} = -\left(\frac{1}{n}\sum_{i=1}^n g_ig_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n g_i\right) + o_p(n^{-1/2}).$$

Combining these results,

$$T_1 = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n g_i\right)'\left(\frac{1}{n}\sum_{i=1}^n g_ig_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n g_i\right) + o_p(1) \xrightarrow{d} \chi_p^2.$$

Finally, by $\max_{1\le i\le n}|-e^{\hat{\lambda}'g_i}+1|\xrightarrow{p}0$, it holds $\hat{M}+1\xrightarrow{p}0$ and then $T_2\xrightarrow{p}0$. Therefore, the conclusion follows.

B.2. **Proof of Theorem 2.** The basic idea of the proof is similar to that of Robinson, Ronchetti and Young (2003, Theorem 1). Let $m(\lambda)=\log E[e^{\lambda'\breve{g}(X)}]$ so that $T_n=2m(\check{\lambda})$. By Assumption 2, the tail probability of T_n is approximated as

$$
\begin{aligned}
&\Pr\{nT_n\ge nt:F\}=\Pr\{2m(\check{\lambda})\ge t:F\}\\
=\;&\int_{\{y:2m(y)\ge t\}}\left(\frac{n}{2\pi}\right)^{p/2}e^{-nh(y)}\frac{\det B(y)}{\sqrt{\det\Sigma(y)}}dy(1+O(n^{-1}))\\
\equiv\;&A(1+O(n^{-1})). \qquad \text{(B.1)}
\end{aligned}
$$

To evaluate the integral A, consider the polar transformation $y\mapsto(r,s)$ (with radius r and angle s) and another transformation $(r,s)\mapsto(u,s)$ with $u=\sqrt{2m(y)}$. The Jacobians of these transformations are $J_1(y)=(y'y)^{\frac{p-1}{2}}$ and $J_2(y)=\frac{\sqrt{y'y}\sqrt{2m(y)}}{m_1(y)'y}$, respectively, where $m_1(y)=dm(y)/dy$. Define the transform $y\mapsto(u,s)$ as $y=\varphi(u,s)$. By the change of variables, the above integral is written as

$$
A=\int_{\sqrt{t}}^{\infty}c_nu^{p-1}e^{-nu^2/2}\left\{\int_S\delta(u,s)ds\right\}du, \qquad \text{(B.2)}
$$

where $c_n=n^{p/2}/(2^{p/2-1}\Gamma(p/2))$, S is the p-dimensional unit sphere, and

$$
\delta(u,s)=\frac{e^{nu^2/2-nh(\varphi(u,s))}\Gamma(p/2)}{2\pi^{p/2}u^{p-1}}\frac{\det B(\varphi(u,s))}{\sqrt{\det\Sigma(\varphi(u,s))}}J_1(\varphi(u,s))J_2(\varphi(u,s)).
$$

We expand each term in $\delta(u,s)$. First, note that

$$
\begin{aligned}
\det B(\varphi(u,s)) &= \det B(0)\{1+r\xi_1(s)+r^2R_1(r,s)\},\\
\frac{1}{\sqrt{\det\Sigma(\varphi(u,s))}} &= \frac{1}{\sqrt{\det\Sigma(0)}}\{1+r\xi_2(s)+r^2R_2(r,s)\},
\end{aligned}
$$

where ξ_1 and ξ_2 are linear combinations of components of s, and R_1 and R_2 are uniformly bounded for r bounded. Due to the normalization $E[\breve{g}(X)\breve{g}(X)']=I$, we have $\frac{\det B(0)}{\sqrt{\det\Sigma(0)}}=1$. Thus, other terms are expanded as

$$
\begin{aligned}
e^{nu^2/2-nh(\varphi(u,s))} &= 1+r^2R_3(r,s),\\
J_1(y) &= r^{p-1},\\
J_2(y) &= 1+r\xi_4(s)+r^2R_4(r,s),\\
u &= r\{1+r\xi_5(s)+r^2R_5(r,s)\},
\end{aligned}
$$

where ξ_4 and ξ_5 are linear combinations of terms of the form $s_is_js_k$, and R_3, R_4 and R_5 are uniformly bounded for r bounded. Combining all these expansions,

$$
\delta(u,s)=\frac{\Gamma(p/2)}{2\pi^{p/2}}\{1+ub(s)+u^2R_6(u,s)\},
$$

where R_6 is uniformly bounded for r bounded, and $b(s)$ is a linear combination of odd functions satisfying $\int_S b(s)ds = 0$. Thus, by taking integral,

$$G(u) \equiv \int_S \delta(u,s)ds = 1 + u^2 k(u), \tag{B.3}$$

for some $k(u)$ bounded over $u \in (0,\varepsilon)$. Also we can see that $dG(u)/du = uk_1(u)$ for some $k_1(u)$ bounded over $u \in (0,\varepsilon)$.

From (B.1)-(B.3),

$$\begin{aligned}
\Pr\{nT_n \geq nt : F\} &= \int_{\sqrt{t}}^{\infty} c_n u^{p-1} e^{-nu^2/2} G(u) du (1 + O(n^{-1})) \\
&= \int_{\sqrt{t}}^{\infty} c_n u^{p-1} e^{-n(u - \log G(u)/(nu))^2/2} du (1 + O(n^{-1})),
\end{aligned}$$

where the second equality follows from boundedness of $k(u)$ and $k_1(u)$. The conclusion follows by the change of variables $v = u - \log G(u)/(nu)$ and boundedness of $k(u)$ and $k_1(u)$.

B.3. **Proof of Theorem 3.** Using (B.1)-(B.3) in the proof of Theorem 2 and integration by parts, we have

$$1 - F_p(n\xi(K(\lambda))) = \{1 - F_p(nK(\lambda))\}(1 + O(n^{-1})) + \frac{c_n}{n} K(\lambda)^{\frac{p}{2}} e^{-\frac{nK(\lambda)}{2}} \left[\frac{G(\sqrt{K(\lambda)}) - 1}{K(\lambda)}\right]. \tag{B.4}$$

For a random variable χ_p^2 following the chi-squared distribution with p degrees of freedom, we have

$$\Pr\{\chi_p^2 \geq u\} \geq Ce^{-u/2} u^{p/2-1}, \tag{B.5}$$

for some constant $C > 0$. By the mean-value theorem,

$$F_p(nK(\lambda)) = F_p(nK^{\text{tet}}(\lambda)) + \frac{e^{-\bar{u}/2} \bar{u}^{p/2-1}}{2^{p/2}\Gamma(p/2)} (nK^{\text{tet}}(\lambda) - nK(\lambda)),$$

for some $\bar{u}$ between $nK(\lambda)$ and $nK^{\text{tet}}(\lambda)$. Also note that $nK^{\text{tet}}(\lambda) - nK(\lambda) = r_n$, where the sample counterpart of r_n is of order $O_p(n^{1/2}|\lambda|^2)$. Combining these results,

$$\frac{1 - F_p(nK(\lambda))}{1 - F_p(nK^{\text{tet}}(\lambda))} = 1 + r_n',$$

where the sample counterpart of r_n' is of order $O_p(n^{1/2}|\lambda|^2)$.

Finally, consider the second term on the right hand side of (B.4). By the definition of $G(u) = 1 + u^2 k(u)$ for some $k(u)$ bounded over $u \in (0,\varepsilon)$, the sample counterpart of $\left[\frac{G(\sqrt{K(\lambda)})-1}{K(\lambda)}\right]$ is of order $O_p(1)$. Also, by (B.5), we have

$$1 - F_p(nK^{\text{tet}}(\lambda)) \geq Ce^{-nK^{\text{tet}}(\lambda)/2} (nK^{\text{tet}}(\lambda))^{p/2-1},$$

for some $C > 0$. Therefore, by the definition of $c_n = n^{p/2}/(2^{p/2-1}\Gamma(p/2))$ and expansions in (2.1), we have

$$\frac{c_n}{n} K(\lambda)^{\frac{p}{2}} e^{-\frac{nK(\lambda)}{2}} \left[\frac{G(\sqrt{K(\lambda)}) - 1}{K(\lambda)}\right] (1 - F_p(nK^{\text{tet}}(\lambda)))^{-1} = r_n''.$$

where the sample counterpart of r''_n is of order $O_p(|\lambda|^2)$. This concludes the proof of Theorem 3.

B.4. **Proof of Theorem 4.**

Proof of Part (i). Let $\Omega = E[g(X,\theta_0)g(X,\theta_0)']$ and $M = \Omega^{-1/2}E[\partial g(X,\theta_0)/\partial\theta_1']$. By the spectral decomposition of the idempotent matrix (Czellar and Ronchetti, 2010), there exists a matrix $C = [C_1 : C_2]$ such that

$$M(M'M)^{-1}M' = C\begin{bmatrix} I_q & 0 \\ 0 & 0_{(p-q)\times(p-q)} \end{bmatrix}C',$$

and $C'C = CC' = I_p$. Based on Newey and Smith (2004, p. 240), we can see that $\sqrt{n}\tilde{\lambda}$ is asymptotically equivalent to $\sqrt{n}\Omega^{-1/2}C_2\tilde{\gamma}$, where $\tilde{\gamma}$ solves

$$\sum_{i=1}^{n} e^{\tilde{\gamma}'\tilde{g}(X_i,\theta_0)}\tilde{g}(X_i,\theta_0) = 0,$$

where

$$\tilde{g}(X,\theta_0) = C_2'\Omega^{-1/2}g(X,\theta_0). \tag{B.6}$$

The saddlepoint density of $\tilde{\gamma}$ is given by (A.1) with replacement of $g(X_i,\theta_0)$ with $\tilde{g}(X_i,\theta_0)$. Let $\tilde{K}(\gamma) = K(\Omega^{-1/2}C_2\gamma, \tilde{\theta}_1(\Omega^{-1/2}C_2\gamma))$. We can also see that

$$\Pr\{nT_{n,c} \geq nt_{n,c} : F\} = \Pr\{2n\tilde{K}(\tilde{\gamma}) \geq n\tilde{K}(\tilde{\gamma}_o) : F\}(1 + O(e^{-n\epsilon})),$$

for any $\epsilon > 0$ small enough. Then the conclusion follows as in the proof of Theorem 2 by replacing $h^{\lambda}(y) = \log E[e^{y'g(X,0)}]$ with $\log E[e^{y'\tilde{g}(X,\theta_0)}]$.

Proof of Part (ii). Using the spectral decomposition of idempotent matrix adopted in the proof of (i), we can show that $nK^{\text{tet}}(\lambda, \tilde{\theta}_1^{\text{tet}}(\lambda)) - nK(\lambda, \tilde{\theta}_1(\lambda)) = O(n^{1/2}|\lambda|^2)$. Therefore, (ii) follows by using the same arguments adopted for the proof of Theorem 3.

B.5. **Proof of Theorem 5.** The proof is similar to that of Theorem 4.

References

[1] Almudevar, A., Field, C., and Robinson, J. (2000). The density of multivariate M-estimates. *Annals of Statistics*, 28, 275-297.

[2] Baggerly, K.A. (1998). Empirical likelihood as a goodness-of-fit measure, *Biometrika*, 85, 535-547.

[3] Brown, B.W., and Newey, W.K. (1998). Efficient semiparametric estimation of expectations, *Econometrica*, 66, 453-464.

[4] Czellar, V., and Ronchetti, E. (2010). Accurate and robust tests for indirect inference, *Biometrika*, 97, 621-630.

[5] Daniels, H.E., and Young, G.A. (1991). Saddlepoint approximation for the studentized mean, with an application to the bootstrap, *Biometrika*, 78, 169-179.

[6] Efron, B. (1981). Nonparametric standard errors and confidence intervals, *Canadian Journal of Statistics*, 9, 139-172.

[7] Efron, B. (1982). Transformation theory: How normal is a family of distributions? *Annals of Statistics*, 10, 323-339.

[8] DiCiccio, T.J., Hall, P., and Romano, J. (1991). Empirical likelihood is Bartlett-correctable, *Annals of Statistics*, 19, 1053-1061.

[9] Field, C.A. (1982). Small sample asymptotic expansions for multivariate M-estimates, *Annals of Statistics*, 10, 672-689.

[10] Imbens, G.W., Spady, R.H., and Johnson, P. (1998). Information theoretic approaches to inference in moment condition models, *Econometrica*, 66, 333-357.

[11] Jensen, J.L., and Wood, A.T.A. (1998). Large deviation and other results for minimum contrast estimators, *Annals of the institute of mathematical statistics*, 50, 673-695.

[12] Jing, B.Y., and Robinson, J. (1994). Saddlepoint approximations for marginal and conditional probabilities of transformed variables, *Annals of Statistics*, 22, 1115-1132.

[13] Jing, B.Y., and Wood, A.T.A. (1996). Exponential empirical likelihood is not Bartlett correctable, *Annals of Statistics*, 24, 365-369.

[14] Kitamura, Y., and Stutzer, M. (1997). An information-theoretic alternative to generalized method of moments estimation, *Econometrica*, 65, 861-874.

[15] Kolassa, J., and Robinson, J. (2011). Saddlepoint approximations for likelihood ratio like statistics with applications to permutation tests, *Annals of Statistics*, 39, 3357-3368.

[16] Ma, Y., and Ronchetti, E. (2011). Saddlepoint test in measurement error models, *Journal of the American Statistical Association*, 106, 147-156.

[17] Newey, W.K., and Smith, R.J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica*, 72, 219-255.

[18] Osipov, L.V. (1981). On large deviations for sums of random vectors in R^k, *Journal of Multivariate Analysis*, 11, 115-126.

[19] Owen, A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, 75, 237-249.

[20] Owen, A.B. (2001). *Empirical Likelihood*, New York, Chapman and Hall.

[21] Robinson, J., Ronchetti, E., and Young, G.A. (2003). Saddlepoint approximations and tests based on multivariate M-estimates, *Annals of Statistics*, 31, 1154-1169.

[22] Skovgaard, I.M. (1990). On the density of minimum contrast estimators, *Annals of Statistics*, 18, 779-789.

[23] Tingley, M.A., and Field, C.A. (1990). Small-sample confidence intervals. *Journal of the American Statistical Association*, 85, 427-434.

School of Economics, University of Surrey, Guildford, Surrey, GU2 7XH, UK.

E-mail address: `l.camponovo@surrey.ac.uk`

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

E-mail address: `t.otsu@lse.ac.uk`