Shrinkage of Variance for Minimum Distance Based Tests

Saraswata Chaudhuri*and Eric Renault†

This version: May 17, 2011.

Abstract

This paper applies the information theoretic approaches to inference in moment condition models to the improvement of score tests about the full parameter vector θ. In the context of optimum minimum distance estimation, the different score test statistics differ only through the way of estimating the variance of the moment conditions. We resort to implied probabilities provided by constrained maximization of general entropy to improve this variance estimator under the null. Our improvement is akin to shrinkage of the variance estimator to efficiently take advantage of the information content of moment conditions. With respect to the recent literature on generalized empirical likelihood (GEL), our use of implied probabilities is new since we compute constrained probabilities under the null while estimation of θ is irrelevant. As a result, these probabilities are even relevant in the case of just-identified moment conditions. We document, both by theoretical higher order expansions and by Monte-Carlo evidence that our improved score tests have better finite sample size properties. Our approach, that does not take any simulation work, can then be seen as a user friendly substitute or complement to more sophisticated bootstrapping approaches. This claim is confirmed by our Monte Carlo experiments on inference on covariance structures as previously studied by Horowitz (1998) with bootstrap methods.

JEL Classification: C12; C13; C30

Keywords: Entropy; GMM; Implied probabilities; Score test; Shrinkage

*Department of Economics, CB 3305, University of North Carolina, Chapel Hill, NC 27519. Telephone: 919-966-3962. Fax: 919-966-4986. Email: saraswata_chaudhuri@unc.edu.

†Department of Economics, CB 3305, University of North Carolina, Chapel Hill, NC 27519. Telephone: 919-966-5326. Fax: 919-966-4986. Email: renault@email.unc.edu.

1 Introduction

The optimal minimum distance (OMD) estimator $\hat{\theta}_n$ of a vector θ of p unknown parameters identified by $K \geq p$ constraints

$$\lambda = g(\theta)$$

is the solution of the minimization problem

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} \; n \left(\hat{\lambda}_n - g(\theta) \right)' V_n^{-1} \left(\hat{\lambda}_n - g(\theta) \right), \tag{1}$$

where $\Theta \subset \mathbb{R}^p$ is the parameter space, $\hat{\lambda}_n$ is a $\sqrt{n}$-consistent asymptotically normal estimator with a positive definite asymptotic variance V and V_n is any consistent estimator of V.

The focus of our interest in this paper is the test of a null hypothesis

$$H_0 : \theta = \theta_0.$$

We study the dependence of the finite-sample properties of such a test on the choice of the asymptotic variance estimator V_n. We recommend a shrinkage estimator that leads to over-all superior performance of the test.

We assume throughout that the consistent estimator $\hat{\lambda}_n$ is a sample mean of some known functions of the observations. For sake of notational simplicity, we can write without loss of generality,

$$\hat{\lambda}_n = \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

In other words, $\hat{\theta}_n$ can also be seen as an efficient Generalized Method of Moments (GMM) estimator associated to the moment conditions

$$E\left(X_i - g(\theta)\right) = 0. \tag{2}$$

The results of this paper could actually be partly extended to general non-separable moment conditions

$$E\psi(X_i, \theta) = 0. \tag{3}$$

While this general case is studied in a companion paper Chaudhuri and Renault (2011), we focus here on the specific conclusions that can be drawn from the particular form (2), especially because, under the null hypothesis $H_0 : \theta = \theta_0$, it makes the expected Jacobian matrix of the moment conditions known

$$E\left[\frac{\partial\psi(X_i,\theta_0)}{\partial\theta'}\right] = -\frac{\partial g(\theta_0)}{\partial\theta'}(\equiv G \text{ say}). \tag{4}$$

We maintain throughout the common assumption for asymptotic distributional theory of OMD estimators that the Jacobian matrix G (under the null) is of full column-rank p. While this Jacobian matrix is key for efficient estimation and score-type tests, its estimation is known to be an important source of poor finite-sample behavior of GMM-based inference due to a perverse correlation with moments conditions that gets worse when the identification of θ_0 is not strong. By contrast, inference in the context of (2) will, in general, involve only the estimation of the asymptotic variance matrix V. For instance, the Newey and West (1987) score test of the null hypothesis $\theta = \theta_0$ will be simply based on the test statistic

$$\xi_n^S = n\bar{\psi}_n'(V_n^0)^{-1}G\left(G'(V_n^0)^{-1}G\right)^{-1}G'(V_n^0)^{-1}\bar{\psi}_n,$$

where $\bar{\psi}_n := \sum_{i=1}^n \psi_i$ and $\psi_i \equiv \psi(X_i,\theta_0) := X_i - g(\theta_0)$ (for notational simplicity). Under the null, the score statistic ξ_n^S will asymptotically follow a $\chi^2(p)$ distribution if V_n^0 is a consistent estimator of the asymptotic variance matrix V. A common practice is to take for V_n^0 a moment-based estimator of V. For instance, in a case with observations without serial correlation, one would typically use a naive estimator like

$$\bar{V}_n := \frac{1}{n}\sum_{i=1}^n(\psi_i - \bar{\psi}_n)\psi_i' \text{ (unconstrained), or } \bar{V}_n := \frac{1}{n}\sum_{i=1}^n \psi_i\psi_i' \text{ (constrained by } H_0).$$

The main thesis of this paper is that, by contrast with this common practice, the size of the score test would be better controlled by using as V_n^0 an estimator of V that is asymptotically efficient under the null hypothesis. We will see that such an efficient estimation of the variance matrix V amounts to a shrinkage of the naive estimator that takes into account the information content of the moment conditions implied by the null hypothesis. The rationale for efficiently estimating V is that we try to mimic the behavior of the infeasible test statistic that would use the knowledge of the unknown asymptotic variance V, i.e., the statistic,

$$\xi_n^{S,I} \quad = \quad n\bar{\psi}_n' V^{-1} G \left(G' V^{-1} G\right)^{-1} G' V^{-1} \bar{\psi}_n.$$

However, it is worth keeping in mind that the ultimate goal is not really to mimic the infeasible test statistic $\xi_n^{S,I}$ but rather to get a test statistic with a finite-sample distribution function that is as close as possible to a $\chi^2(p)$ distribution under the null. With this goal in mind, our methodology for comparing competing tests will be twofold. On the one hand, we will provide some compelling Monte-Carlo evidence that the targeted $\chi^2(p)$ distribution is better tracked by our proposed test statistics with variance shrinkage than by standard test statistics. On the other hand, we will display some rationale for this evidence through the asymptotic expansions of the distribution functions of the various test statistics. We refer to Cavanagh (1983) for technical results used for such expansions. We will typically show that we get with variance shrinkage a more accurate approximation of the distribution function of the infeasible score test statistic. The latter theoretical result is completely derived only in the just-identified case ($K = p$) because in this case the score test statistic can be written in the simpler form of a Studentized sample mean:

$$\xi_n^S = n\bar{\psi}_n' (V_n^0)^{-1} \bar{\psi}_n. \tag{5}$$

However, both the theoretical analysis and the Monte-Carlo experiments confirm the intuition that finite-sample improvements are also present in the over-identified case. The main Monte-Carlo illustration considers the estimation of covariance structures as commonly met in a variety of economic examples. Abowd and Card (1989) and Altonji and Segal (1996) have documented the poor finite sample properties of OMD estimators and inference in this context. Horowitz (1998) have put forward bootstrap methods for finite-sample improvements. It is worth realizing that the shrinkage strategy studied in this paper is not aimed at replacing bootstrap. First, it proposes a simple and user-friendly way to improve finite-sample size properties of score tests, without resorting to any simulation. Second, it could well be coupled with the bootstrap methods if necessary. Our approach is actually quite close in spirit to the bootstrapping for GMM as devised by Brown and Newey (2002). As in their work, we take advantage of the probabilities implied by moment conditions (under the null hypothesis) for a proper re-weighting of the observations at hands. While they do that for the purpose of re-sampling, we just do it to find the efficient estimator V_n^0 of the asymptotic variance matrix V.

The paper is organized as follows. In section 2, we discuss the efficient estimators of the variance matrix that are provided by a generalized maximum entropy approach to the moment conditions under the null. In section 3, score tests statistics are compared through asymptotic expansions of their distribution functions. The toy example of the standard Student test for a population mean allows us to further interpret the finite-sample improvements. An extensive Monte-Carlo illustration is provided in section 4 in the context of OMD inference on covariance structures. Our approach appears in many cases to be competitive with the more involved bootstrap approach of Horowitz (1998). Section 5 is a mathematical appendix that sketches the extension of asymptotic comparisons of distribution functions of score test statistics in the context of over-identified moment restrictions.

2 Efficient estimation of the variance matrix under the null hypothesis

The information theoretic approaches to inference in moment condition models have become popular in econometrics since the seminal paper by Imbens et al. (1998). The idea in the context of general moment conditions (3) is to look simultaneously for an estimator $\hat{\theta}_n^\gamma$ of θ and for the implied probabilities $\hat{\pi}_n^\gamma = (\hat{\pi}_{i,n}^\gamma)_{1\leq i\leq n}$ as solutions of

$$\min_{\theta\in\Theta,\pi} \frac{1}{\gamma(\gamma-1)} \sum_{i=1}^{n} \left[(n\pi_i)^{1-\gamma} - 1\right] \tag{6}$$
$$\text{subject to} \qquad \sum_{i=1}^{n} \pi_i = 1 \text{ and } \sum_{i=1}^{n} \pi_i \psi(X_i, \theta) = 0.$$

The objective function (6) is defined for any real γ, including the two limit cases $\gamma \to 0$ and $\gamma \to 1$. The family of these functions, indexed by γ, is generally referred to as the Cressie-Read family of power divergence statistics (see Imbens et al. (1998) and references therein). It is known that in the case of i.i.d. observations $X_i, i = 1, ...n$, and under standard regularity conditions, the estimator $\hat{\theta}_n^\gamma$ is asymptotically efficient (and asymptotically equivalent to efficient GMM) for any value of γ. In case of serially dependent observations, this result can be extended by applying the above power divergence minimization to properly pre-averaged moments in the spirit of Kitamura and Stutzer (1997). All what is done in the following could be extended like that to time series models but will not be stated explicitly for the sake of expositional simplicity.

It is generally believed that implied probabilities are relevant for inference only in the case of over-identified moment conditions since, when $K = p$, one may generically find a method of moments estimator $\hat{\theta}_n$ such that

$$\sum_{i=1}^{n} \psi(X_i, \hat{\theta}_n) = 0$$

$$\text{and then,} \qquad \hat{\pi}_{i,n}^{\gamma} = \frac{1}{n}, \forall i = 1, ..., n, \forall \gamma \in \mathbb{R}. \tag{7}$$

Our use of implied probabilities in this paper is new since we want to devise the proper shrinkage implied by the null hypothesis $\theta = \theta_0$. In other words, in the context of separable moment conditions (2), we defined implied probabilities $\hat{\pi}_{i,n}^{\gamma,0}$ as solutions of

$$\min_{\pi} \frac{1}{\gamma(\gamma - 1)} \sum_{i=1}^{n} \left[(n\pi_i)^{1-\gamma} - 1 \right] \tag{8}$$

$$\text{subject to} \qquad \sum_{i=1}^{n} \pi_i = 1 \text{ and } \sum_{i=1}^{n} \pi_i \psi(X_i, \theta_0) = 0. \tag{9}$$

(Recall that $\psi_i \equiv \psi(X_i, \theta_0) := X_i - g(\theta_0)$.) As a consequence, even in the just-identified case, implied probabilities do not coincide with the empirical distribution (7) because the null hypothesis is not exactly fulfilled with sample moments. The consistent estimators V_n^0 of the variance matrix we promote in this paper are the ones associated to these implied probabilities,

$$V_n^{\gamma,0} = \sum_{i=1}^{n} \hat{\pi}_{i,n}^{\gamma,0} \psi_i \psi_i'. \tag{10}$$

It is worth comparing the estimators $V_n^{\gamma,0}$ (for any choice of the power-divergence parameter γ) with the naive estimation principle based on the empirical probabilities (1/n) (and mentioned in the Introduction) that, under the null, i.e., when working with the constrained "estimator" of θ, would lead to consider

$$\bar{V}_n = \frac{1}{n} \sum_{i=1}^{n} \psi_i \psi_i'. \tag{11}$$

The key difference between (10) and (11) is that we have replaced the empirical distribution (7) by implied probabilities which make sure that the moment conditions (with the value of θ under the null) are fulfilled in the sample. In yet other words, we have shrunk the variance estimator to take advantage of the information brought by the null hypothesis $H_0 : \theta = \theta_0$. This shrinkage interpretation will be confirmed by the computation of implied probabilities below.

Let us also note that this can be related to a point already made by Hall (2000) who recommends that variances be calculated using the data in mean deviation form. It is precisely a way to acknowledge that the naive estimator must be shrunk in due proportion of the in-sample violation of the moment conditions. Hall's shrinkage would simply lead to replace $\bar{V}_n$ by

$$\bar{V}_n^* = \frac{1}{n}\sum_{i=1}^{n}\psi_i\psi_i' - \bar{\psi}\bar{\psi}'. \tag{12}$$

It is worth realizing that, by contrast with (12), the shrinkage (10) makes an efficient use of the information content of the moment conditions. To see that, first note that the first order conditions of the minimization (8) subject to constraints (9) gives, for a non-zero γ

$$\begin{aligned}\hat{\pi}_{i,n}^{\gamma,0} &\propto \left[1 + \beta_\gamma'(X_i - g(\theta_0)\right]^{-1/\gamma} \\ &= 1 - \frac{1}{\gamma}\beta_\gamma'(X_i - g(\theta_0) + o_p(1/\sqrt{n})\end{aligned}$$

where $\propto$ means "proportional to" and β_γ stands for a vector of Lagrange multipliers. Note that, for the sake of expositional simplicity, we exclude the limit case $\gamma = 0$ which corresponds to the Kullback-Leibler Information Criterion estimator put forward by Kitamura and Stutzer (1997).

Since for any value of γ, the sequences $\sqrt{n}\beta_\gamma$ are asymptotically normal and asymptotically equivalent (see e.g. Imbens et al. (1998)), it is worth interpreting the implied probabilities in the particular case $\gamma = -1$, which corresponds to the Euclidean Empirical Likelihood (EEL) that is extensively documented in Antoine et al. (2007). However, it must be kept in mind that the score test extensively studied in the present paper, based on the estimator $V_n^{-1,0}$ for the variance matrix, is not the score test associated to EEL or used by Kleibergen (2005). Indeed, the first order conditions of EEL, that deliver an estimator of θ numerically equal to the continuously updated GMM estimator of Hansen et al. (1996) do not resort to the efficient estimator for the variance matrix. It is actually the reason why Antoine et al. (2007) had proposed the 3-step EEL. The first two steps are not needed here since an efficient estimator of θ under the null is obviously available.

The advantage of the Euclidean case $\gamma = -1$ is that it provides closed form formulas for

Lagrange multipliers and implied probabilities so that

$$\hat{\pi}_{i,n}^{-1,0} = \frac{1}{n} - \frac{1}{n}\bar{\psi}_n' \left(\bar{V}_n^*\right)^{-1} \psi_i.$$

This closed form formula allows Antoine et al. (2007) to give a control variable interpretation of the constrained estimator of the expectation of any integrable function of the variables ψ_i. Under the null, if the expectation with respect to the empirical distribution in (7) is denoted by $\hat{E}$ and that with respect to the implied probabilities $\hat{\pi}_{i,n}^{\gamma,0}$ by $\hat{E}^{\gamma}$, we get, for any scalar function $h(X_1)$:

$$\begin{aligned}
\hat{E}^{-1}[h(X_1)] &= \sum_{i=1}^{n} \hat{\pi}_i^{-1,0} h(X_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} h(X_i) - \frac{1}{n}\bar{\psi}_n' \left(\bar{V}_n^*\right)^{-1} \sum_{i=1}^{n} \psi_i h(X_i) \\
&= \hat{E}[h(X_1)] - \hat{E}[\psi_1] \left(\bar{V}_n^*\right)^{-1} \hat{E}[\psi_1 h(X_1)].
\end{aligned}$$

Therefore, in order to estimate $E[h(X_1)]$, we compute the sample mean of the residual of the regression of $h(X_i)$ on ψ_i. The control variable principle tells us that it is an efficient way to estimate $E[h(X_1)]$ while taking into account the information that $E[\psi_1] = 0$. In other words, as rigorously proved in Antoine et al. (2007), the estimator $\hat{E}^{-1}[h(X_1)]$ reaches the semi-parametric efficiency bound for estimation of $E[h(X_1)]$ under the null hypothesis $H_0 : \theta = \theta_0$. The aforementioned first order equivalence implies that the semi-parametric efficiency bound is also reached under the null by a bunch of constrained estimators, associated to any value of the power γ:

$$\hat{E}^{\gamma}[h(X_1)] = \sum_{i=1}^{n} \hat{\pi}_i^{\gamma,0} h(X_i).$$

The focus of our interest in this paper will be the use, for the purpose of score testing, of two constrained estimators $V_n^{\gamma,0}$ of the variance matrix associated respectively to the values $\gamma = -1$ and $\gamma = 0$. As explained above, the use of $V_n^{-1,0}$ is in the line of score testing in the context of 3step-EEL. The use of $V_n^{0,0}$ is in the line of score testing with Empirical Likelihood (EL) as studied by Guggenberger and Smith (2005) since the minimization of (6) in the limit case $\gamma \to 0$ amounts to the maximization of the empirical likelihood $\sum_{i=1}^{n} \log(\pi_i)$. The control variable interpretation above allows us to interpret these constrained estimators of the variance matrix as a result of a kind of shrinkage, that is replacing cross products of components of $\psi_i \equiv (X_i - g(\theta_0)$ by the residual of their regression on the moment conditions. This interpretation is exact in the EEL case and asymptotic in the EL case.

Note also that the implied probabilities in the EEL case may take negative values in finite samples. It may be an issue for positive definite estimation of the variance matrix. Antoine et al. (2007) have proposed an additional shrinkage step to get rid of this non-positivity issue. They consider instead the following implied probabilities

$$\hat{\pi}_i^{-1,0,p} \quad = \quad \frac{1}{1+\varepsilon_n}\hat{\pi}_i^{-1,0} + \frac{\varepsilon_n}{1+\varepsilon_n}.\frac{1}{n}, \text{ where } \varepsilon_n = -n \times \min\left\{\min_{1\leq i\leq n} \hat{\pi}_i^{-1,0}, 0\right\}.$$

They show that this additional shrinkage does not prevent from reaching the semi-parametric efficiency bound under the null.

Irrespective of the use of implied probabilities $\hat{\pi}_i^{-1,0}, \hat{\pi}_i^{-0,0}$ or $\hat{\pi}_i^{-1,0,p}$, we end up with an estimator of the variance matrix asymptotically equivalent under the null to

$$V_n^{-1,0} = \bar{V}_n - \bar{\psi}_n' \left(\bar{V}_n^*\right)^{-1} \frac{1}{n}\sum_{i=1}^n \psi_i\psi_i\psi_i'. \tag{13}$$

Formula (13) clearly shows that the proposed improvement for estimation of the covariance matrix will matter when the moment conditions display some kind of multivariate skewness. This will be compellingly illustrated in the Student-type examples in the next sections.

3 Theoretical analysis

For the sake of exposition, let us start with a very simple, albeit representative, example where θ and ψ_i are both scalars, i.e., $p = K = 1$. So test of $H_0 : \theta = \theta_0$ essentially boils down to the test of an univariate population mean for the random variable ψ, i.e., $H_0 : E[\psi] = 0$. In this case the score statistics, as well as the the Wald and QLR statistics, can be represented by the so called t-ratio. Extension to vector-valued ψ for just-identified model is straightforward and hence, for now, we focus just on the t-ratio because this is simple and ubiquitous enough to explain the main idea of the paper in a way that could be of interest to many.

Denoting V by σ^2 (to adhere to the convention), the t-ratios based on $\psi_1, \ldots, \psi_n$ and the variance forms of variance estimators – infeasible, naive (see (11)) and modified (see (13)) – for testing $H_0 : E[\psi] = 0$ are defined as:

$$\begin{aligned}
\text{infeasible:} \quad t_n &:= \sqrt{n}\frac{\bar{\psi}_n}{\sigma}, & (14)\\
\text{naive:} \quad \bar{t}_n &:= \sqrt{n}\frac{\bar{\psi}_n}{\bar{\sigma}_n} \text{ where } \bar{\sigma}_n^2 := \frac{1}{n}\sum_{i=1}^{n}\psi_i^2 (\equiv \bar{V}_n), & (15)\\
\text{modified:} \quad \widehat{t}_n &:= \sqrt{n}\frac{\bar{\psi}_n}{\widehat{\sigma}_n} \text{ where } \widehat{\sigma}_n^2 := \frac{1}{n}\sum_{i=1}^{n}\left[1 - \frac{\bar{\psi}_n\psi_i}{\bar{\sigma}_n^2}\right]\psi_i^2 (\equiv V_n^{-1,0}). & (16)
\end{aligned}$$

Since we are only concerned with the behavior of the tests when the null hypothesis is assumed to be true, the precise measure of cost due to the unknown σ^2 considered here is $|P[t_n \leq \tau] - P[\bar{t}_n \leq \tau]|$ or $|P[t_n \leq \tau] - P[\widehat{t}_n \leq \tau]|$. The main result, in terms of absolute difference in p-values, is given in the following proposition. This result heavily draws from Lemma A.1. of Cavanagh (1983). Our Monte-Carlo simulations focus on the size of the tests, i.e., τ is taken as the critical values for the concerned tests.

Proposition 3.1 *Assume that*

(i) $E[\psi_i^s] = \mu_s$ *for* $s = 1, \ldots, r$ *exists for* r *sufficiently large such that (iii) holds,*

(ii) $Z_{1n} := \sum_{i=1}^{n}\psi_i/n$, $Z_{2n} := \sum_{i=1}^{n}\psi_i^2/n$ *and* $Z_{3n} := \sum_{i=1}^{n}\psi_i^3/n$ *are such that*

$$\sqrt{n}\begin{pmatrix} Z_{1n} - \mu_1 \\ Z_{2n} - \mu_2 \\ Z_{3n} - \mu_3 \end{pmatrix} \xrightarrow{d} N\left(0, \Sigma := \begin{bmatrix} \mu_2 - \mu_1^2 & & \\ \mu_3 - \mu_2\mu_1 & \mu_4 - \mu_2^2 & \\ \mu_4 - \mu_3\mu_1 & \mu_5 - \mu_3\mu_2 & \mu_6 - \mu_3^2 \end{bmatrix}\right),$$

(iii) the t-ratios t_n, $\bar{t}_n$ *and* $\widehat{t}_n$ *defined in* (14), (15) *and* (16) *can be expressed as*

$$\begin{aligned}
t_n &:= \sqrt{n}\frac{\bar{\psi}_n}{\sigma} = \sqrt{n}\frac{Z_{1n}}{\sigma}\\
\bar{t}_n &:= \sqrt{n}\frac{\bar{\psi}_n}{\bar{\sigma}_n} = t_n + \sum_{j=1}^{\infty} a_j\left[1 - \frac{\bar{\sigma}_n^2}{\sigma^2}\right]^j t_n = t_n + \frac{\bar{A}_n}{\sqrt{n}} + o_p(1/\sqrt{n})\\
\widehat{t}_n &:= \sqrt{n}\frac{\bar{\psi}_n}{\widehat{\sigma}_n} = t_n + \sum_{j=1}^{\infty} a_j\left[1 - \frac{\widehat{\sigma}_n^2}{\sigma^2}\right]^j t_n = t_n + \frac{\widehat{A}_n}{\sqrt{n}} + o_p(1/\sqrt{n})
\end{aligned}$$

where $\sigma^2 := \mu_2 - \mu_1^2$, $\bar{\sigma}_n^2 := \sum_{i=1}^{n}\psi_i^2/n = Z_{2n}$, $\widehat{\sigma}_n^2 := \sum_{i=1}^{n}\left[1 - \bar{\psi}_n\psi_i/\bar{\sigma}_n^2\right]\psi_i^2/n = Z_{2n} - Z_{3n}Z_{1n}/Z_{2n}$ *and* $a_j := \prod_{l=1}^{j}(2l-1)/2$ *for* $j = 1, 2, \ldots$. *The functions* $\bar{A}_n$, $\widehat{A}_n$ *are smooth functions of* t_n *and* $\bar{Y}_n := \sqrt{n}(Z_{2n} - \sigma^2)$, *and of* t_n, $\widehat{Y}_n := (\sqrt{n}(Z_{2n} - \sigma^2), \sqrt{n}(Z_{3n} - \mu_3)))$

respectively and are given by

$$
\begin{aligned}
\bar{A}_n \equiv \bar{A}_n(t_n, \bar{Y}_n) &:= -\frac{a_1}{\sigma^2}\sqrt{n}\left(\bar{\sigma}_n^2 - \sigma^2\right) t_n = -\frac{a_1}{\sigma^2}\sqrt{n}\left(Z_{2n} - \sigma^2\right) t_n \\
\widehat{A}_n \equiv \bar{A}_n(t_n, \widehat{Y}_n) &:= -\frac{a_1}{\sigma^2}\sqrt{n}\left(\widehat{\sigma}_n^2 - \sigma^2\right) t_n = \bar{A}_n + \frac{a_1}{\sigma^2}\frac{\sqrt{n}Z_{1n}Z_{3n}}{\sigma^2[1-(1-\bar{\sigma}_n^2/\sigma^2)]} t_n \\
&= \bar{A}_n + a_1 t_n^2 \frac{Z_{3n}}{\sigma^3}\sum_{j=0}^{\infty}\left(1 - \frac{\bar{\sigma}_n^2}{\sigma^2}\right)^j = \bar{A}_n + a_1 t_n^2\frac{Z_{3n}}{\sigma^3} + o_p(1/\sqrt{n}).
\end{aligned}
$$

Then under assumptions (i), (ii) and (iii), and for any τ at which the density $f_{t_n}(\tau)$ of t_n exists and the conditional distributions $\bar{A}_n|t_n$ and $\widehat{A}_n|t_n$ are defined, the following results will hold:

$$
P\left[\bar{t}_n \le \tau\right] = P\left[t_n \le \tau\right] - \frac{1}{\sqrt{n}}E[\bar{A}_n|t_n = \tau]f_{t_n}(\tau) + o_p(1/\sqrt{n}) \quad (17)
$$

$$
P\left[\widehat{t}_n \le \tau\right] = P\left[t_n \le \tau\right] - \frac{1}{\sqrt{n}}E[\widehat{A}_n|t_n = \tau]f_{t_n}(\tau) + o_p(1/\sqrt{n}). \quad (18)
$$

Proof: Follows by construction from the proof of Lemma A.1. (pages 20-21) of Cavanagh (1983) and ignoring the $o_p(1/\sqrt{n})$ terms. ■

The improvement in approximation of the p-values in Proposition 3.1 due to the use of the modified estimator of σ^2 can be explicitly viewed by computing $E[\bar{A}_n|t_n]$ and $E[\widehat{A}_n|t_n]$ from Corollary 3.2 (proof is obvious).

Corollary 3.2 *Under assumptions (ii) and (iii) of Proposition 3.1,*

$$
\begin{aligned}
E[\bar{A}_n|t_n] &= -a_1 t_n^2 \frac{\mu_3 - \mu_2\mu_1}{\sigma^3} \\
&= -a_1 t_n^2 \frac{\mu_3}{\sigma^3} \text{ under } H_0 : E[\psi](\equiv \mu_1) = 0, \\
E[\widehat{A}_n|t_n] &= E[\bar{A}_n|t_n] + \frac{a_1 t_n^2}{\sigma^3}E[Z_{3n}|t_n] = -a_1 t_n^2\frac{\mu_3}{\sigma^3} + a_1 t_n^2 \frac{\mu_3}{\sigma^3} - \frac{1}{\sqrt{n}}a_1 t_n^3 \frac{\mu_4 - \mu_3\mu_1}{\sigma^4} \\
&= -\frac{1}{\sqrt{n}}a_1 t_n^3\frac{\mu_4}{\sigma^4} \text{ under } H_0 : E[\psi](\equiv \mu_1) = 0.
\end{aligned}
$$

Remarks:

(i) Gains in terms of order of magnitude is obtained if the underlying distribution of ψ is skewed.[1]

(ii) We stress again that the use of the modified estimator does not correct for all the skewness-

[1]Even otherwise, in unreported simulations similar to (iii) but with $\psi_i \sim N(0,1)$, we found some gain by the use of the modified estimator.

related errors of approximation of the exact distribution of the t-ratio by the first-order asymptotics, i.e., $N(0,1)$. As can be seen from a formal Edgeworth expansion of the infeasible statistic t_n, the effect of skewness is still present in the first-order approximation error of its exact distribution. To see this, note that under the assumptions – $E[|\psi|^{j+2}] < \infty$ and $\limsup_{s\to\infty} E[exp(is\psi)] < 1$ (Cramer's condition) –

$$P(t_n \le \tau) = \Phi(\tau) + \frac{1}{6\sqrt{n}}\frac{\mu_3}{\sigma^3}(2\tau^2+1)\phi(\tau) + o_p(1/\sqrt{n})$$

where $\phi(\tau)$ and $\Phi(\tau)$ are respectively the pdf and cdf of a N(0,1) distribution, $i := \sqrt{-1}$ (only here). Typically re-sampling methods like bootstrap can be useful for completely removing the effect of skewness up to order $o_p(1/\sqrt{n})$. Other methods of modifying the t-ratio that are similar in spirit to our proposal, such as those proposed by Johnson (1978), Lyon et al. (1999) and Yanagihara and Yuan (2005), also do not achieve this.

(iii) In Figure-1 we plot the empirical size of the t-tests of $H_0 : E[\psi] = 0$ against $H_1 : E[\psi] > 0$ based on the infeasible, naive and modified t-ratios against the nominal level varied from 0 to 1. Results are reported based on 5000 Monte-Carlo trials. The underlying distributions of ψ_i, viz. $\chi_1^2 - 1$ and $LN(0,1) - exp(.5)$, are skewed. As can be seen from the plots, for small sample size $n = 50$, there is considerable difference between the naive and the infeasible t-tests; and this is significantly mitigated by the use of the modified t-ratios. This difference, as is common with the first-order asymptotics, is less pronounced in large samples where the sample size dominates the effect of skewness in order of magnitudes.

We conjecture without a formal proof that extension to over-identified models will be similar in nature, albeit introducing additional issues. The relevant asymptotic expansions are given in details in appendix (section 5) as well as a discussion of the additional issues raised for performance comparisons. The key point is that, even though the first order conditions of efficient GMM amount to picking a subset of just-identified moment conditions, the efficient estimation of the variance matrix under the null will take advantage of the whole set of moment conditions. In fact, simulation results reported below show that the empirical size can be made closer to the nominal level (based on the first-order asymptotics) by the use of the score statistics that involve, respectively, the modified estimators $V_n^{-1,0}$ (i.e., EEL) and $V_n^{-0,0}$ (i.e., EL) of the asymptotic variance.

4 Covariance structure model: A Monte-Carlo experiment

The data generating process is taken from Horowitz (1998) and was originally proposed by Altonji and Segal (1996). For $i = 1, \ldots, n$ and $j = 1, \ldots, J+1$ we generate $Y_{i,j} \overset{\text{i.i.d.}}{\sim} f(y, \theta)$ where f is: uniform, normal, t_{10}, exponential and log-normal, all of them standardized to have expectation 0 and variance θ. We further generate $Z_{i,j} = (Y_{i,j} + .5 \times Y_{i,j+1})/\sqrt{1+(.5)^2}$ for $j = 1, \ldots, J$ and $i = 1, \ldots, n$. Therefore, for each $i = 1, \ldots, n$, and $j = 1, \ldots, J$ we have $E[Z_{i,j}] = 0$, $V(Z_{i,j}) = \theta$, $Cov(Z_{i,j}, Z_{i,j+1}) = .5 \times \theta/(1+(.5)^2)$ and $Cov(Z_{i,j}, Z_{i,j+s}) = 0$ for $s \geq 2$. Therefore, taking $X_i = (Z_{i,1}, \ldots, Z_{i,J})$ and $K = 2J - 1$, we have

$$\begin{aligned}
\psi'(X_i, \theta) \quad = \quad & \left[Z_{i,1}^2 - \theta, Z_{i,2}^2 - \theta, \ldots, Z_{i,J}^2 - \theta, Z_{i,1}Z_{i,2} - .5 \times \theta/(1+(.5)^2),\right. \\
& \left. Z_{i,2}Z_{i,3} - .5 \times \theta/(1+(.5)^2), \ldots, Z_{i,J-1}Z_{i,J} - .5 \times \theta/(1+(.5)^2)\right].
\end{aligned}$$

The true value $\theta = 1$ in our simulations and is considered unknown for testing $H_0 : \theta = \theta_0$. Since we are interested in studying the finite-sample gains due to the use of the modified variances under the null hypothesis, we also take the hypothesized value $\theta_0 = 1$. We consider $J = 4, 6, 8, 10$ (i.e., $k = 7, 11, 15, 19$) in the above setup, and perform the score test for $H_0 : \theta = \theta_0$ based on the three score statistics that use for V_n^0, respectively the naive estimator $\bar{V}_n$ and the modified estimators $V_n^{-1,0}$ (i.e., EEL) and $V_n^{-0,0}$ (i.e., EL) of the asymptotic variance V.

The results are reported based on 5000 Monte-Carlo trials. Finite Sample rejection rates for the score tests with nominal size 1%, 5% and 10% are reported in Tables 1 (for sample size $n = 500$) and 2 (for sample size $n = 500$). Naive score test performs poorly for $n = 500$. When $n = 5000$, i.e., sample size is increases without any change in the 3rd (population) moment of the elements of the moment vector, improvements are noticed. Also, not surprisingly, use of the implied probabilities (i.e., EEL or EL) provides significant improvement over the use of the naive empirical probabilities. The ordering of the three methods are similar across different distributional assumptions, but the quality of approximation gets worse with thicker tails of the moment vector.

To study the properties of the three score statistics more thoroughly, we plot in Figures 2-6 the empirical cdf of the score statistics and compare them with the cdf of the χ_1^2 distribution (first-order asymptotic limit). The top panel corresponds to $n = 500$ while the bottom to $n = 5000$. The closeness of the empirical cdfs to the χ_1^2 cdf (not only at the tails) when the score statistic uses variance estimators based on EEL and EL (especially the latter) is a major improvement over the use of the naive estimator of asymptotic variance.

Number of Moments $= 2J-1$		7			11			15			19		
DGP	Nominal level	Naive	EEL	EL	Naive	EEL	EL	Naive	EEL	EL	Naive	EEL	EL
Uniform	1	1.36	1.14	0.92	1.84	1.46	1.08	1.9	1.4	0.98	2	1.78	1.02
	5	5.9	5.5	4.7	7.26	6.5	5.32	7.52	6.66	5.08	7.82	6.82	4.96
	10	11.12	10.58	9.74	13.86	12.48	10.52	13.56	12.56	10.34	14.5	12.86	10.28
Normal	1	2.5	1.64	1.1	3.74	2.2	1.3	4.96	2.18	1.2	5.76	2.58	1.34
	5	9.14	6.64	5.28	11.1	7.74	5.7	12.84	8.6	6.24	14.44	8.72	6.24
	10	15.1	12.46	10.66	18.22	13.98	11.72	20.4	14.56	11.8	22.42	14.52	11.46
t: df 10	1	4.62	2.88	1.3	6.72	3.44	1.6	8.74	3.72	1.82	10.14	4.28	1.84
	5	11.76	8.7	6.56	16.22	10.32	7	19.28	10.6	6.78	22.54	11.7	7.02
	10	18.56	14.32	11.76	23.98	17.36	13.32	27.46	17.06	12.58	31.42	18.5	13.16
Exponential	1	13.74	5.96	2.42	21.38	8.6	4.42	27.88	9.04	3.82	33.8	10.48	5.06
	5	24.58	14.54	9.64	34.78	17.34	11.86	44.16	19.26	12.3	50.42	20.64	13.78
	10	32.8	21.44	16.2	43.64	24.38	18.28	52.94	26.76	20.04	59.9	29.02	21.56
Lognormal	1	60.98	33.48	23.8	78.9	47.48	36.68	89.52	59.94	49.98	94.2	68.44	59.6
	5	71.74	47.42	39.34	87.3	61.84	53.82	94.36	71.86	65.96	97.62	78.92	74.92
	10	77.14	55.36	49.2	90.64	70.28	62.68	96.12	76.94	73.38	98.76	83.02	81.1

Table 1: Finite-sample rejection rate (in %) of the true parameter value by score tests with nominal size 1%, 5% and 10%. Naive, EEL and EL correspond to the score tests based on the statistics using $\bar{V}_n$, $V_n^{-1,0}$ and $V_n^{0,0}$ respectively for V_n^0. Number of observations $n = 500$. Number of Monte-Carlo trials is 5000.

Number of Moments = $2J-1$		7			11			15			19		
DGP	Nominal level	Naive	EEL	EL	Naive	EEL	EL	Naive	EEL	EL	Naive	EEL	EL
Uniform	1	0.9	0.7	0.68	1.18	1.02	0.94	0.94	0.84	0.8	1.1	1.16	1.08
	5	5.02	4.82	4.78	5.08	4.92	4.8	5.12	5.1	5.02	5.64	5.36	5.18
	10	9.9	9.66	9.48	10.1	10.24	9.98	10.68	10.3	9.94	10.88	10.7	10.44
Normal	1	1.02	1.04	0.96	1.08	1	0.9	1.72	1.52	1.36	1.4	1.16	1.04
	5	5.44	5.18	5.06	5.34	5.04	4.78	6.2	5.6	5.48	5.68	5.26	4.92
	10	10.5	9.94	9.76	10.34	9.7	9.46	11	10.38	10.12	10.94	9.8	9.48
t: df 10	1	1.04	1.04	0.9	1.58	1.44	1.2	1.36	1.2	0.86	1.78	1.08	0.88
	5	5.52	5.6	5.02	5.94	5.66	5.1	6.36	5.24	4.7	6.82	5.66	5.14
	10	11.18	10.68	9.88	11.78	10.56	9.78	11.68	10.84	9.96	13.3	10.82	10.24
Exponential	1	1.92	1.22	1	2.36	1.42	1.1	3.08	1.44	1.1	3.74	1.72	1.24
	5	7.72	5.84	4.98	9.1	5.96	4.94	10.26	6.2	5.26	11.94	6.62	5.46
	10	13.46	11.58	10.6	15.68	12.1	10.8	16.92	11.08	9.82	19.24	12.34	10.7
Lognormal	1	23.02	10.56	5.26	34.98	13.54	7.28	44.5	16.5	9.48	54.86	20.98	12.46
	5	36.92	21.46	14.38	51.8	25.94	18.32	61.18	29.68	21.56	70.68	34.28	26.3
	10	45.86	29.36	22.36	60.2	34.1	26.8	69.06	38.26	30.58	77.9	43.98	35.96

Table 2: Finite-sample rejection rate (in %) of the true parameter value by score tests with nominal size 1%, 5% and 10%. Naive, EEL and EL correspond to the score tests based on the statistics using $\bar{V}_n$, $V_n^{-1,0}$ and $V_n^{0,0}$ respectively for V_n^0. Number of observations $n = 5000$. Number of Monte-Carlo trials is 5000.

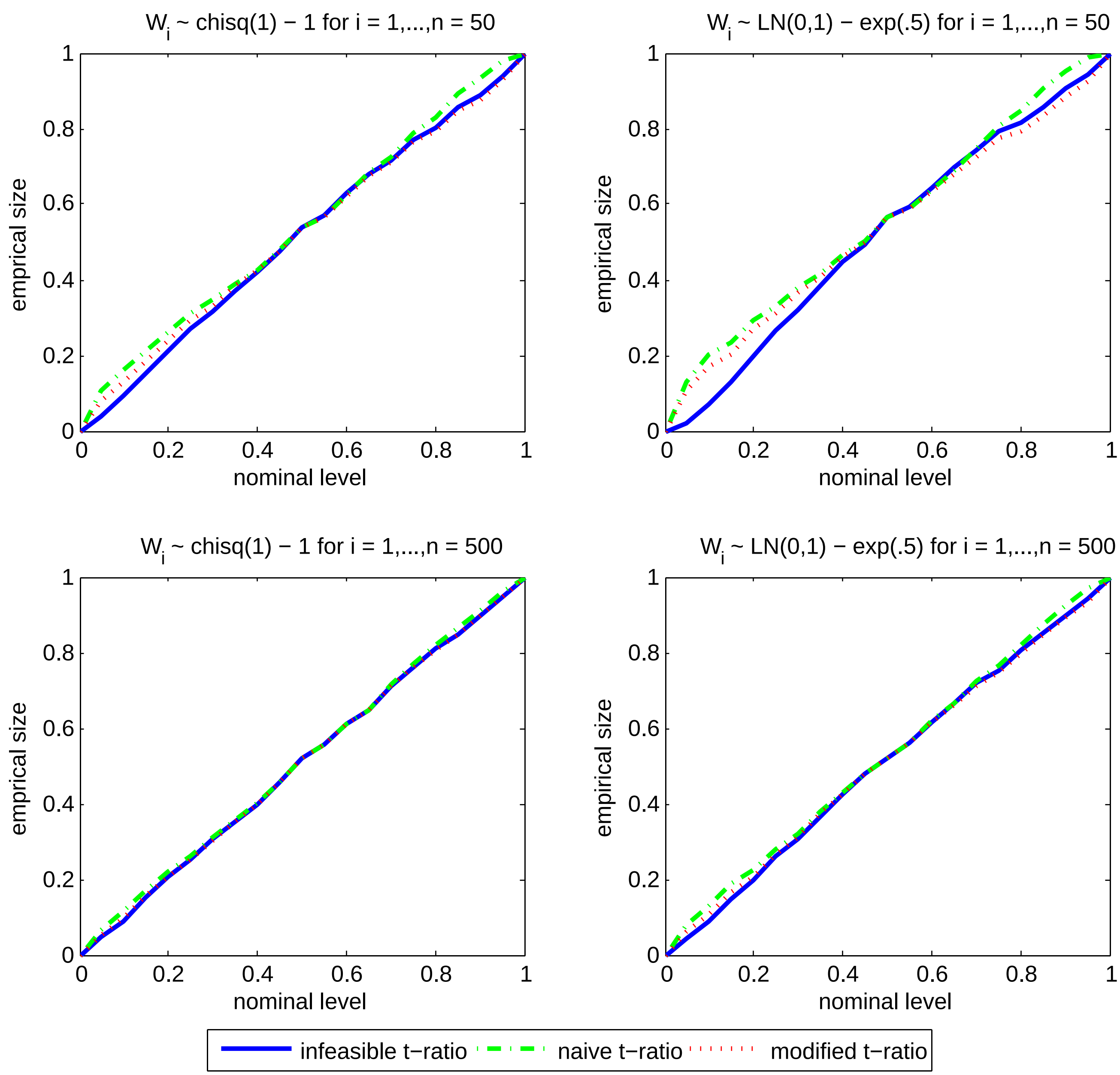

Figure 1: Empirical size v/s nominal level of t-tests for population mean (against alternative that true value is greater than hypothesized value) based on the infeasible t-ratio (blue), naive t-ratio (green) and the modified t-ratio (red). Number of Monte-Carlo Trials is 5000.

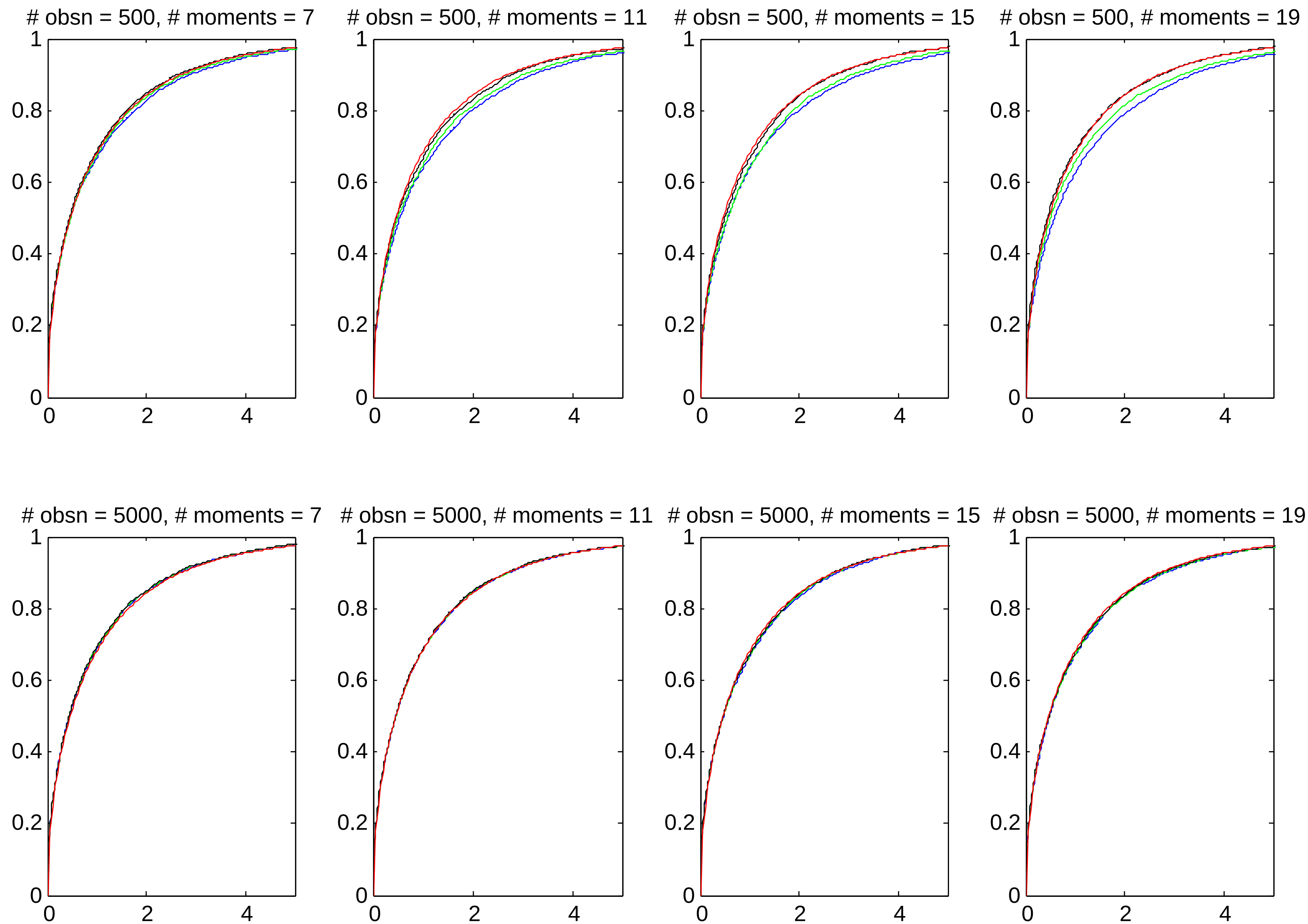

Figure 2: DGP- Uniform: Empirical cdf of score statistics using Naive (blue), EEL (green) and EL (black). CDF of χ^2_1 (red). Naive, EEL and EL correspond to the score tests based on the statistics using $\bar{V}_n$, $V_n^{-1,0}$ and $V_n^{0,0}$ respectively for V^0, n. Number of Monte-Carlo Trials is 5000.

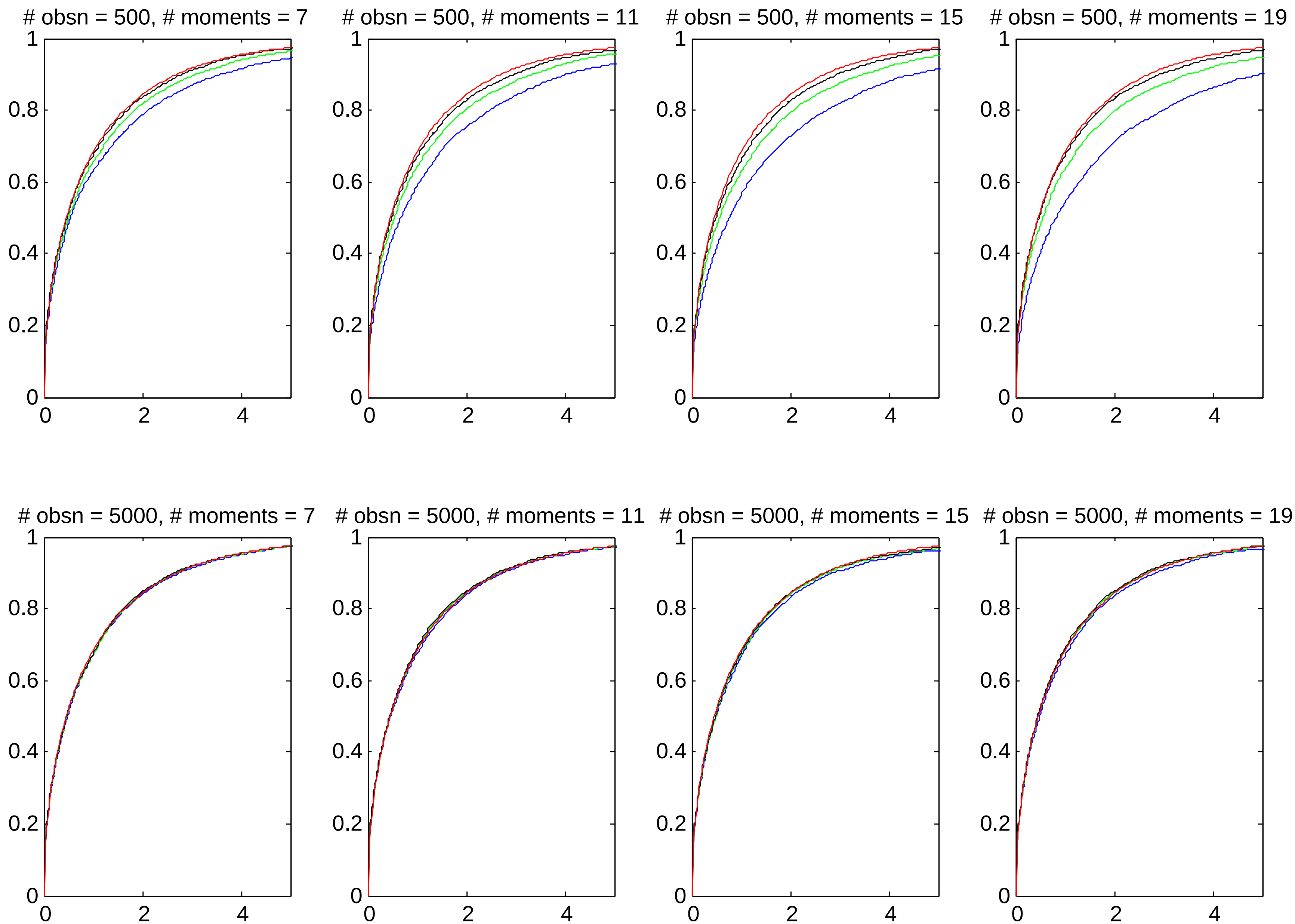

Figure 3: DGP- Normal: Empirical cdf of score statistics using Naive (blue), EEL (green) and EL (black). CDF of χ^2_1 (red). Naive, EEL and EL correspond to the score tests based on the statistics using $\bar{V}_n$, $V_n^{-1,0}$ and $V_n^{0,0}$ respectively for V^0, n. Number of Monte-Carlo Trials is 5000.

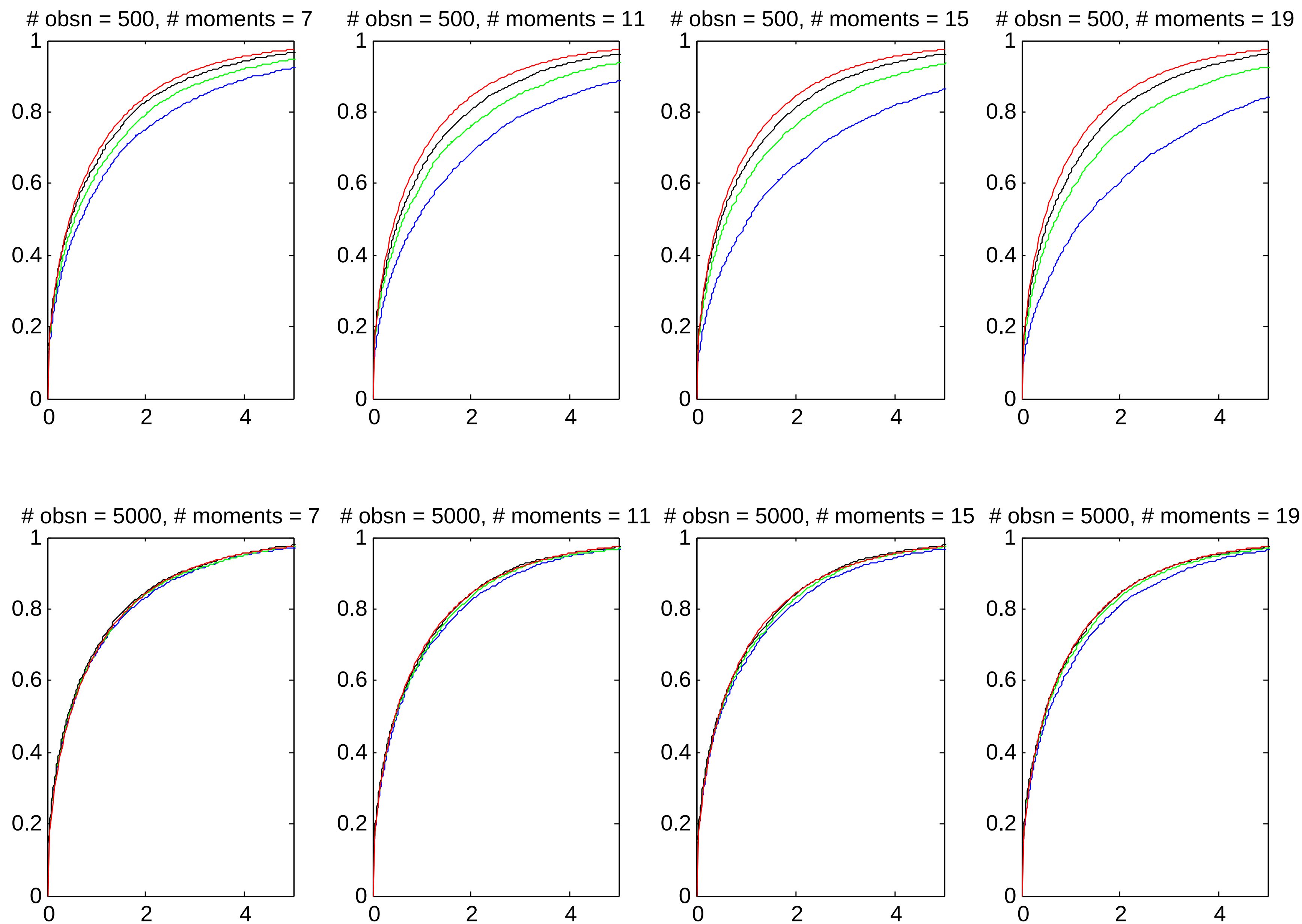

Figure 4: DGP- t with d.f. 10: Empirical cdf of score statistics using Naive (blue), EEL (green) and EL (black). CDF of χ_1^2 (red). Naive, EEL and EL correspond to the score tests based on the statistics using $\bar{V}_n$, $V_n^{-1,0}$ and $V_n^{0,0}$ respectively for V_n^0. Number of Monte-Carlo Trials is 5000.

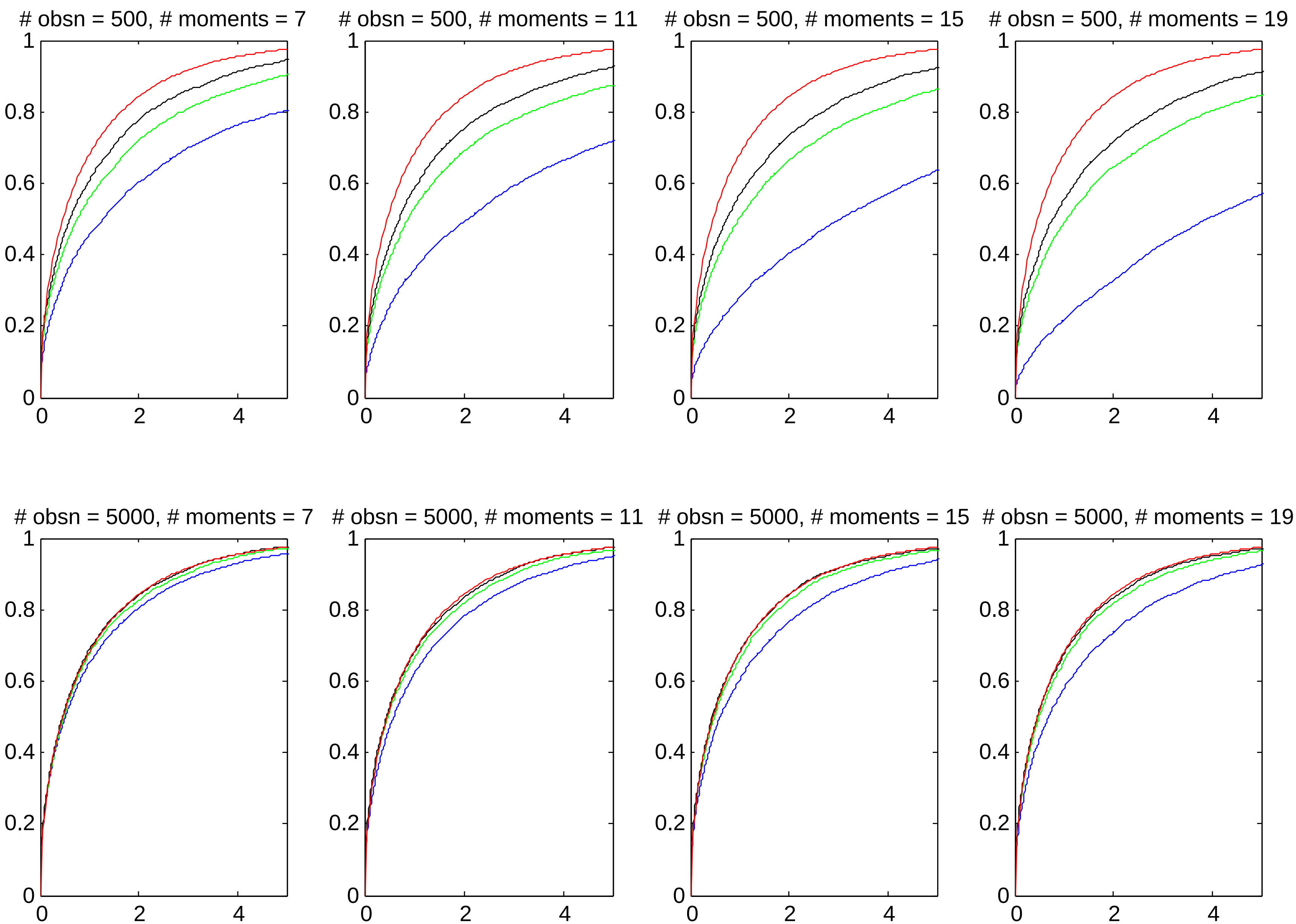

Figure 5: DGP- Exponential: Empirical cdf of score statistics using Naive (blue), EEL (green) and EL (black). CDF of χ_1^2 (red). Naive, EEL and EL correspond to the score tests based on the statistics using $\bar{V}_n$, $V_n^{-1,0}$ and $V_n^{0,0}$ respectively for V_n^0. Number of Monte-Carlo Trials is 5000.

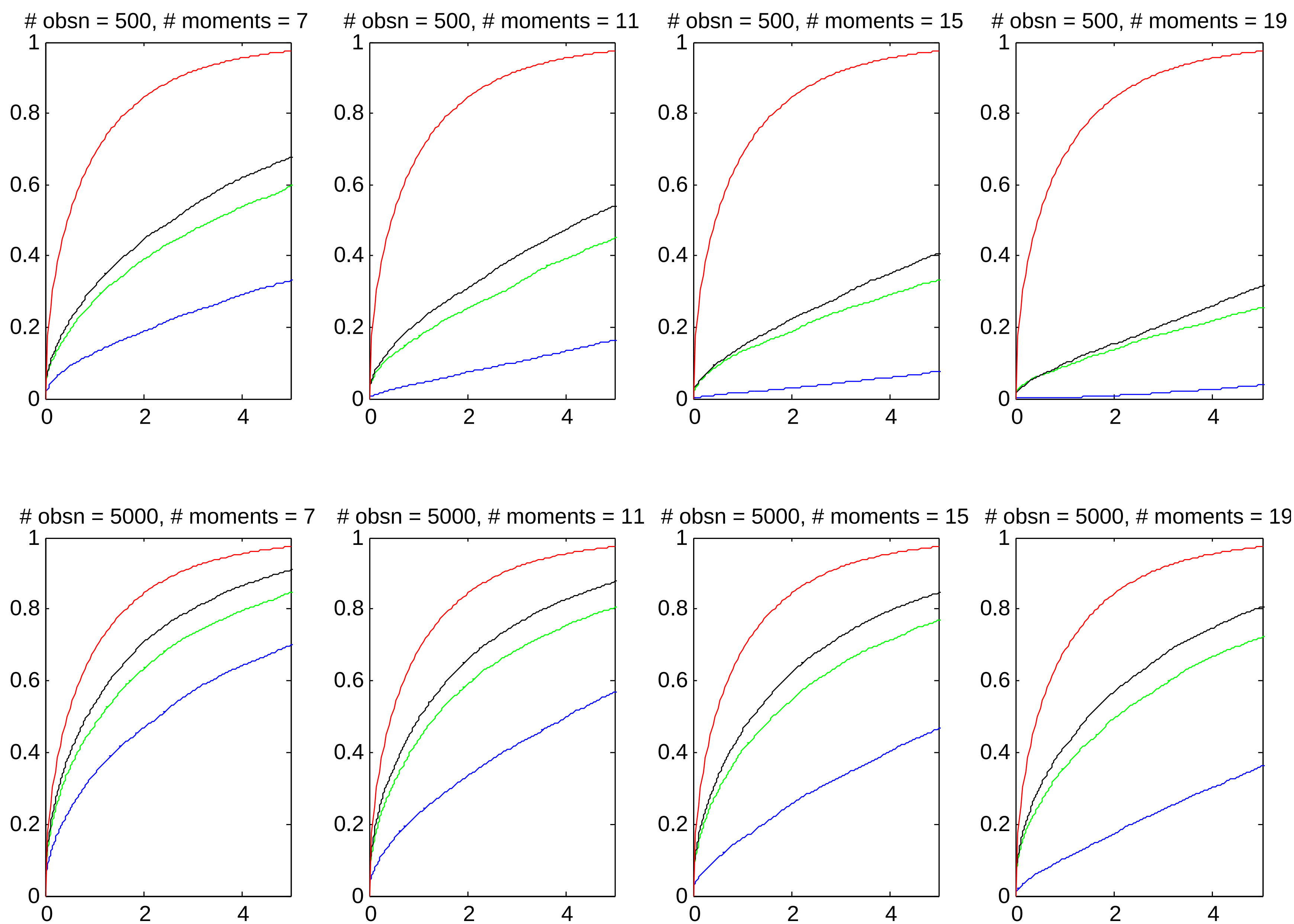

Figure 6: DGP- Log Normal: Empirical cdf of score statistics using Naive (blue), EEL (green) and EL (black). CDF of χ^2_1 (red). Naive, EEL and EL correspond to the score tests based on the statistics using $\bar{V}_n$, $V_n^{-1,0}$ and $V_n^{0,0}$ respectively for V_n^0. Number of Monte-Carlo Trials is 5000.

5 Appendix: Why difficult to extend to over-identified models

Consider an over-identified model similar to our Covariance structure simulation example where θ is a scalar and G is a $K \times 1$ vector of constants. Consider any $K \times K$ matrix V_n^0 such that $\sqrt{n}(V_n^0 - V) = \mathcal{O}_p(1)$. Typically $\bar{V}_n$, $V_n^{-1,0}$ and $V_n^{-1,0}$ will satisfy this condition for V_n^0. Now, define

$$t_n^O(V_n^0) := \frac{G'(V_n^0)^{-1}\sqrt{n}\bar{\psi}_n}{\sqrt{G'(V_n^0)^{-1}G}},$$

where the superscript O in t_n is to signify an over-identified model and the dependence on V_n^0 is made explicit because the theme of the paper is the choice of V_n^0. Assume that sufficient moments of ψ_i exist for the following approximation of $t_n^O(V_n^0)$ to be valid.

$$\begin{aligned}
& t_n^O(V_n^0) \\
= & \left[G'V^{-1/2}\left(I_k + V^{-1/2'}(V - V_n^0)V^{-1/2}\right)V^{-1/2'}G + o_p\left(\frac{1}{\sqrt{n}}\right)\right]^{-1/2} \\
& \times \left[G'V^{-1/2}\left(I_k + V^{-1/2'}(V - V_n^0)V^{-1/2}\right)V^{-1/2'}\sqrt{n}\bar{\psi}_n + o_p\left(\frac{1}{\sqrt{n}}\right)\right] \\
= & \left[1 + \left(1 - \frac{G'V^{-1}V_n^0V^{-1}G}{G'V^{-1}G} + o_p\left(\frac{1}{\sqrt{n}}\right)\right)\right]^{-1/2}\left[\frac{G'V^{-1}\sqrt{n}\bar{\psi}_n}{\sqrt{G'V^{-1}G}} + \frac{G'V^{-1}(V - V_n^0)V^{-1}\sqrt{n}\bar{\psi}_n}{\sqrt{G'V^{-1}G}}\right] + o_p\left(\frac{1}{\sqrt{n}}\right) \\
= & \left[1 - \frac{1}{2}\left(1 - \frac{G'V^{-1}V_n^0V^{-1}G}{G'V^{-1}G}\right) + o_p\left(\frac{1}{\sqrt{n}}\right)\right]\left[\frac{G'V^{-1}\sqrt{n}\bar{\psi}_n}{\sqrt{G'V^{-1}G}} + \frac{G'V^{-1}(V - V_n^0)V^{-1}\sqrt{n}\bar{\psi}_n}{\sqrt{G'V^{-1}G}}\right] + o_p\left(\frac{1}{\sqrt{n}}\right) \\
= & \frac{G'V^{-1}\sqrt{n}\bar{\psi}_n}{\sqrt{G'V^{-1}G}} - \frac{G'V^{-1}\sqrt{n}(V_n^0 - V)V^{-1}G}{2\sqrt{n}G'V^{-1}G}\frac{G'V^{-1}\sqrt{n}\bar{\psi}_n}{\sqrt{G'V^{-1}G}} \\
& - \frac{G'V^{-1}\sqrt{n}(V_n^0 - V)V^{-1/2}M(V^{-1/2'}G)V^{-1/2'}\sqrt{n}\bar{\psi}_n}{\sqrt{n}\sqrt{G'V^{-1}G}} + o_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

where for any $r \times s$ matrix C of full column-rank, $P(C) := C(C'C)^{-1}C'$ and $M(C) := I_r - P(C)$. Therefore,

$$t_n^O(V_n^0) = t_n^O + \frac{B_{1n}(V_n^0)}{\sqrt{n}} + \frac{B_{2n}(V_n^0)}{\sqrt{n}} + o_p\left(\frac{1}{\sqrt{n}}\right) \text{ where} \tag{19}$$

$$B_{1n}(V_n^0) := -\frac{G'V^{-1}\sqrt{n}(V_n^0 - V)V^{-1}G}{2G'V^{-1}G}t_n^O \tag{20}$$

$$B_{2n}(V_n^0) := -\frac{G'V^{-1}\sqrt{n}(V_n^0 - V)V^{-1/2}M(V^{-1/2'}G)V^{-1/2'}\sqrt{n}\bar{\psi}_n}{\sqrt{G'V^{-1}G}}, \tag{21}$$

$$\text{and } t_n^O \equiv t_n^O(V_n^0 = V) = \frac{G'V^{-1}\sqrt{n}\bar{\psi}_n}{\sqrt{G'V^{-1}G}}. \tag{22}$$

To see the connection with the t-ratios as defined in the statement of Proposition 3.1 in the context of a just-identified model, let us perform the following exercise. This will help us to see the effect of over-identifying restrictions.

(i) INFEASIBLE – $V_n^0 = V$**:**

Define the sequence of i.i.d. random variables $W_i := G'V^{-1}\psi_i$ with variance $\sigma^2 := G'V^{-1}G$ so that $\sqrt{n}\bar{W}_n/\sigma = t_n^O$, the infeasible t-ratio, where $\bar{W}_n := \sum_{i=1}^n W_i/n$. This shows the similarity of the infeasibles, t_n and t_n^O, and implies that there is no effect of (a finite number of) over-identifying restrictions if V is known.

(ii) NAIVE – $V_n^0 = \bar{V}_n$**:**

Define the sequence of i.i.d. random variables $W_i := G'V^{-1}\psi_i$ with variance $\sigma^2 := G'V^{-1}G$ and the naive estimate of σ^2 as $\bar{\sigma}_n^2 := \sum_{i=1}^n W_i^2/n$. Therefore, $B_{1n}(V_n^0 = \bar{V}_n) = -\frac{1}{2}\sqrt{n}(\bar{\sigma}_n^2 - \sigma^2)t_n^O$, which is same as $\bar{A}_n$ in the statement of Proposition 3.1. Now, recalling from (17) that

$$P\left[\bar{t}_n^O \leq \tau\right] \equiv P\left[t_n^O(V_n^0 = \bar{V}_n) \leq \tau\right] = P\left[t_n^O \leq \tau\right] - \frac{1}{\sqrt{n}}E[\bar{A}_n|t_n^O = \tau]f_{t_n^O}(\tau) + o_p(1/\sqrt{n}),$$

with $\bar{A}_n = B_{1n}(V_n^0 = \bar{V}_n) + B_{2n}(V_n^0 = \bar{V}_n)$, it is clear that in this case, the difference due to over-identifying restrictions arises from the term $E[B_{2n}(V_n^0 = \bar{V}_n)|t_n^O]$. We analyze this term now. To avoid notational clutter define the rotated vectors $\breve{\psi}_i := V^{-1/2'}\psi_i$ and $\breve{G} := V^{-1/2'}G$, which give $t_n^O = \breve{G}'\sqrt{n}\bar{\breve{\psi}}_n/\sqrt{\breve{G}'\breve{G}}$ where $\bar{\breve{\psi}}_n := \sum_{i=1}^n \breve{\psi}_i/n$ and, similarly, also give

$$B_{2n}(\bar{V}_n) \equiv B_{2n}(V_n^0 = \bar{V}_n) = -\frac{\breve{G}'}{\sqrt{\breve{G}'\breve{G}}}\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \breve{\psi}_i\breve{\psi}_i' - I_K\right)M(\breve{G})\sqrt{n}\bar{\breve{\psi}}_n.$$

Therefore, noting that $E[B_{2n}(\bar{V}_n)|t_n^O] = E\left(E[B_{2n}(\bar{V}_n)|\sqrt{n}\bar{\breve{\psi}}_n, t_n^O]|t_n^O\right) = E\left(E[B_{2n}(\bar{V}_n)|\sqrt{n}\bar{\breve{\psi}}_n]|t_n^O\right)$, we obtain

$$E[B_{2n}(\bar{V}_n)|t_n^O] = -\frac{\breve{G}'}{\sqrt{\breve{G}'\breve{G}}}E\left(E\left[\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \breve{\psi}_i\breve{\psi}_i' - I_k\right)\Big|\sqrt{n}\bar{\breve{\psi}}_n\right]M(\breve{G})\sqrt{n}\bar{\breve{\psi}}_n\,|t_n^O\right). \quad (23)$$

(iii) MODIFIED – $V_n^0 = V_n^{-1,0}$**:**

For the same exercise with $V_n^0 = V_n^{-1,0}$ (or $V_n^0 = V_n^{0,0}$), define $V_n^{-1,0,I} := \sum_{i=1}^n \frac{1}{n}[1 - \bar{\psi}_n'V^{-1}\psi_i]\psi_i\psi_i'$ that is based on the infeasible V, and then note that $V_n^{-1,0,I} = V_n^{-1,0} + o_p(1/\sqrt{n})$. In addition, under standard assumptions, note that $B_{jn}(V_n^0 = V_n^{-1,0,I}) = B_{jn}(V_n^0 = V_n^{-1,0}) + \mathcal{O}_p(1/\sqrt{n})$ for $j = 1, 2$. Now, defining the sequence of i.i.d. random variables $W_i := G'V^{-1}\psi_i$

with variance $\sigma^2 := G'V^{-1}G$, we obtain

$$
\begin{aligned}
B_{1n}(V_n^{-1,0,I}) &\equiv B_{1n}(V_n^0 = V_n^{-1,0,I}) \\
&:= -\frac{G'V^{-1}\sqrt{n}(V_n^{-1,0,I} - V)V^{-1}G}{2G'V^{-1}G} t_n^O \\
&= B_{1n}(\bar{V}_n) + \frac{1}{2G'V^{-1}G} G'V^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\bar{\psi}_n' V^{-1} \psi_i) \psi_i \psi_i' V^{-1} G t_n^O \\
&= B_{1n}(\bar{V}_n) + \frac{1}{2n} \sum_{i=1}^{n} \frac{W_i^3}{\sigma^3} \frac{\sqrt{n}\bar{W}_n}{\sigma} t_n^O + \frac{1}{2\sqrt{n}} \sum_{i=1}^{n} \left[\bar{\psi}_n' V^{-1/2} M(V^{-1/2'} G) V^{-1/2'} \psi_i \right] \frac{W_i^2}{\sigma^2} t_n^O \\
&= B_{1n}(\bar{V}_n) + \frac{(t_n^O)^2}{2} \left(\frac{1}{n} \sum_{i=1}^{n} \frac{W_i^3}{\sigma^3} \right) + \frac{t_n^O}{2\sqrt{n}} \sum_{i=1}^{n} \left[\bar{\breve{\psi}}_n' M(\breve{G}) \breve{\psi}_i \right] \left[\breve{\psi}_i' P(\breve{G}) \breve{\psi}_i \right] \qquad (24)
\end{aligned}
$$

because $W_i^2/\sigma^2 = \breve{\psi}_i' P(\breve{G}) \breve{\psi}_i$, where, as before, $\breve{\psi}_i := V^{-1/2'} \psi_i$ and $\breve{G} := V^{-1/2'} G$. Therefore, from Proposition 3.1 and Corollary 3.2, it is clear that the second term on the right-hand side of (24) corrects for the $\mathcal{O}_p(1/\sqrt{n})$-approximation error due to the use of the naive variance $\bar{V}_n$ that happens through $B_{1n}(\bar{V}_n)$. The question is: What does the third term on the right-hand side of (24) do? Note that the contribution of the third term to the $\mathcal{O}_p(1/\sqrt{n})$-approximation error due to the use of the modified variance estimator is proportional to

$$
\begin{aligned}
& E\left[\frac{t_n^O}{2\sqrt{n}} \sum_{i=1}^{n} \left[\bar{\breve{\psi}}_n' M(\breve{G}) \breve{\psi}_i \right] \left[\breve{\psi}_i' P(\breve{G}) \breve{\psi}_i \right] | t_n^O \right] \\
= \; & \frac{t_n^O}{2n\sqrt{n}} \sum_{i=1}^{n} E\left[\left[\breve{\psi}_i' M(\breve{G}) \breve{\psi}_i \right] \left[\breve{\psi}_i' P(\breve{G}) \breve{\psi}_i \right] | t_n^O \right] + \frac{t_n^O}{2n\sqrt{n}} \sum_{i=1}^{n} \sum_{j \neq i} E\left[\left[\breve{\psi}_j' M(\breve{G}) \breve{\psi}_i \right] \left[\breve{\psi}_i' P(\breve{G}) \breve{\psi}_i \right] | t_n^O \right].
\end{aligned}
$$

this makes the extension of the analysis of the simple t-ratios to over-identified models difficult. Note that this was not a problem in a just-identified model because, in that case, the third term on the right-hand side of (24) is identically equal to zero since $M(\breve{G})$ is a zero-matrix of dimension same as the dimension of θ.

Next let us consider $B_{2n}(V_n^{-1,0,I})$. At the outset, let us announce that Proposition 3.1 and Corollary 3.2 do not really give any immediate guidance to deal with this term. This is another part of the difficulty of extending the analysis of the simple t-ratios to over-identified models.

$$\begin{aligned} B_{2n}(V_n^{-1,0,I}) &\equiv B_{2n}(V_n^0 = V_n^{-1,0,I}) \\ &:= -\frac{G'V^{-1}\sqrt{n}(V_n^0 - V)V^{-1/2}M(V^{-1/2'}G)V^{-1/2'}\sqrt{n}\bar{\psi}_n}{\sqrt{G'V^{-1}G}} \\ &= B_{2n}(\bar{V}_n) + \frac{\breve{G}'}{\sqrt{\breve{G}'\breve{G}}}\sqrt{n}\bar{\breve{\psi}}_n'\left(\frac{1}{n}\sum_{i=1}^{n}\breve{\psi}_i\breve{\psi}_i\breve{\psi}_i'\right)M(\breve{G})\sqrt{n}\bar{\breve{\psi}}_n. \end{aligned} \tag{25}$$

Therefore, from Proposition 3.1 and Corollary 3.2, (24) and (25), it follows that the contribution of $B_{2n}(V_n^{-1,0,I})$ to the $\mathcal{O}_p(1/\sqrt{n})$-approximation error due to the use of the modified variance estimator is

$$\begin{aligned} & E[B_{2n}(V_n^{-1,0,I})|t_n^O] \\ = \; & E\left(E[B_{2n}(V_n^{-1,0,I})|\sqrt{n}\bar{\breve{\psi}}_n, t_n^O]|t_n^O\right) \\ = \; & E\left(E[B_{2n}(V_n^{-1,0,I})|\sqrt{n}\bar{\breve{\psi}}_n]|t_n^O\right) \\ = \; & -\frac{\breve{G}'}{\sqrt{\breve{G}'\breve{G}}}E\left(E\left[\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\breve{\psi}_i\breve{\psi}_i' - I_K\right) - \sqrt{n}\bar{\breve{\psi}}_n'\left(\frac{1}{n}\sum_{i=1}^{n}\breve{\psi}_i\breve{\psi}_i\breve{\psi}_i'\right)\Big|\sqrt{n}\bar{\breve{\psi}}_n\right]M(\breve{G})\sqrt{n}\bar{\breve{\psi}}_n\,\Big|t_n^O\right) \\ = \; & -\frac{\breve{G}'}{\sqrt{\breve{G}'\breve{G}}}E\left(E\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\breve{\psi}_i\breve{\psi}_i'(1-\bar{\breve{\psi}}_n'\breve{\psi}_i) - I_K\right\}\Big|\sqrt{n}\bar{\breve{\psi}}_n\right]M(\breve{G})\sqrt{n}\bar{\breve{\psi}}_n\,\Big|t_n^O\right). \end{aligned} \tag{26}$$

Now we will point out the improvement in approximation due to the modified variance estimator as opposed to the naive variance estimator by showing that the conditional expectation in the middle the expression for the right-hand side of (26) is of smaller order than that of (23). In that endeavor, let us make the following assumptions on the existence of moments of sufficient orders and a CLT. This is related to assumptions (i) and (ii) in Proposition 3.1. To make the notation simpler, let us work with the rotated vectors $\breve{\psi}_i$ and under the null hypothesis that $E[\breve{\psi}_i] = 0$.

Assumption: Let moments of sufficient order for $g_{i,j}$ for $j = 1, \ldots, K$ exist, and for all $i = 1, \ldots, n$ denote $E[\prod_{l=1}^{r} g_{i,j_l}] = \mu_{j_1,\ldots,j_r}$ such that $1 \le j_1 \le \ldots \le j_r \le K$. There exists a CLT such that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} ((\breve{\psi}_{i,j}))_{1\le j\le K} \\ ((\breve{\psi}_{i,j_1}\breve{\psi}_{i,j_2} - \mu_{j_1,j_2}))_{1\le j_1\le j_2\le K} \\ ((\breve{\psi}_{i,j_1}\breve{\psi}_{i,j_2}\breve{\psi}_{i,j_3} - \mu_{j_1,j_2,j_3}))_{1\le j_1\le j_2\le j_3\le K} \end{pmatrix} \xrightarrow{d} N\left(0, \begin{bmatrix} I_k & \Omega_{12} & \Omega_{13} \\ \Omega_{12}' & \Omega_{22} & \Omega_{23} \\ \Omega_{13}' & \Omega_{23}' & \Omega_{33} \end{bmatrix}\right),$$

where the terms in the covariance matrices are constrained to take values listed below. Just to explain the seemingly messy notation, let us, for the purpose of simplicity, take $K = 2$. (The results continue to hold for a general K to which we revert after this explanation.) Then the dimensions and the actual indices (in parentheses) with respect to the j's in three blocks of the column vectors stacked vertically on the left-hand side are 2×1 (1,2), 3×1 (11, 12, 22) and 4×1 (111, 112, 122, 222) respectively. The right-hand side should, then, be read accordingly. We only provide below explicit expressions for the terms that are actually required for our purpose. Accordingly, the explicit expressions assumed above are

$$\mu_{j_1,j_2} = \begin{cases} 1 \text{ if } j_1 = j_2 \\ 0 \text{ otherwise} \end{cases}, \Omega_{12} = \begin{pmatrix} \mu_{1,1,1} & \mu_{1,1,2} & \mu_{1,2,1} \\ \mu_{1,1,2} & \mu_{1,2,2} & \mu_{2,2,2} \end{pmatrix}, \Omega_{13} = \begin{pmatrix} \mu_{1,1,1,1} & \mu_{1,1,1,2} & \mu_{1,1,2,2} & \mu_{1,2,2,2} \\ \mu_{1,1,1,2} & \mu_{1,1,2,2} & \mu_{1,2,2,2} & \mu_{2,2,2,2} \end{pmatrix}.$$

Now, reverting to the general K, we will consider the right-hand side of (23) and (26) respectively. For our purpose, it is sufficient to consider the inner expectations of the $K\times K$ symmetric matrices

$$\text{for (23):} \qquad E\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\breve{\psi}_i\breve{\psi}_i' - I_K\right\}\Big|\sqrt{n}\bar{\breve{\psi}}_n\right] \tag{27}$$

$$\text{for (26):} \qquad E\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\breve{\psi}_i\breve{\psi}_i'(1-\bar{\breve{\psi}}_n'\breve{\psi}_i) - I_K\right\}\Big|\sqrt{n}\bar{\breve{\psi}}_n\right]. \tag{28}$$

The (s,t)-th element, for $1 \le s \le t \le K$, in (27) is

$$\sqrt{n}\sum_{j=1}^{K}\mu_{s,t,j}\bar{\breve{\psi}}_{n,j}, \tag{29}$$

with a slight abuse of notation in terms of proper placement of j in the subscript of $\mu_{s,t,j}$. On the other hand, the (s,t)-th element, for $1 \le s \le t \le K$, in (28) is

$$\sqrt{n}\sum_{j=1}^{K}\mu_{s,t,j}\bar{\breve{\psi}}_{n,j} - \left[\sqrt{n}\sum_{j=1}^{K}\mu_{s,t,j}\bar{\breve{\psi}}_{n,j} + \sqrt{n}\sum_{j=1}^{K}\sum_{l=1}^{K}\mu_{s,t,j,l}\bar{\breve{\psi}}_{n,j}\bar{\breve{\psi}}_{n,l}\right] = -\frac{1}{\sqrt{n}}\sum_{j=1}^{K}\sum_{l=1}^{K}\mu_{s,t,j,l}(\sqrt{n}\bar{\breve{\psi}}_{n,j})(\sqrt{n}\bar{\breve{\psi}}_{n,l}), \tag{30}$$

again, with a slight abuse of notation in terms of proper placement of j and l in the subscript of $\mu_{s,t,j,l}$. Comparing (29) and (30), it is clear that the former is $\mathcal{O}_p(1)$ while the latter is $\mathcal{O}_p(1/\sqrt{n})$. This is the correction in approximation achieved by the use of the modified variance estimator.

References

Abowd, J. M. and Card, D. (1989). On the Covariance Structure of Earnings and Hours Changes. *Econometrica*, 57: 411–445.

Altonji, J. and Segal, L. M. (1996). Small Sample Bias in GMM Estimation of Covariance Structures. *Journal of Business and Economic Statistics*, 14: 353–366.

Antoine, B., Bonnal, H., and Renault, E. (2007). On the efficient use of the informational content of estimating equations: Implied probabilities and Euclidean empirical likelihood. *Journal of Econometrics*, 138: 461–487.

Brown, B. and Newey, W. (2002). Generalised method of moments, efficient bootstrapping, and improved inference. *Journal of Business and Economic Statistics*, 20: 507–517.

Cavanagh, C. (1983). *Hypothesis Testing in Models with Discrete Dependent Variable*. PhD thesis, University of California, Berkeley.

Chaudhuri, S. and Renault, E. (2011). Finite-sample improvements of score tests by the use of implied probabilities from Generalized Empirical Likelihood. Technical report, University of North Carolina, Chapel Hill.

Guggenberger, P. and Smith, R. (2005). Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification. *Econometric Theory*, 21: 667–709.

Hall, A. R. (2000). Covariance Matrix Estimation and the Power of the Overidentifying Restrictions Test. *Econometrica*, 68: 1517–1527.

Hansen, L. P., Heaton, J., and Yaron, A. (1996). Finite-Sample Properties of Some Alternative GMM Estimators. *Journal of Business and Economic Statistics*, 14: 262–280.

Horowitz, J. (1998). Bootstrap Methods for Covariance Structures. *The Journal of Human Resources*, 33: 39–61.

Imbens, G. W., Spady, R. H., and Johnson, P. (1998). Information Theoretic Approaches to Inference in Moment Condition Models. *Econometrica*, 66: 333–357.

Johnson, N. J. (1978). Modified t Tests and Confidence Intervals for Asymmetrical Populations. *Journal of the American Statistical Association*, 74: 536–544.

Kitamura, Y. and Stutzer, M. (1997). An Information-Theoretic Alternative to Generalized Method of Moments Estimation. *Econometrica*, 65: 861–874.

Kleibergen, F. (2005). Testing Parameters In GMM Without Assuming That They Are Identified. *Econometrica*, 73: 1103–1123.

Lyon, J. D., Barber, B. M., and Tsai, C.-L. (1999). Improved Methods for Tests of Long-Run Abnormal Stock Returns. *Journal of Finance*, LIV: 165–201.

Newey, W. K. and West, K. D. (1987). Hypothesis Testing with Efficient Method of Moments Estimation. *International Economic Review*, 28: 777–787.

Yanagihara, H. and Yuan, K.-H. (2005). Four improved statistics for contrasting means by correcting skewness and kurtosis. *British Journal of Mathematical and Statistical Psychology*, 58: 209–237.