Working Paper
Series

Saddlepoint Approximations and Test Statistics for Accurate Inference in Overidentified Moment Conditions Models

Elvezio Ronchetti   Fabio Trojani

Saddlepoint Approximations and Test Statistics for Accurate Inference in Overidentified Moment Conditions Models

Elvezio Ronchetti[a], Fabio Trojani[b]

[a]Dept. of Econometrics, University of Geneva,* and
Institute of Finance, University of Southern Switzerland

[b]Institute of Finance, University of Southern Switzerland†

Abstract

We propose a new class of test statistics inducing accurate dual likelihood ratio tests of parametric constraints in overidentified moment conditions models. These statistics are derived from the dual likelihood implied by the exponent in the saddlepoint approximation of a general GMM estimator and are shown to be asymptotically chi-squared distributed to higher order, with a relative error of order $O(1/n)$. Since these statistics require the knowledge of the moment generating function of the given orthogonality function we introduce an empirical likelihood version of these tests which can be applied in the fully nonparametric setting and which only requires a preliminary GMM parameter estimation to be computed. Therefore, it can be also easily incorporated into available GMM estimation packages. Finally, we provide some numerical Monte Carlo evidence on the accuracy of the new statistics. In these experiments we find that empirical dual likelihood ratio tests provide a higher accuracy than standard GMM test statistics and some recent information theoretic alternatives for a broad class of GMM models.

Keywords: Dual Likelihood, Empirical Likelihood, Generalized Method of Moments, Higher Order Asymptotics, Moment Condition Models, Relative Errors, Saddlepoint Approximations.

*Correspondence address: Dept. of Econometrics, Bld Pont d'Arve 40, 1211 Geneva, Switzerland, e-mail: Elvezio.Ronchetti@metri.unige.ch.

†Correspondence address: Institute of Finance, University of Southern Switzerland, Via Buffi 13, 6900 Lugano, Switzerland, e-mail: Fabio.Trojani@lu.unisi.ch. Fabio Trojani gratefully acknowledges the financial support of the Swiss National Science Foundation (grant 12-65196.01 and NCCR FINRISK).

Contents

1   Introduction

The finite sample properties of tests implied by Generalized Method of Moments (GMM, Hansen (1982)) based models have been the object of a large literature in the last fifteen years. Several papers report important differences between the finite sample distribution of several corresponding GMM statistics and their (first order) asymptotic chi-squared distribution (cf. for instance the papers by Altonji and Segal (1996), Burnside and Eichenbaum (1996) and Hansen, Heaton and Yaron (1996) in the special issue of the Journal of Business and Economic Statistics). Attempts to improve the finite sample properties of these tests have followed primarily two directions.

A first approach is based on a refinement of the asymptotic distribution of a GMM statistic. In this case, finite sample accuracy is improved using bootstrap techniques applied to some asymptotically (first order) chi-squared distributed pivotal statistics (cf. for instance Brown and Newey (1998), Hall and Horowitz (1996), Hall and Presnell (1999) and Hansen (1999)).

A second direction adopts test statistics that are derived either from the estimated objective function or the studentized parameter estimates in some Generalized Empirical Likelihood (GEL) estimation of moment condition models. These estimators minimize an information-theoretic concept of closeness between the empirical distribution and an estimated least favorable distribution (cf. Di Ciccio and Romano (1990)) that matches exactly the given moment conditions. Members of this class are the Empirical Likelihood estimator (EL, Qin and Lawless (1994), Imbens (1997), Imbens, Spady and Johnson (ISJ, 1998)), the Exponential Tilting estimator (ET, Kitamura and Stutzer (1997), ISJ (1998), Imbens and Spady (2002)), the Euclidean Likelihood Estimator (EU, see for instance Owen (2001) and Back and Brown (1993)), and the continuously updated GMM estimator of Hansen, Heaton and Yaron (1996). To first order, standard GMM and GEL statistics are asymptotically equivalent. To higher order, GEL estimators remove the bias of GMM that is associated with the Jacobian term appearing in the optimal linear combination of instruments. In particular, EL estimators also eliminate the bias due to estimation of the weighting matrix

in an higher order bias approximation based on stochastic Taylor expansions (Newey and Smith (2001)). For the exactly identified case some higher order direct Edgeworth expansions have been used by Bravo (1999) to analyze the higher oder properties of empirical likelihood-ratio tests by means of Mykland's (1995) dual likelihood theory. In that paper empirical likelihood-ratio tests are shown to be accurate up to an absolute order $o\left(1/n\right)$, this accuracy being improvable to an order $O\left(1/n^2\right)$ by means of a scale correction, as in standard parametric theory[1] .

This paper bridges the gap between these two approaches to inference in moment conditions models by making use of saddlepoint techniques applied to overidentified moment conditions models. On the one side, saddlepoint techniques improve on Edgeworth expansions and related techniques by providing very accurate approximations to the distribution of a statistic even in small samples and in the tails of the distribution. On the other hand, these techniques motivate and justify a new class of accurate dual likelihood statistics for testing parametric hypotheses. For the new statistics proposed in this paper we apply saddlepoint techniques to derive higher order asymptotics that show formally their higher finite sample accuracy.

The basic idea behind saddlepoint approximations goes back to Daniels (1954) seminal paper. These methods provide approximations of the density of general M-estimators having relative error of order $O(1/n)$ (see for example Field and Hampel (1982), Field (1982), Tingley and Field (1990), Field and Ronchetti (1990), Daniels and Young (1991), Jing and Robinson (1994), Fan and Field (1995), Davison, Hinkley and Worton (1995), Gatto and Ronchetti (1996), Almudevar, Field and Robinson (2001)). Notice that when using saddlepoint techniques, the accuracy of the corresponding approximations is measured by relative errors, which are a more relevant and more stringent measure of accuracy if we are interested in the tail of the distribution of a statistic. To our knowledge, such relative error approximation results can be obtained only by means of saddlepoint methods. By contrast, techniques based on a stochastic Taylor or a direct Edgeworth

---

[1] General extensions of these results to the overidentified situation or to situations with nuisance parameters are difficult because for these cases the higher order asymptotics of the empirical dual likelihood ratio statistics changes dramatically as recently shown by Lazar and Mykland (1999).

expansion (cf. for instance Bravo (1999), Hansen (1999) and Newey and Smith (2001)) can provide approximations with absolute errors only.

The specific contributions of the paper to the literature are the following. First, we derive a saddlepoint approximation to the density of a GMM estimator. This takes into account the second order effects due to the estimation of (i) the Jacobian term appearing in the optimal linear combination of instruments and (ii) the weighting matrix of a GMM estimator, leading to an approximation with relative error of order $O(1/n)$. This result provides powerful analytical approximations to the finite sample density of a GMM estimator. Second, we use the saddlepoint approximation for the density of a GMM estimator to define a new dual likelihood ratio test statistic in overidentified moment condition models which is asymptotically chi-squared distributed[2] up to a relative error of order $O(1/n)$. As a consequence, accurate new statistics for testing parametric constraints in moment conditions models are obtained. To our knowledge, higher order asymptotic results based on relative errors are not available so far within overidentified moment conditions models. Since the new statistics require the moment generating function of the given orthogonality function, we introduce an empirical likelihood version of these tests which can be applied in the fully nonparametric setting and which only requires a preliminary GMM parameter estimation to be computed. Therefore, it can be also easily incorporated into available GMM estimation packages. Finally, we provide some numerical Monte Carlo evidence on the accuracy of the new statistics. In these experiments we find that empirical dual likelihood ratio tests provide a higher accuracy than standard GMM test statistics and some recent information theoretic alternatives for a broad class of models.

The approach pursued in this paper is based on an interpretation of exponentially tilted empirical likelihoods as dual likelihoods in the given tilting parameter. Testing parametric hypotheses in moment condition models can be interpreted as a dual (exponentially tilted) likelihood ratio test of

---

[2] The higher order asymptotic pivotality of the new statistic is particularly desirable to obtain improved finite sample critical values when using nonparametric bootstrap.

the hypothesis of a zero tilting parameter (similarly to the dual empirical likelihood characterization proposed by Mykland (1995)). Hence, we define a class of dual (exponentially tilted) empirical likelihood ratio tests which are the nonparametric analogous to the standard log likelihood ratio test in the parametric case. Precisely, we show that the exponentially tilted log dual likelihood is in fact the exponent in the saddlepoint approximation of the density function of a (fully iterated) GMM estimator. The first step to achieve the high accuracy of our statistics consists therefore in modifying the results on saddlepoint density approximations for standard M-estimators to the case of an overidentified moment conditions model. This provides saddlepoint approximations to the finite sample density of a GMM estimator which are accurate up to a relative error of order $O(1/n)$. Extending results in Robinson, Ronchetti and Young (2003) to take overidentifying moment conditions into account, we are then in a position to derive a saddlepoint approximation for the distribution of our dual likelihood ratio test statistics. These approximations mantain the accuracy up to a relative error of order $O(1/n)$ and are shown to yield an asymptotic chi-squared distribution up to the same order. From these results, an empirical likelihood version of the new statistics which can be expected to maintain high accuracy also in a general nonparametric GMM setting is derived in a natural way.

The remainder of the paper is organized as follows. Section 2 introduces the standard (first order) GMM setting. Section 3 derives saddlepoint approximations for the finite sample density of a general GMM estimator. In Section 4, dual likelihood ratio test statistics are proposed. These statistics are shown to be asymptotically chi-squared distributed to higher order. An empirical likelihood version of our dual likelihood ratio test is then proposed. Section 5 analyzes the finite sample accuracy of the new statistics by Monte Carlo simulation, while Section 6 concludes and summarizes.

2 First Order GMM Setting

Let $(X_i)_{i=1}^n$ be a sequence of i.i.d. random variables with values in $\mathbb{R}^N$ and distribution function F. We are interested in estimating a parameter $\vartheta_0 \in \Theta \subset \mathbb{R}^k$ defined as the unique solution of a set of orthogonality conditions

$$E\left[\psi\left(X_1;\vartheta\right)\right] = 0 \quad , \tag{1}$$

given an orthogonality function $\psi : \mathbb{R}^N \times \Theta \to \mathbb{R}^H$, and an open set Θ such that $\overline{\Theta}$ is compact. We focus on the overidentified case $H > k$.

GMM estimators $\left(\widehat{\vartheta}_n\right)_{n \in \mathbb{N}}$ for ϑ_0 (cf. Hansen (1982)) are defined as the solutions of the sequence of minimization problems

$$\widehat{\vartheta}_n = \underset{\vartheta}{\arg\inf} \left[\frac{1}{n}\sum_{i=1}^n \psi\left(X_i;\vartheta\right)\right]' W \left[\frac{1}{n}\sum_{i=1}^n \psi\left(X_i;\vartheta\right)\right] \quad , \quad n \in \mathbb{N} \quad , \tag{2}$$

where W is a positive definite deterministic matrix. An efficient GMM estimator is obtained when setting $W = V_0^{-1}$ and in this case under standard conditions $\left(\widehat{\vartheta}_n\right)_{n \in \mathbb{N}}$ is consistent and asymptotically first order normally distributed

$$\sqrt{n}\left(\widehat{\vartheta}_n - \vartheta_0\right) \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}\left(0, \left(\Gamma_0' V_0^{-1}\Gamma_0\right)^{-1}\right) \quad , \tag{3}$$

where

$$\Gamma_0 = E\left[\nabla_{\vartheta'}\psi\left(X;\vartheta_0\right)\right] \ , \ V_0 = E\left[\psi\left(X;\vartheta_0\right)\psi\left(X;\vartheta_0\right)'\right] \ . \tag{4}$$

A feasible version of an efficient GMM estimator is based on a two step procedure where in the first step a consistent GMM estimator $\overline{\vartheta}_n$ is obtained based on an arbitrary weighting matrix such as for instance $\overline{W} = id_{\mathbb{R}^H \times \mathbb{R}^H}$. $\overline{\vartheta}_n$ is used to consistently estimate V_0^{-1} through an optimal weighting sequence $(W_n)_{n \in \mathbb{N}}$ defined by

$$W_n = \left[\frac{1}{n}\sum_{i=1}^n \psi\left(X_i;\overline{\vartheta}_n\right)\psi\left(X_i;\overline{\vartheta}_n\right)'\right]^{-1} \quad . \tag{5}$$

In Section 3 we will develop a saddlepoint approximation for the finite sample density of the fully

iterated version of a two step GMM estimator $\widehat{\vartheta}_n$ (cf. Hansen, Heaton and Yaron (1996))[3] .

For first order asymptotic testing purposes it is important that the normalized objective function of an efficient GMM estimator is first order asymptotically chi-squared distributed when evaluated at the parameter estimate $\widehat{\vartheta}_n$ and at the correct parameter value ϑ_0:

$$\widehat{\xi}_n\left(\widehat{\vartheta}_n\right) := n \cdot \left[\frac{1}{n}\sum_{i=1}^{n}\psi\left(X_i;\widehat{\vartheta}_n\right)\right]' W_n \left[\frac{1}{n}\sum_{i=1}^{n}\psi\left(X_i;\widehat{\vartheta}_n\right)\right] \xrightarrow[n\to\infty]{d} \mathcal{X}_{H-k}^2 \quad , \qquad (6)$$

and

$$\widehat{\xi}_n\left(\vartheta_0\right) := n \cdot \left[\frac{1}{n}\sum_{i=1}^{n}\psi\left(X_i;\vartheta_0\right)\right]' W_n \left[\frac{1}{n}\sum_{i=1}^{n}\psi\left(X_i;\vartheta_0\right)\right] \xrightarrow[n\to\infty]{d} \mathcal{X}_{H}^2 \quad . \qquad (7)$$

In this paper we consider tests of parametric hypotheses of the form

$$\mathcal{H}_0 \qquad : \qquad g\left(\vartheta\right) = 0 \qquad , \qquad\qquad (8)$$

for some smooth function $g : \mathbb{R}^k \to \mathbb{R}^q$, $q \leq k$. Several Likelihood-type statistics are available to test (8) in a GMM setting. For instance, it is well known that a likelihood ratio-type GMM test can be based on the statistic

$$\widehat{\xi}_n^{LR} = \widehat{\xi}_n\left(\widehat{\vartheta}_n^c\right) - \widehat{\xi}_n\left(\widehat{\vartheta}_n\right) \qquad ,$$

where $\widehat{\vartheta}_n^c$ is a constrained GMM estimator of ϑ_0 under the null hypothesis (8). Under $\mathcal{H}_0$ the statistic $\widehat{\xi}_n^{LR}$ is asymptotically first order chi-squared distributed:

$$\widehat{\xi}_n^{LR} \xrightarrow[n\to\infty]{d} \mathcal{X}_q^2 \qquad . \qquad\qquad (9)$$

Thus, an asymptotic test of the parametric hypothesis (8) can be based on the empirical quantiles of a $\mathcal{X}_q^2$. Similar results hold for a Wald-type or a Lagrange multiplier-type GMM statistics for testing (8) (see also Newey and West (1987) and Gourieroux and Monfort (1989)).

Asymptotic approximations of the form (3), (6), (7), (9) are implied by the central limit theorem. By the Berry-Esseen Theorem the difference between the exact distribution G_n (say) of

---

[3] Similar saddlepoint approximations can be obtained with the same methodology also for a two step GMM estimator.

the standardized statistic

$$W_n^{-\frac{1}{2}} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi \left(X_i; \widehat{\vartheta}_n \right) \right] \quad ,$$

and its asymptotic standard normal limit Φ is uniformly bounded by the inequality

$$\sup_{x \in \mathbb{R}^N} \left| G_n \left(x \right) - \Phi \left(x \right) \right| \leq \frac{C}{\sqrt{n}} \quad , \tag{10}$$

where C is some positive constant that depends on $\widehat{\vartheta}_n$ and F but not on n. Therefore, the asymptotic approximation produces errors of order $O \left(1/\sqrt{n} \right)$ that are absolute in nature.

In the next sections we construct GMM test statistics for testing (8) whose finite sample distribution can be approximated by a chi-squared distribution up to a relative error of order $O \left(1/n \right)$.

3   Second Order GMM Estimation Setting

Powerful approximations of the finite sample density of a general M-estimator have been derived by Field (1982) using saddlepoint techniques, extending Daniels (1954) original idea. The basic idea behind these approximations is to recenter the finite sample density of an M-estimator at a given approximating point by means of a so-called conjugate density. In a second step a multivariate Edgeworth expansion in used locally to obtain an approximation of the underlying density at this point. By contrast with direct global Edgeworth expansions this approach yields accurate approximations with relative errors of order $O \left(1/n \right)$ over a broad support of the finite sample density of an M-estimator. In this section we use these techniques to derive saddlepoint approximations for the finite sample density of a general GMM estimator. Based on these approximations, we define in the next section a test statistic for testing (8) which is asymptotically chi-squared distributed up to relative errors of order $O \left(1/n \right)$.

3.1   GMM Estimators as Extended M-Estimators

We embed efficient, fully iterated, GMM estimators in the M-estimation setting (cf. Huber (1981) for an overview), by means of an appropriate extended score function Ψ. The system of GMM

estimating equations for ϑ_0, Γ_0, V_0, is given by

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{\Gamma}'_n\widehat{V}_n^{-1}\psi\left(X_i,\widehat{\vartheta}_n\right) = 0 \quad,$$

$$\frac{1}{n}\sum_{i=1}^{n}\left[\nabla_{\vartheta'}\psi\left(X_i,\widehat{\vartheta}_n\right)-\widehat{\Gamma}_n\right] = 0 \quad,$$

$$\frac{1}{n}\sum_{i=1}^{n}\left[\psi\left(X_i,\widehat{\vartheta}_n\right)\psi'\left(X_i,\widehat{\vartheta}_n\right)-\widehat{V}_n\right] = 0 \quad,$$

see also Imbens (1997). Thus, we can interpret an efficient, fully iterated, GMM estimator of the parameters ϑ_0, Γ_0, W_0, as an M-estimator defined by the estimating equations

$$E\left[\Gamma'V^{-1}\psi\left(X,\vartheta\right)\right] = 0 \quad, \tag{11}$$

$$E\left[\nabla_{\vartheta'}\psi\left(X,\vartheta\right)-\Gamma\right] = 0 \quad, \tag{12}$$

$$E\left[\psi\left(X,\vartheta\right)\psi'\left(X,\vartheta\right)-V\right] = 0 \quad. \tag{13}$$

We will obtain a saddlepoint approximation for the marginal density of $\widehat{\vartheta}_n$ in two steps:

(i) we first consider the joint estimation of ϑ_0, Γ_0 and V_0, by means of $\widehat{\vartheta}_n$, $\widehat{\Gamma}_n$ and $\widehat{V}_n$, respectively, and define an appropriate exponentially tilted conjugate density;

(ii) we then marginalize $\widehat{\Gamma}_n$, $\widehat{V}_n$, and obtain a saddlepoint approximation for the finite sample density of $\widehat{\vartheta}_n$.

Let us introduce some notation, in order to take into account separately the different blocks of parameters and estimating equations in (11)-(13).

Notation: *We define an extended parameter vector* $\theta=\left(\theta'_1,\theta'_2,\theta'_3\right)'$ *by*

$$\theta_1=\vartheta \quad,\quad \theta_2=vec\left(\Gamma\right) \quad,\quad \theta_3=vech\left(V\right) \quad,$$

and introduce *an extended score function* $\Psi = (\Psi'_1, \Psi'_2, \Psi'_3)' : \Theta_\Psi := \Theta \times \mathbb{R}^{\frac{k(3k+1)+H(H-k)}{2}} \times$

$\mathbb{R}^{\frac{H(H+1)}{2}} \to \mathbb{R}^{\frac{3k(k+1)+H(2H-(k-1))}{2}}$ *by*

$$\Psi_1(X, \theta) = \Gamma' V^{-1} \psi(X, \theta_1) \quad , \tag{14}$$

$$\Psi_2(X, \theta) = vec\left[\nabla_{\vartheta'} \psi(X, \theta_1)\right] - \theta_2 \tag{15}$$

$$\Psi_3(X, \theta) = vech\left[\psi(X, \theta_1)\psi'(X, \theta_1)\right] - \theta_3 \quad . \tag{16}$$

Finally, let $\widehat{\theta}_n = \left(\widehat{\theta}'_{1n}, \widehat{\theta}'_{2n}, \widehat{\theta}'_{3n}\right)'$ *be the M-estimator associated with* Ψ, *that is the solution to the system of estimating equations*

$$\frac{1}{n}\sum_{i=1}^{n} \Psi\left(X_i, \widehat{\theta}_n\right) = 0 \quad . \tag{17}$$

By construction, $\widehat{\theta}_{1n} = \widehat{\vartheta}_n$, *the (fully iterated) efficient GMM estimator of* ϑ_0 *associated to the weighting matrix* V_0^{-1}.

To derive a saddlepoint approximation for the finite sample density of $\widehat{\theta}_{1n}$, we start from a tilting procedure that recenters the density of $\widehat{\theta}_n$ at the point where it has to be approximated. This is obtained by embedding the joint density of $X_1, .., X_n$, into a particular exponential family, which is defined through a tilting parameter that satisfies the given centering requirement.

3.2   Conjugate Densities

We make use of conjugate densities to tilt the finite sample density of the GMM estimator $\widehat{\theta}_n$. This allows, in a second step, to develop Edgeworth approximations for the tilted marginal distribution of $\widehat{\theta}_{1n}$ which have relative approximation error of order $O\left(\frac{1}{n}\right)$.

We tilt the distribution of $\widehat{\theta}_n$ in a way that recenters only the first marginal component $\widehat{\theta}_{1n}$. This is a necessary procedure in order to make use of tilting families than can be numerically computed in applications. Indeed, extending the tilting procedure also to $\widehat{\theta}_{2n}$ and $\widehat{\theta}_{3n}$ would require solving high dimensional systems of implicit equations already for low dimensional models. Moreover, we show below that the chosen tilting procedure is a natural one because it allows to reinterpret the asymptotically optimal GMM estimating equations (17) under the initial measure

F as a set of asymptotically optimal GMM estimating equations also under the corresponding tilted distribution. In the sequel we denote by

$$M_Y(\mu) = E\left[\exp\left(i\mu'Y\right)\right] \quad, \quad K_Y(\mu) = \log\left(M_Y(-i\mu)\right) \quad,$$

where $i^2 = 1$, the characteristic and the cumulant generating function of a random vector Y (provided the latter exists), respectively.

The exponential family defining a tilting procedure for the M-estimator defined by (11)-(13) is introduced in the next definition.

Definition 1 *Let $dF^{(n)}(x_1,..,x_n) := \prod_{i=1}^{n} dF(x_i)$ be the product measure of F and define for fix $\theta_1 \in \Theta$ and $\mu \in \mathbb{R}^H$ the tilted measure $H_{\theta_1,\mu}^{(n)}$ by*

$$dH_{\theta_1,\mu}^{(n)}(x_1,..,x_n) = \prod_{i=1}^{n} \exp\left[\mu'\psi(x_i,\theta_1) - K_{\psi(X,\theta_1)}(\mu)\right] \cdot dF^{(n)}(x_1,..,x_n) \quad. \tag{18}$$

Assumption 1 in the Appendix implies the existence of a well defined tilting procedure (18). Precisely, in Assumption 1, (i), we assume the existence of the cumulant generating function of $\psi(X,\theta_1)$. This condition is generally satisfied by GMM estimators based on a bounded orthogonality function ψ (cf. for instance Singleton (2000) and Ronchetti and Trojani (2001)) and ensures the existence of an equivalent measure $H_{\theta_1,\mu}^{(n)}$. Assumption 1, (ii), can be weakened to handle multiple roots. It simplifies the arguments in the proof of Proposition 1 below. For a discussion on how to handle multiple roots see Field (1982), Section 3. Finally, Assumption 1, (iii), ensures the existence of a well-defined Fourier inversion, under $F^{(n)}$ and $H_{\theta_1,\mu}^{(n)}$ as well.

Under Assumption 1, we can now express the density $p^{(n)}$ of $\widehat{\theta}_n$ under $F^{(n)}$ in terms of the density $q_{\theta_1,\mu}^{(n)}$ of $\widehat{\theta}_n$ under $H_{\theta_1,\mu}^{(n)}$. This result is given in the next proposition and is the starting point for defining a tilting procedure that recenters the density of $\widehat{\theta}_n$ at the point where it has to be approximated. All proofs of the paper are given in the Appendix.

Proposition 1 *Let Assumption 1 be satisfied and fix $\theta_1 \in \Theta$ and $\mu \in \mathbb{R}^H$. It then follows for any $\theta^* \in \Theta_\Psi$ such that $\theta_1^* = \theta_1$:*

$$p^{(n)}(\theta^*) = C_\mu(\theta_1)^{-n} D_\mu(\theta^*) \cdot q_{\theta_1,\mu}^{(n)}(\theta^*) \quad, \tag{19}$$

11

where

$$C_\mu(\theta_1) = \exp\left(-K_{\psi(X,\theta_1)}(\mu)\right)$$

$$D_\mu(\theta^*) = E_{H_{\theta_1},\mu}\left[\exp\left(-\sum_{i=1}^{n}\mu'\psi(X_i;\theta_1)\right)\bigg|\,\widehat{\theta}_n = \theta^*\right] \quad .$$

We make use of Proposition 1 to determine a specific tilting parameter $\mu(\theta_1)$ that can be used to recenter $q^{(n)}_{\theta_1,\mu}$ at a specific point θ^* where an approximation for $p^{(n)}$ has to be produced. In this case, the factors $C_{\mu(\theta_1)}(\theta_1)$, $D_{\mu(\theta_1)}(\theta^*)$ define the mapping by which we can switch from $p^{(n)}$ to $q^{(n)}_{\theta_1,\mu(\theta_1)}$ and vice-versa. Moreover, performing an Edgeworth expansion of $q^{(n)}_{\theta_1,\mu(\theta_1)}$ at $\theta = \theta^*$ will typically produce a very good local approximation because we are approximating the density $q^{(n)}_{\theta_1,\mu(\theta_1)}$ at its center.

In principle, we could produce a saddlepoint approximation for the density of the whole estimator $\widehat{\theta}_n$ by defining a corresponding recentering procedure with respect to all components θ_1, θ_2, θ_3. However, our main focus is clearly on the first marginal $\widehat{\theta}_{1n}$, given the nuisance parameters $\widehat{\theta}_{2n}$, $\widehat{\theta}_{3n}$. Therefore, we make use of tilting procedures that recenter the underlying density with respect to the first component θ_1 only. Intuitively, this can be obtained by selecting $\mu(\theta_1)$ so that under the distribution $H^{(n)}_{\theta_1,\mu(\theta_1)}$ the GMM orthogonality condition (1) has expectation zero.

Definition 2 *Define the saddlepoint $\mu(\theta_1)$ as the implicit solution of the saddlepoint equation*

$$E\left[\psi(X,\theta_1)\exp\left(\mu'\psi(X,\theta_1)\right)\right] = 0 \quad . \tag{20}$$

We call the density given by $q^{(n)}_{\theta_1} := q^{(n)}_{\theta_1,\mu(\theta_1)}$ the tilted conjugate density of the GMM estimator $\widehat{\theta}_n$ implied by the extended score function Ψ.

In the sequel we assume uniqueness of the saddlepoint defining the tilted conjugate density in Definition 2 (cf. Assumption 2 in the Appendix).

By construction, under the tilted distribution $H^{(n)}_{\theta_1} := H^{(n)}_{\theta_1,\mu(\theta_1)}$ the GMM estimating equation induced by Ψ_1 has expectation zero, so that under $q^{(n)}_{\theta_1}$ the GMM estimator $\widehat{\theta}_{1n}$ has been recentered at θ_1. Precisely, under $q^{(n)}_{\theta_1}$ one can associate to the GMM estimator induced by (11)-(13) the tilted estimating equations

$$E_{H^{(n)}_{\theta_1}}\left[\Psi(X,\theta)\right] = 0 \quad . \tag{21}$$

The finite sample distribution of the extended GMM estimator induced by (21) is precisely $q_{\theta_1}^{(n)}$. Moreover, notice that under $H_{\theta_1}^{(n)}$ the GMM estimators $\widehat{V}_n$ and $\widehat{\Gamma}_n$ have been recentered at

$$V_{\theta_1} := E_{H_{\theta_1}^{(n)}} \left[\psi \left(X, \theta_1 \right) \psi' \left(X, \theta_1 \right) \right] \quad , \quad \Gamma_{\theta_1} := E_{H_{\theta_1}^{(n)}} \left[\nabla_\vartheta \psi \left(X, \theta_1 \right) \right] \quad , \tag{22}$$

respectively. Therefore, under the tilted distribution $H_{\theta_1}^{(n)}$ the score function Ψ defines an efficient optimal GMM estimator of θ_1.

In the next section we develop accurate saddlepoint approximations for the finite sample density of $\widehat{\theta}_{1n}$.

3.3 Saddlepoint Approximations for the Finite Sample Density of a GMM Estimator

We now derive a local approximation of the GMM estimator $\widehat{\theta}_n$ under the tilted conjugate measure $H_{\theta_1}^{(n)}$ in a neighborhood of a point $\theta^* \in \Theta_\Psi$ given by

$$\theta_1^* = \theta_1 \quad , \quad \theta_2^* = vec \left(\Gamma_{\theta_1} \right) \quad , \quad \theta_3^* = vech \left(V_{\theta_1} \right) \quad . \tag{23}$$

We then compute an Edgeworth expansion for the marginal density[4] $q_{\theta_1}^{(n)}$ of $\widehat{\theta}_{1n}$. Since, by construction, the distribution of $\widehat{\theta}_n$ has been recentered at θ^*, i.e., $E_{H_{\theta_1}^{(n)}} \left(\widehat{\theta}_n \right) = \theta^*$, a local (low order) approximation of the density of $\widehat{\theta}_{1n}$ under $H_{\theta_1}^{(n)}$ will produce accurate results. An accurate saddlepoint approximation of the finite sample density of $\widehat{\theta}_{1n}$ under $F^{(n)}$ is then obtained using Proposition 1.

Assumption 3 in the Appendix mimics the set of assumptions in Bhattacharya and Ghosh (1978) and Field (1982) for the existence of a formal Edgeworth expansion for the density of the estimator $\widehat{\theta}_n$ under an arbitrary tilted distribution $H_{\theta_1}^{(n)}$, $\theta_1 \in \Theta$. It implies the existence of a formal Edgeworth expansion for $q_{\theta_1}^{(n)}$ with errors that are uniform over compact subsets of Θ and is generally satisfied by robust GMM models based on a bounded orthogonality function ψ (cf. Ronchetti and Trojani (2001)).

---

[4] By a slight abuse of notations whe use the symbols $p^{(n)}$ and $q_{\theta_1}^{(n)}$ to denote at the same time joint and marginal densities under $F^{(n)}$ and $H_{\theta_1}^{(n)}$, respectively. No confusion should arise.

To obtain the desired saddlepoint approximation for the finite sample density of $\widehat{\theta}_{1n}$ we need a preliminary technical, but important, Lemma.

Lemma 1 *Under Assumption 1-3 we have*

$$D_{\mu(\theta_1)}\left(\theta^*\right) = 1 + O\left(\frac{1}{n}\right) \quad , \tag{24}$$

with $D_\mu\left(\theta^*\right)$ defined in Proposition 1.

Lemma 1 enables us to approximate the mapping (19) between densities under $F^{(n)}$ and $H_{\theta_1}^{(n)}$, respectively, without having to calculate exactly the conditional expectation $D_{\mu(\theta_1)}\left(\theta^*\right)$, an unattractive task in applications. The approximation (24) maintains relative errors of order $O\left(\frac{1}{n}\right)$ in the final approximation of the finite sample density of $\widehat{\theta}_{1n}$ under $F^{(n)}$. After integrating over θ_2^* and θ_3^*, Proposition 1 and Lemma 1 give the relation between the finite sample densities of $\widehat{\theta}_{1n}$ under $F^{(n)}$ and $H_{\theta_1}^{(n)}$, respectively.

Proposition 2 *Under Assumption 1-3 we have for any $\theta_1 \in \Theta$*

$$p^{(n)}\left(\theta_1\right) = C_{\mu(\theta_1)}\left(\theta_1\right)^{-n} q_{\theta_1}^{(n)}\left(\theta_1\right)\left(1 + O\left(\frac{1}{n}\right)\right) \quad , \tag{25}$$

with the constant $C_{\mu(\theta_1)}\left(\theta_1\right)$ defined in Proposition 1.

We now compute a local Edgeworth approximation for $q_{\theta_1}^{(n)}\left(\theta_1\right)$, which will produce relative errors of order $O\left(\frac{1}{n}\right)$ because $q_{\theta_1}^{(n)}$ has been centered at θ_1. Combining this approximation with (25) we obtain the desired saddlepoint approximation for the density $p^{(n)}$ of an efficient GMM estimator. The next proposition gives the corresponding result in terms of the GMM estimator $\widehat{\vartheta}_n$ for the initial parameter of interest ϑ.

Proposition 3 *Under Assumption 1-3 it follows:*

(i) An asymptotic expansion for the density of $\widehat{\vartheta}_n$ under $H_\vartheta^{(n)}$ is given by

$$q_\vartheta^{(n)}\left(\vartheta\right) = \left(\frac{n}{2\pi}\right)^{\frac{k}{2}}\left|\det\left(\Gamma_\vartheta' V_\vartheta^{-1} \Gamma_\vartheta\right)\right|^{\frac{1}{2}}\left(1 + O\left(\frac{1}{n}\right)\right) \quad ,$$

where

$$\Gamma_\vartheta = E_{H_\vartheta^{(n)}}\left[\nabla_{\vartheta'}\psi\left(X, \vartheta\right)\right] \quad , \quad V_\vartheta = E_{H_\vartheta^{(n)}}\left[\psi\left(X, \vartheta\right)\psi\left(X, \vartheta\right)'\right] \quad .$$

(ii) An asymptotic expansion for the density of $\widehat{\vartheta}_n$ under $F^{(n)}$ is given by

$$p^{(n)}(\vartheta) = \left(\frac{n}{2\pi}\right)^{\frac{k}{2}} C_{\mu(\vartheta)}(\vartheta)^{-n} \left|\det\left(\Gamma'_\vartheta V_\vartheta^{-1}\Gamma_\vartheta\right)\right|^{\frac{1}{2}} \left(1 + O\left(\frac{1}{n}\right)\right) \quad . \tag{26}$$

Even if this is not the primary goal of the paper, the approximation (26) can be used to approximate the finite sample density of a GMM estimator and to compute corresponding tail area approximations. In this case, the integrating constant $\left(\frac{n}{2\pi}\right)^{\frac{k}{2}}$ will not be used in numerical work. Instead, a normalization constant will be computed numerically, following the practice advocated by Hampel (1973).

Under a compactness assumption on $\overline{\Theta}$ the approximation (26) is uniform in ϑ; if $\overline{\Theta}$ is not compact, the approximation is locally uniform on compact subsets of $\overline{\Theta}$. This implies, that tail areas approximations over compact sets will maintain $O\left(\frac{1}{n}\right)$ relative errors, under a continuity assumption on the function $C_{\mu(\cdot)}(\cdot)$; cf. Field (1982).

The next section makes use of Proposition 3 to develop versions of a GMM statistic for testing parametric hypotheses, which under the null hypothesis are $\mathcal{X}_q^2$ distributed up to relative errors of order $O\left(\frac{1}{n}\right)$.

4 Second Order GMM Testing Setting

We now consider the problem of testing a simple parametric hypothesis of the form

$$\mathcal{H}_0 \quad : \quad \vartheta = \vartheta_{\mathcal{H}_0} \quad , \tag{27}$$

for some $\vartheta_{\mathcal{H}_0} \in \Theta$, based on a statistic whose finite sample distribution can be accurately approximated by a $\mathcal{X}_k^2$ distribution up to relative errors of order $O\left(\frac{1}{n}\right)$. We focus on a simple hypothesis $\mathcal{H}_0$ only for brevity. Using the saddlepoint approximation for the density of a GMM estimator in the last section, also the general nuisance parameters case can be treated along the same lines as in Robinson, Ronchetti and Young (2003) within overidentified moment conditions models.

4.1 Dual Likelihoods and Dual Hypotheses Associated with GMM Estimators

We obtain an accurate test of (27) by reinterpreting (27) as a null hypothesis with respect to a dual parameter implied by the natural dual likelihood (cf. Mykland (1995)) associated to a score process $M^{\vartheta} := \left(M_i^{\vartheta}\right)_{i \in \mathbb{N}} := (\psi\left(X_i; \vartheta\right))_{i \in \mathbb{N}}$, $\vartheta \in \Theta$. Dual likelihoods are likelihood-type objects that are associated with martingale score statistics. By construction, dual likelihoods induce the same efficiency properties as those implied by the original score statistics. Moreover, as conjectured by Mykland (1995), they should produce higher accuracy in finite samples because of their link to saddlepoint approximations. For the given overidentified GMM setting we provide in the next section a proof of this conjecture by definining a class of dual likelihood test statistics that are implied by the family of score processes $\left\{M^{\vartheta}; \vartheta \in \Theta\right\}$. These tests are related to the saddlepoint approximation (26) and induce statistics that are $\mathcal{X}_k^2$ distributed up to relative errors of order $O\left(\frac{1}{n}\right)$.

Definition 3 *A dual likelihood associated with the score processes $\left\{M^{\vartheta}; \vartheta \in \Theta\right\}$, is a function $L : \mathcal{M} \subset \mathbb{R}^H \times \Theta \times \mathbb{R}^N \to \mathbb{R}^+$ such that*

  (i) For any $\vartheta \in \Theta$ the function $l_{\vartheta}\left(\mu; x\right) := \ln L\left(\mu, \vartheta, x\right)$ is a log likelihood in the dual parameter μ,

  (ii) For any $\vartheta \in \Theta$

$$\psi\left(X; \vartheta\right) = \nabla_{\mu} l_{\vartheta}\left(0; X\right) \quad , \quad F - a.s. \quad . \tag{28}$$

In our setting, a test of the null hypothesis (27) can be interpreted as a test of the martingale difference property for M^{ϑ_0}. Thus, $\mathcal{H}_0$ can be also equivalently expressed as

$$\mathcal{H}_0 \quad : \quad E\left[\nabla_{\mu} l_{\vartheta_{\mathcal{H}_0}}\left(0; X\right)\right] = 0 \quad . \tag{29}$$

More specifically, given a GMM estimator $\widehat{\vartheta}_n$ converging a.s. under $\mathcal{H}_0$ to $\vartheta_{\mathcal{H}_0} = \vartheta_0$, a test of (27) is equivalent to a test of the dual parametric hypothesis

$$\mathcal{H}_0 \quad : \quad \mu\left(\vartheta_0\right) = 0 \quad , \tag{30}$$

where $\mu(\vartheta_0)$ is the dual parameter associated to ϑ_0. Therefore, we can treat ϑ_0 as a nuisance parameter and test by a version of a dual quasi likelihood ratio test based on the dual quasi likelihood $L_{\widehat{\vartheta}_n}(\cdot; x)$ the parametric dual hypothesis (30).

4.2 Saddlepoint Approximations for Dual Quasi Likelihood Ratio Tests Induced by GMM Estimators

In the GMM setting a natural dual likelihood associated to $\left\{M^{\vartheta}; \vartheta \in \Theta\right\}$ is provided by the tilted exponential family (18) in Definition 1, yielding

$$L(\mu, \vartheta; x) = \exp\left(\mu' \psi(x; \vartheta) - K_{\psi(X;\vartheta)}(\mu)\right) \quad . \tag{31}$$

The natural dual parameter associated with ϑ_0 is the maximizer of the expected dual log likelihood objective

$$\mu(\vartheta_0) = \arg\sup_{\mu}\left\{\mu' E\left[\psi(X; \vartheta_0)\right] - K_{\psi(X;\vartheta_0)}(\mu)\right\} \underset{\mathcal{H}_0}{=} \arg\sup_{\mu}\left\{-K_{\psi(X;\vartheta_0)}(\mu)\right\} \quad , \tag{32}$$

which for the specific dual likelihood (31) is the saddlepoint solution in the saddlepoint approximation of the density of $\widehat{\vartheta}_n$ (cf. (20) and (26)). Since $K_{\psi(X;\vartheta)}(0) = 0$, the dual likelihood ratio statistic for testing (30) is

$$LR_{\vartheta}^{D} := -2 \cdot K_{\psi(X;\vartheta)}(\mu(\vartheta)) \quad . \tag{33}$$

Therefore, the dual quasi likelihood ratio statistic associated with a GMM estimator $\widehat{\vartheta}_n$ is

$$LR_{\widehat{\vartheta}_n}^{D} = -2 \cdot K_{\psi\left(X;\widehat{\vartheta}_n\right)}\left(\mu\left(\widehat{\vartheta}_n\right)\right) \quad .$$

The next proposition derives accurate $\mathcal{X}_k^2$ approximations for the finite sample distribution of $n \cdot LR_{\widehat{\vartheta}_n}^{D}$, having relative errors of order $O\left(\frac{1}{n}\right)$.

Proposition 4 *Under Assumption 1-3 an asymptotic expansion for the tail probabilities of the statistic $LR_{\widehat{\vartheta}_n}^{D}$ under the null hypothesis (27) (or equivalently either (29) or (30)) is given by*

$$P_{\mathcal{H}_0}\left(n \cdot LR_{\widehat{\vartheta}_n}^{D} > a\right) = \left(1 - \mathcal{X}_k^2(a)\right)\left(1 + O\left(\frac{1}{n}\right)\right) \quad , \tag{34}$$

where $\mathcal{X}_k^2(\cdot)$ is the distribution function of a chi-squared distributed random variable with k degrees of freedom.

The computation of $LR_{\widehat{\vartheta}_n}^D$ requires the cumulant generating function $K_{\psi(X;\vartheta)}(\mu)$, which in the parametric setting can be computed either explicitly (rarely) or using an indirect inference approach. In the fully nonparametric GMM setting empirical $p-$values of $LR_{\widehat{\vartheta}_n}^D$ can be computed by some version of a nonparametric bootstrap. Notice that by Proposition 4, $LR_{\widehat{\vartheta}_n}^D$ is pivotal up to relative errors of order[5] $O\left(\frac{1}{n}\right)$. Therefore, we can expect bootstrap approximations of the asymptotic distribution of $LR_{\widehat{\vartheta}_n}^D$ to be very accurate as well.

4.3 Empirical Dual Likelihood Tests

In practice, the distribution F underlying the data sample $X_1, .., X_n$ is tipically unknown. In this case an empirical exponential likelihood may be used to provide empirical versions of the dual likelihood ratio test based on $LR_{\widehat{\vartheta}_n}^D$. To do this for the simple hypothesis $\mathcal{H}_0$ we need to consider the tilted empirical moment generating function of ψ under $\mathcal{H}_0$:

$$\widehat{K}_{\psi(X;\vartheta)}(\mu) := \log\left[\sum_{i=1}^{n} \exp\left(\mu'\psi(X_i;\vartheta)\right) \cdot \left(\frac{\exp\left(\mu_n(\vartheta_0)'\psi(X_i;\vartheta_0)\right)}{\sum_{i=1}^{n}\exp\left(\mu_n(\vartheta_0)'\psi(X_i;\vartheta_0)\right)}\right)\right] \qquad ,$$

where $\mu_n(\vartheta_0)$ is the solution of the empirical saddlepoint equation,

$$\sum_{i=1}^{n}\psi(X_i;\vartheta_0)\exp\left(\mu'\psi(X_i;\vartheta_0)\right) = 0 \qquad , \qquad (35)$$

under the null hypothesis $\mathcal{H}_0$. Thus, $\widehat{K}_{\psi(X;\vartheta)}(\mu)$ is the empirical cumulant generating function of ψ under the tilted empirical distribution $\widehat{F}^{\mathcal{H}_0}$ defined by

$$d\widehat{F}^{\mathcal{H}_0}(X_i) = \frac{\exp\left(\mu_n(\vartheta_0)'\psi(X_i;\vartheta_0)\right)}{\sum_{i=1}^{n}\exp\left(\mu_n(\vartheta_0)'\psi(X_i;\vartheta_0)\right)}dF_n(X_i) \qquad ,$$

where F_n is the empirical distribution of $X_1, .., X_n$. Precisely, $\widehat{F}^{\mathcal{H}_0}$ is the (exponentially tilted) empirical likelihood estimator of the underlying distribution F under the null hypothesis $\mathcal{H}_0$. Thus, the empirical dual likelihood ratio statistic for testing (29) is

$$\widetilde{LR_{\widehat{\vartheta}_n}^D} = -2 \cdot \widehat{K}_{\psi(X;\widehat{\vartheta}_n)}\left(\widehat{\mu}\left(\widehat{\vartheta}_n\right)\right) \qquad ,$$

---

[5] To our knowledge, no other statistic for testing $\mathcal{H}_0$ has such an higher order pivotality property.

where $\widehat{\mu}\left(\widehat{\vartheta}_n\right)$ is the solution of the tilted empirical saddlepoint equation

$$\sum_{i=1}^{n}\left(\psi\left(X_i;\widehat{\vartheta}_n\right)\exp\left(\mu'\psi\left(X_i;\widehat{\vartheta}_n\right)\right)\cdot\exp\left(\mu_n\left(\vartheta_0\right)'\psi\left(X_i;\vartheta_0\right)\right)\right)=0 \quad,$$

or, equivalently:

$$\widehat{\mu}\left(\widehat{\vartheta}_n\right)=\arg\sup_{\mu}\left\{-\widehat{K}_{\psi\left(X;\widehat{\vartheta}_n\right)}\left(\mu\right)\right\} \quad.$$

The bootstrap $p-$value of this empirical dual likelihood statistic is given by

$$p^*=P\left(\widehat{LR}^D_{\widehat{\vartheta}^*_n}>\widehat{LR}^D_{\widehat{\vartheta}_n}\right) \quad,$$

where $\widehat{\vartheta}^*_n$ if the GMM estimator of ϑ implied by a bootstrap sample $X_1^*,..,X_n^*$ from $\widehat{F}^{\mathcal{H}_0}$; cf. Brown and Newey (1998) for a similar bootstrap procedure. The $p-$value p^* can be estimated in a Monte Carlo simulation by bootstrapping the original sample $X_1,..,X_n$. However, we can expect the results of Theorem 4 to yield accurate estimates of p^* also by means of a direct approximation based on a $\mathcal{X}^2_k$ distribution, as demonstrated for instance in Example 1 of Robinson, Ronchetti and Young (2003) for the exactly identified case. In fact, in all Monte Carlo simulations in the next section empirical dual likelihood tests are shown to provide accurate finite sample inferences also in overidentified moment conditions models.

5   Monte Carlo Investigation

In this section we analyze the accuracy of empirical dual likelihood ratio tests based on the asymptotic approximation (34) in some Monte Carlo simulations. We compute empirical rejection rates based on a $\mathcal{X}^2_k$ approximation for $\widehat{LR}^D_{\widehat{\vartheta}_n}$ and for likelihood-ratio$-$ and Wald$-$type GMM statistics in the relevant model settings. We also compute empirical rejection rates using a $\mathcal{X}^2_k$ approximation for one of the statistics proposed in ISJ (1998) and adapted by Imbens and Spady (2002) to obtain accurate interval estimation procedures. These statistics are based on the value of the tilting parameter in an empirical likelihood estimation of the unknown parameter ϑ_0. We make use of an easily computable version of such statistics (that avoids as for $\widehat{LR}^D_{\widehat{\vartheta}_n}$ a complete

19

empirical likelihood estimation of ϑ_0) which computes the desired tilting parameter with respect to the GMM estimate $\widehat{\vartheta}_n$. Tests based on statistics of this type are asymptotically first order $\mathcal{X}_k^2$ distributed and have been shown to provide accurate finite sample inference in several Monte Carlo experiments; see for instance ISJ (1998) and Imbens and Spady (2002). The ISJ statistic used in our experiments is defined by

$$\widehat{\xi}_n^{ISJ} := n \left(\mu_n \left(\widehat{\vartheta}_n \right) - \mu_n \left(\vartheta_{\mathcal{H}_0} \right) \right)' A_n B_n^{-1} A_n \left(\mu_n \left(\widehat{\vartheta}_n \right) - \mu_n \left(\vartheta_{\mathcal{H}_0} \right) \right) \xrightarrow[n \to \infty]{d} \mathcal{X}_k^2 \quad,$$

where $\mu_n \left(\widehat{\vartheta}_n \right)$ and $\mu_n \left(\vartheta_{\mathcal{H}_0} \right)$ are solutions of the empirical saddlepoint equation (35) for $\widehat{\vartheta}_n$ and $\vartheta_{\mathcal{H}_0}$, respectively, and $A_n B_n^{-1} A_n$ is a sandwich covariance matrix estimator as given for instance in Imbens and Spady (2002), p. 92.

We consider five basic simulation settings. The first three settings imply the existence of the moments generating function of ψ, an assumption required to prove both the higher order asymptotic properties of $LR_{\widehat{\vartheta}_n}^D$ and the first order asymptotic properties of $\widehat{\xi}_n^{ISJ}$; cf. Assumption 1 in the Appendix and Theorem 3.1 in Newey and Smith (2001) for the first and the second statistic, respectively. In the last two Monte Carlo experiments the moment generating function of ψ does not exist. Even if this fact does not seem to be crucial for the empirical version of our dual likelihood ratio statistic (in this case the underlying distribution is a discrete one), these last examples illustrate the accuracy of tests based on $\widehat{LR}_{\widehat{\vartheta}_n}^D$ when the moment generating function of ψ does not exist. For brevity we denote in the sequel by ξ^{LR}, ξ^W, ξ^{ISJ} and ξ^{RT} the different statistics under scrutiny.

5.1 Model 1: Burnside-Eichenbaum

The first Monte Carlo simulation focuses on a model setting of the type considered by Burnside and Eichenbaum (1996), ISJ (1998) and Imbens and Spady (2002) among others. The orthogonality function is given by

$$\psi\left(X, \vartheta \right) = \left(\begin{array}{ccc} X_1 - \vartheta & X_2 - \vartheta & X_3 - \vartheta \end{array} \right)' \quad .$$

20

For all simulations $(X_1, .., X_3)'$ is a vector of independent $\mathcal{X}_1^2$ distributed random variables. We test the correct hypothesis $\vartheta = 1$. Therefore $H = 3$ and $k = 1$. Table Ia reports results for sample sizes $n = 100, 400$.

<center>**Insert Table Ia about here**</center>

In this example ξ^{ISJ} and ξ^{RT} perform quite similarly (with a slight advantage for ξ^{ISJ}) and better than the classical GMM tests, especially in the lowest quantiles of the distribution. Notice, that in this example the numerical differences between ξ^{LR} and ξ^{W} are very small. This causes the corresponding empirical quantiles to coincide.

An alternative way of testing the hypothesis $\vartheta = 1$ in this setting consists in treating ϑ as a scale, rather than a location, parameter. We do this by means of an orthogonality function given by

$$\psi(X, \vartheta) = \left(\begin{array}{ccc} X_1/\vartheta - 1 & X_2/\vartheta - 1 & X_3/\vartheta - 1 \end{array} \right)' \quad .$$

The results obtained by using this orthogonality function are collected in Table Ib.

<center>**Insert Table Ib about here**</center>

In this second setting, the GMM Likelihood ratio type test based on ξ^{LR} is the most accurate one. The statistic ξ^{RT} performs slightly better than ξ^{W}, while ξ^{ISJ} produces some excess oversize in the low tails of the distribution even when compared with ξ^{W}.

5.2 Model 2: Overidentified Linear Regression Model

The second Monte Carlo simulation is based on a linear regression model of the form

$$Y = \vartheta X + U \quad ,$$

with U standard normal distributed and X beta distributed with parameters α and β. We estimate ϑ by means of an overidentified orthogonality function given by

$$\psi\left(Y, X, \vartheta\right) = \begin{pmatrix} Y - \vartheta X \\ \left(Y - \vartheta X\right) X \\ \left(Y - \vartheta X\right)^2 - 1 \end{pmatrix} . \qquad (36)$$

Thus, $H = 3$ and $k = 1$. The simulations are performed for sample sizes $n = 80, 140$ with parameter choices $\vartheta_0 = 1$, $\alpha = 1$, $\beta = 1$ (uniform distribution for X). We test the parametric hypothesis $\vartheta = 1$.

<center>**Insert Table II about here**</center>

From Table II we see that the accuracy of ξ^{RT} is slightly better than the one of ξ^{LR} in the lowest tails of the distribution's support. The statistics ξ^W and ξ^{ISJ} are oversized, especially in the lowest distribution quantiles. Similar patterns arise for the case $\alpha = 2$, $\beta = 2$ (symmetric distribution for X with mode at 0.5) and $\alpha = 2$, $\beta = 3$ (asymmetric distribution for X).

5.3 Model 3: Overidentified Nonlinear Regression Model

The third Monte Carlo simulation is based on a nonlinear regression model of the form

$$Y = X^\vartheta + U \qquad ,$$

with U standard normal distributed and X beta distributed with parameters α and β. We estimate ϑ by means of an overidentified orthogonality function given by

$$\psi\left(Y, X, \vartheta\right) = \begin{pmatrix} Y - X^\vartheta \\ \left(Y - X^\vartheta\right) X^\vartheta \ln\left(X\right) \end{pmatrix} .$$

Thus, $H = 2$ and $k = 1$. The simulations are performed for sample sizes $n = 500, 1000$, with parameter choices $\vartheta_0 = 0$, $\alpha = 1$, $\beta = 1$ (uniform distribution for X). We test the parametric hypothesis $\vartheta = 0$.

<center>**Insert Table III about here**</center>

<center>22</center>

From Table III we see that the accuracy of ξ^{RT} is very similar to the one of ξ^W. The statistics ξ^{ISJ} and ξ^{LR} are clearly oversized, especially in the lowest distribution quantiles. Similar patterns arise for the cases $\alpha = 2$, $\beta = 2$.

5.4   Model 4: Exponential Distribution with two Moments

The fourth Monte Carlo experiment focuses again on a low-dimensional model with two orthogonality conditions and one parameter. The orthogonality function is

$$\psi\left(X, \vartheta\right) = \left(\begin{array}{c} X - \vartheta \\ \\ X^2 - 2\vartheta^2 \end{array} \right) \quad,$$

and in all simulations X is exponentially distributed with mean $\vartheta_0 = 1$. We test the hypothesis $\vartheta = 1$. This simulation setting induces an asymmetric finite sample distribution of the standard GMM estimators through a very high skewness of the second component of ψ (cf. Imbens and Spady (2002)). Table IVa reports results for sample sizes $n = 200, 300$.

<div align="center">

Insert Table IVa about here

</div>

In this model ξ^{ISJ} gives the most accurate empirical sizes, followed by ξ^{RT}. By contrast, both likelihood-type GMM statistics are strongly oversized.

An alternative way of testing the hypothesis $\vartheta = 1$ also in this setting (cf. Model 1 above) consists in treating ϑ as a scale, rather than a location, parameter. We do this by means of an orthogonality function given by

$$\psi\left(X, \vartheta\right) = \left(\begin{array}{c} X/\vartheta - 1 \\ \\ \left(X/\vartheta\right)^2 - 2 \end{array} \right) \quad.$$

The results obtained by using this orthogonality function are collected in Table IVb.

<div align="center">

Insert Table IVb about here

</div>

In this second setting the statistic ξ^{LR} provides the most accurate inferences, followed by ξ^{RT}. By contrast, both ξ^W and ξ^{ISJ} yield strongly oversized inferences.

5.5 Model 5: Linear Regression Model with Squared Normal Regressors

The last Monte Carlo simulation is based on a linear regression model of the form

$$Y = \vartheta X^2 + U \qquad ,$$

with U and X both standard normally distributed . We estimate ϑ by means of an overidentified orthogonality function given by

$$\psi\left(Y, X, \vartheta\right) = \begin{pmatrix} Y - \vartheta X^2 \\ \left(Y - \vartheta X^2\right) X^2 \\ \left(Y - \vartheta X^2\right)^2 - 1 \end{pmatrix} \qquad . \tag{37}$$

Thus, $H = 3$ and $k = 1$. The simulations are performed using a parameter choice $\vartheta_0 = 1$ and we test the parametric hypothesis $\vartheta = 1$. Table V reports results for sample sizes $n = 200, 300$.

Insert Table V about here

In this last setting the statistic ξ^{LR} provides quite accurate inferences, while the other tests under scrutiny are strongly oversized.

6   Conclusion

We computed a class of saddlepoint approximations of a general GMM estimator which motivate a new set of accurate dual likelihood ratio tests of parametric constraints in overidentified moment conditions models. By means of saddlepoint techniques, these statistics are shown to be asymptotically chi-squared distributed to higher order, with a relative error of order $O(1/n)$. Since they require the knowledge of the moment generating function of the given orthogonality function we introduced an empirical likelihood version of the new tests which is can be applied to the fully nonparametric setting and which only requires a preliminary GMM parameter estimation to be computed. Monte Carlo evidence shows that the new empirical dual likelihood ratio tests pro-

vide a higher accuracy than standard GMM test statistics and some recent information theoretic

alternatives for a broad class of GMM models.

7 Appendix

Assumption 1 *We assume the following conditions to be satisfied:*

(i) $K_{\psi(X,\theta_1)}(\cdot)$ *exists for any $\theta_1 \in \Theta$.*

(ii) *The system of equations*

$$\sum_{i=1}^{n} \Psi(X_i, \theta) = 0 \quad ,$$

has a unique solution $F^{(n)}-$ almost surely.

(iii) *For any $\theta \in \Theta_\Psi$ the joint density of the random vector*

$$S := (S_1', S_2')' := \left(\sum_{i=1}^{n} \psi(X_i, \theta_1)', \widehat{\theta}_n' \right)'$$

with respect to $dF^{(n)}$ exists. Furthermore, for any $\theta_1 \in \Theta$ and $\mu \in \mathbb{R}^H$ the Fourier transform of S under $F^{(n)}$ and under $H_{\theta_1,\mu}^{(n)}$, respectively, is integrable with respect to Lebesgue measure, that is:

$$\int |E\left[\exp\left(iu'S\right)\right]| \, du < \infty \;, \quad \int \left| E_{H_{\theta_1,\mu}^{(n)}} \left[\exp\left(iu'S\right)\right] \right| du < \infty \;. \tag{38}$$

Assumption 2 *For any $\theta_1 \in \Theta$ the saddlepoint equation (20) has a unique solution $\mu(\theta_1)$.*

Let be a multi-index $\alpha = (\alpha_1', \alpha_2', \alpha_3')$ such that α_i is a multi index of the same dimension as θ_i.

The following multi-index notation is used to express Assumption 3 below:

$$
\begin{aligned}
|\alpha| &= |\alpha_1| + |\alpha_2| + |\alpha_3| = \sum_{i=1}^{3} \sum_{j=1}^{\dim(\theta_i)} \alpha_{ij} \quad, \\
\alpha! &= \alpha_1! \cdot \alpha_2! \cdot \alpha_3! = \prod_{i=1}^{3} \prod_{j=1}^{\dim(\theta_i)} \alpha_{ij}! \quad, \\
\partial^\alpha \Psi(x;\theta) &= \frac{\partial^\alpha \Psi(x;\theta)}{\partial \theta_1^{\alpha_1} \partial \theta_2^{\alpha_2} \partial \theta_3^{\alpha_3}} = \frac{\partial^\alpha \Psi(x;\theta)}{\prod_{i=1}^{3} \prod_{j=1}^{\dim(\theta_i)} \partial \theta_i^{\alpha_{ij}}} \quad,
\end{aligned}
$$

for $s = 1,..,\dim(\Theta_\Psi)$.

Assumption 3 *We assume the following conditions to be satisfied.*

(i) *The derivatives*

$$\partial^\alpha \Psi(X;\theta) \quad ; \quad |\alpha| \leq 4, \theta \in \Theta_\Psi \quad,$$

exist $F-$almost surely. Moreover, for any compact set $K \subset \Theta_\Psi$, any $\theta \in \Theta_\Psi$ and some $\varepsilon > 0$

$$\sup_{|\alpha|<4} \sup_{\theta_0 \in K} E_{H_{\theta_0}^{(n)}} \left[\|\partial^\alpha \Psi_s(X;\theta)\|^4 \right] < \infty \quad. \tag{39}$$

$$\sup_{|\alpha|=4} \sup_{\theta_0 \in K} E_{H_{\theta_0}^{(n)}} \left[\left(\sup_{\|\theta^*-\theta_0\|<\varepsilon} \|\partial^\alpha \Psi_s(X;\theta^*)\| \right)^4 \right] < \infty \quad. \tag{40}$$

(ii) For any $\theta \in \Theta_\Psi$ the matrices

$$\Gamma_\Psi(\theta) := E_{H^{(n)}_{\theta_1}}\left[\nabla_{\theta'}\Psi(X;\theta)\right] \quad, \quad V_\Psi(\theta) = E_{H^{(n)}_{\theta_1}}\left[\Psi(X;\theta)\Psi(X;\theta)'\right]$$

are non singular.

(iii) For any $|\alpha|, |\alpha^*| < 4$ the functions

$$\theta \longmapsto \Gamma_\Psi(\theta) \quad, \tag{41}$$

$$\theta \longmapsto E_{H^{(n)}_{\theta_1}}\left[\partial^\alpha \Psi_s(X;\theta)\partial^{\alpha^*}\Psi_s(X;\theta)'\right] \quad, \tag{42}$$

are continuous.

(iv) For any $|\alpha| < 4$ the function $\partial^\alpha \Psi(\cdot;\theta)$ is continuously differentiable with respect to the first argument, $F-$ almost surely.

Proof of Proposition 1. Denote the density of

$$Z := (Z_1', Z_2', Z_3')' = (Z_1', S')' = \left(\sum_{i=1}^{n}\Psi_1(X_i, \theta), S'\right)'$$

under $F^{(n)}$ and $H^{(n)}_{\theta_1,\mu}$, by $g^{(n)}$ and $g^{(n)}_{\theta_1,\mu}$, respectively, and define $\widetilde{k} = k + H + \dim(\Theta_\Psi)$. By Fourier inversion we have

$$
\begin{aligned}
g^{(n)}(z) &= (2\pi)^{-\widetilde{k}}\int \exp(-iu'z)\, M_Z(u)\, du \\
&= (2\pi)^{-\widetilde{k}} C_\mu(\theta_1)^{-n}\exp(-\mu'z_2)\int \exp(-iu'z)\, M_Z(u)\exp\left(\mu'z_2 - nK_{\psi(X,\theta_1)}(\mu)\right)du \\
&= (2\pi)^{-\widetilde{k}} C_\mu(\theta)^{-n}\exp(-\mu'z_2) \\
&\quad \times \int \exp(-iu'z)\, M_Z(u_1, -i\mu + u_2, u_3)\exp\left(-nK_{\psi(X,\theta_1)}(\mu)\right)du
\end{aligned}
$$

Now,

$$
\begin{aligned}
M_Z(u_1, -i\mu + u_2, u_3) &= \int \exp(iu'z + \mu'z_2)\, g^{(n)}(z)\, dz \\
&= \int \exp(iu'z)\exp(\mu'z_2)\, g^{(n)}(z)\, dz \\
&= \exp\left(nK_{\psi(X,\theta_1)}(\mu)\right)\int \exp(iu'z)\, g^{(n)}_{\theta_1,\mu}(z)\, dz \\
&= \exp\left(nK_{\psi(X,\theta_1)}(\mu)\right) M_Z^{\theta_1,\mu}(u) \quad,
\end{aligned}
$$

where $M_Z^{\theta_1,\mu}$ is the Fourier transform of Z under $H^{(n)}_{\theta_1,\mu}$. By Fourier inversion this gives

$$g^{(n)}(z) = C_\mu(\theta_1)^{-n}\, g^{(n)}_{\theta_1,\mu}(z)\exp(-\mu'z_2) \quad.$$

Integrating the RHS and the LHS with respect to z_1, z_2, on the set $z_3 = \theta^*$, proves the proposition.

∎

Proof of Lemma 1. Under the given assumptions, we have from the proof of Lemma 1 in Fan and Field (1995) for our case

$$\widehat{\theta}_n - \theta^* = -A\left(\theta^*\right)^{-1} \overline{\Psi} + O_{H_\theta^{(n)}}\left(\frac{1}{n}\right) \quad,$$

where

$$\overline{\Psi} = \frac{1}{n}\sum_{i=1}^{n}\Psi\left(X_i; \theta^*\right) \quad, \quad A\left(\theta^*\right) = E_{H_{\theta_1}^{(n)}}\left[\frac{1}{n}\sum_{i=1}^{n}\nabla_{\theta'}\Psi\left(X_i;\theta^*\right)\right] \quad.$$

Computing $A\left(\theta^*\right)$ explicitly, it first follows

$$\nabla_{\theta'}\Psi\left(X,\theta^*\right) = \begin{bmatrix} \nabla_{\theta'_1}\Psi_1\left(X,\theta^*\right) & \nabla_{\theta'_2}\Psi_1\left(X,\theta^*\right) & \nabla_{\theta'_3}\Psi_1\left(X,\theta^*\right) \\[2mm] \nabla_{\theta'_1}\Psi_2\left(X,\theta^*\right) & \nabla_{\theta'_2}\Psi_2\left(X,\theta^*\right) & \nabla_{\theta'_3}\Psi_2\left(X,\theta^*\right) \\[2mm] \nabla_{\theta'_1}\Psi_3\left(X,\theta^*\right) & \nabla_{\theta'_2}\Psi_3\left(X,\theta^*\right) & \nabla_{\theta'_3}\Psi_3\left(X,\theta^*\right) \end{bmatrix}$$

where

$$\nabla_{\theta'_1}\Psi_1\left(X,\theta^*\right) = \Gamma'_{\theta_1}V_{\theta_1}^{-1}\nabla_{\theta'_1}\psi\left(X,\theta_1\right) \quad,$$

$$\nabla_{\theta'_2}\Psi_1\left(X,\theta^*\right) = \left[I_k \otimes V_{\theta_1}^{-1}\psi\left(X,\theta_1\right)\right]\frac{\partial vec\left(\Gamma_{\theta_1}\right)}{\partial\theta'_2} \quad,$$

$$\nabla_{\theta'_3}\Psi_1\left(X,\theta^*\right) = \left[\psi'\left(X,\theta_1\right)\otimes\Gamma'_{\theta_1}\right]\frac{\partial vech\left(V_{\theta_1}^{-1}\right)}{\partial\theta'_3} \quad,$$

$$\nabla_{\theta'_2}\Psi_2\left(X,\theta^*\right) = -id_{\dim(\theta_2)\times\dim(\theta_2)} \quad,$$

$$\nabla_{\theta'_3}\Psi_2\left(X,\theta^*\right) = 0_{\dim(\theta_2)\times\dim(\theta_3)} \quad,$$

$$\nabla_{\theta'_3}\Psi_3\left(X,\theta^*\right) = -id_{\dim(\theta_3)\times\dim(\theta_3)} \quad.$$

Since by construction under the tilted distribution $H_{\theta_1}^{(n)}$ the orthogonality function $\psi\left(X,\theta_1\right)$ has expectation 0 we obtain

$$E_{H_{\theta_1}^{(n)}}\left[\nabla_{\theta'}\Psi\left(X,\theta^*\right)\right] = \begin{bmatrix} E_{H_{\theta_1}^{(n)}}\left[\nabla_{\theta'_1}\Psi_1\left(X,\theta^*\right)\right] & 0 & 0 \\[2mm] 0 & E_{H_{\theta_1}^{(n)}}\left[\nabla_{\theta'_2}\Psi_2\left(X,\theta^*\right)\right] & 0 \\[2mm] 0 & 0 & E_{H_{\theta_1}^{(n)}}\left[\nabla_{\theta'_3}\Psi_3\left(X,\theta^*\right)\right] \end{bmatrix} \quad,$$

where

$$E_{H_{\theta_1}^{(n)}} \left[\nabla_{\theta_1'} \Psi_1 \left(X, \theta^* \right) \right] = \Gamma_{\theta_1}' V_{\theta_1}^{-1} E_{H_{\theta_1}^{(n)}} \left[\nabla_{\theta_1'} \psi \left(X, \theta_1 \right) \right] = \Gamma_{\theta_1}' V_{\theta_1}^{-1} \Gamma_{\theta_1} \quad ,$$

$$E_{H_{\theta_1}^{(n)}} \left[\nabla_{\theta_2'} \Psi_2 \left(X, \theta^* \right) \right] = -id_{\dim(\theta_2) \times \dim(\theta_2)} \quad ,$$

$$E_{H_{\theta_1}^{(n)}} \left[\nabla_{\theta_3'} \Psi_3 \left(X, \theta^* \right) \right] = -id_{\dim(\theta_3) \times \dim(\theta_3)} \quad ,$$

implying

$$\begin{pmatrix} \widehat{\theta}_{1n} - \theta_1 \\ \\ \widehat{\theta}_{2n} - vech\left(\Gamma_{\theta_1}\right) \\ \\ \widehat{\theta}_{3n} - vech\left(V_{\theta_1}\right) \end{pmatrix} = - \begin{pmatrix} \left(\Gamma_{\theta_1}' V_{\theta_1}^{-1} \Gamma_{\theta_1}\right)^{-1} \overline{\Psi}_1 \\ \\ \overline{\Psi}_2 \\ \\ \overline{\Psi}_3 \end{pmatrix} + O_{H_{\theta_1}^{(n)}} \left(\frac{1}{n} \right)$$

$$= - \begin{pmatrix} \left(\Gamma_{\theta_1}' V_{\theta_1}^{-1} \Gamma_{\theta_1}\right)^{-1} \Gamma_{\theta_1}' V_{\theta_1}^{-1} \overline{\psi} \\ \\ \overline{\Psi}_2 \\ \\ \overline{\Psi}_3 \end{pmatrix} + O_{H_{\theta_1}^{(n)}} \left(\frac{1}{n} \right) \quad ,$$

where

$$\overline{\Psi}_k = \frac{1}{n} \sum_{i=1}^{n} \Psi_k \left(X_i, \theta^* \right) \quad , \quad \overline{\psi} = \frac{1}{n} \sum_{i=1}^{n} \psi \left(X_i, \theta_1 \right) \quad .$$

Therefore we get

$$\Gamma_{\theta_1}' V_{\theta_1}^{-1} \overline{\psi} = - \left(\Gamma_{\theta_1}' V_{\theta_1}^{-1} \Gamma_{\theta_1} \right) \left(\widehat{\theta}_{1n} - \theta_1 \right) + O_{H_{\theta_1}^{(n)}} \left(\frac{1}{n} \right) \quad ,$$

that is

$$\widehat{\theta}_{1n} - \theta_1^* = - \left(\Gamma_{\theta_1}' V_{\theta_1}^{-1} \Gamma_{\theta_1} \right)^{-1} \Gamma_{\theta_1}' V_{\theta_1}^{-1} \overline{\psi} + O_{H_{\theta_1}^{(n)}} \left(\frac{1}{n} \right) \quad .$$

This last result implies that for some vector $M_{\theta_1} \in \mathbb{R}^k$ we can have

$$\mu' \left(\theta_1 \right) \sum_{i=1}^{n} \psi \left(X_i, \theta_1 \right) = M_{\theta_1}' \left(\widehat{\theta}_{1n} - \theta_1^* \right) + O_{H_{\theta_1}^{(n)}} \left(\frac{1}{n} \right) \quad .$$

Indeed, by writing the system

$$\mu' \left(\theta_1 \right) = -\frac{1}{n} M_{\theta_1}' \left(\Gamma_{\theta_1}' V_{\theta_1}^{-1} \Gamma_{\theta_1} \right)^{-1} \Gamma_{\theta_1}' V_{\theta_1}^{-1} \quad ,$$

it follows

$$M_{\theta_1} = -n \cdot \Gamma_{\theta_1}' \mu \left(\theta_1 \right) \quad .$$

29

This last result finally implies

$$
\begin{aligned}
D_\mu\left(\theta^*\right) &= E_{H_{\theta_1},\mu}\left[\exp\left(-\sum_{i=1}^n \mu'\psi\left(X_i;\theta_1\right)\right)\bigg|\,\widehat{\theta}_n = \theta^*\right] \\
&= E_{H_{\theta_1},\mu}\left[\exp\left(-M_{\theta_1}'\left(\widehat{\theta}_{1n} - \theta_1^*\right) - O_{H_{\theta_1}^{(n)}}\left(\frac{1}{n}\right)\right)\bigg|\,\widehat{\theta}_n = \theta^*\right] \\
&= 1 + O\left(\frac{1}{n}\right) \quad,
\end{aligned}
$$

concluding the proof of the Lemma. ∎

Proof of Proposition 3. From the proof of Lemma 1 we have

$$
\widehat{\theta}_{1n} = \theta_1 - \left(\Gamma_{\theta_1}' V_{\theta_1}^{-1}\Gamma_{\theta_1}\right)^{-1}\Gamma_{\theta_1}' V_{\theta_1}^{-1}\overline{\psi} + O_{H_{\theta_1}^{(n)}}\left(\frac{1}{n}\right) \quad, \tag{43}
$$

where

$$
\overline{\psi} = \frac{1}{n}\sum_{i=1}^n \psi\left(X_i,\theta_1\right) \quad.
$$

Notice that under $H_{\theta_1}^{(n)}$ equation (43) is centered at θ_1, up to the first order. Thus, we can apply

an Edgeworth expansion to approximate the $H_{\theta_1}^{(n)}-$ density of $\widehat{\theta}_{1n}$, to obtain (cf. also Field (1982),

Theorem 1)

$$
q_{\theta_1}\left(\theta_1\right) = \left(\frac{n}{2\pi}\right)^{\frac{k}{2}}\left|\det\left(\Gamma_{\theta_1}' V_{\theta_1}^{-1}\Gamma_{\theta_1}\right)\right|^{\frac{1}{2}}\left(1 + O\left(\frac{1}{n}\right)\right) \quad,
$$

using the relation

$$
E_{H_{\theta_1}^{(n)}}\left[\nabla_{\theta_1'}\Gamma_{\theta_1}' V_{\theta_1}^{-1}\psi\left(X,\theta_1\right)\right] = \Gamma_{\theta_1}' V_{\theta_1}^{-1}\Gamma_{\theta_1} = E_{H_{\theta_1}^{(n)}}\left[\Gamma_{\theta_1}' V_{\theta_1}^{-1}\psi\left(X,\theta_1\right)\psi\left(X,\theta_1\right)' V_{\theta_1}^{-1}\Gamma_{\theta_1}\right]
$$

obtained in the proof of Lemma 1. The second statement in the proposition now follows from

Lemma 1. ∎

Proof of Theorem 4.

The proof uses arguments along the lines of Robinson, Ronchetti and Young (2003), applied

to the saddlepoint approximation for the densitiy of a GMM estimator in Proposition 3. Let

$$
p = P_{\mathcal{H}_0}\left(n \cdot LR_{\widehat{\vartheta}_n}^D > a\right)
$$

be the p−value of the test under scrutiny and define

$$h\left(\vartheta\right) = -K_{\psi(X,\vartheta)}\left(\mu\left(\vartheta\right)\right) \ ,$$

where LR_ϑ^D is given in (32) and (33), and $\mu\left(\vartheta\right)$ is defined by (20). Without loss of generality we can assume the existence of $\vartheta_0 = 0$ such that $\mathcal{H}_0$ is satisfied and $\frac{\partial^2 h}{\partial\vartheta\partial\vartheta'}\left(\vartheta_0\right) = id_k$. Denote by A the set

$$A = \left\{\vartheta \ \mid \ n \cdot LR_\vartheta^D > a\right\} = \left\{\vartheta \ \mid \ 2n \cdot h\left(\vartheta\right) > a\right\}.$$

Then, by means of the saddlepoint approximation of the density of $\widehat{\vartheta}_n$ given by Proposition 3 (ii), we obtain

$$
\begin{aligned}
p &= \int_A \left(\frac{n}{2\pi}\right)^{\frac{k}{2}} C_{\mu(\vartheta)}\left(\vartheta\right)^{-n} \sqrt{\det\left(\Gamma_\vartheta' V_\vartheta^{-1}\Gamma_\vartheta\right)} \left(1 + O\left(\frac{1}{n}\right)\right) d\vartheta \\
&= \int_A \left(\frac{n}{2\pi}\right)^{\frac{k}{2}} \exp\left(-nh\left(\vartheta\right)\right) \sqrt{\det\left(\widetilde{D}\left(\vartheta\right)\right)} \left(1 + O\left(\frac{1}{n}\right)\right) d\vartheta \\
&= \int_{\widetilde{A}} c_n n^{-\frac{1}{2}} \exp\left(-nh\left(zn^{-\frac{1}{2}}\right)\right) \sqrt{\det\left(\widetilde{D}\left(zn^{-\frac{1}{2}}\right)\right)} \left(1 + O\left(\frac{1}{n}\right)\right) dz \qquad (44)
\end{aligned}
$$

where $\widetilde{D}\left(\vartheta\right) = \Gamma_\vartheta' V_\vartheta^{-1}\Gamma_\vartheta$, $\widetilde{A} = \left\{z \ \mid \ 2n \cdot h\left(zn^{-\frac{1}{2}}\right) > a\right\}$, and c_n is the normalizing constant (to make the integrand a proper density). We now perform two transformations

$$z \overset{f_1}{\longmapsto} \begin{pmatrix} r \\ s \end{pmatrix} \overset{f_2}{\longmapsto} \begin{pmatrix} u \\ s \end{pmatrix}.$$

f_1 is the polar transformation defined by $r = \|z\|$ (the radial component) and $s \in S_k$, the k−dimensional unit sphere. f_2 is defined by $u = 2nh\left(zn^{-\frac{1}{2}}\right)$. The Jacobians J_1, J_2, respectively, of these two transformations are $J_1 = r^{k-1}$ and

$$J_2 := \left|\frac{\partial u}{\partial r}\right|^{-1} = \frac{r}{2n^{\frac{1}{2}}\frac{\partial h}{\partial\vartheta'}\left(zn^{-\frac{1}{2}}\right)z} \ ,$$

respectively. Therefore, we can rewrite (44) as

$$
\begin{aligned}
p &= \int_a^\infty \int_{S_k} c_n n^{-\frac{1}{2}} e^{-\frac{u}{2}} \sqrt{\det\left(\widetilde{D}\left(zn^{-\frac{1}{2}}\right)\right)} J_1 J_2 \left(1 + O\left(\frac{1}{n}\right)\right) ds\, du \\
&= \int_a^\infty c_n e^{-\frac{u}{2}} \left[\int_{S_k} \delta\left(u, s\right) \left(1 + O\left(\frac{1}{n}\right)\right) ds\right] du
\end{aligned}
$$

where

$$\delta\left(u,s\right)=\frac{r^{k}\sqrt{\det\left(\widetilde{D}\left(zn^{-\frac{1}{2}}\right)\right)}}{2n\frac{\partial h}{\partial\vartheta'}\left(zn^{-\frac{1}{2}}\right)z}=\frac{1}{n}\frac{\left(z'z\right)^{\frac{k}{2}}\sqrt{\det\left(\widetilde{D}\left(zn^{-\frac{1}{2}}\right)\right)}}{2\frac{\partial h}{\partial\vartheta'}\left(zn^{-\frac{1}{2}}\right)z}=:\Delta\left(z\right)\quad.$$

We now first expand $\Delta\left(z\right)$ around $z=0$. For any component $\widetilde{D}_{ij}$ of $\widetilde{D}$ it follows

$$\widetilde{D}_{ij}\left(zn^{-\frac{1}{2}}\right)=\widetilde{D}_{ij}\left(0\right)+n^{-\frac{1}{2}}\frac{\partial\widetilde{D}_{ij}\left(0\right)}{\partial\vartheta'}z+O\left(n^{-1}\right)\quad.$$

Therefore, when computing the determinant of the matrix

$$\left[\widetilde{D}_{ij}\left(0\right)\right]_{1\leq i,j\leq k}+n^{-\frac{1}{2}}\left[\frac{\partial\widetilde{D}_{ij}\left(0\right)}{\partial\vartheta'}\right]_{1\leq i,j\leq k}z$$

we see that up to orders $O\left(n^{-1}\right)$ we obtain

$$\det\left(\widetilde{D}\left(zn^{-\frac{1}{2}}\right)\right)=\det\left(\widetilde{D}\left(0\right)\right)\left(1+n^{-\frac{1}{2}}\xi_{1}\left(z\right)+O\left(n^{-1}\right)\right)$$

and

$$\sqrt{\det\left(\widetilde{D}\left(zn^{-\frac{1}{2}}\right)\right)}=\sqrt{\det\left(\widetilde{D}\left(0\right)\right)}\left(1+n^{-\frac{1}{2}}\xi_{2}\left(z\right)+O\left(n^{-1}\right)\right)$$

where $\xi_{1}\left(z\right)$ and $\xi_{2}\left(z\right)$ are linear functions of z, implying $\int_{S_{k}}\xi_{i}dz=0$, $i=1,2$. We further expand u in $\Delta\left(z\right)$ at $z=0$. This gives,

$$\begin{aligned}u&=2nh\left(zn^{-\frac{1}{2}}\right)\\&=2n\left(h\left(0\right)+n^{-\frac{1}{2}}\frac{\partial h}{\partial\vartheta}\left(0\right)z+\frac{1}{2}n^{-1}z'\frac{\partial^{2}h}{\partial\vartheta\partial\vartheta'}\left(0\right)z+\frac{1}{6}n^{-\frac{3}{2}}\rho\left(z\right)+O\left(n^{-2}\right)\right)\\&=z'z\left(1+\frac{1}{3}n^{-\frac{1}{2}}\frac{\rho\left(z\right)}{z'z}+O\left(n^{-1}\right)\right),\end{aligned}$$

and

$$z'z=u\left(1-\frac{1}{3}n^{-\frac{1}{2}}\frac{\rho\left(z\right)}{z'z}+O\left(n^{-1}\right)\right)\quad,$$

since $h\left(0\right)=0$, $\frac{\partial h}{\partial\vartheta}\left(0\right)=0$ and, by assumption,

$$\frac{\partial^{2}h}{\partial\vartheta\partial\vartheta'}\left(0\right)=\Gamma_{0}'V_{0}^{-1}\Gamma_{0}=id_{k}\quad.$$

Notice, that $\rho\left(z\right)$ is a linear combination of terms of the form $z_{i}z_{j}z_{w}$, $1\leq i,j,w\leq k$, where $z=\left(z_{1},z_{2},..,z_{k}\right)'$, so that $\int_{S_{k}}\rho\left(z\right)dz=0$. In a similar way as for the expansion of u, it also

32

follows

$$2\frac{\partial h}{\partial \vartheta'}\left(zn^{-\frac{1}{2}}\right)z = 2n^{-\frac{1}{2}}z'z + O\left(n^{-1}\right) = 2n^{-\frac{1}{2}}u\left(1 - \frac{1}{3}n^{-\frac{1}{2}}\frac{\rho(z)}{z'z} + O\left(n^{-1}\right)\right) \quad .$$

Collecting these results, we thus obtain

$$\Delta(z) = n^{-\frac{1}{2}}\sqrt{\det\left(\widetilde{D}(0)\right)}u^{\frac{k}{2}-1}\left(1 + n^{-\frac{1}{2}}b(z) + O\left(n^{-1}\right)\right) \quad ,$$

where b is an odd function of z, implying $\int_{S_k} b(z)\, dz$. When inserting this last result in the expression for p we finally get

$$\begin{aligned}
p &= \int_a^\infty c_n e^{-\frac{u}{2}}\left[\int_{S_k}\delta(u,s)\left(1 + O\left(\frac{1}{n}\right)\right)ds\right]du \\
&= c_n n^{-\frac{1}{2}}\sqrt{\det\left(\widetilde{D}(0)\right)}C_k\int_a^\infty e^{-\frac{u}{2}}u^{\frac{k}{2}-1}\left(1 + O\left(\frac{1}{n}\right)\right)du \quad ,
\end{aligned}$$

where C_k is the surface of S_k. The statement of the theorem now follows by noting that $c_n n^{-\frac{1}{2}}\sqrt{\det\left(\widetilde{D}(0)\right)}C_k$ has to be the normalizing constant of the density of a $\mathcal{X}_k^2$ distribution.

∎

8 Tables

Table Ia: Empirical Size of the Tests: Model 1 (Burnside-Eichenbaum) with $\mathcal{X}_1^2$-distribution, $H = 3$, $k = 1$, sample sizes $n = 100, 400$. 10000 replications.

Size $(n = 100)$	ξ^{LR}	ξ^{W}	ξ^{ISJ}	ξ^{RT}	Size $(n = 400)$	ξ^{LR}	ξ^{W}	ξ^{ISJ}	ξ^{RT}
0.100	0.1778	0.1778	0.1696	0.1707	0.100	0.1264	0.1264	0.1202	0.1220
0.050	0.1127	0.1127	0.0978	0.1009	0.050	0.0676	0.0676	0.0618	0.0642
0.025	0.0727	0.0727	0.0609	0.0631	0.025	0.0373	0.0373	0.0334	0.0343
0.010	0.0443	0.0443	0.0340	0.0359	0.010	0.0190	0.0190	0.0142	0.0153
0.005	0.0326	0.0326	0.0212	0.0230	0.005	0.0113	0.0113	0.0070	0.0085
0.001	0.0149	0.0149	0.0071	0.0089	0.001	0.0025	0.0025	0.0017	0.0020

Table Ib: Empirical Size of the Tests: Model 1 (Burnside-Eichenbaum) with scaled

$\mathcal{X}_1^2$-distribution, $H = 3$, $k = 1$, sample sizes $n = 100, 400$. 10000 replications.

Size $(n = 100)$	ξ^{LR}	ξ^W	ξ^{ISJ}	ξ^{RT}	Size $(n = 400)$	ξ^{LR}	ξ^W	ξ^{ISJ}	ξ^{RT}
0.100	0.1312	0.1436	0.1477	0.1425	0.100	0.1120	0.1143	0.1124	0.1133
0.050	0.0735	0.0833	0.0895	0.0819	0.050	0.0550	0.0598	0.0605	0.0568
0.025	0.0415	0.0489	0.0588	0.0473	0.025	0.0300	0.0324	0.0325	0.0309
0.010	0.0183	0.0264	0.0337	0.0260	0.010	0.0129	0.0139	0.0154	0.0135
0.005	0.0116	0.0167	0.0235	0.0156	0.005	0.0053	0.0083	0.0095	0.0066
0.001	0.0027	0.0078	0.0114	0.0055	0.001	0.0013	0.0017	0.0023	0.0017

Table II: Empirical Size of the Tests: Model 2, regression model with, $H = 3$, $k = 1$,

sample sizes $n = 80, 140$. $\alpha = 1$, $\beta = 1$ (uniform distribution). 10000 replications.

Size $(n = 80)$	ξ^{LR}	ξ^{W}	ξ^{ISJ}	ξ^{RT}	Size $(n = 140)$	ξ^{LR}	ξ^{W}	ξ^{ISJ}	ξ^{RT}
0.100	0.1138	0.1286	0.1366	0.1173	0.100	0.1046	0.1123	0.1150	0.1083
0.050	0.0598	0.0686	0.0768	0.0612	0.050	0.0544	0.0601	0.0622	0.0554
0.025	0.0334	0.0398	0.0479	0.0320	0.025	0.0300	0.0328	0.0343	0.0288
0.010	0.0169	0.0191	0.0262	0.0136	0.010	0.0129	0.0146	0.0155	0.0117
0.005	0.0097	0.0118	0.0167	0.0075	0.005	0.0070	0.0080	0.0089	0.0061
0.001	0.0044	0.0047	0.0082	0.0023	0.001	0.0012	0.0017	0.0031	0.0010

Table III: Empirical Size of the Tests: Model 3, nonlinear regression model with,

$H = 2$, $k = 1$, sample sizes $n = 500, 1000$, $\alpha = 1$, $\beta = 1$. 10000 replications.

Size $(n = 500)$	ξ^{LR}	ξ^{W}	ξ^{ISJ}	ξ^{RT}	Size $(n = 1000)$	ξ^{LR}	ξ^{W}	ξ^{ISJ}	ξ^{RT}
0.100	0.1233	0.1144	0.1142	0.1182	0.100	0.1035	0.0991	0.1014	0.1008
0.050	0.0778	0.0596	0.0657	0.0656	0.050	0.0576	0.0518	0.0538	0.0511
0.025	0.0479	0.0353	0.0422	0.0346	0.025	0.0351	0.0260	0.0299	0.0278
0.010	0.0297	0.0187	0.0266	0.0178	0.010	0.0171	0.0133	0.0159	0.0127
0.005	0.0222	0.0121	0.0190	0.0122	0.005	0.0117	0.0076	0.0105	0.0063
0.001	0.0129	0.0073	0.0103	0.0063	0.001	0.0050	0.0025	0.0040	0.0020

Table IVa: Empirical Size of the Tests: Model 4, Exponential Distribution with two

Moments, $H = 2$, $k = 1$, sample sizes $n = 200, 300$. 10000 replications.

Size $(n = 200)$	ξ^{LR}	ξ^{W}	ξ^{ISJ}	ξ^{RT}	Size $(n = 300)$	ξ^{LR}	ξ^{W}	ξ^{ISJ}	ξ^{RT}
0.100	0.1660	0.1523	0.1257	0.1366	0.100	0.1505	0.1391	0.1158	0.1273
0.050	0.1101	0.0951	0.0652	0.0781	0.050	0.0933	0.0814	0.0608	0.0683
0.025	0.0782	0.0616	0.0347	0.0447	0.025	0.0611	0.0495	0.0309	0.0380
0.010	0.0526	0.0348	0.0139	0.0215	0.010	0.0378	0.0281	0.0139	0.0196
0.005	0.0398	0.0248	0.0078	0.0130	0.005	0.0289	0.0201	0.0078	0.0105
0.001	0.0225	0.0109	0.0022	0.0043	0.001	0.0158	0.0087	0.0022	0.0027

Table IVb: Empirical Size of the Tests: Model 4, Scaled Exponential Distribution

with two Moments, $H = 2$, $k = 1$, sample sizes $n = 200, 300$. 10000 replications.

Size $(n = 200)$	ξ^{LR}	ξ^{W}	ξ^{ISJ}	ξ^{RT}	Size $(n = 300)$	ξ^{LR}	ξ^{W}	ξ^{ISJ}	ξ^{RT}
0.100	0.1115	0.1341	0.1419	0.1215	0.100	0.1069	0.1248	0.1301	0.1146
0.050	0.0581	0.0813	0.0932	0.0684	0.050	0.0541	0.0673	0.0755	0.0603
0.025	0.0325	0.0501	0.0630	0.0378	0.025	0.0274	0.0400	0.0462	0.0320
0.010	0.0145	0.0271	0.0403	0.0179	0.010	0.0122	0.0226	0.0284	0.0148
0.005	0.0088	0.0187	0.0307	0.0111	0.005	0.0071	0.0143	0.0206	0.0093
0.001	0.0040	0.0088	0.0182	0.0035	0.001	0.0021	0.0060	0.0103	0.0023

Table V: Empirical Size of the Tests: Model 5, linear regression model with $H = 3$, $k = 1$, sample sizes $n = 200, 300$. 10000 replications.

Size $(n = 200)$	ξ^{LR}	ξ^{W}	ξ^{ISJ}	ξ^{RT}	Size $(n = 300)$	ξ^{LR}	ξ^{W}	ξ^{ISJ}	ξ^{RT}
0.100	0.1222	0.1276	0.1320	0.1072	0.100	0.1144	0.1174	0.1182	0.1063
0.050	0.0699	0.0731	0.0744	0.0512	0.050	0.0642	0.0662	0.0649	0.0514
0.025	0.0394	0.0417	0.0449	0.0246	0.025	0.0329	0.0348	0.0348	0.0232
0.010	0.0196	0.0214	0.0246	0.0092	0.010	0.0144	0.0150	0.0168	0.0078
0.005	0.0115	0.0118	0.0157	0.0044	0.005	0.0087	0.0090	0.0107	0.0039
0.001	0.0042	0.0048	0.0056	0.0005	0.001	0.0026	0.0030	0.0035	0.0006

References

Almudevar, A., C. Field, and J. Robinson (2000) The Density of Multivariate M-Estimates, *Annals of Statistics, 28, 275-297.*

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

Back, K. and D. Brown (1993) Estimating Distributions from Moment Restrictions, Manuscript.

Battacharya, R. N. and J. K. Gosh (1978) On the Validity of the Formal Edgeworth Expansion, *Annals of Statistics*, 6, 434-451.

Bravo, F. (1999) Empirical Likelihood Based Inference with Applications to Some Econometric Models, Manuscript, University of York.

Brown, B. and W. Newey (1998) GMM, Efficient Bootstrapping and Improved Inference, Mimeo, Massachusetts Institute of Technology.

Burnside, C. and M. Eichenbaum (1996) Small Sample Properties of Generalized Method of Moments Wald Tests, *Journal of Business and Economic Statistics*, 14, 294-308.

Daniels, H. E. (1954) Saddlepoint Approximations in Statistics, *Annals of Mathematical Statistics*, 25, 631-650.

Daniels, H. E. and G. A. Young (1991) Saddlepoint Approximations fo the Studentized Mean, with an Application to the Bootstrap, *Biometrika*, 78, 169-179.

Davison, A. C., D. V. Hinkley and B. J. Worton (1995) Accurate and Efficient Construction of Bootstrap Likelihoods, *Statistics and Computing*, 5, 257-264.

Di Ciccio, T. J., and J. P. Romano (1990) Nonparametric Confidence Limits by Resampling Methods and Least Favourable Distributions, *International Statistical Review*, 58, 59-76.

Fan, R. Y. K. and C. A. Field (1995) Approximations for Marginal Densities of M-Estimators, *Canadian Journal of Statistics*, 23, 185-197.

Field, C. A. and F. Hampel (1982) Small Sample Asymptotic Distributions of M-Estimators of Location, *Biometrika*, 69, 29-46.

Field, C. A. (1982) Small Sample Asymptotics for Multivariate M-Estimates, *Annals of Statistics*, 10, 672-689.

Field, C. A. and E. Ronchetti (1990) *Small Sample Asymptotics*, Hayward CA, IMS Monograph Series.

Gatto, R. and E. Ronchetti (1996) General Saddlepoint Approximations of Marginal Densities and Tail Probabilities, *Journal of the American Statistical Association*, 91, 666-673.

Gourieroux, C. and A. Monfort (1989) A General Framework for Testing a Null Hypothesis in a Mixed Form, *Econometric Theory*, 5, 63-82.

Hall, P. and J. Horowitz (1996) Bootstrap Critical Values for Tests Based on Generalized Method of Moments Estimators, *Econometrica*, 64, 891-916.

Hall, P. and B. Presnell (1999) Intentionally Biased Bootstrap Methods, *Journal of the Royal Statistical Society, Series B*, 61, 143-158.

Hampel, F. (1973) Some Small Sample Asymptotics, *Proceedings of the Second Prague Symposium on Asymptotic Statistics*, Prague, 173-179.

Hansen, B. (1999) Non Parametric Dependent Data Bootstrap for Conditional Moment Models, Mimeo, University of Wisconsin-Madison.

Hansen, L. P. (1982) Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 50, 1029-1054.

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

Huber (1981) *Robust Statistics*, Wiley, New York.

Imbens, G. W. (1997) One-step Estimators in Overidentified Generalized Method of Moments estimator. *Review of Economic Studies*, 64, 359–383.

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

Imbens, G. W., and R. H. Spady (2002) Confidence Intervals in Generalized Method of Moments Models, *Journal of Econometrics*, 107, 87-98.

Jing B. and J. Robinson (1994) Saddlepoint Approximations for Marginal and Conditional Probabilities of Transformed Variables. *Annals of Statistics*, 22, 1115-1132.

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

Lazar, N. and P. Mykland (1999) Empirical Likelihood in the Presence of Nuisance Parameters, *Biometrika*, 86, 203-211.

Mykland, P. (1995) Dual Likelihood, *Annals of Statistics*, 23, 396-421.

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

Newey, W. K. and R. J. Smith (2001) Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators, Mimeo, Massachusetts Institute of Technology.

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

Qin, J. and J. Lawless (1994) Empirical Likelihood and General Estimating Equations, *Annals of Statistics*, 22, 300-325.

Robinson, J. E. Ronchetti and G. A. Young (2003) Saddlepoint Approximations and Tests Based on Multivariate M-Estimates, *Annals of Statistics*, forthcoming.

Ronchetti, E., and F. Trojani (2001) Robust Inference with GMM Estimators, *Journal of Econometrics* 101, 37-69.

Singleton, K. (2000) Estimation of Affine Asset Pricing Models Using the Empirical Characteristic function, Manuscript.

Tingley, M. A. and C. A. Field (1990) Small Sample Confidence Intervals, *Journal of the American Statistical Society*, 85, 427-434.