Econometric Theory, **37**, 2021, 1214–1237.
doi:10.1017/S0266466621000074

RELATIVE ERROR ACCURATE STATISTIC BASED ON NONPARAMETRIC LIKELIHOOD*

LORENZO CAMPONOVO
University of Applied Sciences and Arts of Italian Switzerland

YUKITOSHI MATSUSHITA
Hitotsubashi University

TAISUKE OTSU
London School of Economics

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 functions 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 admits a Lugannani–Rice style adjustment with relative errors 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 overidentified moment condition models. A limitation of our analysis is that the theoretical approximation results are exclusively for the infeasible saddlepoint statistic, and closeness of the p-values for the infeasible statistic to the ones for the feasible TET statistic is only numerically assessed.

1. INTRODUCTION

This paper is concerned with inference on moment condition models and proposes a new test statistic that shows desirable tail behaviors in terms of relative errors. For this problem, there are various test statistics available in the literature, such as the Wald, empirical likelihood (Owen, 1988), exponential tilting (ET; Efron, 1981, 1982; Kitamura and Stutzer, 1997; Imbens, Spady, and Johnson, 1998), power divergence (Baggerly, 1998), and saddlepoint statistics (Robinson, Ronchetti, and Young, 2003; Ma and Ronchetti, 2011), among others. In particular, it is known that the empirical likelihood statistic admits the Bartlett correction, a higher-order

*We gratefully acknowledge support from the JSPS KAKENHI (18K01541; Matsushita) and the ERC Consolidator Grant (SNP 615882; Otsu). We would like to thank Elvezio Ronchetti and Richard Spady for their helpful discussions and seminar participants at the University of Geneva for their comments. We would also like to thank the Editor, a Co-Editor, and anonymous referees for their helpful comments. Address correspondence to Taisuke Otsu, Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, UK; e-mail: t.otsu@lse.ac.uk.

© The Author(s), 2021. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

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 ET (Jing and Wood, 1996; 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 delivers 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 et al. 2003; Kolassa and Robinson, 2011, among others). In particular, Robinson et al. (2003) considered the situation where the cumulant generating function is known to the researcher and developed a novel saddlepoint statistic that is asymptotically chi-square 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 the knowledge of the cumulant generating function, Robinson et al. (2003) and Ma and Ronchetti (2011) proposed some feasible versions of the saddlepoint statistic by using the ET weights (Efron, 1981, 1982).

The basic idea of our statistic is to note that the conventional ET statistic is constructed from estimating the cumulant generating function by the sample average, and to modify the cumulant estimation by using the ET weights instead of the uniform weights. In other words, we tilt the ET statistic. Thus, the new statistic is called the tilted exponential tilting (TET) statistic.

We show that the TET statistic is asymptotically chi-square distributed, and demonstrate that its asymptotic p-value provides an accurate approximation to the p-value of some ideal (but infeasible) saddlepoint statistic, which admits a Lugannani–Rice style adjustment with a relative error of order n^{-1} both in normal and large deviation regions. Thus, as far as the p-value for the TET statistic is well approximated by the one for the ideal statistic, we can argue that the asymptotic p-value approximation for the TET statistic using the chi-square distribution is very accurate even in the tails. Due to technical difficulty, however, this paper only establishes the theoretical approximation results for the ideal saddlepoint statistic, and admittedly, the approximation property between the ideal saddlepoint and feasible TET statistics is only numerically investigated. 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 overidentified moment condition models.

We emphasize that accuracy of p-values in tail areas for testing moment condition models is a substantial issue in applied research. In practice, researchers are concerned with accuracy of relatively small p-values (say, less than 10%) which will be typical borderlines to decide whether the null should be rejected or not. As our simulation studies below indicate, the empirical quantiles of the TET statistic are very close to the ones of the limiting chi-square distribution for the tail areas. Furthermore, we note that moment condition models are widely used in empirical economic analyses, and it is well known that the (first-order) asymptotic

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

approximations for those models may not provide accurate inference methods in small to moderate samples (see, e.g., the special issue in the *Journal of Business & Economic Statistics*, vol. 14).

Finally, we study through Monte Carlo simulations the accuracy of the proposed TET statistic. In addition to a benchmark setup for inference on means, we consider both just- and overidentified 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 small sample sizes.

This paper is organized as follows. In Section 2, we consider a benchmark case, simple parameter hypothesis testing for just-identified moment conditions, and present basic theoretical results illustrated by a small simulation study. Section 3 extends our method to testing problems for composite hypotheses (Section 3.1), and overidentifying restrictions (Section 3.2). Finally, Section 4 presents simulation results for the general case. All proofs are presented in Appendix A. Appendix B contains figures for simulation results in Section 4.

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 discuss 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. Let F be the distribution function of X and $\mathbb{E}[\cdot]$ be the expectation under F. Consider the (just-identified) moment conditions

$$\mathbb{E}[g(X, \theta_0)] = 0,$$

where g is a p-dimensional vector of moment functions and θ_0 is a p-dimensional vector of true parameter values. In this section, we focus on hypothesis testing for the null $H_0 : \theta_0 = 0$ against the two-sided alternative $H_1 : \theta_0 \neq 0$. 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 ET statistic (Efron, 1981, 1982; Kitamura and Stutzer, 1997; Imbens et al. 1998) under the ET weights

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

based on the restriction $\mathbb{E}[g(X,\theta_0)] = 0$ under the null $H_0 : \theta_0 = 0$. Let $\hat{\lambda}$ be a solution of $\sum_{i=1}^n e^{\hat{\lambda}' g(X_i,0)} g(X_i,0) = 0$. The conventional ET statistic for $H_0 : \theta_0 = 0$ is written as

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

It is known that nT_n^{et} converges in distribution to the χ_p^2 distribution under H_0. This statistic is obtained by minimization of the empirical relative entropy

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

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 $i = 1,\ldots,n$. 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 $\mathbb{E}[g(X,\theta_0)] = 0$ subject to the null $H_0 : \theta_0 = 0$ (Brown and Newey, 1998).

Intuitively, the ET statistic T_n^{et} is constructed by taking expectation of $e^{\hat{\lambda}' g(X_i,0)}$ under the empirical distribution with weights $\frac{1}{n}$. 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 $\{\hat{\pi}_i\}_{i=1}^n$, i.e.,

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

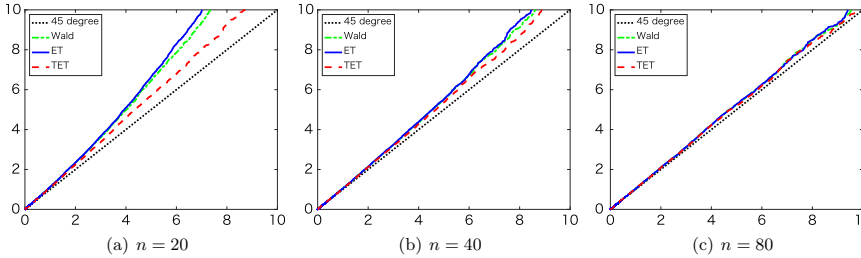
We call this statistic the TET statistic. As shown in the proof of Theorem 1, the reason for the positive sign of T_n^{tet} can be seen from a second-order expansion around $\hat{\lambda}' g(X_i,0) = 0$,

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

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 χ_p^2 distribution under H_0. To make the argument rigorous, we impose the following assumption.

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

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press



(a) $n = 20$ (b) $n = 40$ (c) $n = 80$

FIGURE 1. Empirical quantiles of the TET statistic (dashed red line), ET statistic (solid blue line), and Wald statistic (dash-dotted green line) against quantiles of the χ^2_1 distribution for (a) $n = 20$, (b) $n = 40$, and (c) $n = 80$.

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* χ^2_p *distribution.*

Therefore, under H_0, the TET statistic nT_n^{tet} is asymptotically equivalent to the ET 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 a preliminary simulation result. We generate random samples $\{X_i\}_{i=1}^n$ of sizes $n = 20, 40$, and 80 according to $X_i \sim N(0, 1)$. The parameter θ_0 is defined by the moment condition $\mathbb{E}[g(X, \theta_0)] = \mathbb{E}[X - \theta_0] = 0$. 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 proposed TET statistic nT_n^{tet} with the ET statistic nT_n^{et} and the conventional Wald statistic. All the statistics converge in distribution to the χ^2_1 distribution under H_0. Figure 1 reports the q–q plots of the empirical quantiles of these test statistics against those of the χ^2_1 distribution. The number of Monte Carlo replications is 20,000.

As this figure shows, the empirical quantiles of the TET statistic are closer to those of the limiting χ^2_1 distribution even for relatively small sample sizes. The accuracies of the ET and Wald statistics increase as the sample size increases, although the TET statistic outperforms for all cases. In the next subsection, we provide some explanation for this excellent finite sample performance of the TET statistic.

2.3. Relative Error Properties

2.3.1. *Overview.* The main purpose of this paper is to provide some theoretical justification for the desirable accuracy of the TET statistic nT_n^{tet} using χ^2 asymp-

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

totic approximation particularly in the tail area as observed in the q–q plots in the last subsection. Based on the statistics literature of saddlepoint approximations (see, e.g., Jensen, 1995, for a review), an ideal goal for this purpose is to evaluate the order (say, ϱ_n) in the relative error for the χ^2 asymptotic approximation:

$$\Pr\{nT_n^{\text{tet}} \geq nt : F\} = \{1 - F_p(nt)\}(1 + O(\varrho_n)), \tag{3}$$

and compare the approximation error $O(\varrho_n)$ with the ones for other statistics, where F_p is the cumulative distribution function of the χ_p^2 distribution. However, the existing techniques for saddlepoint approximations are mostly confined to smooth function models for finite dimensional sample means and parametric models, and it is difficult to establish the approximation result for the TET statistic as in (3) under moment condition models. To the best of our knowledge, there is no relative error approximation results for the Generalized Method of Moments (GMM) or generalized empirical likelihood-based statistic under general moment condition models.

To circumvent this technical difficulty, we consider an infeasible version of the TET statistic, called T_n in (5) below, by replacing the weighted average in (2) using $\{\hat{\pi}_i\}_{i=1}^n$ with the population mean. Although T_n is infeasible, it is possible to establish certain relative error approximation results for $\Pr\{nT_n \geq nt : F\}$ as in (3) (see, Theorem 2 below).

Based on these considerations, we assume that the p-value $p^{\text{tet}} = \Pr\{nT_n^{\text{tet}} \geq nt^{\text{tet,o}} : F\}$ for the TET statistic is well approximated by its infeasible version $p^{\text{ideal}} = \Pr\{nT_n \geq nt^{\text{o}} : F\}$, where $t^{\text{tet,o}}$ and t^{o} are observed values of T_n^{tet} and T_n, respectively. Then, under this assumption, we concentrate on the relative error properties to approximate the p-value of the proxy p^{ideal} for p^{tet}. Our main results (Theorems 2 and 3 below) establish the relation between p^{ideal} and the asymptotic p-value $p^{\text{asy}} = 1 - F_p(nt^{\text{tet,o}})$ for our TET test as

$$p^{\text{ideal}} = p^{\text{asy}}(1 + r_n(\hat{\lambda}^{\text{o}}))(1 + O(n^{-1})), \tag{4}$$

where $\hat{\lambda}^{\text{o}}$ is an observed value of $\hat{\lambda}$ and $r_n(\cdot)$ is a function defined in (19). Based on this relative error approximation, we argue that $r_n(\cdot)$ exhibits a desirable property not shared by the conventional ET test.

A major limitation of this paper is that due to technical difficulty in analyzing T_n^{tet}, we cannot provide any theory to characterize the quality of the approximation for p^{tet} by p^{ideal}. The original object of interest, the ratio of the p-value p^{tet} of the TET statistic to its asymptotic approximation p^{asy}, can be decomposed as

$$\frac{p^{\text{tet}}}{p^{\text{asy}}} = (1 + \rho_1)(1 + \rho_2),$$

where $\rho_1 = (p^{\text{tet}} - p^{\text{ideal}})/p^{\text{ideal}}$ and $\rho_2 = (p^{\text{ideal}} - p^{\text{asy}})/p^{\text{asy}}$. Our result in (4) only characterizes ρ_2. For the benchmark simulation in Section 2.2, we evaluate the magnitudes of the component ρ_1 by the ratios $\left|\frac{\rho_1}{\rho_2}\right|$ for 1,000 Monte Carlo replications, and find that their interquartile ranges are: $[0.49, 1.71]$ for $n = 20$,

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

$[0.53, 1.47]$ for $n = 40$, and $[0.57, 1.45]$ for $n = 80$. Thus, at least for the benchmark simulation setup, the approximation error ρ_1 is comparable to ρ_2, which will be characterized by our theorems. Theoretical analysis on ρ_1 is a substantial challenge and should be left for future research.

2.3.2. *Ideal Statistic T_n.* We now introduce the ideal statistic T_n. In the definition of the TET statistic (2), we propose to take expectation of $e^{\hat{\lambda}' g(X_i,0)}$ under the tilted weights $\{\hat{\pi}_i\}_{i=1}^n$ satisfying $\sum_{i=1}^n \hat{\pi}_i g(X_i,0) = 0$. The ideal (but infeasible) statistic is given by evaluating this expectation under the population:

$$T_n = 2K(\hat{\lambda}), \tag{5}$$

where $K(\lambda) = \log \mathbb{E}[e^{\lambda' g(X,0)}]$ 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 the expectation $\mathbb{E}[\cdot]$ to evaluate the cumulant. Furthermore, T_n is different from the saddlepoint statistic proposed in Robinson et al. (2003), because it does not involve any estimators of θ_0.

To study the relative error property of T_n, we introduce further notation. Let $\check{g}(x) = \mathbb{E}[g(X,0)g(X,0)']^{-1/2} g(x,0)$ be a normalized counterpart of $g(x,0)$, so that $Var(\check{g}(X)) = I$, and define $\check{\lambda} = \mathbb{E}[g(X,0)g(X,0)']^{1/2}\hat{\lambda}$. Based on the definition of $\hat{\lambda}$, $\check{\lambda}$ can be written as a solution of $\sum_{i=1}^n \psi(X_i, \check{\lambda}) = 0$, where $\psi(x,y) = -e^{y'\check{g}(x)}\check{g}(x)$. Thus, $\psi(x,y)$ may be interpreted as an estimating function for $\check{\lambda}$. Also, let $K(t,y) = \log \mathbb{E}[e^{t'\psi(X,y)}]$ and $t(y)$ be a solution of $\frac{\partial K(t(y),y)}{\partial t} = 0$. We impose the following high-level assumption.

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

$$f_{\check{\lambda}}(y) = \left(\frac{n}{2\pi}\right)^{p/2} e^{-nh(y)} \frac{\det B(y)}{\sqrt{\det \Sigma(y)}} (1 + O(n^{-1})), \tag{6}$$

where

$$h(y) = K(t(y), y),$$
$$B(y) = e^{nK(t(y),y)} \mathbb{E}\left[e^{t(y)'\psi(X,y)} \frac{\partial \psi(X,y)}{\partial y} \right],$$
$$\Sigma(y) = e^{nK(t(y),y)} \mathbb{E}[e^{t(y)'\psi(X,y)} \psi(X,y)\psi(X,y)'].$$

Assumption 2 requires existence of the usual form of the saddlepoint approximation of the density of $\check{\lambda}$. We note that the saddlepoint approximation error in (6) is typically of relative order $O(n^{-1})$. For primitive conditions under which this assumption holds, we refer to Field (1982); Skovgaard (1990); Jensen and Wood (1998); Almudevar, Field, and Robinson (2000). For example, suppose $\mathbb{E}[g(X,0)g(X,0)']$ is invertible and all components in $\psi(x,y)$ and their first four derivatives with respect to y are bounded and continuous as functions of x at each y. Then, certain smoothness conditions that enable the development of an Edgeworth

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

expansion are sufficient to guarantee Assumption 2 (see, Ma and Ronchetti, 2011, pp. 154–155).[1,2]

Let F_p be the cumulative distribution function of the χ^2_p distribution. 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 hold. Then, under* $H_0 : \theta_0 = 0$,

$$\Pr\{nT_n \geq nt : F\} = \{1 - F_p(n\xi(t))\}(1 + O(n^{-1})), \tag{7}$$

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* (17) *in Appendix A.*

Theorem 2 provides a Lugannani–Rice style accurate approximation formula for the tail area probability of T_n. This approximation holds not only for the normal region (i.e., nt is bounded) but also for the large deviation region (i.e., t is bounded). This theorem shows that the ideal statistic T_n admits a relative error of order $O(n^{-1})$ up to the large deviation region. Note that relative error orders typically deliver a more useful measure of the quality of tail area approximation compared to absolute ones. Robinson et al. (2003) established an analogous desirable relative error property for their saddlepoint statistic, which is also infeasible in the present setup. Also, Huber and Ronchetti (2009, Chap. 14.6) argued that for parametric models the three classical hypothesis tests (i.e., likelihood ratio, Wald, and score) do not satisfy the relative error property as in (7).

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

[1]More precisely, let $U^y = (U_1^{y\prime}, U_2^{y\prime}, U_2^{y\prime})$, where $U_1^y = \psi(X, y)$, U_2^y is the vector formed by the elements of $\partial\psi(X, y)/\partial y'$ and $\psi(X, y)\psi(X, y)'$, and U_3^y is the vector formed by the elements of $\partial U_2^y/\partial y'$. Then, let F_U^y be the distribution function of U^y and define a random vector $\tilde{U}^y$ with the distribution function

$$F_{\tilde{U}}^y(u_1, u_2, u_3) = \int_{(a_1,a_2,a_3)\leq(u_1,u_2,u_3)} e^{t(y)'a_1 - K(t(y),y)} dF_U^y(a_1, a_2, a_3),$$

and the characteristic function $\phi^y(\xi) = \mathbb{E}[e^{i\xi'\tilde{U}^y}]$. Based on this notation, the smoothness condition is stated as: for each y, there exist positive constants c, C, and c_1 such that (i) $c < \det(Var(\tilde{U}^y)) < C$, and (ii) $|\phi^y(\xi)| \leq 1 - c_1$, for all $c < |\xi| < Cn^{(\dim \tilde{U}^y + 1)/2}$.

[2]To be specific, consider the (just-identified) instrumental variable regression model, where $g(X, Y, Z, \theta_0) = Z(Y - X'\theta_0)$ for a dependent variable Y, endogenous regressors X, and instrumental variables Z. Assumptions 1 and 2 are satisfied if $\{X_i, Y_i, Z_i\}_{i=1}^n$ is i.i.d., $\mathbb{E}[g(X, Y, Z, \theta_0)g(X, Y, Z, \theta_0)']$ invertible, and the random variable $W = g(X, Y, Z, \theta_0)$ is continuously distributed on some compact support $\mathcal{W}$, which is typically guaranteed by the compact support of the observables (Y, X', Z'). Although this is a rather restrictive sufficient condition, our simulation studies in Sections 2.2 and 4 do not impose compact support for the data and we can still observe desirable performances of the TET statistic.

We note that most existing results using the saddlepoint approximation require finite moment generating functions (around zero) for the estimating equations (i.e., $\psi(X, y)$ in our notation). A notable exception is Jing, Shao, and Zhou (2004) who established the saddlepoint approximation for tail probabilities of the t statistic with no moment conditions. Although it is beyond the scope of this paper, it is interesting to see whether their theoretical development can be adapted to our context.

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

sphere of dimension p, but a Monte Carlo approximation to any degree of accuracy required can be obtained; see Kolassa and Robinson (2011) for details.

2.3.3. *Relative Error Property of the TET Statistic.* Motivated by the desirable relative error property of the ideal statistic T_n, we now establish the result in (4) and 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

$$\begin{aligned} K(\lambda) &= \frac{1}{2}\lambda' \mathbb{E}[g(X,0)g(X,0)']\lambda + O(|\lambda|^3), \\ K^{\text{tet}}(\lambda) &= \frac{1}{2}\lambda' \left(\sum_{i=1}^{n} \hat{\pi}_i^{\text{o}} g(x_i^{\text{o}},0)g(x_i^{\text{o}},0)'\right)\lambda + O(|\lambda|^3). \end{aligned} \tag{8}$$

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 ET statistic so that $t_n^{\text{et}} = 2K^{\text{et}}(\hat{\lambda}^{\text{o}})$, then an expansion yields

$$K^{\text{et}}(\lambda) = -\lambda'\bar{g}^{\text{o}} - \frac{1}{2}\lambda'\hat{V}^{\text{o}}\lambda + O(|\lambda|^3), \tag{9}$$

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 $O_p(\max\{n^{-1/2}|\lambda|, |\lambda|^2\})$. Also, we can see that the same comment applies to the function $K^{\text{cu}}(\lambda) = -\lambda'\bar{g}^{\text{o}} - \frac{1}{2}\lambda'\hat{V}^{\text{o}}\lambda$ for the continuous updating GMM, i.e., the sample counterpart of the difference $K(\lambda) - K^{\text{cu}}(\lambda)$ is of order $O_p(\max\{n^{-1/2}|\lambda|, |\lambda|^2\})$. Note that $K^{\text{cu}}(\lambda)$ evaluated at $\lambda = (\hat{V}^{\text{o}})^{-1}\bar{g}^{\text{o}}$ is the conventional Wald statistic.[3]

[3] Similarly, even if we consider the class of generalized empirical likelihood criterion by Newey and Smith (2004) (i.e., $K^{\text{gel}}(\lambda) = \frac{1}{n}\sum_{i=1}^n \rho(\lambda' g(x_i^{\text{o}},0))$ for a concave function ρ), we can see that the sample counterpart of the difference $K(\lambda) - K^{\text{gel}}(\lambda)$ is of order $O_p(\max\{n^{-1/2}|\lambda|, |\lambda|^2\})$. However, we conjecture that a "tilted" version of the generalized empirical likelihood $K^{\text{tgel}}(\lambda) = \sum_{i=1}^n \hat{\pi}_{\rho,i}^{\text{o}} \rho(\lambda' g(x_i^{\text{o}},0))$ will exhibit a similar property as $K^{\text{tet}}(\lambda)$ in Theorem 3, where $\hat{\pi}_{\rho,i}^{\text{o}} = \rho_1(\hat{\lambda}_\rho^{\text{o}\prime} g(x_i^{\text{o}},0))/\sum_{i=1}^n \rho_1(\hat{\lambda}_\rho^{\text{o}\prime} g(x_i^{\text{o}},0))$ is the generalized empirical likelihood implied probability with $\rho_1(v) = \partial\rho(v)/\partial v$ and $\hat{\lambda}_\rho^{\text{o}} = \arg\max_\lambda \sum_{i=1}^n \rho(\lambda' g(x_i^{\text{o}},0))$. We note that in contrast to the (observed) TET statistic

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

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 hold true. Then, under* $H_0 : \theta_0 = 0$*, it holds*

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

for each realization $\{x_i^{\text{o}}\}_{i=1}^n$*, where* $r_n(\lambda)$ *is a nonrandom function of* λ *defined in (19) and its sample counterpart, obtained by replacing* $\{x_i^{\text{o}}\}_{i=1}^n$ *with* $\{X_i\}_{i=1}^n$*, 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 in (7). The error of this approximation is relative and of (stochastic) order $n^{1/2}|\lambda|^2$. Therefore, in the normal region, for $\lambda = O(n^{-1/2})$, the relative error is of order $n^{-1/2}$. Beyond the normal region, e.g., $\lambda = O(n^{-1/3})$, the relative error approximation is of order $n^{-1/6}$. On the other hand, it is clear from (9) that the function $K^{\text{et}}(\lambda)$ for ET does not have such a relative error property.

Combining Theorems 2 and 3 (evaluated at $\lambda = \hat{\lambda}^{\text{o}}$), we can establish the equality in (4). Therefore, as far as the p-value p^{tet} for the TET statistic is well approximated by the one p^{ideal} for the ideal statistic T_n, we can argue that the asymptotic approximation for the TET tends to be accurate particularly in the tail area as illustrated in Figure 1 for the benchmark simulation.

3. GENERAL CASE

In this section, we generalize the theoretical results obtained in the last section for testing composite hypotheses (Section 3.1) and overidentifying restrictions (Section 3.2).

3.1. Composite Hypothesis Test

In this subsection, we extend the results for the benchmark case to composite hypothesis testing. We first consider just-identified moment conditions, and then discuss how to adapt our results for overidentified models.

3.1.1. *Just-Identified Case.* Consider just-identified moment conditions $\mathbb{E}[g(X, \theta_{10}, \theta_{20})] = 0$, where θ_{10} and θ_{20} are p_1- and p_2-dimensional parameters, respectively, and g is a $p = p_1 + p_2$ dimensional vector of moment functions. 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 ET statistic may be written

$t_n^{\text{tet}} = 2K^{\text{tet}}(\hat{\lambda}^{\text{o}})$, its generalized empirical likelihood counterpart $t_n^{\text{gel}} = 2K^{\text{tgel}}(\hat{\lambda}^{\text{o}})$ involves two Lagrange multipliers, $\hat{\lambda}_\rho^{\text{o}}$ and $\hat{\lambda}^{\text{o}}$. Full investigation of such an extension is left for future research.

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

as

$$T_{n,c}^{\text{et}} = -2 \max_{\theta_1 \in \Theta_1} \log \left(\frac{1}{n} \sum_{i=1}^{n} e^{\hat{\lambda}(\theta_1)' g(X_i, \theta_1, 0)} \right), \tag{10}$$

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 $\chi_{p_2}^2$ distribution under $H_0 : \theta_{20} = 0$. Let $\tilde{\theta}_1$ be a 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

$$T_{n,c}^{\text{tet}} = 2 \log \left(\sum_{i=1}^{n} \tilde{\pi}_i e^{\tilde{\lambda}' g(X_i, \tilde{\theta})} \right) = 2 \left[\log \left(\sum_{i=1}^{n} e^{2\tilde{\lambda}' g(X_i, \tilde{\theta})} \right) - \log \left(\sum_{i=1}^{n} e^{\tilde{\lambda}' g(X_i, \tilde{\theta})} \right) \right], \tag{11}$$

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 $K(\lambda, \theta_1) = \log \mathbb{E}[e^{\lambda' g(X, \theta_1, 0)}]$ and $\tilde{\theta}_1(\lambda)$ solves $\mathbb{E}\left[e^{\lambda' g(X, \tilde{\theta}_1(\lambda), 0)} \lambda' \left(\frac{\partial g(X, \tilde{\theta}_1(\lambda), 0)}{\partial \theta_1'} \right) \right] = 0$, 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 3. $\{X_i\}_{i=1}^{n}$ is i.i.d., $\theta_{10} \in \text{int}\Theta_1$ is the unique solution of $\mathbb{E}[g(X, \theta_{10}, \theta_{20})] = 0$, 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, $\mathbb{E}[\sup_{\theta_1 \in \Theta_1} |g(X, \theta_1, \theta_{20})|^{\zeta}] < \infty$, for some $\zeta > 2$, $\mathbb{E}[\sup_{\theta_1 \in \mathcal{N}} |\partial g(X, \theta_1, \theta_{20}) / \partial \theta_1'|] < \infty$, $\mathbb{E}[\partial g(X, \theta_{10}, \theta_{20}) / \partial \theta_1']$ is full column rank, and $\mathbb{E}[g(X, \theta_0) g(X, \theta_0)']$ 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 (21) instead of $g(X, 0)$. Then, under $H_0 : \theta_{20} = 0$,*

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

$$\Pr\{nT_{n,c} \geq nt : F\} = \{1 - F_{p_2}(n\xi_c(t))\}(1 + O(n^{-1})),$$

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 $\tilde{g}(X, \theta_0)$),

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

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

$$1-F_{p_2}(n\xi_c(K(\lambda,\tilde{\theta}_1(\lambda))))$$
$$=\{1-F_{p_2}(nK^{\mathrm{tet}}(\lambda,\tilde{\theta}_1^{\mathrm{tet}}(\lambda)))\}(1+r_{n,c}(\lambda))(1+O(n^{-1})),$$

for each realization $\{x_i^{\mathrm{o}}\}_{i=1}^n$, *where* $r_{n,c}(\lambda)$ *is a nonrandom function of* λ *and its sample counterpart, obtained by replacing* $\{x_i^{\mathrm{o}}\}_{i=1}^n$ *with* $\{X_i\}_{i=1}^n$, *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 be expected to provide accurate approximations of the tail area probabilities of the ideal statistic $T_{n,c}$.

We can compare 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 $\sum_{i=1}^n g(X_i,\hat{\theta})=0$, and $\hat{\mu}$ 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 $nT_{n,c}^{\mathrm{tet}}$ only requires to solve for $\tilde{\theta}$ and $\tilde{\lambda}$.[4]

3.1.2. *Overidentified Case.* The results in Theorem 4 can be adapted to overidentified models. For overidentified moment conditions $\mathbb{E}[g(X,\theta_{10},\theta_{20})]=0$, where θ_{10} and θ_{20} are p_1- and p_2-dimensional, respectively, and g is $d>p_1+p_2$ dimensional, suppose we wish to test the null hypothesis $H_0:\theta_{20}=0$ against the two-sided alternative $H_1:\theta_{20}\neq 0$. Indeed, this testing problem can be written as the one for $H_0:\theta_{20}=0$ in the following augmented moment conditions:

$$\mathbb{E}[h(X,\mu_0,\theta_{10},\theta_{20})]=\mathbb{E}\left[\begin{array}{c} e^{\mu_0'g(X,\theta_{10},\theta_{20})}g(X,\theta_{10},\theta_{20}) \\ e^{\mu_0'g(X_i,\theta_{10},\theta_{20})}\left(\frac{\partial g(X,\theta_{10},\theta_{20})}{\partial(\theta_1',\theta_2')}\right)'\mu_0 \end{array}\right]=0.$$

Note that due to the additional d-dimensional parameters μ_0, this moment condition model is just-identified. Therefore, the results in Section 3.1.1 for the just-identified case can be adapted by replacing the moment function $g(\cdot)$ with $h(\cdot)$, and nuisance parameters θ_{10} with $\vartheta_{10}=(\theta_{10}',\mu_0')'$. Let $\tilde{\vartheta}_1=$

[4]Although the TET statistic $nT_{n,c}^{\mathrm{tet}}$ takes a simpler form than Ma and Ronchetti's (2011), we still need to compute $\tilde{\theta}=(\tilde{\theta}_1',0')'$ and $\tilde{\lambda}$ to implement our test. The computation of $\tilde{\lambda}$ (for a given $\tilde{\theta}$) can be formulated as a convex optimization problem, so that Newton iterations can be applied. We use the MATLAB codes provided by Kirill Evdokimov and Yuichi Kitamura (available at https://www.kirillevdokimov.com/EL_codes.zip) for our numerical studies. On the other hand, the computation of $\tilde{\theta}$ typically requires a nested optimization algorithm, where the inner loop computes $\hat{\lambda}(\theta_1)$, for each θ_1, and the outer loop computes $\tilde{\theta}_1$ in (10). Similar to $\tilde{\lambda}$, the inner loop is typically a convex optimization problem which can be implemented by Newton iterations. The outer loop is a general nonlinear optimization problem, and we employ a quasi-Newton method, which works well in our numerical studies. See Kitamura (2007) for a detailed discussion on the nested optimization algorithm.

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

$\arg\max_{\vartheta_1} \log\left(\frac{1}{n}\sum_{i=1}^n e^{\hat{\lambda}(\vartheta_1)'h(X_i,\vartheta_1,0)}\right)$, where $\hat{\lambda}(\vartheta_1)$ solves $\sum_{i=1}^n e^{\hat{\lambda}(\vartheta_1)'h(X_i,\vartheta_1,0)}h(X_i,\vartheta_1,0)=0$, for each ϑ_1. The TET statistic is defined as

$$T_{n,c1}^{\text{tet}} = 2\log\left(\sum_{i=1}^n \tilde{\pi}_i e^{\tilde{\lambda}'h(X_i,\tilde{\vartheta})}\right),$$

where $\tilde{\vartheta}=(\tilde{\vartheta}_1',0')'$, $\tilde{\pi}_i=\frac{e^{\tilde{\lambda}'h(X_i,\tilde{\vartheta})}}{\sum_{j=1}^n e^{\tilde{\lambda}'h(X_j,\tilde{\vartheta})}}$, and $\tilde{\lambda}$ solves $\sum_{i=1}^n e^{\tilde{\lambda}'h(X_i,\tilde{\vartheta})}h(X_i,\tilde{\vartheta})=0$. By applying the results in Theorem 4, analogous relative error properties for $T_{n,c1}^{\text{tet}}$ can be derived.

3.2. Overidentifying Restriction Test

In this subsection, we consider the case of overidentifying moment restrictions $\mathbb{E}[g(X,\theta_0)]=0$, 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., $H_0:\mathbb{E}[g(X,\theta)]=0$, for some θ, against $H_1:\mathbb{E}[g(X,\theta)]\neq 0$, for any θ. This is a specification testing problem for the model specified by moment restrictions. In this case, the conventional ET statistic may be written as

$$T_{n,v}^{\text{et}} = -2\max_{\theta}\log\left(\frac{1}{n}\sum_{i=1}^n e^{\hat{\lambda}(\theta)'g(X_i,\theta)}\right), \tag{12}$$

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 χ^2_{d-p} distribution 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

$$T_{n,v}^{\text{tet}} = 2\log\left(\sum_{i=1}^n \bar{\pi}_i e^{\bar{\lambda}'g(X_i,\bar{\theta})}\right) = 2\left[\log\left(\sum_{i=1}^n e^{2\bar{\lambda}'g(X_i,\bar{\theta})}\right) - \log\left(\sum_{i=1}^n e^{\bar{\lambda}'g(X_i,\bar{\theta})}\right)\right], \tag{13}$$

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 $\mathbb{E}\left[e^{\lambda'g(X,\bar{\theta}(\lambda))}\lambda'\left(\frac{\partial g(X,\bar{\theta}(\lambda))}{\partial\theta'}\right)\right]=0$. 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

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

$\frac{1}{n}\sum_{i=1}^{n} e^{\lambda' g(x_i^{\mathrm{o}},\bar{\theta}^{\mathrm{tet}}(\lambda))}\lambda'\left(\frac{\partial g(x_i^{\mathrm{o}},\bar{\theta}^{\mathrm{tet}}(\lambda))}{\partial\theta'}\right) = 0$, for each λ. To analyze the properties of the TET statistic $T_{n,v}^{\mathrm{tet}}$, we modify Assumption 1 as follows.

Assumption 1″. $\{X_i\}_{i=1}^n$ is i.i.d., $\theta_0 \in \mathrm{int}\Theta$ is the unique solution of $\mathbb{E}[g(X,\theta_0)] = 0$, 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, $\mathbb{E}[\sup_{\theta\in\Theta}|g(X,\theta)|^{\zeta}] < \infty$, for some $\zeta > 2$, $\mathbb{E}[\sup_{\theta\in\mathcal{N}}|\partial g(X,\theta)/\partial\theta'|] < \infty$, $\mathbb{E}[\partial g(X,\theta_0)/\partial\theta']$ is full column rank, and $\mathbb{E}[g(X,\theta_0)g(X,\theta_0)']$ is nonsingular.

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

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

(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) $K^{\mathrm{tet}}(\lambda,\theta)$ *satisfies*

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

for each realization $\{x_i^{\mathrm{o}}\}_{i=1}^n$, *where* $r_{n,v}(\lambda)$ *is a nonrandom function of* λ *and its sample counterpart, obtained by replacing* $\{x_i^{\mathrm{o}}\}_{i=1}^n$ *with* $\{X_i\}_{i=1}^n$, *is of order* $O_p(n^{1/2}|\lambda|^2)$.

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

4. SIMULATION FOR GENERAL CASE

In this section, we evaluate the finite sample performance of the TET statistic for the general case considered in the last section. We generate random samples $\{W_i\}_{i=1}^n = \{Y_i, X_i, Z_i'\}_{i=1}^n$ of size $n = 40$ according to

$$Y_i = \theta_1 + \theta_2 X_i + U_i,$$
$$X_i = Z_i'\pi + V_i, \tag{14}$$

where $\pi = (c,c)'$ and $Z_i = (1, Z_{2i})'$ with $Z_{2i} \sim N(0,1)$. For each Monte Carlo replication, we set the value of c to fix the value of the concentration parameter $\delta^2 = \pi'(\sum_{i=1}^n Z_i Z_i')\pi$ (given the realized values of Z_i). We set $(\theta_1,\theta_2) = (0,0)$ for the true parameter vector. The error terms are generated as $(U_i, V_i) = (\epsilon_{1i}, 0.8\epsilon_{1i} + \sqrt{1 - 0.8^2}\epsilon_{2i})$, where ϵ_{1i} and ϵ_{2i} are independent. The number of Monte Carlo

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

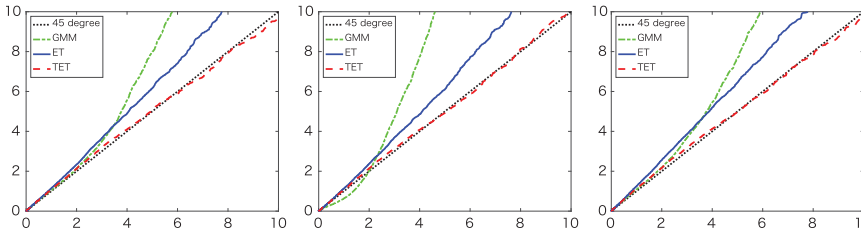


FIGURE 2. q–q plots for $H_0 : \theta_2 = 0$ with $n = 40$: (a) Normal, $\delta^2 = 50$, (b) Normal, $\delta^2 = 10$, and (c) χ^2_3, $\delta^2 = 50$.

replications is 10,000 for all cases. All figures reporting the simulation results are presented in Appendix B.

4.1. Composite Hypothesis Testing for $H_0 : \theta_2 = 0$

First, as an illustration for the composite hypothesis testing in Section 3.1, we consider testing the null hypothesis $H_0 : \theta_2 = 0$. For the model in (14), we consider three cases: (i) $\epsilon_{1i}, \epsilon_{2i} \sim N(0, 1)$ and $\delta^2 = 50$, (ii) $\epsilon_{1i}, \epsilon_{2i} \sim N(0, 1)$ and $\delta^2 = 10$, and (iii) $\epsilon_{1i}, \epsilon_{2i} \sim (\chi^2_3 - 3)/\sqrt{6}$ (standardized χ^2_3) and $\delta^2 = 50$. Case (i) is a baseline setup, (ii) is for investigating the effect of weaker instruments, and (iii) is for the effect of asymmetric distributions of the error terms. The cases of even weaker instruments with a weakly identified nuisance parameter are investigated in Section 4.3.

We compare the proposed TET statistic $T^{\text{et}}_{n,c}$ in (11) with the Wald statistic based on the GMM estimator and ET statistic $T^{\text{et}}_{n,c}$ in (10). All the statistics converge in distribution to the χ^2_1 distribution under H_0. Figure 2 reports the q–q plots of the empirical quantiles of these test statistics against those of the χ^2_1 distribution for the cases (i)–(iii).

Overall, the empirical quantiles of the TET statistic are very close to those of the limiting distribution. These q–q plots of the TET statistics are not sensitive to weaker instruments (as shown in Figure 2(b)) or asymmetric error distributions (as shown in Figure 2(c)). On the other hand, the quantiles of the conventional Wald and ET statistics tend to be larger than the ones of the limiting distribution. In particular, the Wald statistic is worse for all the cases. Such patterns of the q–q plots for the Wald and ET statistics are commonly observed in the literature (e.g., Imbens et al. 1998).[5] Interestingly, the weaker instruments induce more severe size distortions for the Wald test (see Figure 2(b)).

[5] In the working paper version (available at: http://sticerd.lse.ac.uk/dps/em/em593.pdf), we also report the results for the case of $n = 80$ and the empirical cumulative distribution functions (ECDFs) of the p-values for all the tests considered in Figures 2 and 4 below. The accuracy of the ET statistic improves as the sample size increases. However, the TET always outperforms the ET and Wald. Furthermore, based on the ECDFs for the p-values, we can see the

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

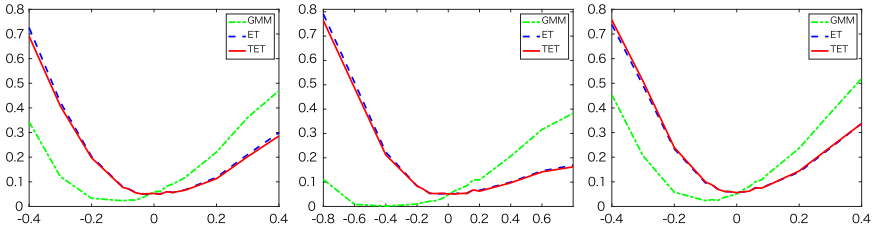


FIGURE 3. Calibrated powers for $H_0 : \theta_2 = 0$ with $n = 40$: (a) Normal, $\delta^2 = 50$, (b) Normal, $\delta^2 = 10$, and (c) χ^2_3, $\delta^2 = 50$.

We also investigate the power properties of the tests for $H_0 : \theta_2 = 0$ under the alternative hypotheses $H_1 : \theta_2 = 0 + \Delta$ for different values of Δ. Figures 3 displays the calibrated power curves of all the tests at 5% significance level (i.e., the rejection frequencies of these tests, where the critical values are given by the Monte Carlo 95th percentiles of these test statistics under H_0) for the cases (i)–(iii). The results suggest that the proposed TET tests exhibit good calibrated power. Indeed, the calibrated power curves of the TET are very close to those of the ET, and do not show declining power for some negative values of Δ as in the Wald test.

4.2. Specification Testing

Second, as an illustration for specification testing in Section 3.2, we consider testing the overidentifying restrictions $H_0 : \mathbb{E}[Z(Y - \theta X)] = 0$, for some θ, against $H_1 : \mathbb{E}[Z(Y - \theta X)] \neq 0$, for any θ. We compare the proposed TET statistic $T^{\text{tet}}_{n,v}$ in (13) with the J statistic based on GMM,[6] and ET statistic $T^{\text{et}}_{n,v}$ in (12). All the statistics converge in distribution to the χ^2_1 distribution under H_0.

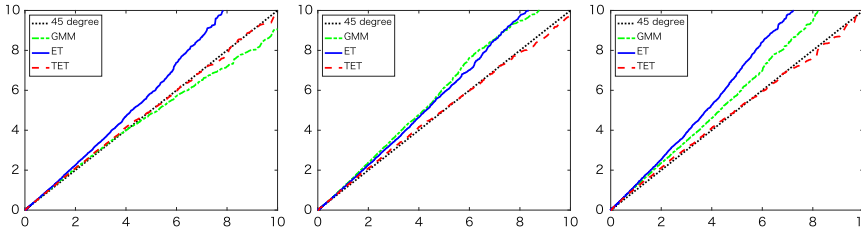
In addition to the cases (i) and (ii) above, we consider the case of heteroskedastic error terms (iv) $Y_i = \theta_1 + \theta_2 X_i + |Z_i| U_i$, $\epsilon_{1i}, \epsilon_{2i} \sim N(0, 1)$, and $\delta^2 = 50$. Figure 4 reports the q–q plots of the empirical quantiles of these test statistics against those of the χ^2_1 distribution for the normal disturbance case.

Although the patterns of the plots are different, we obtain similar conclusions as the simulations for composite hypothesis testing. Overall, the empirical quantiles of the TET statistic are very close to those of the limiting distribution compared to the ET and J statistics. The performances of the ET and J statistics are comparable. It is interesting to note that the q–q plots of the TET statistic are not sensitive to heteroskedastic error terms. On the other hand, we can see that heteroskedastic errors deteriorate the size properties of the ET and J statistics.

degrees of size distortions of these statistics. As consistent with the q–q plots, the asymptotic test by the TET shows better size properties (particularly for the region of nominal sizes less than 0.10) than other asymptotic tests.

[6]Letting $g(W_i, \theta) = Z_i(Y_i - \theta X_i)$ and $\bar{g}(\theta) = n^{-1} \sum_{i=1}^n g(W_i, \theta)$, the version of the J-test statistic considered here is $J = n \min_\theta \bar{g}(\theta)' [n^{-1} \sum_{i=1}^n g(W_i, \check{\theta}) g W_i, \check{\theta})]^{-1} \bar{g}(\theta)$, where $\check{\theta}$ is the two-stage least-square estimator.

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press



(a) Normal, $\delta^2 = 50$ (b) Normal, $\delta^2 = 10$ (c) Normal, $\delta^2 = 50$, heteroskedastic

FIGURE 4. q–q plots for overidentification with $n = 40$: (a) Normal, $\delta^2 = 50$, (b) Normal, $\delta^2 = 10$, and (c) Normal, $\delta^2 = 50$, heteroskedastic.

Overall, our simulation results are encouraging: the TET statistic performs excellently compared to the existing statistics.

4.3. Additional Results: Weak Instruments

Finally, we consider the situation where the instrumental variable Z is very weak (i.e., $\delta^2 = 5$ and 1). There is no guarantee that the proposed TET statistic works well under partial or weak identification; see Phillips (1980, 1984, 1985, 1986, 1989); Staiger and Stock (1997). In particular, the full rank conditions for the Jacobians of g (with respect to partially or weakly identified parameters) in Assumptions 1' and 1" typically fail, and our asymptotic approximations will be invalid. Although its theoretical analysis is beyond the scope of this paper, we here illustrate finite sample performances of the TET statistic under weak instruments.

Specifically, we consider three cases: (i') $\epsilon_{1i}, \epsilon_{2i} \sim N(0,1)$ and $\delta^2 = 5$, (ii') $\epsilon_{1i}, \epsilon_{2i} \sim N(0,1)$ and $\delta^2 = 1$, and (iii') $\epsilon_{1i}, \epsilon_{2i} \sim (\chi^2_3 - 3)/\sqrt{6}$ (standardized χ^2_3) and $\delta^2 = 5$. Figures 5 and 6 report the q–q plots of the empirical quantiles of the test statistics against those of the χ^2_1 distribution for testing $H_0 : \theta_2 = 0$ and $H_0 : \theta_1 = 0$, respectively. We note that for testing $H_0 : \theta_1 = 0$, the nuisance parameter θ_2 is weakly identified and we conjecture that the TET test will be asymptotically invalid based on the existing literature of weak instruments (e.g., Andrews and Stock (2007), for a survey).

For testing $H_0 : \theta_2 = 0$, where the nuisance parameter θ_1 is strongly identified, the q–q plots in Figure 5 are similar to the ones in Figure 2, and the TET statistics are not sensitive to very weak instruments. However, for testing $H_0 : \theta_1 = 0$, where the nuisance parameter θ_2 is weakly identified, the q–q plots in Figure 6 suggest that the quantiles of the TET statistics deviate from the ones of the χ^2 limiting distribution.[7] In particular, the TET is worse than other statistics for the case (ii').

[7] In the preliminary simulation study, we considered overidentification testing under the cases (i')–(iii'). The q–q plots also indicate that the TET statistic for overidentifying restrictions is not robust to very weak instruments.

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

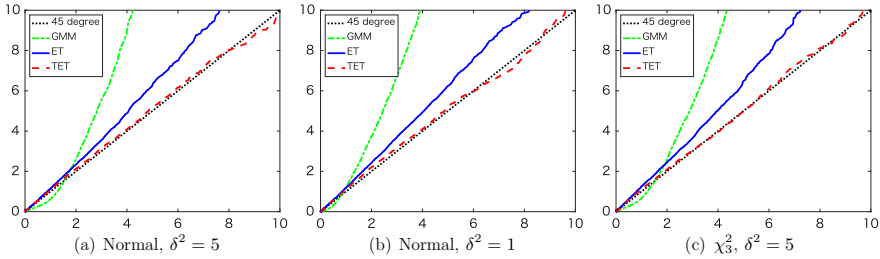


FIGURE 5. q–q plots for $H_0 : \theta_2 = 0$ with $n = 40$: (a) Normal, $\delta^2 = 5$, (b) Normal, $\delta^2 = 1$, and (c) χ^2_3, $\delta^2 = 5$.

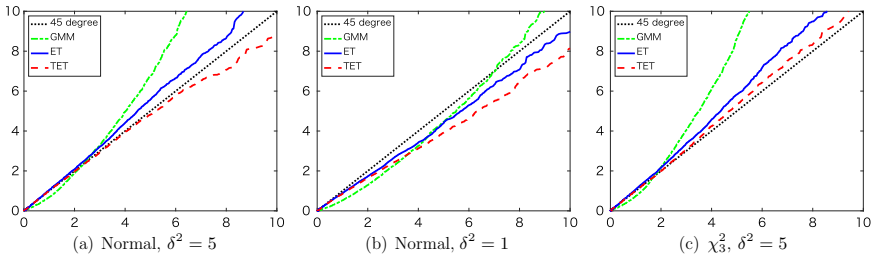


FIGURE 6. q–q plots for $H_0 : \theta_1 = 0$ with $n = 40$: (a) Normal, $\delta^2 = 5$, (b) Normal, $\delta^2 = 1$, and (c) χ^2_3, $\delta^2 = 5$.

In order to understand the behavior of the TET statistic in cases of weak instruments and instrument failure, it is intriguing to apply the general asymptotic theory under instrument failure developed in the seminal work of Phillips (1989). In particular, the framework developed in Phillips (1989, Sects. 2 and 4) would be general enough to cover our TET statistic. As illustrated in Phillips (1989, Sect. 2.4) for the Wald statistics, we expect that Phillips (1989) asymptotics will lead to a very different limiting behavior for the TET statistic in comparison to the usual strong instrument case. The analysis under weak instrument asymptotics as in Staiger and Stock (1997) is followed by adding a noncentral term to Phillips (1989) asymptotic framework. The discrepancies among those sets of asymptotic results explain the poor performance of the TET statistic in this context. Another interesting (though technically more involved) direction of future research is to explore the finite sample theory for the TET statistic under Gaussian errors. For this direction, we need to extend the general exact theory in Phillips (1980, 1984, 1985, 1986) for the instrumental variable, limited information maximum likelihood, and full information maximum likelihood estimators to our TET statistic. Even for the benchmark case in Section 2, such an exact analysis would be highly beneficial to understand the desirable tail behavior of the TET statistic. Also, we note that the exact analyses in Phillips (1984, 1986) shed light on instrument failure cases with many interesting phenomena being reported. In particular, it is interesting to

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

investigate the effect of the sample size on the exact distribution of the TET statistic under instrument failure cases.[8]

5. CONCLUSION

In this paper, we modify the ET statistic by introducing additional tilting weights to estimate cumulant functions, and propose a novel statistic, called the TET statistic. Simulation studies show that the asymptotic p-values of the TET statistic are highly accurate in the tails, and we provide a theoretical explanation for this accuracy by analyzing its relative error properties for the approximation of tail area probabilities. In particular, the proposed p-values accurately approximate those of an infeasible saddlepoint statistic, which has relative errors of order n^{-1} both in normal and large deviation regions. Our TET test allows both just- and overidentified moment condition models, and can be applied for both (possibly composite) parameter hypotheses and overidentifying restriction testing problems.

There are at least two important directions for future research. First, as repeatedly mentioned, the present analysis lacks theoretical evaluations of the closeness of the p-values for the infeasible statistic compared to those for the proposed TET statistic. Although it is a substantial challenge given the existing results on saddlepoint approximations, this theoretical gap should be addressed. Another interesting direction is to investigate the theoretical properties of the TET statistic under weak or partial identification, as indicated in the last section, and to develop a modified statistic that is not only accurate in the tails but also robust to weak or partial identification.

APPENDIX

A. Proofs

A.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

$$nT_n^{\text{tet}} = -\hat{M}^{-1} n\hat{\lambda}' \left[\frac{1}{n} \sum_{i=1}^n e^{(\hat{\lambda}+\bar{\lambda})' g_i} g_i g_i' \right] \hat{\lambda}$$

[8]The authors are grateful to Peter Phillips for suggesting references in this paragraph and directions for future research.

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

$$= n\hat{\lambda}' \left[\frac{1}{n} \sum_{i=1}^{n} e^{(\hat{\lambda}+\bar{\lambda})' g_i} g_i g_i' \right] \hat{\lambda} + \bar{M}^{-2} n\hat{\lambda}' \left[\frac{1}{n} \sum_{i=1}^{n} e^{(\hat{\lambda}+\bar{\lambda})' g_i} g_i g_i' \right] \hat{\lambda}(\hat{M}+1)$$
$$= T_1 + T_2,$$

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

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

Combining these results,

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

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.

A.2. Proof of Theorem 2

The basic idea of the proof is similar to that of Robinson et al. (2003, Thm. 1). Let $m(\lambda) = \log \mathbb{E}[e^{\lambda' g(X)}]$, so that $T_n = 2m(\check{\lambda})$. By Assumption 2, the tail probability of T_n is approximated as

$$\Pr\{nT_n \ge nt : F\} = \Pr\{2m(\check{\lambda}) \ge t : F\}$$
$$= \int_{\{y:2m(y)\ge t\}} \left(\frac{n}{2\pi} \right)^{p/2} e^{-nh(y)} \frac{\det B(y)}{\sqrt{\det \Sigma(y)}} dy (1 + O(n^{-1}))$$
$$\equiv A(1 + O(n^{-1})). \tag{15}$$

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

$$A = \int_{\sqrt{t}}^{\infty} c_n u^{p-1} e^{-nu^2/2} \left\{ \int_S \delta(u, s) ds \right\} du, \tag{16}$$

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

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

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

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

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

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 $\mathbb{E}[\check{g}(X)\check{g}(X)'] = I$, we have $\frac{\det B(0)}{\sqrt{\det \Sigma(0)}} = 1$. Thus, other terms are expanded as

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

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

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

where R_6 is uniformly bounded for r bounded, and $b(s)$ is a linear combination of odd functions satisfying $\int_S b(s)ds = 0$. Integrating over the sphere gives

$$G(u) \equiv \int_S \delta(u,s)ds = 1 + u^2 k(u), \tag{17}$$

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 (15) to (17),

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

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 $k_1(u)$.

A.3. Proof of Theorem 3

Using (15)–(17) in the proof of Theorem 2 and integration by parts, we have

$$1 - F_p(n\xi(K(\lambda))) = \{1 - F_p(nK(\lambda))\}(1 + O(n^{-1})) + \frac{c_n}{n} K(\lambda)^{\frac{p}{2}} e^{-\frac{nK(\lambda)}{2}} \left[\frac{G(\sqrt{K(\lambda)}) - 1}{K(\lambda)}\right]. \tag{18}$$

For the first term on the right-hand side of (18), the mean-value theorem yields

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

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

where $r_{1n}(\lambda) = \frac{e^{-\bar{u}_\lambda/2}\bar{u}_\lambda^{p/2-1}}{2^{p/2}\Gamma(p/2)} \frac{nK^{\text{tet}}(\lambda)-nK(\lambda)}{1-F_p(nK^{\text{tet}}(\lambda))}$, for some $\bar{u}_\lambda$, between $nK(\lambda)$ and $nK^{\text{tet}}(\lambda)$. For the second term on the right-hand side of (18), let

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

Thus, letting

$$r_n(\lambda) = r_{1n}(\lambda) + r_{2n}(\lambda), \tag{19}$$

(18) can be written as

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

Let $R_{1n}(\lambda)$, $R_{2n}(\lambda)$, and $R_n(\lambda)$ be the sample counterparts of $r_{1n}(\lambda)$, $r_{2n}(\lambda)$, and $r_n(\lambda)$, respectively, obtained by replacing $K^{\text{tet}}(\lambda)$ with $\hat{K}^{\text{tet}}(\lambda)$. It remains to characterize the stochastic order of $R_n(\lambda)$. Note that the χ^2_p distribution satisfies

$$\Pr\{\chi^2_p \geq u\} = 1-F_p(u) \geq Ce^{-u/2}u^{p/2-1}, \tag{20}$$

for some constant $C > 0$.

For $R_{1n}(\lambda)$, by using (20), we have

$$|R_{1n}(\lambda)| \leq O_p(1)\{n\hat{K}^{\text{tet}}(\lambda) - nK(\lambda)\} = O_p(n^{1/2}|\lambda|^2),$$

where $\hat{K}^{\text{tet}}(\lambda) = \log\left(\sum_{i=1}^n \hat{\pi}_i e^{\lambda' g(X_i,0)}\right)$ is the sample counterpart of $K^{\text{tet}}(\lambda)$, and the equality follows from (8) and $\sum_{i=1}^n \hat{\pi}_i g(X_i,0)g(X_i,0)' - \mathbb{E}[g(X,0)g(X,0)'] = O_p(n^{1/2})$.

We now consider $R_{2n}(\lambda)$. By the definition of $G(u) = 1+u^2k(u)$, for some $k(u)$, bounded over $u \in (0,\varepsilon)$, $\left[\frac{G(\sqrt{K(\lambda)})-1}{K(\lambda)}\right]$ is also bounded. Thus, by using (20) and the definition of $c_n = n^{p/2}/(2^{p/2-1}\Gamma(p/2))$, we have

$$|R_{2n}(\lambda)| \leq O_p(1)\hat{K}^{\text{tet}}(\lambda) = O_p(|\lambda|^2).$$

Combining these results, we obtain the conclusion $R_n(\lambda) = O_p(n^{1/2}|\lambda|^2)$.

A.4. Proof of Theorem 4

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

$$M(M'M)^{-1}M' = C\begin{bmatrix} I_{p_1} & 0 \\ 0 & 0_{p_2\times p_2} \end{bmatrix} C',$$

and $C'C = CC' = I_p$ with $p = p_1 + p_2$. 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

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

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

where

$$\tilde{g}(X,\theta_0) = C_2'\Omega^{-1/2}g(X,\theta_0). \tag{21}$$

The saddlepoint density of $\tilde{\gamma}$ is given by (6) 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

$$\Pr\{nT_{n,c} \geq nt_{n,c} : F\} = \Pr\{2n\tilde{K}(\tilde{\gamma}) \geq n\tilde{K}(\tilde{\gamma}_0) : F\}(1 + O(e^{-n\epsilon})),$$

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

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. ■

A.5. Proof of Theorem 5

The proof is similar to that of Theorem 4.

REFERENCES

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

Andrews, D.W.K. & J.H. Stock (2007) Inference with weak instruments. In R. Blundell, W. Newey, & T. Persson (eds.), *Advances in Economics and Econometrics, Theory and Applications. Ninth World Congress*, vol. 3, pp. 122–173. Cambridge University Press.

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

Brown, B.W. & W.K. Newey (1998) Efficient semiparametric estimation of expectations. *Econometrica* 66, 453–464.

Czellar, V. & E. Ronchetti (2010) Accurate and robust tests for indirect inference. *Biometrika* 97, 621–630.

Daniels, H.E. & G.A. Young (1991) Saddlepoint approximation for the studentized mean, with an application to the bootstrap. *Biometrika* 78, 169–179.

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

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

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

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

Huber, P.J. & E.M. Ronchetti (2009) *Robust Statistics, 2nd Edition*. Wiley.

Imbens, G.W., R.H. Spady, & P. Johnson (1998) Information theoretic approaches to inference in moment condition models. *Econometrica* 66, 333–357.

Jensen, J.L. (1995) *Saddlepoint Approximations*. Oxford University Press.

Jensen, J.L. & A.T.A. Wood (1998) Large deviation and other results for minimum contrast estimators. *Annals of the Institute of Mathematical Statistics* 50, 673–695.

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

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press

Jing, B.Y., Q.M. Shao, & W. Zhou (2004) Saddlepoint approximation for student's *t*-statistic with no moment conditions. *Annals of Statistics* 32, 2697–2711.

Jing, B.Y. & A.T.A. Wood (1996) Exponential empirical likelihood is not Bartlett correctable. *Annals of Statistics* 24, 365–369.

Kitamura, Y. (2007) Empirical likelihood methods in econometrics: Theory and practice. In R. Blundell, W.K. Newey, & T. Persson (eds.), *Advances in Economics and Econometrics*, vol. III, pp. 174–237. Cambridge University Press.

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

Kolassa, J. & J. Robinson (2011) Saddlepoint approximations for likelihood ratio like statistics with applications to permutation tests. *Annals of Statistics* 39, 3357–3368.

Ma, Y. & E. Ronchetti (2011) Saddlepoint test in measurement error models. *Journal of the American Statistical Association* 106, 147–156.

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

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

Owen, A.B. (2001) *Empirical Likelihood*. Chapman and Hall.

Phillips, P.C.B. (1980) The exact distribution of instrumental variable estimators in an equation containing $n+1$ endogenous variables. *Econometrica* 48, 861–878.

Phillips, P.C.B. (1984) The exact distribution of LIML: I. *International Economic Review* 25, 249–261.

Phillips, P.C.B. (1985) The exact distribution of LIML: II. *International Economic Review* 26, 21–36.

Phillips, P.C.B. (1986) The distribution of FIML in the leading case. *International Economic Review* 27, 239–243.

Phillips, P.C.B. (1989) Partially identified econometric models. *Econometric Theory* 5, 181–240.

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

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

Staiger, D. & J.H. Stock (1997) Instrumental variables regression with weak instruments. *Econometrica* 65, 557–586.

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

https://doi.org/10.1017/S0266466621000074 Published online by Cambridge University Press