

A New Accurate Test for Indirect Inference

Veronika Czellar and Elvezio Ronchetti

Abstract In this paper we summarize the results of Czellar and Ronchetti (2010), where accurate parameter and overidentification tests for indirect inference are derived. Under the null hypothesis the new tests are asymptotically χ^2 -distributed with a *relative error* of order n^{-1} . They exhibit a better finite sample accuracy than classical tests for indirect inference which have the same asymptotic distribution but with an absolute error of order $n^{-1/2}$. Robust versions of the tests are also provided. Their accuracy is illustrated in various models, including nonlinear regression, Poisson regression with overdispersion, and diffusion models.

Key words: indirect inference, M -estimator, nonlinear regression, overdispersion, parameter test, robust estimator, saddlepoint test, sparsity, test for overidentification

In complex models where the likelihood function is not available in closed form, indirect inference provides a broad class of estimators which includes the method of moments and the generalized method of moments estimators (Hansen, 1982) with additively separable orthogonality functions, and the simulated method of moments estimators (Lee and Ingram, 1991; Duffie and Singleton, 1993). Indirect inference estimators were introduced by Smith (1993) and Gouriéroux, Monfort, Renault (1993), and have now been applied in a variety of fields, including finance and biostatistics. Good surveys on indirect inference are provided by Gouriéroux and Monfort (1996), Heggland and Frigessi (2004) and Jiang and Turnbull (2004).

Veronika Czellar

Department of Economics and Decision Sciences, HEC Paris, 1 rue de la Libération, 78351 Jouy en Josas, France, e-mail: czellar@hec.fr

Elvezio Ronchetti

Department of Econometrics, University of Geneva, Blv. Pont d'Arve 40, 1211 Geneva, Switzerland, e-mail: Elvezio.Ronchetti@unige.ch

In this paper we focus on testing for indirect inference. Let $y_1, \dots, y_n$ be a random sample from a distribution F_θ , where $\theta \in \mathfrak{R}^p$ is an unknown parameter. If the direct estimation of θ is unfeasible, choose a simple(r) auxiliary estimator $\tilde{\mu} \in \mathfrak{R}^r$, $r \geq p$ which can be written as an M -estimator defined by

$$\sum_{i=1}^n \Psi(y_i; \tilde{\mu}) = 0, \quad (1)$$

where Ψ is an appropriate function. Then for a given one-to-one function $\mu : \mathfrak{R}^p \rightarrow \mathfrak{R}^r$ and a given symmetric, positive definite matrix $\Omega \in \mathfrak{R}^r \times \mathfrak{R}^r$, the indirect inference estimator $\hat{\theta}$ is defined by

$$\hat{\theta} = \arg \min_{\theta} \{ \tilde{\mu} - \mu(\theta) \}^T \Omega \{ \tilde{\mu} - \mu(\theta) \}, \quad (2)$$

i.e. $\hat{\theta}$ is the value of the parameter θ such that $\mu(\hat{\theta})$ is closest to $\tilde{\mu}$ as measured by the metric defined by the matrix Ω . The function $\mu(\theta)$ is typically chosen as follows.

1. The function $\mu(\theta) = E_{F_\theta}(\tilde{\mu})$, or $\text{plim}_{F_\theta} \tilde{\mu}$, is deterministic and is often called binding function, or bridge relationship, or target function.
2. The function $\mu(\theta) = m^{-1} \sum_{s=1}^m \tilde{\mu}^{(s)}(\theta)$ is obtained by simulation, where for each $s = 1, \dots, m$, $\tilde{\mu}^{(s)}(\theta)$ is the auxiliary estimate calculated using simulated data $\{y_i^{(s)}(\theta)\}$ generated from F_θ , and $m \geq 1$ is fixed.
3. An alternative choice obtained by simulation is $\mu(\theta) = \arg \min_{\mu} m^{-1} \sum_{s=1}^m \mathcal{L}[\mu; \{y_i^{(s)}(\theta)\}]$, where $\mathcal{L}$ is the objective function associated with the auxiliary estimator, i.e. $\partial \mathcal{L} / \partial \mu[\tilde{\mu}; \{y_i\}] = 0$.

Assume that the parameter θ is partitioned into $(\theta_1^T, \theta_2^T)^T$, $\theta_1 \in \mathfrak{R}^{p_1}$ and $\theta_2 \in \mathfrak{R}^{p-p_1}$ and consider the null hypothesis $H_0 : \theta_1 = \theta_{10}$. In the indirect inference setting, likelihood ratio-type tests are based on the optimal value of the objective function in (2), i.e.

$$W(\hat{\theta}, \hat{\theta}^c, \Omega^*) = \{ \tilde{\mu} - \mu(\hat{\theta}^c) \}^T \Omega^* \{ \tilde{\mu} - \mu(\hat{\theta}^c) \} - \{ \tilde{\mu} - \mu(\hat{\theta}) \}^T \Omega^* \{ \tilde{\mu} - \mu(\hat{\theta}) \}, \quad (3)$$

where Ω^* is the inverse of the asymptotic variance-covariance matrix of the auxiliary estimator under the true data-generating model, and $\hat{\theta}^c$ is the constrained indirect inference estimator under H_0 . In contrast with standard likelihood ratio tests, the factor 2 does not appear in front. Under the null hypothesis $nW(\hat{\theta}, \hat{\theta}^c, \Omega^*)$ is asymptotically $\chi_{p_1}^2$ -distributed.

The classical test for the validity of the overidentifying conditions ($r > p$) is given by the J test:

$$J(\hat{\theta}, \Omega^*) = \{ \tilde{\mu} - \mu(\hat{\theta}) \}^T \Omega^* \{ \tilde{\mu} - \mu(\hat{\theta}) \},$$

where $nJ(\hat{\theta})$ is asymptotically χ_{r-p}^2 -distributed. For the simulation-based indirect inference, the W and J statistics are premultiplied by $m/(m+1)$. In general, the optimal weighting matrix Ω^* is unknown and needs to be estimated by a consistent estimator, for instance $\tilde{\Omega}^* = M_\mu I^{-1} M_\mu$, with $M_\mu = n^{-1} \sum_{i=1}^n \partial \Psi / \partial \mu^T(y_i; \tilde{\mu})$ and

$I = n^{-1} \sum_{i=1}^n \{\Psi(y_i; \tilde{\mu}) \Psi^T(y_i; \tilde{\mu})\}$. The χ^2 -approximations to the distributions of the W and J tests have an absolute error of order $n^{-1/2}$ and their finite sample accuracy can be very poor and lead to misleading inference; see e.g. Michaelides and Ng (2000) and Duffee and Stanton (2008). To improve this, we introduce saddlepoint tests based on indirect inference estimators.

In the just-identified case ($r = p$) the saddlepoint test statistic is defined by

$$S(\hat{\theta}_1) = \inf_{\theta_2} \sup_{\lambda} \left\{ -\log E_{F_{(\theta_1, \theta_2)}} \left(\exp[\lambda^T \Psi\{y_i; \mu(\hat{\theta}_1, \theta_2)\}] \right) \right\}, \quad (4)$$

where $\hat{\theta} = (\hat{\theta}_1^T, \hat{\theta}_2^T)^T$ is the indirect inference estimator given by (2). Then, under H_0 , $2nS(\hat{\theta}_1)$ is asymptotically $\chi_{p_1}^2$ -distributed with relative error of order n^{-1} . This corresponds to the test introduced by Robinson, Ronchetti, Young (2003) for M -estimators and has the additional advantage that it does not require the knowledge of Ω^* .

A nonparametric version of the test is also considered when the expectation in (4) does not have a closed form. Moreover, an important extension to the overidentified case ($r > p$) is provided, which plays a crucial role in econometrics, where it is well-known that the accuracy of the χ^2 -approximation of overidentification tests can be very poor.

We show by simulation in various models, including nonlinear regression and Poisson regression with overdispersion that saddlepoint tests are very accurate down to small sample sizes. They outperform in accuracy the classical tests with both nonsimulation-based and simulation-based indirect inference estimators.

Finally, we introduce robust versions of the saddlepoint tests. They provide accurate inference in the presence of local deviations from the stochastic assumptions of the model and we illustrate their performance in diffusion models. In some cases they outperform nonrobust saddlepoint and classical tests even when there are no deviations from the assumed model.

Our paper is related to several earlier contributions, in particular work by Robinson, Ronchetti, Young (2003), Lô and Ronchetti (2009), Kitamura and Stutzer (1997), and Sowell (1996, 2009).

References

1. Czellar, V., Ronchetti, E.: Accurate and robust tests for indirect inference. *Biometrika*, to appear (2010)
2. Duffee, D., Singleton, K. J.: Simulated moments estimation of Markov models of asset prices. *Econometrica* **61**, 929–952 (1993)
3. Duffee, G. R., Stanton, R.: Evidence on simulation inference for near unit-root processes with implications for term structure estimation. *Journal of Financial Econometrics* **6**, 108–142 (2008)
4. Gouriéroux, C., Monfort, A.: Simulation-based econometric methods. Oxford University Press (1996)
5. Gouriéroux, C., Monfort, A., Renault, E.: Indirect inference. *Journal of Applied Econometrics* **8**, S85–S118 (1993)

6. Hansen, L. P.: Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029–1054 (1982)
7. Heggland, K., Frigessi, A. : Estimating functions in indirect inference. *Journal of the Royal Statistical Society, Series B* **66**, 447–462 (2004)
8. Jiang, W., Turnbull, B. : The indirect method: inference based on intermediate statistics — A synthesis and examples. *Statistical Science* **19**, 239–263 (2004)
9. Kitamura, Y., Stutzer, M. : An information-theoretic alternative to generalized method of moments estimation. *Econometrica* **65**, 861–874 (1997)
10. Lee, B., Ingram, B. : Simulation estimation of time-series models. *Journal of Econometrics* **47**, 197–205 (1991)
11. Lô, S., Ronchetti, E. : Robust and accurate inference for generalized linear models. *Journal of Multivariate Analysis* **100**, 2126–2136 (2009)
12. Michaelides, A., Ng, S. : Estimating the rational expectations model of speculative storage: A Monte Carlo comparison of three simulation estimators. *Journal of Econometrics* **96**, 231–266 (2000)
13. Robinson, J., Ronchetti, E., Young, G.A. : Saddlepoint approximations and tests based on multivariate M-estimates. *The Annals of Statistics* **31**, 1154–1169 (2003)
14. Smith, A. A. : Estimating nonlinear time series models using simulated vector autoregressions. *Journal of Applied Econometrics* **8**, S63–S84 (1993)
15. Sowell, F. B. : Optimal tests for parameter instability in the generalized method of moments framework. *Econometrica* **64**, 1085–1107 (1996)
16. Sowell, F. B. : The empirical saddlepoint likelihood estimator applied to two-step GMM. Technical Report, Tepper School of Business, Carnegie Mellon University (2009)