UNIVERSITÀ DI SIENA 1240

QUADERNI DEL DIPARTIMENTO DI ECONOMIA POLITICA E STATISTICA

Federico Crudu
Felipe Osorio

Bilinear form test statistics for extremum estimation

n. 804 – Giugno 2019

Bilinear form test statistics for extremum estimation

Federico Crudu[a,1,*], Felipe Osorio[b,1]

[a] *Department of Economics and Statistics, University of Siena, Italy*
[b] *Departamento de Matemática, Universidad Técnica Federico Santa María, Chile*

Abstract

This paper develops a set of test statistics based on bilinear forms in the context of the extremum estimation framework. We show that the proposed statistic converges to a conventional chi-square limit. A Monte Carlo experiment suggests that the test statistic works well in finite samples.

Keywords: Extremum estimation, Gradient statistic, Bilinear form test, Nonlinear hypothesis.
JEL: C12, C14, C69.

1. Introduction

The purpose of this paper is to introduce a novel test statistic for extremum estimation (EE). In this very general setting (see for instance Gourieroux and Monfort, 1995; Hayashi, 2000), conventional test statistics are defined either in terms of differences (pseudo likelihood ratio or distance statistic) or in terms of quadratic forms (Wald, Lagrange multiplier). The test proposed in this paper is defined in terms of a bilinear form (BF). This approach is not entirely new as a bilinear form test for maximum likelihood was introduced by Terrell (2002) (see also the monograph by Lemonte, 2016). Our test statistic has a conventional chi-square limit and, similarly to the Wald test, it is generally not invariant to the definition of the null hypothesis. It is, though, easy to see that in the context of linear models the BF test is equal to the distance statistic, which is, on the other hand, invariant. Furthermore, when nonlinear models are involved our Monte Carlo simulations suggest that the discrepancy induced by equivalent definitions of the null hypothesis is relatively small when compared, e.g., to the Wald test. To the best of our knowledge this is the first paper that deals with this problem in the context of EE.

The remainder of the paper unfolds as follows. Section 2 contains the description of the test statistics for a generic, potentially nonlinear, null hypothesis and their asymptotic properties; the asymptotic results and the corresponding proofs are presented in a concise fashion and are mostly based on the results in Gourieroux and Monfort (1995). In Section 3 we study, via Monte Carlo experiments, the finite sample properties of the test in comparison with other more conventional EE test statistics. Section 4 offers some conclusions while the appendices contain the proofs of the asymptotic results.

2. A bilinear form test statistic

Let us consider a scalar objective function $Q_n(\boldsymbol{\beta})$ that depends on a set of data $\boldsymbol{w}_i, i = 1, \dots, n$ with $\boldsymbol{w}_i \in \mathbb{R}^k$ and $\boldsymbol{\beta} \in \mathcal{B} \subset \mathbb{R}^p$ where $\mathcal{B}$ is compact. The EE for our objective function can be defined as

$$\widehat{\boldsymbol{\beta}}_n = \underset{\beta \in \mathcal{B}}{\operatorname{argmax}}\, Q_n(\boldsymbol{\beta}). \tag{1}$$

Let us now suppose that we want to test the following null hypothesis

$$H_0 : \boldsymbol{g}(\boldsymbol{\beta}_0) = \boldsymbol{0} \tag{2}$$

given that $\boldsymbol{g} : \mathbb{R}^p \to \mathbb{R}^q$ is a continuously differentiable function and $\boldsymbol{G}(\boldsymbol{\beta}) = \partial \boldsymbol{g}(\boldsymbol{\beta})/\partial \boldsymbol{\beta}^\top$ is a $q \times p$ matrix with $\mathrm{rk}(\boldsymbol{G}(\boldsymbol{\beta})) = q$. The resulting constrained estimator is defined as the solution of the Lagrangian problem

$$L_n(\boldsymbol{\beta}, \boldsymbol{\lambda}) = Q_n(\boldsymbol{\beta}) - \boldsymbol{g}^\top(\boldsymbol{\beta})\boldsymbol{\lambda}, \tag{3}$$

where $\boldsymbol{\lambda}$ denotes a vector of Lagrange multipliers. Hence,

$$\widetilde{\boldsymbol{\beta}}_n = \underset{\beta \in \{\mathcal{B}: g(\beta)=0\}}{\operatorname{argmax}}\, L_n(\boldsymbol{\beta}, \boldsymbol{\lambda}). \tag{4}$$

The null hypothesis in Equation (2) can be tested, for example, by means of the simple Wald (W) test, that only requires the unconstrained estimator or either the Lagrange multiplier (LM) test or the distance metric (D) statistic that both require the constrained estimator in Equation (4). The BF tests that we propose are generalizations of

☆This research was supported by Fondecyt grant 11140433, Regione Autonoma della Sardegna Master and Back grant PRR-MAB-A2011-24192 (F. Crudu) and Fondecyt grant 1140580 (F. Osorio).

*Correspondence to: Department of Economics and Statistics, University of Siena, Piazza San Francesco, 7/8 53100 Siena, Italy.

Email addresses: `federico.crudu@unisi.it` (Federico Crudu), `felipe.osorios@usm.cl` (Felipe Osorio)

Terrell's gradient statistic (Terrell, 2002) to the EE context.[1] Let us first define $\boldsymbol{A}_n(\boldsymbol{\beta}_0) := \partial^2 Q_n(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top$ and assume that $\boldsymbol{A}_n(\boldsymbol{\beta}_0) \overset{\text{a.s.}}{\to} \boldsymbol{A}$ uniformly. Let us also assume that

$$\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}} \overset{\mathsf{D}}{\to} \mathsf{N}_p(\mathbf{0}, \boldsymbol{B}).$$

Furthermore, let $\boldsymbol{G} := \boldsymbol{G}(\boldsymbol{\beta}_0)$, $\boldsymbol{S} = \boldsymbol{G}\{-\boldsymbol{A}\}^{-1}\boldsymbol{G}^\top$ and $\boldsymbol{\Omega} = \boldsymbol{G}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{G}^\top$. Then,

$$BF_1 := n\widetilde{\boldsymbol{\lambda}}_n^\top \boldsymbol{S}\boldsymbol{\Omega}^{-1}\boldsymbol{g}(\widehat{\boldsymbol{\beta}}_n) \tag{5}$$

where $\widetilde{\boldsymbol{\lambda}}_n$ is the solution for $\boldsymbol{\lambda}$ in the Lagrangian problem defined by Equation 3. The BF statistic also has the following alternative formulations. Let $\boldsymbol{G}^+ = \boldsymbol{G}^\top\{\boldsymbol{G}\boldsymbol{G}^\top\}^{-1}$ denote the Moore-Penrose inverse of $\boldsymbol{G}$ (see, for instance, Magnus and Neudecker, 2007, p. 38). Then,

$$BF_2 := n\frac{\partial Q_n(\widetilde{\boldsymbol{\beta}}_n)}{\partial\boldsymbol{\beta}^\top}\boldsymbol{G}^+\boldsymbol{S}\boldsymbol{\Omega}^{-1}\boldsymbol{g}(\widehat{\boldsymbol{\beta}}_n) \tag{6}$$

$$BF_3 := n\frac{\partial Q_n(\widetilde{\boldsymbol{\beta}}_n)}{\partial\boldsymbol{\beta}^\top}\boldsymbol{G}^+\boldsymbol{S}\boldsymbol{\Omega}^{-1}\boldsymbol{G}(\widehat{\boldsymbol{\beta}}_n - \widetilde{\boldsymbol{\beta}}_n) \tag{7}$$

Let us define $\boldsymbol{P}_G := \boldsymbol{G}^+\boldsymbol{G}$ and assume that $\boldsymbol{B} = -\boldsymbol{A}$, which leads to $\boldsymbol{S} = \boldsymbol{\Omega}$. We then obtain the following specifications:

$$BF_4 := n\widetilde{\boldsymbol{\lambda}}_n^\top \boldsymbol{g}(\widehat{\boldsymbol{\beta}}_n) \tag{8}$$

$$BF_5 := n\frac{\partial Q_n(\widetilde{\boldsymbol{\beta}}_n)}{\partial\boldsymbol{\beta}^\top}\boldsymbol{G}^+\boldsymbol{g}(\widehat{\boldsymbol{\beta}}_n) \tag{9}$$

$$BF_6 := n\frac{\partial Q_n(\widetilde{\boldsymbol{\beta}}_n)}{\partial\boldsymbol{\beta}^\top}\boldsymbol{P}_G(\widehat{\boldsymbol{\beta}}_n - \widetilde{\boldsymbol{\beta}}_n) \tag{10}$$

$$BF_7 := n\frac{\partial Q_n(\widetilde{\boldsymbol{\beta}}_n)}{\partial\boldsymbol{\beta}^\top}(\widehat{\boldsymbol{\beta}}_n - \widetilde{\boldsymbol{\beta}}_n). \tag{11}$$

The assumption that $\boldsymbol{B} = -\boldsymbol{A}$ is not very restrictive as it may include as special cases maximum likelihood and GMM statistics (see Hayashi, 2000, Chapter 7). The following theorem shows that the BF tests are asymptotically equivalent and have a conventional chi-square limit.

Proposition 1. *Under the assumptions of Property 24.16 and Property 24.10 in Gourieroux and Monfort (1995), with $\boldsymbol{g}: \mathbb{R}^p \to \mathbb{R}^q$ being a continuously differentiable function and $\boldsymbol{G}(\boldsymbol{\beta}) = \partial\boldsymbol{g}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^\top$ a $q\times p$ matrix with $\operatorname{rk}(\boldsymbol{G}(\boldsymbol{\beta})) = q$,*

$$BF_k \overset{\mathsf{D}}{\to} \chi^2_q, \qquad k = 1, 2, 3.$$

If, in addition, $\boldsymbol{B} = -\boldsymbol{A}$ holds, then

$$BF_k \overset{\mathsf{D}}{\to} \chi^2_q, \qquad k = 4, 5, 6, 7.$$

Proof. See Appendix A.

Remark 1. When $Q_n(\boldsymbol{\beta}) = \bar{\ell}_n(\boldsymbol{\beta})$ is the log-likelihood function we obtain that the BF statistic is given by

$$BF = \boldsymbol{U}_n^\top(\widetilde{\boldsymbol{\beta}}_n)\boldsymbol{G}^+\boldsymbol{g}(\widehat{\boldsymbol{\beta}}_n), \tag{12}$$

where $\boldsymbol{U}_n(\boldsymbol{\beta}) = \partial\bar{\ell}_n(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$ denotes the score function. We must highlight that (12) is an extension of the test proposed by Terrell (2002) to tackle nonlinear hypotheses.

Remark 2. It is interesting to see that in the case of the linear model, D and BF are equal. Let us consider, the example in Hansen (2006). The BF statistic is

$$BF = (\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}_n)^\top\boldsymbol{X}\boldsymbol{B}^{-1}\boldsymbol{X}^\top\boldsymbol{X}(\widehat{\boldsymbol{\beta}}_n - \widetilde{\boldsymbol{\beta}}_n).$$

Since $\widehat{\boldsymbol{\beta}}_n = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$ and $\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_n) = \mathbf{0}$, it follows immediately that $BF = D$.

Proposition 2. *The BF test statistic in Equation (5) and the Lagrange multiplier test statistic*

$$LM := n\widetilde{\boldsymbol{\lambda}}_n^\top\boldsymbol{S}\boldsymbol{\Omega}^{-1}\boldsymbol{S}\widetilde{\boldsymbol{\lambda}}_n,$$

are asymptotically equivalent under $H_0 : \boldsymbol{g}(\boldsymbol{\beta}_0) = \mathbf{0}$. Their common asymptotic distribution is χ^2_q.

Proof. See Appendix B.

3. Monte Carlo simulations

To study the finite sample properties of the BF statistic we consider two equivalent nonlinear null hypotheses, as in Gregory and Veall (1985) (see also Hansen, 2006; Lafontaine and White, 1986). The BF test, which is not invariant to the specification of the null, is compared against the W, LM and D statistics. While the first two tests are known to be not invariant, the last test is invariant and works well in finite samples (see also Hansen, 2006). The performance of the tests is measured in terms of how close the empirical size is to the 5% nominal size and in terms of the discrepancy between the empirical sizes produced by competing equivalent hypotheses.

3.1. Setup

We consider the model specification

$$\boldsymbol{y} = \mathbf{1}_n\beta_1 + \boldsymbol{x}_2\beta_2 + \exp(\boldsymbol{x}_3\beta_3) + \boldsymbol{\varepsilon},$$

where $\mathbf{1}_n$ is a n-vector of ones, $\boldsymbol{x}_j \sim \mathsf{N}_n(\mathbf{0}, 0.16\,\boldsymbol{I})$, $j = 2, 3$ and $\boldsymbol{\varepsilon} \sim \mathsf{N}_n(\mathbf{0}, 0.16\,\boldsymbol{I})$. Moreover, we consider the following combinations of parameters

$$(\beta_1, \beta_2, \beta_3) \in \{(1, 10, 0.1), (1, 5, 0.2), (1, 2, 0.5), (1, 1, 1)\},$$

and sample sizes $n \in \{20, 50, 100, 500\}$. We test two equivalent null hypotheses

$$H_0^A : \beta_2 - \frac{1}{\beta_3} = 0, \tag{13}$$

and

$$H_0^B : \beta_2\beta_3 - 1 = 0. \tag{14}$$

The number of Monte Carlo replications is set to 5000.

[1]Sometimes the term *gradient statistic* is used to indicate the LM test for GMM (see for example Chapter 22 in Ruud, 2000). To avoid confusion we prefer the expression *bilinear form test* and the corresponding abbreviation BF.

Table 1: Empirical size for a 5% test. The superscripts A and B refer to the fact that W, LM and BF are computed using the null hypotheses in Equations (13) and (14), respectively.

(β_2, β_3)	n	W^A	W^B	BF^A	BF^B	LM^A	LM^B	D
(10,0.1)	20	0.420	0.176	0.067	0.064	0.087	0.084	0.087
	50	0.282	0.106	0.065	0.059	0.074	0.068	0.074
	100	0.197	0.077	0.059	0.059	0.061	0.061	0.061
	500	0.104	0.052	0.048	0.048	0.049	0.049	0.049
(5,0.2)	20	0.277	0.178	0.068	0.065	0.086	0.083	0.086
	50	0.171	0.108	0.058	0.057	0.068	0.067	0.068
	100	0.127	0.078	0.058	0.058	0.062	0.061	0.062
	500	0.070	0.052	0.047	0.047	0.048	0.048	0.048
(2,0.5)	20	0.145	0.175	0.066	0.067	0.082	0.082	0.082
	50	0.096	0.113	0.062	0.057	0.075	0.070	0.075
	100	0.078	0.082	0.056	0.056	0.062	0.062	0.062
	500	0.049	0.055	0.045	0.045	0.050	0.050	0.050
(1,1)	20	0.140	0.170	0.084	0.070	0.101	0.086	0.101
	50	0.095	0.108	0.062	0.062	0.070	0.070	0.070
	100	0.074	0.080	0.066	0.066	0.067	0.067	0.067
	500	0.055	0.055	0.056	0.056	0.061	0.061	0.061

3.2. Comments on the simulations

The results in Table 1 suggest that the BF test works well in finite samples even when the sample size is as small as $n = 20$. In most of the considered cases the BF test outperforms the distance statistic D as well as the LM test. Finally, it is worth noticing that, unlike W, the BF test is not very sensitive to the specification of the null hypothesis.

4. Concluding remarks

In this paper we introduced a set of bilinear form tests for EE that may be considered as a generalization of Terrell's gradient statistics (Terrell, 2002). The asymptotic distribution of the proposed tests is chi-square with degrees of freedom equal to the number of restrictions. A Monte Carlo experiment shows that the BF test works well in finite samples and that it generally outperforms its competitors. Furthermore, while the BF test is not generally invariant to the specification of the null, its finite sample performance seems to be only marginally affected by such a property.

Appendix A. Proof of Proposition 1

Following Property 24.16 in Gourieroux and Monfort (1995), we know that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{\mathsf{D}} \mathsf{N}_p(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}). \tag{A.1}$$

Then, by the delta method, we find that under $H_0 : \boldsymbol{g}(\boldsymbol{\beta}_0) = \mathbf{0}$

$$\sqrt{n}\,\boldsymbol{g}(\widehat{\boldsymbol{\beta}}_n) \xrightarrow{\mathsf{D}} \mathsf{N}_q(\mathbf{0}, \boldsymbol{\Omega}). \tag{A.2}$$

From Property 24.10 in Gourieroux and Monfort (1995), we have that $\widetilde{\boldsymbol{\beta}}_n$ and $\widetilde{\boldsymbol{\lambda}}_n$ are the solutions of the first order conditions of the Lagrangian problem in Equation (3):

$$\frac{\partial Q_n(\widetilde{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}} - \boldsymbol{G}^\top(\widetilde{\boldsymbol{\beta}}_n)\widetilde{\boldsymbol{\lambda}}_n = \mathbf{0} \tag{A.3}$$

$$\boldsymbol{g}(\widetilde{\boldsymbol{\beta}}_n) = \mathbf{0} \tag{A.4}$$

and $\widetilde{\boldsymbol{\beta}}_n$ is consistent. A Taylor expansion argument applied to $\partial Q_n(\widehat{\boldsymbol{\beta}}_n)/\partial \boldsymbol{\beta}$ and $\partial Q_n(\widetilde{\boldsymbol{\beta}}_n)/\partial \boldsymbol{\beta}$ around $\boldsymbol{\beta}_0$, $\boldsymbol{A}_n(\boldsymbol{\beta}_0) \xrightarrow{\mathsf{a.s.}} \boldsymbol{A}$ uniformly and simple calculations yield

$$\sqrt{n}\,\boldsymbol{g}(\widehat{\boldsymbol{\beta}}_n) = \boldsymbol{G}\{-\boldsymbol{A}\}^{-1}\sqrt{n}\,\frac{\partial Q_n(\widetilde{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}} + o_{\mathsf{a.s.}}(1). \tag{A.5}$$

From the first order condition (A.3),

$$\sqrt{n}\,\frac{\partial Q_n(\widetilde{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}} = \boldsymbol{G}^\top(\widetilde{\boldsymbol{\beta}}_n)\sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n, \tag{A.6}$$

we obtain that

$$\sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n = [\boldsymbol{G}\{-\boldsymbol{A}\}^{-1}\boldsymbol{G}^\top]^{-1}\sqrt{n}\,\boldsymbol{g}(\widehat{\boldsymbol{\beta}}_n) + o_{\mathsf{a.s.}}(1).$$

Then, using (A.2), we find

$$\sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n \xrightarrow{\mathsf{D}} \mathsf{N}_q(\mathbf{0}, \boldsymbol{S}^{-1}\boldsymbol{\Omega}\boldsymbol{S}^{-1}). \tag{A.7}$$

Let $\boldsymbol{\Omega} = \boldsymbol{R}\boldsymbol{R}^\top$ where $\boldsymbol{R}$ is a nonsingular $q \times q$ matrix. Then, using standardized versions of (A.2) and (A.7), it follows that

$$\begin{aligned} BF_1 &= \{\boldsymbol{R}^{-1}\boldsymbol{S}\sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n\}^\top \boldsymbol{R}^{-1}\sqrt{n}\,\boldsymbol{g}(\widehat{\boldsymbol{\beta}}_n) \\ &= n\widetilde{\boldsymbol{\lambda}}_n^\top \boldsymbol{S}\boldsymbol{\Omega}^{-1}\boldsymbol{g}(\widehat{\boldsymbol{\beta}}_n) \xrightarrow{\mathsf{D}} \chi^2_q. \end{aligned}$$

The proof for BF_2 and BF_3 follows from the equivalences

$$\sqrt{n}\,\boldsymbol{g}(\widehat{\boldsymbol{\beta}}_n) = \boldsymbol{G}\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \widetilde{\boldsymbol{\beta}}_n) + o_{\mathsf{a.s.}}(1)$$

and

$$\sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n = \sqrt{n}\{\boldsymbol{G}^+\}^\top \frac{\partial Q_n(\widetilde{\boldsymbol{\beta}}_n)}{\partial \boldsymbol{\beta}}.$$

The proof for BF_4, BF_5 and BF_6 follows by additionally assuming $\boldsymbol{B} = -\boldsymbol{A}$. Finally, the proof for BF_7 uses the fact that $\boldsymbol{P}_G \boldsymbol{\Omega} \boldsymbol{P}_G = \boldsymbol{\Omega}$.

Appendix B. Proof of Proposition 2

From (A.5) and (A.6), we have that

$$\begin{aligned}\sqrt{n}\,\boldsymbol{g}(\widehat{\boldsymbol{\beta}}_n) &= \boldsymbol{G}\{-\boldsymbol{A}\}^{-1}\boldsymbol{G}^\top \sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n + o_{\mathsf{a.s.}}(1)\\ &= \boldsymbol{S}\,\sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n + o_{\mathsf{a.s.}}(1),\end{aligned}$$

and this implies that,

$$\begin{aligned}BF &= \sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n^\top \boldsymbol{S}\boldsymbol{\Omega}^{-1}\sqrt{n}\,\boldsymbol{g}(\widehat{\boldsymbol{\beta}}_n)\\ &= \sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n^\top \boldsymbol{S}\boldsymbol{\Omega}^{-1}\boldsymbol{S}\,\sqrt{n}\,\widetilde{\boldsymbol{\lambda}}_n + o_{\mathsf{a.s.}}(1).\end{aligned}$$

By using the asymptotic distribution given in Equation (A.7), the proposition is verified.

References

Gourieroux, C., Monfort, A., 1995. Statistics and Econometrics Models: Testing, Confidence Regions, Model Selection and Asymptotic Theory. Vol. 2. Cambridge University Press.

Gregory, A.W., Veall, M.R., 1985. Formulating Wald tests of nonlinear restrictions. Econometrica 53, 1465-1468.

Hansen, B.E., 2006. Edgeworth expansions for the Wald and GMM statistics for nonlinear restrictions. In: Corbae, D., Durlauf, S.N., Hansen, B.E. (Eds.), Econometric Theory and Practice: Frontiers of Analysis and Applied Research. Cambridge, 9-35.

Hayashi, F., 2000. Econometrics. Princeton University Press.

Lafontaine, F., White, K.J., 1986. Obtaining any Wald statistic you want. Economics Letters 21, 35-40.

Lemonte, A., 2016. The Gradient Test: Another Likelihood-based Test. Academic Press, Amsterdam.

Magnus, J.R., Neudecker, H., 2007. Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley, New York.

Ruud, P.A., 2000. An Introduction to Classical Econometric Theory. Oxford University Press, New York.

Terrell, G.R., 2002. The gradient statistic. Computing Science and Statistics 34, 206-215.