Invariant tests based on M-estimators, estimating functions, and the generalized method of moments*

Jean-Marie Dufour † McGill University

Alain Trognon ‡ CREST-ENSAE, Paris

Purevdorj Tuvaandorj § CREST-ENSAI, Rennes

First version: May 1997
Revised: April 2001, September 2008, February 2013, February 2015, July 2015
This version: May 2016
Compiled: June 22, 2016, 5:55

A shorter version of this paper is forthcoming in *Econometric Reviews*. The present version includes a few minor editorial corrections.

*The authors thank Marine Carrasco, Jean-Pierre Cotton, Russell Davidson, Abdeljelil Farhat, V. P. Godambe, Christian Gouriéroux, Stéphane Grégoir, Hervé Mignon, Denis Pelletier, Mohamed Taamouti, Pascale Valéry, three anonymous referees, and the Editor Esfandiar Maasoumi for several useful comments. Earlier versions of this paper were presented at the annual meeting Statistical Society of Canada and at INSEE (CREST, Paris). This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

† William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 4400 ext. 09156; FAX: (1) 514 398 4800; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

‡ Member, CREST-ENSAE (Centre de recherche en économie et statistique) and Associate Professor, University Paris 1. Mailing address: CREST PARIS, Timbre J310, 15 Boulevard Gabriel Péri, 92254 MALAKOFF CEDEX, France. TEL.: 33 1 41175131; FAX: 33 1 41176480. e-mail: trognon@ensae.fr . Web page: www.crest.fr/component/

§ CREST-ENSAI - Campus de Ker-Lann, Rue Blaise Pascal - BP 37203, 35712 BRUZ cedex, France. TEL.: 33 (0)2 99 05 32 63; FAX: 33 (0)2 99 05 32 05. e-mail: purevdorj.tuvaandorj@ensai.fr .

ABSTRACT

We study the invariance properties of various test criteria which have been proposed for hypothesis testing in the context of incompletely specified models, such as models which are formulated in terms of estimating functions (Godambe, 1960, *Ann. Math. Stat.*) or moment conditions and are estimated by generalized method of moments (GMM) procedures (Hansen, 1982, *Econometrica*), and models estimated by pseudo-likelihood (Gouriéroux, Monfort and Trognon, 1984, *Econometrica*) and M-estimation methods. The invariance properties considered include invariance to (possibly nonlinear) hypothesis reformulations and reparameterizations. The test statistics examined include Wald-type, LR-type, LM-type, score-type, and $C(\alpha)-$type criteria. Extending the approach used in Dagenais and Dufour (1991, *Econometrica*), we show first that all these test statistics except the Wald-type ones are invariant to equivalent hypothesis reformulations (under usual regularity conditions), but all five of them are *not generally invariant* to model reparameterizations, including measurement unit changes in nonlinear models. In other words, testing two equivalent hypotheses in the context of equivalent models may lead to completely different inferences. For example, this may occur after an apparently innocuous rescaling of some model variables. Then, in view of avoiding such undesirable properties, we study restrictions that can be imposed on the objective functions used for pseudo-likelihood (or M-estimation) as well as the structure of the test criteria used with estimating functions and GMM procedures to obtain invariant tests. In particular, we show that using linear exponential pseudo-likelihood functions allows one to obtain invariant score-type and $C(\alpha)-$type test criteria, while in the context of estimating function (or GMM) procedures it is possible to modify a LR-type statistic proposed by Newey and West (1987, *Int. Econ. Rev.*) to obtain a test statistic that is invariant to general reparameterizations. The invariance associated with linear exponential pseudo-likelihood functions is interpreted as a strong argument for using such pseudo-likelihood functions in empirical work.

Key words: Testing; Invariance; Hypothesis reformulation; Reparameterization; Measurement unit; Estimating function; Generalized method of moment (GMM); Pseudo-likelihood; M-estimator; Linear exponential model; Nonlinear model; Wald test; Likelihood ratio test; score test; Lagrange multiplier test; $C(\alpha)$ test

Journal of Economic Literature classification: C3; C12

Contents

List of Tables

List of Figures

List of propositions and theorems

1. Introduction

Model and hypothesis formulation in econometrics and statistics typically involve a number of arbitrary choices, such as the labelling of i.i.d. observations or the selection of measurement units. Further, in hypothesis testing, these choices often do not affect the interpretation of the null and the alternative hypotheses. When this is the case, it appears desirable that statistical inference remain *invariant* to such choices; see Hotelling (1936), Pitman (1939), Lehmann (1983, Chapter 3), Lehmann (1986, Chapter 6) and Ferguson (1967). Among other things, when the way a null hypothesis is written has no particular interest or when the parameterization of a model is largely arbitrary, it is natural to require that the results of test procedures do not depend on such choices. This holds, for example, for standard t and F tests in linear regressions under *linear* hypothesis reformulations and reparameterizations. In *nonlinear* models, however, the situation is more complex.

It is well known that Wald-type tests are not invariant to equivalent hypothesis reformulations and reparameterizations; see Cox and Hinkley (1974, p. 302), Burguete, Gallant and Souza (1982, p. 185), Gregory and Veall (1985), Vaeth (1985), Lafontaine and White (1986), Breusch and Schmidt (1988), Phillips and Park (1988), and Dagenais and Dufour (1991). For general possibly nonlinear likelihood models (which are treated as correctly specified), we showed in previous work [Dagenais and Dufour (1991, 1992), Dufour and Dagenais (1992)] that very few test procedures are invariant to general hypothesis reformulations and reparameterizations. The invariant procedures essentially reduce to likelihood ratio (LR) tests and certain variants of score [or Lagrange multiplier (LM)] tests where the information matrix is estimated with either an exact formula for the (expected) information matrix or an outer product form evaluated at the restricted maximum likelihood (ML) estimator. In particular, score tests are not invariant to reparameterizations when the information matrix is estimated using the Hessian matrix of the log-likelihood function evaluated at the restricted ML estimator. Further, $C(\alpha)$ tests are not generally invariant to reparameterizations unless special equivariance properties are imposed on the restricted estimators used to implement them. Among other things, this means that measurement unit changes with no incidence on the null hypothesis tested may induce dramatic changes in the conclusions obtained from the tests and suggests that invariant test procedures should play a privileged role in statistical inference.

In this paper, we study the invariance properties of various test criteria which have been proposed for hypothesis testing in the context of incompletely specified models, such as models which are formulated in terms of estimating functions [Godambe (1960)] – or moment conditions – and are estimated by generalized method of moments (GMM) procedures [Hansen (1982)], and models estimated by M-estimation [Huber (1981)] or pseudo-likelihood methods [Gouriéroux, Monfort and Trognon (1984*b*, 1984*c*), Gouriéroux and Monfort (1993)]. For general discussions of inference in such models, the reader may consult White (1982), Newey (1985), Gallant (1987), Newey and West (1987), Gallant and White (1988), Gouriéroux and Monfort (1989, 1995), Godambe (1991), Davidson and MacKinnon (1993), Newey and McFadden (1994), Hall (1999) and Mátyás (1999); for studies of the performance of some test procedures based on GMM estimators, see also Burnside and Eichenbaum (1996) and Podivinsky (1999).

The invariance properties we consider include invariance to (possibly nonlinear) hypothesis reformulations and reparameterizations. The test statistics examined include Wald-type, LR-type, LM-type, score-type, and $C(\alpha)$-type criteria. Extending the approach used in Dagenais and Dufour (1991) and Dufour and Dagenais (1992) for likelihood models, we show first that all these test statistics except the Wald-type ones are invariant to equivalent hypothesis reformulations (under usual regularity conditions), but all five of them are *not generally invariant* to model reparameterizations, including measurement unit changes in nonlinear models. In other words, testing two equivalent hypotheses in the context of equivalent models may lead to completely different inferences. For example, this may occur after an apparently innocuous rescaling of some model variables.

In view of avoiding such undesirable properties, we study restrictions that can be imposed on the objective functions used for pseudo-likelihood (or M-estimation) as well as the structure of the test criteria used with estimating functions and GMM procedures to obtain invariant tests. In particular, we show that using linear exponential pseudo-likelihood functions allows one to obtain invariant score-type and $C(\alpha)$-type test criteria, while in the context of estimating function (or GMM) procedures it is possible to modify a LR-type statistic proposed by Newey and West (1987) to obtain a test statistic that is invariant to general reparameterizations. The invariance associated with linear exponential pseudo-likelihood functions can be viewed as a strong argument for using such pseudo-likelihood functions in empirical work. Of course, the fact that Wald-type tests are not invariant to both hypothesis reformulations and reparameterizations is by itself a strong argument to avoid using this type of procedure (when they are not equivalent to other procedures) and suggest as well that Wald-type tests can be quite unreliable in finite samples; for further arguments going in the same direction, see Burnside and Eichenbaum (1996), Dufour (1997), and Dufour and Jasiak (2001).

In Section 2, we describe the general setup considered, while the test statistics studied are defined in Section 3. The invariance properties of the available test statistics are studied in Section 4. In Section 5, we make suggestions for obtaining tests that are invariant to general hypothesis reformulations and reparameterizations. Numerical illustrations of the invariance (and noninvariance) properties discussed are provided in Section 6. We conclude in Section 8. The regularity conditions under which distributional results for test statistics are obtained appear in Appendix.

2. Framework

We consider an inference problem about a parameter of interest $\theta \in \Theta \subseteq \mathbb{R}^p$. This parameter appears in a model which is not fully specified. In order to identify θ, we assume there exist a $m \times 1$ vector score-type function $D_n(\theta; Z_n)$ where $Z_n = [z_1, z_2, \ldots, z_n]'$ is a $n \times k$ stochastic matrix such that

$$D_n(\theta; Z_n) \xrightarrow[n\to\infty]{\mathrm{p}} D_\infty(\theta; \theta_0)\ . \tag{2.1}$$

$D_\infty(\cdot; \theta_0)$ is a mapping from Θ onto $\mathbb{R}^m$ such that:

$$D_\infty(\theta; \theta_0) = 0 \Longleftrightarrow \theta = \theta_0 \tag{2.2}$$

so the value of θ is uniquely determined by $D_\infty(\theta; \theta_0)$. Furthermore, we assume:

$$\sqrt{n} D_n(\theta_0; Z_n) \xrightarrow[n\to\infty]{L} N[0, I(\theta_0)] \tag{2.3}$$

$$H_n(\theta_0; Z_n) = \frac{\partial D_n(\theta_0; Z_n)}{\partial \theta'} \xrightarrow[n\to\infty]{\mathrm{p}} J(\theta_0) \tag{2.4}$$

where $I(\theta_0)$ and $J(\theta_0)$ are $m \times m$ and $m \times p$ full-column rank matrices.

Typically, such a model is estimated by minimizing with respect to θ an expression of the form

$$M_n(\theta) = D_n(\theta; Z_n)' W_n D_n(\theta; Z_n) \tag{2.5}$$

where W_n is a symmetric positive definite matrix. The method of estimating functions [Durbin (1960), Godambe (1960, 1991), Basawa, Godambe and Taylor (1997)], the generalized method of moments [Hansen (1982), Hall (2004)], maximum likelihood, pseudo-maximum likelihood, M-estimation and instrumental variable methods may all be cast in this setup. Under general regularity conditions, the estimator $\hat{\theta}_n$ so obtained has a normal asymptotic distribution:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n\to\infty]{L} N[0, \Sigma(W_0)] \tag{2.6}$$

where

$$\Sigma(W_0) = \left(J_0'W_0J_0\right)^{-1} J_0'W_0I_0W_0J_0 \left(J_0'W_0J_0\right)^{-1}, \tag{2.7}$$

$J_0 = J(\theta_0)$, $I_0 = I(\theta_0)$, $W_0 = \operatorname*{plim}_{n\to\infty} W_n$, $\det(W_0) \neq 0$; see Gouriéroux and Monfort (1995, Ch. 9). Note also that "asymptotic estimation efficiency" arguments suggest one to use $W_n = I_n^{-1}$ as weighting matrix, where I_n is consistent estimator of I_0.[1]

If we assume that the number of equations is equal to the number of parameters $(m = p)$, a general method for estimating θ also consists in finding an estimator $\hat{\theta}_n$ which satisfies the equation

$$D_n(\hat{\theta}_n; Z_n) = 0\,. \tag{2.8}$$

Typically, in such cases, $D_n(\theta; Z_n)$ is the derivative of an objective function $S_n(\theta; Z_n)$, which is maximized (or minimized) to obtain $\hat{\theta}_n$, so that

$$D_n(\theta; Z_n) = \frac{\partial S_n(\theta; Z_n)}{\partial \theta}\,,\; H_n(\theta; Z_n) = \frac{\partial^2 S_n(\theta; Z_n)}{\partial \theta \partial \theta'}\,. \tag{2.9}$$

In this case, $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically normal with zero mean and asymptotic variance

$$\Sigma_D(\theta_0) = \left[J(\theta_0)' I(\theta_0)^{-1} J(\theta_0)\right]^{-1} = \left(J_0' I_0^{-1} J_0\right)^{-1}. \tag{2.10}$$

Obviously, condition (2.8) is entailed by the minimization of $M_n(\theta)$ when $m = p$. It is also interesting to note that problems with $m > p$ can be reduced to cases with $m = p$ through an appropriate redefinition of the score-type function $D_n(\theta; Z_n)$, so that the characterization (2.8) also covers most

[1] This "optimal" choice may be infeasible (or far from "efficient") in finite samples when I_0 (or I_n) is not invertible or "ill-conditioned" (close to non-invertibility). For this reason, we consider here the general formulation in (2.5), though the weighting matrix I_n^{-1} is allowed as a special case. Note also that "efficiency" from the estimation viewpoint is not in general equivalent to efficiency from the testing viewpoint (in terms of power), so it is not clear $W_n = I_n^{-1}$ is an optimal choice for the purpose of hypothesis testing.

classical asymptotic estimation methods. A typical list of methods is the following.

a) *Maximum likelihood.* In this case, the model is fully specified with log-likelihood function $L_n(\theta; Z_n)$ and score function

$$D_n(\theta; Z_n) = \frac{1}{n}\frac{\partial L_n(\theta; Z_n)}{\partial \theta}\ . \tag{2.11}$$

b) *Generalized method of moments* (*GMM*). θ is identified through a $m \times 1$ vector of conditions of the form: $\mathsf{E}[h_t(\theta; z_t)] = 0\ ,\quad t = 1, \ldots, n$. Then one considers the sample analogue of this mean,

$$\bar{h}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} h_t(\theta; z_t)\ , \tag{2.12}$$

and the quadratic form

$$M_n(\theta) = \bar{h}_n(\theta)' W_n \bar{h}_n(\theta) \tag{2.13}$$

where W_n is a symmetric positive definite matrix. In this case, the score-type function is:

$$D_n(\theta; Z_n) = 2\frac{\partial \bar{h}_n(\theta)'}{\partial \theta} W_n \bar{h}_n(\theta)\ . \tag{2.14}$$

c) *M-estimator.* $\hat{\theta}_n$ is defined by minimizing (or maximizing) an objective function $\bar{M}_n$ of the form:

$$\bar{M}_n(\theta; Z_n) = \frac{1}{n}\sum_{t=1}^{n} \xi(\theta; z_t)\,. \tag{2.15}$$

The score function has the following form:

$$D_n(\theta; Z_n) = \frac{\partial \bar{M}_n(\theta; Z_n)}{\partial \theta} = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial \xi(\theta; z_t)}{\partial \theta}. \tag{2.16}$$

3. Test statistics

Consider now the problem of testing

$$H_0 : \psi(\theta) = 0 \tag{3.1}$$

where $\psi(\theta)$ is a $p_1 \times 1$ continuously differentiable function of θ, $1 \le p_1 \le p$ and the $p_1 \times p$ matrix

$$P(\theta) = \frac{\partial \psi}{\partial \theta'} \tag{3.2}$$

has full row rank (at least in an open neighborhood of θ_0). Let $\hat{\theta}_n$ be the unrestricted estimator obtained by minimizing $M_n(\theta)$, and $\hat{\theta}_n^0$ the corresponding constrained estimator under H_0.

At this stage, it is not necessary to specify closely the way the matrices $I(\theta_0)$ and $J(\theta_0)$ are estimated. We will denote by $\hat{I}_0$ and $\hat{J}_0$ or by $\hat{I}$ and $\hat{J}$ the corresponding estimated matrices depending on whether they are obtained with or without the restriction $\psi(\theta) = 0$. In particular, if

$$D_n(\theta; Z_n) = \frac{1}{n} \sum_{t=1}^{n} h_t(\theta; z_t), \tag{3.3}$$

standard definitions of $\hat{I}(\theta)$ and $\hat{J}(\theta)$ would be

$$\hat{I}(\theta) = \frac{1}{n} \sum_{t=1}^{n} h_t(\theta; z_t) h_t(\theta; z_t)', \quad \hat{J}(\theta) = \frac{\partial D_n(\theta)}{\partial \theta'} = H_n(\theta; Z_n), \tag{3.4}$$

where θ can be replaced by an appropriate estimator. For M-estimators, we have $h_t(\theta; z_t) = \dfrac{\partial \xi(\theta; z_t)}{\partial \theta}$ the derivative of the (pseudo-)likelihood associated with an individual observation.

For $\hat{I}(\theta)$, other estimators are also widely used. Here, we shall consider general estimators of the form

$$\hat{I}(\theta) = \sum_{s=1}^{n} \sum_{t=1}^{n} w_{st}(n) h_s(\theta; z_s) h_t(\theta; z_t)' = h(\theta; Z_n) W_I(n) h(\theta; Z_n)' \tag{3.5}$$

where $W_I(n) = [w_{st}(n)]$ is a $n \times n$ matrix of weights (which depend of the sample size n and, possibly, on the data) and

$$h(\theta; Z_n) = [h_1(\theta; z_1), h_2(\theta; z_2), \ldots, h_n(\theta; z_n)]. \tag{3.6}$$

For example, a "mean corrected" version of $\hat{I}(\theta)$ may be obtained on taking $W_I(n) = \frac{1}{n}(I_n - \frac{1}{n}\iota_n \iota_n')$, where I_n is the identity matrix of order n and $\iota_n = (1, 1, \ldots, 1)'$, which yields

$$\hat{I}(\theta) = \frac{1}{n} \sum_{t=1}^{n} [h_t(\theta; z_t) - \overline{h}(\theta)][h_t(\theta; z_t) - \overline{h}(\theta)]' \tag{3.7}$$

where $\overline{h}(\theta) = \frac{1}{n} \sum_{t=1}^{n} h_t(\theta; z_t)$. Similarly, so-called "heteroskedasticity-autocorrelation consistent (HAC)" covariance matrix estimators can usually be rewritten in the form (3.5). In most cases, such estimators are defined by a formula of the type:

$$\hat{I}(\theta) = \sum_{j=-n+1}^{n-1} \bar{k}(j/B_n) \hat{\Gamma}(j, \theta) \tag{3.8}$$

where $\bar{k}(\cdot)$ is a kernel function, B_n is a bandwidth parameter (which depends on the sample size and, possibly, on the data), and

$$\hat{\Gamma}(j, \theta) = \begin{cases} \frac{1}{n} \sum_{t=j+1}^{n} h_t(\theta; z_t) h_{t-j}(\theta; z_{t-j})', & \text{if } j \geq 0, \\ \frac{1}{n} \sum_{t=-j+1}^{n} h_{t+j}(\theta; z_{t+j}) h_t(\theta; z_t)', & \text{if } j < 0. \end{cases} \tag{3.9}$$

For further discussion of such estimators, the reader may consult Newey and West (1987), Andrews (1991), Andrews and Monahan (1992), Hansen (1992), and Cushing and McGarvey (1999).

In this context, analogues of the Wald, LM, score and $C(\alpha)$ test statistics can be shown to have asymptotic null distributions without nuisance parameters, namely $\chi^2(p_1)$ distributions. On assuming that the referenced inverse matrices do exist, these test criteria can be defined as follows: (a) the *Wald-type statistic,*

$$W(\psi) = n\,\psi(\hat{\theta}_n)'\big[\hat{P}\big(\hat{J}'\hat{I}^{-1}\hat{J}\big)^{-1}\hat{P}'\big]^{-1}\psi(\hat{\theta}_n) \tag{3.10}$$

where $\hat{P} = P(\hat{\theta}_n)$, $\hat{I} = \hat{I}(\hat{\theta}_n)$ and $\hat{J} = \hat{J}(\hat{\theta}_n)$;
(b) the *score-type statistic,*

$$S(\psi) = n\,D_n(\hat{\theta}_n^0; Z_n)'\hat{I}_0^{-1}\hat{J}_0\big(\hat{J}_0'\hat{I}_0^{-1}\hat{J}_0\big)^{-1}\hat{J}_0'\hat{I}_0^{-1}D_n(\hat{\theta}_n^0; Z_n) \tag{3.11}$$

where $\hat{I}_0 = \hat{I}(\hat{\theta}_n^0)$ and $\hat{J}_0 = \hat{J}(\hat{\theta}_n^0)$;
(c) the *Lagrange-multiplier-type (LM-type) statistic,*

$$LM(\psi) = n\,\hat{\lambda}_n'\hat{P}_0\big(\hat{J}_0'\hat{I}_0^{-1}\hat{J}_0\big)^{-1}\hat{P}_0'\hat{\lambda}_n \tag{3.12}$$

where $\hat{P}_0 = P(\hat{\theta}_n^0)$ and $\hat{\lambda}_n$ is the Lagrange multiplier in the corresponding constrained optimization problem;
(d) the $C(\alpha)$*-type statistic,*

$$PC(\tilde{\theta}_n^0; \psi) = n\,D_n\big(\tilde{\theta}_n^0; Z_n\big)'\tilde{W}_0\,D_n\big(\tilde{\theta}_n^0; Z_n\big) \tag{3.13}$$

where $\tilde{\theta}_n^0$ is any root-n consistent estimator of θ that satisfies $\psi(\tilde{\theta}_n^0) = 0$, and

$$\tilde{W}_0 \equiv \tilde{I}_0^{-1}\tilde{J}_0\left(\tilde{J}_0'\tilde{I}_0^{-1}\tilde{J}_0\right)^{-1}\tilde{P}_0'\left[\tilde{P}_0\big(\tilde{J}_0'\tilde{I}_0^{-1}\tilde{J}_0\big)^{-1}\tilde{P}_0'\right]^{-1}\tilde{P}_0\left(\tilde{J}_0'\tilde{I}_0^{-1}\tilde{J}_0\right)^{-1}\tilde{J}_0'\tilde{I}_0^{-1}$$

with $\tilde{P}_0 = P(\tilde{\theta}_n^0)$, $\tilde{I}_0 = \hat{I}(\tilde{\theta}_n^0)$ and $\tilde{J}_0 = \hat{J}(\tilde{\theta}_n^0)$.

The above Wald-type and score-type statistics were discussed by Newey and West (1987) in the context of GMM estimation, and for pseudo-maximum likelihood estimation by Trognon (1984). The $C(\alpha)$-type statistic is given by Davidson and MacKinnon (1993, p. 619). Of course, LR-type statistics based on the difference of the maxima of the objective function $S_n(\theta; Z_n)$ have also been considered in such contexts:

$$LR(\psi) = S_n\big(\hat{\theta}_n; Z_n\big) - S_n\big(\hat{\theta}_n^0; Z_n\big)\,. \tag{3.14}$$

It is well known that, in general, this difference is distributed as a mixture of independent chi-square with coefficients depending upon nuisance parameters [see, for example, Trognon (1984) and Vuong (1989)]. Nevertheless, there is one “LR-type” test statistic whose distribution is asymptotically pivotal with a chi-square distribution, namely the D statistic suggested by Newey and West (1987):

$$D_{NW}(\psi) = n\left[M_n\left(\hat{\theta}_n^0; \tilde{I}_0\right) - M_n\left(\hat{\theta}_n; \tilde{I}_0\right)\right] \tag{3.15}$$

where

$$M_n\left(\theta; \tilde{I}_0\right) = D_n(\theta; Z_n)' \tilde{I}_0^{-1} D_n(\theta; Z_n) , \tag{3.16}$$

$\tilde{I}_0$ is a consistent estimator of $I(\theta_0)$, $\hat{\theta}_n$ minimizes $M_n(\theta; \tilde{I}_0)$ without restriction and $\hat{\theta}_n^0$ minimizes $M_n(\theta; \tilde{I}_0)$ under the restriction $\psi(\theta) = 0$. Note, however, that this "LR-type" statistic is more accurately viewed as a score-type statistic: if D_n is the derivative of some other objective function (*e.g.*, a log-likelihood function), the latter is not used as the objective function but replaced by a quadratic function of the "score" D_n.

Using the constrained minimization condition,

$$H_n\left(\hat{\theta}_n^0; Z_n\right)' \tilde{I}_0^{-1} D_n\left(\hat{\theta}_n^0; Z_n\right) = P\left(\hat{\theta}_n^0\right)' \hat{\lambda}_n , \tag{3.17}$$

we see that

$$S(\psi) = LM(\psi) , \tag{3.18}$$

i.e., the score and LM statistics are identical in the present circumstances. Further, it is interesting to observe that the score, LM and $C(\alpha)$-type statistics given above may all be viewed as special cases of a more general $C(\alpha)$-type statistic obtained by considering the generalized "score-type" function:

$$s\left(\tilde{\theta}_n^0, W_n\right) = \sqrt{n}\, \tilde{Q}[W_n]\, D_n\left(\tilde{\theta}_n^0; Z_n\right) \tag{3.19}$$

where $\tilde{\theta}_n^0$ is consistent restricted estimate of θ_0 such that $\psi(\tilde{\theta}_n^0) = 0$ and $\sqrt{n}(\tilde{\theta}_n^0 - \theta_0)$ is asymptotically bounded in probability,

$$\tilde{Q}[W_n] \equiv \tilde{P}_0 (\tilde{J}_0' W_n \tilde{J}_0)^{-1} \tilde{J}_0' W_n , \tag{3.20}$$

$\tilde{P}_0 = P\left(\tilde{\theta}_n^0\right)$, $\tilde{J}_0 = \hat{J}\left(\tilde{\theta}_n^0\right)$, and W_n is a symmetric positive definite (possibly random) $m \times m$ matrix such that

$$\operatorname*{plim}_{n\to\infty} W_n = W_0 , \quad \det(W_0) \neq 0. \tag{3.21}$$

Under standard regularity conditions [see Appendix A], we have:

$$s\left(\tilde{\theta}_n^0; Z_n\right) \xrightarrow[n\to\infty]{L} N\left[0, Q(\theta_0)\, I(\theta_0)\, Q(\theta_0)'\right] \tag{3.22}$$

where

$$Q(\theta_0) = \operatorname*{plim}_{n\to\infty} \tilde{Q}[W_n] = P(\theta_0)\left[J(\theta_0)' W_0 J(\theta_0)\right]^{-1} J(\theta_0)' W_0 \tag{3.23}$$

and $\operatorname{rank}[Q(\theta_0)] = p_1$. This suggests the following generalized $C(\alpha)$ criterion:

$$PC\left(\tilde{\theta}_n^0; \psi, W_n\right) = n D_n\left(\tilde{\theta}_n^0; Z_n\right)' \tilde{Q}[W_n]' \left\{\tilde{Q}[W_n]\, \tilde{I}_0 \tilde{Q}[W_n]'\right\}^{-1} \tilde{Q}[W_n]\, D_n\left(\tilde{\theta}_n^0; Z_n\right) \tag{3.24}$$

where $\tilde{I}_0 = \hat{I}(\tilde{\theta}_n^0)$. Under general regularity conditions, the asymptotic distribution of $PC\left(\tilde{\theta}_n^0; \psi, W_n\right)$

is $\chi^2(p_1)$ under H_0.[2] It is clear that $PC(\tilde{\theta}_n^0;\, \psi, W_n)$ includes as special cases various other $C(\alpha)$-type statistics proposed in the statistical and econometric literatures.[3] On taking $W_n = \tilde{I}_0^{-1}$, as suggested by efficiency arguments, $PC(\tilde{\theta}_n^0;\, \psi, W_n)$ reduces to $PC(\tilde{\theta}_n^0;\, \psi)$ in (3.13). When the number of equations equals the number of parameters $(m = p)$, we have $\tilde{Q}[W_n] = \tilde{P}_0 \tilde{J}_0^{-1}$ and $PC(\tilde{\theta}_n^0;\, \psi, W_n)$ does not depend on the choice of W_n:

$$PC(\tilde{\theta}_n^0;\, \psi, W_n) = PC\big(\tilde{\theta}_n^0;\, \psi\big) = D_n\big(\tilde{\theta}_n^0;\, Z_n\big)'(\tilde{J}_0^{-1})'\tilde{P}_0'\big[\tilde{P}_0\big(\tilde{J}_0'\tilde{I}_0^{-1}\tilde{J}_0\big)^{-1}\tilde{P}_0'\big]^{-1}\tilde{P}_0\tilde{J}_0^{-1}D_n\big(\tilde{\theta}_n^0;\, Z_n\big)\,.$$

In particular, this will be the case if $D_n(\theta;\, Z_n)$ is the derivative vector of a (pseudo) log-likelihood function. Finally, for $m \geq p$, when $\tilde{\theta}_n^0$ is obtained by minimizing $M_n(\theta) = D_n(\theta;\, Z_n)'\,\tilde{I}_0^{-1}D_n(\theta;\, Z_n)$ subject to $\psi(\theta) = 0$, we can write $\tilde{\theta}_n^0 \equiv \hat{\theta}_n^0$ and $PC\big(\tilde{\theta}_n^0;\, \psi, W_n\big)$ is identical to the score (or LM)-type statistic suggested by Newey and West (1987). Since the statistic $PC\big(\tilde{\theta}_n^0;\, \psi, W_n\big)$ is quite comprehensive, it will be convenient for establishing general invariance results.

4. Invariance

Following Dagenais and Dufour (1991), we will consider two types of invariance properties: (1) invariance with respect to the formulation of the null hypothesis, and (2) invariance with respect to reparameterizations.

4.1. Hypothesis reformulation

Let

$$\Theta_0 = \{\theta \in \Theta \mid \psi(\theta) = 0\} \tag{4.1}$$

and $\boldsymbol{\Psi}$ be the set of differentiable functions $\bar{\psi} : \Theta \to \mathbb{R}^m$ such that

$$\{\theta \in \Theta \mid \bar{\psi}(\theta) = 0\} = \Theta_0\,. \tag{4.2}$$

A test statistic is invariant with respect to $\boldsymbol{\Psi}$ if it is the same for all $\psi \in \boldsymbol{\Psi}$. It is obvious the LR-type statistics $LR(\psi)$ and $D_{NW}(\psi)$ (when applicable) are invariant to such hypothesis reformulations because the optimal values of the objective function (restricted or unrestricted) do not depend on the way the restrictions are written. Now, a reformulation does not affect $\hat{I}, \hat{J}, \hat{I}_0$ and $\hat{J}_0$. The same holds for $\tilde{I}_0$ and $\tilde{J}_0$ provided the restricted estimator $\tilde{\theta}_n^0$ used with $C(\alpha)$ tests does not depend on which function $\psi \in \boldsymbol{\Psi}$ is used to obtain it. However, $\hat{P}$, $\hat{\lambda}_n$ and $\psi\big(\hat{\theta}_n\big)$ change. Following Dagenais

[2] A rigorous proof of the latter assertion appears in Dufour, Trognon and Tuvaandorj (2015). See also earlier versions of this paper [Dufour, Trognon and Tuvaandorj (2013*a*)].

[3] For further discussion of $C(\alpha)$ tests, the reader may consult Basawa (1985), Ronchetti (1987), Smith (1987), Berger and Wallenstein (1989), Dagenais and Dufour (1991), Davidson and MacKinnon (1991, 1993) and Kocherlakota and Kocherlakota (1991)

and Dufour (1991), if $\bar{\psi} \in \Psi$, we have:

$$\bar{P}(\theta) = \frac{\partial \bar{\psi}}{\partial \theta'} = \bar{P}_1(\theta)\, G(\theta)\ , \quad P(\theta) = \frac{\partial \psi}{\partial \theta'} = P_1(\theta)\, G(\theta)\ , \tag{4.3}$$

where $\bar{P}_1$ and P_1 are two $p_1 \times p_1$ invertible functions and $G(\theta)$ is a $p_1 \times p$ full row-rank matrix. Since $\bar{P}_1^{0\prime}\, \bar{\lambda}_n = \hat{P}_1^{0\prime}\, \hat{\lambda}_n$ where $\bar{P}_1^0 = \bar{P}_1(\hat{\theta}_n^0), \hat{P}_1^0 = P_1(\hat{\theta}_n^0)$ and $\bar{\lambda}_n$ is the Lagrange multiplier associated with $\bar{\psi}$, we deduce that all the statistics, except the Wald-type statistics, are invariant with respect to a reformulation. This leads to the following proposition.

Proposition 4.1 INVARIANCE TO HYPOTHESIS REFORMULATIONS. *Let Ψ be a family of $p_1 \times 1$ continuously differentiable functions of θ such that $\frac{\partial \psi}{\partial \theta'}$ has full row rank when $\psi(\theta) = 0$ $(1 \le p_1 \le p)$, and*

$$\psi(\theta) = 0 \Longleftrightarrow \bar{\psi}(\theta) = 0, \forall \psi,\ \bar{\psi} \in \Psi\,. \tag{4.4}$$

Then, $T(\psi) = T(\bar{\psi})$ where T stands for any one of the test statistics $S(\psi)$, $LM(\psi)$, $PC(\tilde{\theta}_n^0; \psi)$, $LR(\psi)$, $D_{NW}(\psi)$ and $PC(\tilde{\theta}_n^0; \psi, W_n)$ defined in (3.11) - (3.15) *and* (3.24).

Note that the invariance of the $S(\psi)$, $LM(\psi)$, $LR(\psi)$ and $D_{NW}(\psi)$ statistics to hypothesis reformulations has been pointed out by Gouriéroux and Monfort (1989) for mixed-form hypotheses.

4.2. Reparameterization

Let $\bar{g}$ be a one-to-one differentiable transformation from $\Theta \subseteq \mathbb{R}^p$ to $\Theta_* \subseteq \mathbb{R}^p : \theta_* = \bar{g}(\theta)$. $\bar{g}$ represents a reparameterization of the parameter vector θ to a new one θ_*. The latter is often determined by a one-to-one transformation of the data $Z_{n*} = g(Z_n)$, as occurs for example when variables are rescaled (measurement unit changes). But it may also represent a reparameterization without any variable transformation. Let $k = \bar{g}^{-1}$ be the inverse function associated with $\bar{g}$:

$$k(\theta_*) = \bar{g}^{-1}(\theta_*) = \theta\ . \tag{4.5}$$

Set

$$\bar{G}(\theta) = \frac{\partial \bar{g}'}{\partial \theta} \text{ and } K(\theta_*) = \frac{\partial k}{\partial \theta_*'}\ . \tag{4.6}$$

Since $k[\bar{g}(\theta)] = \theta$ and $\bar{g}[k(\theta_*)] = \theta_*$, we have by the chain rule of differentiation:

$$K[\bar{g}(\theta)]\, \bar{G}(\theta) = I_p \text{ and } \bar{G}[k(\theta_*)]\, K(\theta_*) = I_p\ , \forall \theta_* \in \Theta_*\ , \forall \theta \in \Theta\ . \tag{4.7}$$

Let

$$\psi^*(\theta_*) = \psi\left[\bar{g}^{-1}(\theta_*)\right]\ . \tag{4.8}$$

Clearly,

$$\psi^*(\theta_*) = 0 \Leftrightarrow \psi(\theta) = 0\,, \tag{4.9}$$

and $H_0^* : \psi^*(\theta_*) = 0$ is an equivalent reformulation of $H_0 : \psi(\theta) = 0$ in terms of θ_*. We shall call $\psi^*(\theta_*) = 0$ the *canonical reformulation* of $\psi(\theta) = 0$ in terms of θ_*. Other (possibly more "natural") reformulations are of course possible, but the latter has the convenient property that $\psi^*(\theta_*) = \psi(\theta)$. If a test statistic is invariant to reparameterizations when the null hypothesis is reformulated as $\psi^*(\theta_*) = 0$, we will say it is *canonically invariant*.

By the invariance property of Proposition **4.1**, it will be sufficient for our purpose to study invariance to reparameterizations for any given reformulation of the null hypothesis in terms of θ_*. From the above definition of $\psi^*(\theta_*)$, it follows that

$$P_*(\theta_*) \equiv \frac{\partial \psi^*}{\partial \theta_*'} = \frac{\partial \psi}{\partial \theta'} \frac{\partial \theta}{\partial \theta_*'} = P[k(\theta_*)]\, K(\theta_*) = P(\theta)\, K[\bar{g}(\theta)] \ . \tag{4.10}$$

We need to make an assumption on the way the score-type function $D_n(\theta; Z_n)$ changes under a given reparameterization. We will consider two cases. The first one consists in assuming that $D_n(\theta; Z_n) = \sum_{t=1}^{n} h_t(\theta; z_t)/n$ as in (3.3) where the values of the scores are unaffected by the reparameterization, but are simply reexpressed in terms of θ_* and z_{t*} (*invariant scores*):

$$h_t(\theta_*; z_{t*}) = h_t(\theta; z_t) \ , \quad t = 1, \ldots, n \, , \tag{4.11}$$

where $Z_{n*} = g(Z_n)$ and $\theta_* = \bar{g}(\theta)$. The second one is the one where $D_n(\theta; Z_n)$ can be interpreted as the derivative of an objective function.

Under condition (4.11), we see easily that

$$H_{n*}(\theta_*; Z_{n*}) = \frac{\partial D_{n*}(\theta_*; Z_{n*})}{\partial \theta_*'} = H_n(\theta; Z_n)\, K(\theta_*) = H_n(\theta; Z_n)\, K[\bar{g}(\theta)] \ . \tag{4.12}$$

Further the functions $\hat{I}(\theta)$ and $\hat{J}(\theta)$ in (3.4) are then transformed in the following way :

$$\hat{I}_*(\theta_*) = \hat{I}(\theta) \, , \; \hat{J}_*(\theta_*) = \hat{J}(\theta)\, K[\bar{g}(\theta)] \ . \tag{4.13}$$

If $\hat{I}(\theta)$ and $\hat{J}(\theta)$ are defined as in (3.4), $W_{n*} = W_n$ and $\tilde{\theta}_n^0$ is equivariant with respect to $\bar{g}$ [*i.e.*, $\tilde{\theta}_{n*}^0 = \bar{g}(\tilde{\theta}_n^0)$], it is easy to check that the generalized $C(\alpha)$ statistic defined in (3.24) is invariant to the reparameterization $\theta_* = \bar{g}(\theta)$. This suggests the following general sufficient condition for the invariance of $C(\alpha)$ statistics.

Proposition 4.2 $C(\alpha)$ CANONICAL INVARIANCE TO REPARAMETERIZATIONS: INVARIANT SCORE CASE. *Let* $\psi^*(\theta_*) = \psi\left[\bar{g}^{-1}(\theta_*)\right]$, *and suppose the following conditions hold :*

(a) $\tilde{\theta}_{n*}^0 = \bar{g}(\tilde{\theta}_n^0)$,

(b) $D_{n*}(\tilde{\theta}_{n*}^0; Z_{n*}) = D_n(\tilde{\theta}_n^0; Z_n)$,

(c) $\tilde{I}_{0*} = \tilde{I}_0$ *and* $\tilde{J}_{0*} = \tilde{J}_0 \tilde{K}$,

(d) $W_{n*} = W_n$,

where $\tilde{I}_0$, $\tilde{J}_0$ *and* W_n *are defined as in* (3.24), *and* $\tilde{K} = K(\tilde{\theta}^0_{n*})$ *is invertible. Then*

$$PC_*(\tilde{\theta}^0_{n*};\, \psi^*, W_{n*}) \equiv n\tilde{D}'_{n*}\tilde{Q}'_{0*}\big(\tilde{Q}'_{0*}\tilde{I}_{0*}\tilde{Q}_{0*}\big)^{-1}\tilde{Q}_{0*}\tilde{D}_{n*} = PC(\tilde{\theta}^0_n;\, \psi, W_n)$$

where $\tilde{D}_{n*} = D_{n*}(\tilde{\theta}^0_{n*}; Z_{n*})$, $\tilde{Q}_{0*} = \tilde{P}_{0*}\big(\tilde{J}'_{0*}W_{n*}\tilde{J}_{0*}\big)^{-1}\tilde{J}'_{0*}W_{n*}$, $\tilde{P}_{0*} = P_*(\tilde{\theta}^0_{n*})$ *and* $P_*(\theta_*) = \partial\psi^*/\partial\theta'_*$.

It is clear that the estimators $\hat{\theta}_n$ and $\hat{\theta}^0_n$ satisfy the equivariance condition, i.e., $\hat{\theta}_{n*} = \bar{g}(\hat{\theta}_n)$ and $\hat{\theta}^0_{n*} = \bar{g}(\hat{\theta}^0_n)$. Consequently, the above invariance result also applies to score (or LM) statistics. It is also interesting to observe that $W_*(\psi^*) = W(\psi)$. This holds, however, only for the special reformulation $\psi^*(\theta_*) = \psi\left[\bar{g}^{-1}(\theta_*)\right] = 0$, not for all equivalent reformulations $\psi_*(\theta_*) = 0$. On applying Proposition **4.1**, this type of invariance holds for the other test statistics. These observations are summarized in the following proposition.

Theorem 4.3 TEST INVARIANCE TO REPARAMETERIZATIONS AND GENERAL HYPOTHESIS REFORMULATIONS: INVARIANT SCORE CASE. *Let* $\psi_* : \Theta_* \to \Theta$ *be any continuously differentiable function of* $\theta_* \in \Theta_*$ *such that* $\psi_*(\bar{g}(\theta)) = 0 \Leftrightarrow \psi(\theta) = 0$, *let* $m = p$ *and suppose*

(*a*) $D_{n*}(\bar{g}(\theta); Z_{n*}) = D_n(\theta; Z_n)$,

(*b*) $\hat{I}_*[\bar{g}(\theta)] = \hat{I}(\theta)$ *and* $\hat{J}_*[\bar{g}(\theta)] = \hat{J}(\theta)K[\bar{g}(\theta)]$,

where $K(\theta_*) = \partial\bar{g}^{-1}(\theta_*)/\partial\theta'_*$. *Then, provided the relevant matrices are invertible, we have*

$$T(\psi) = T_*(\psi_*) \tag{4.14}$$

where T *stands for any one of the test statistics* $S(\psi)$, $LM(\psi)$, $LR(\psi)$ *and* $D_{NW}(\psi)$. *If* $\hat{\theta}^0_{n*} = \bar{g}(\hat{\theta}^0_n)$, *we also have*

$$PC_*(\tilde{\theta}^0_{n*};\, \psi_*) = PC(\tilde{\theta}^0_n;\, \psi)\,. \tag{4.15}$$

If $\psi_*(\theta) = \psi\left[\bar{g}^{-1}(\theta)\right]$, *the Wald statistic is invariant :* $W_*(\psi_*) = W(\psi)$.

Cases where (4.12) holds only have limited interest because they do not cover problems where D_n is the derivative of an objective function, as occurs for example when M-estimators or (pseudo) maximum likelihood methods are used :

$$D_n(\theta; Z_n) = \frac{\partial S_n(\theta; Z_n)}{\partial\theta}\,. \tag{4.16}$$

In such cases, one would typically have :

$$S_{n*}(\theta_*; Z_{n*}) = S_n(\theta; Z_n) + \kappa(Z_{n*})$$

where $\kappa(Z_{n*})$ may be a function of the Jacobian of the transformation $Z_{n*} = g(Z_n)$. To deal with such cases, we thus assume that $m = p$, and

$$D_{n*}(\theta_*; Z_{n*}) = K(\theta_*)'D_n(\theta; Z_n) = K[\bar{g}(\theta)]'D_n(\theta; Z_n)\,. \tag{4.17}$$

From (2.3) and (4.17), it then follows that

$$\sqrt{n}D_{n*}(\theta_{0*};Z_{n*}) \xrightarrow[n\to\infty]{L} N[0,I_*(\theta_{0*})] \tag{4.18}$$

where $\theta_{0*} = \bar{g}(\theta_0)$ and

$$I_*(\theta_*) = K(\theta_*)' I[k(\theta_*)] K(\theta_*) = K[\bar{g}(\theta)]' I(\theta) K[\bar{g}(\theta)] \,. \tag{4.19}$$

Further,

$$H_{n*}(\theta_*;Z_n) = K[\bar{g}(\theta)]' H_n(\theta;Z_n) K[\bar{g}(\theta)] + \sum_{i=1}^{p} D_{ni}(\theta;Z_n) K_{i\cdot}^{(1)}[\bar{g}(\theta)] \tag{4.20}$$

where $D_{ni}(\theta;Z_n)$, $i=1,\ldots,p$, are the coordinates of $D_n(\theta;Z_n)$ and

$$K_{i\cdot}^{(1)}(\theta_*) = \frac{\partial^2\theta_i}{\partial\theta_*\partial\theta_*'}(\theta_*) = \frac{\partial^2 k_i}{\partial\theta_*\partial\theta_*'}(\theta_*) \,. \tag{4.21}$$

By a set of arguments analogous to those used in Dagenais and Dufour (1991), it appears that all the statistics [except the LR-type statistic] are based upon H_n and so they are sensitive to a reparameterization, unless some specific estimator of J is used. At this level of generality, the following results can be presented using the following notations : $\hat{I},\hat{J},\hat{P}$ are the estimated matrices for a parameterization in θ and $\hat{I}_*,\hat{J}_*,\hat{P}_*$ are the estimated matrices for a parameterization in θ_*. The first proposition below provides an auxiliary result on the invariance of generalized $C(\alpha)$ statistics for the canonical reformulation $\psi^*(\theta_*)=0$, while the following one provides the invariance property for all the statistics considered and general equivalent reparameterizations and hypothesis reformulations.

Proposition 4.4 $C(\alpha)$ CANONICAL INVARIANCE TO REPARAMETERIZATIONS. *Let $\psi^*(\theta_*) = \psi\left[\bar{g}^{-1}(\theta_*)\right]$, and suppose the following conditions hold:*

(a) $\tilde{\theta}_{n*}^{0} = \bar{g}(\tilde{\theta}_{n}^{0})$,

(b) $D_{n*}(\tilde{\theta}_{n*}^{0};Z_{n*}) = K\left[\tilde{\theta}_{n*}^{0}\right]' D(\tilde{\theta}_{n}^{0};Z_n)$,

(c) $\tilde{I}_{0*} = \tilde{K}'\tilde{I}_0\tilde{K}$, $\tilde{J}_{0*} = \tilde{K}'\tilde{J}_0\tilde{K}$,

(d) $W_{n*} = \tilde{K}^{-1}W_n\left(\tilde{K}^{-1}\right)'$,

where $\tilde{I}_0$, $\tilde{J}_0$ and W_n are defined as in (3.24)*, and $\tilde{K} = K(\tilde{\theta}_{n*}^{0})$. Then, provided the relevant matrices are invertible,*

$$PC_*(\tilde{\theta}_{n*}^{0};\psi^*,W_{n*}) = PC(\tilde{\theta}_{n}^{0};\psi,W_n) \,.$$

Theorem 4.5 TEST INVARIANCE TO REPARAMETERIZATIONS AND GENERAL EQUIVALENT HYPOTHESIS REFORMULATIONS. *Let $\psi_*:\Theta_*\to\Theta$ be any continuously differentiable function of $\theta_*\in\Theta_*$ such that $\psi_*[\bar{g}(\theta)]=0 \Leftrightarrow \psi(\theta)=0$, let $m=p$ and suppose :*

(a) $D_{n*}(\bar{g}(\theta);Z_{n*}) = K[\bar{g}(\theta)]' D_n(\theta;Z_n)$,

(b) $\hat{I}_*[\bar{g}(\theta)] = K[\bar{g}(\theta)]' \hat{I}(\theta) K[\bar{g}(\theta)]$,

$(c) \quad \hat{J}_*[\bar{g}(\theta)] = K[\bar{g}(\theta)]' \hat{J}(\theta) K[\bar{g}(\theta)]$,
where $K(\theta_*) = \partial \bar{g}^{-1}(\theta) / \partial \theta_*'$ *. Then, provided the relevant matrices are invertible, we have*

$$T(\psi) = T_*(\psi_*) \tag{4.22}$$

where T *stands for any one of the test statistics* $S(\psi)$, $LM(\psi)$, $LR(\psi)$ *and* $D_{NW}(\psi)$ *. If* $\tilde{\theta}^0_{n*} = \bar{g}(\tilde{\theta}^0_n)$, *we also have*

$$PC_*(\tilde{\theta}^0_{n*}; \psi_*) = PC(\tilde{\theta}^0_n; \psi), \tag{4.23}$$

and, in the case where $\psi_*(\theta) = \psi[\bar{g}^{-1}(\theta)]$,

$$W_*(\psi_*) = W(\psi).$$

It is of interest to note here that condition (a) and (b) of the latter theorem will be satisfied if $D_n(\theta; Z_n) = \frac{1}{n}\sum_{t=1}^{n} h_t(\theta; z_t)$ and each individual "score" gets transformed after reparameterization according to the equation

$$h_{t*}(\bar{g}(\theta); z_{t*}) = K[\bar{g}(\theta)]' h_t(\theta; z_t), \; t = 1, , \ldots, n, \tag{4.24}$$

where $D_{n*}(\bar{g}(\theta); Z_{n*}) = \frac{1}{n}\sum_{t=1}^{n} h_{t*}(\bar{g}(\theta); z_{t*})$. Consequently, in such a case, any estimator $\hat{I}(\theta)$ of the general form (3.5) will satisfy (b) provided the matrix $W_I(n)$ remains invariant under reparameterizations. This will be the case, in particular, for most HAC estimators of the form (3.8) as soon as the bandwidth parameter B_n only depends on the sample size n. However, this may not hold if B_n is data-dependent [as considered in Andrews and Monahan (1992)].

Despite the apparent "positive nature" of the invariance results presented in this section, the main conclusion is that none of the proposed test statistics is invariant to general reparameterizations, especially when the score-type function is derived from an objective function. This is due, in particular, to the behaviour of moment (or estimating function) derivatives under nonlinear reparameterizations. As shown in Dagenais and Dufour (1991), this type of problem is already apparent in fully-specified likelihood models where LM statistics are not invariant to general reparameterizations when the covariance matrix is estimated through the Hessian of the log-likelihood function (*i.e.*, derivatives of the score function). When the true likelihood is not available, test statistics must be modified to control the asymptotic level of the test. Reparameterizations involve derivatives of score-type function (or pseudo-likelihood second derivatives), even in the case of LR-type statistics (see Theorem **4.5**). In other words, the adjustments required to deal with an incompletely specified model (no likelihood function) make invariance more difficult to achieve, and building valid invariant test procedures becomes a challenge.

5. Invariant test criteria

In this section, we propose two ways of building invariant test statistics. The first one is based on modifying the LR-type statistics proposed by Newey and West (1987) for GMM setups, while the second one exploits special properties of the linear exponential family in pseudo-maximum likelihood models.[4]

5.1. Modified Newey–West LR-type statistic

Consider the LR-type statistic

$$D_{NW}(\psi) = n\left[M_n\left(\hat{\theta}_n^0; \tilde{I}_0\right) - M_n\left(\hat{\theta}_n; \tilde{I}_0\right)\right]$$

where $M_n(\theta; \tilde{I}_0) = D_n(\theta; Z_n)' \tilde{I}_0^{-1} D_n(\theta; Z_n)$, proposed by Newey and West (1987, hereafter NW). In this statistic, $\tilde{I}_0$ is any consistent estimator of the covariance matrix $I(\theta_0)$ which is typically a function of a "preliminary" estimator $\bar{\theta}_n$ of θ : $\tilde{I}_0 = \hat{I}\left(\bar{\theta}_n\right)$. The minimized value of the objective function $M_n(\theta; \tilde{I}_0)$ is not invariant to general reparameterizations unless special restrictions are imposed on the covariance matrix estimator $\tilde{I}_0$.

However, there is a simple way of creating the appropriate invariance as soon as the function $\hat{I}(\theta)$ is a reasonably smooth function of θ. Instead of estimating θ by minimizing $M_n(\theta; \tilde{I}_0)$, estimate θ by minimizing $M_n\left(\theta; \hat{I}(\theta)\right)$. For example, such an estimation method was studied by Hansen, Heaton and Yaron (1996). When the score vector D_n and the parameter vector θ have the same dimension $(m = p)$, the unrestricted objective function will typically be zero $[D_n(\hat{\theta}_n; Z_n) = 0]$, so the statistic reduces to $D_{NW}(\psi) = nM_n(\hat{\theta}_n^0, \tilde{I}_0)$. When $m > p$, this will typically not be the case.

Suppose now the following conditions hold :

$$D_{n*}\left(\bar{g}(\theta), Z_{n*}\right) = K\left[\bar{g}(\theta)\right]' D_n(\theta; Z_n), \tag{5.1}$$

$$\hat{I}_*\left(\bar{g}(\theta)\right) = K\left[\bar{g}(\theta)\right]' \hat{I}(\theta) K\left[\bar{g}(\theta)\right]. \tag{5.2}$$

Then, for $\theta_* = \bar{g}(\theta)$,

$$\begin{aligned} M_{n*}\left(\theta_*; \hat{I}_*(\theta_*)\right) &\equiv D_{n*}\left(\bar{g}(\theta), Z_{n*}\right)' \hat{I}_*\left(\bar{g}(\theta)\right)^{-1} D_{n*}\left(\bar{g}(\theta), Z_{n*}\right) \\ &= D_n(\theta; Z_n)' \hat{I}(\theta)^{-1} D_n(\theta; Z_n). \end{aligned} \tag{5.3}$$

Consequently, the unrestricted minimal value $M_n(\hat{\theta}_n; \hat{I}(\hat{\theta}_n))$ and the restricted one $M_n(\hat{\theta}_n^0; I(\hat{\theta}_n^0))$ so obtained will remain unchanged under the new parameterization, and the corresponding J and the LR-type statistics, *i.e.*

$$J = nM_n\left(\hat{\theta}_n; \hat{I}(\hat{\theta}_n)\right), \tag{5.4}$$

$$\bar{D}(\psi) = n\left[M_n\left(\hat{\theta}_n^0; \hat{I}(\hat{\theta}_n^0)\right) - M_n\left(\hat{\theta}_n; \hat{I}(\hat{\theta}_n)\right)\right], \tag{5.5}$$

[4]The reader may note that further insight can be gained on the invariance properties of test statistics by using differential geometry arguments; for some applications to statistical problems, see Bates and Watts (1980), Amari (1990), Kass and Vos (1997), and Marriott and Salmon (2000). Such arguments may allow one to propose reparameterizations and "invariant Wald tests"; see, for example, Bates and Watts (1981), Hougaard (1982), Le Cam (1990), Critchley, Marriott and Salmon (1996), and Larsen and Jupp (2003) in likelihood models. As of now, such procedures tend to be quite difficult to design and implement, and GMM setups have not been considered. Even though this is an interesting avenue for future research, simplicity and generality considerations have led us to focus on procedures which do not require adopting a specific parameterization.

are invariant to reparameterizations of the type considered in (4.17) - (4.19). Under standard regularity conditions stated in Appendix B on the convergence of $D_n(\theta; Z_n)$ and $\hat{I}(\theta)$ as $n \to \infty$ (continuity, uniform convergence), it is easy to see that $\bar{D}$ and D_{NW} are asymptotically equivalent (at least under the null hypothesis) and so have the same asymptotic $\chi^2(p_1)$ distribution. The proof of this claim is available in Dufour, Trognon and Tuvaandorj (2013*b*).

5.2. Pseudo-maximum likelihood methods

5.2.1. PML methods

Consider the problem of making inference on the parameter which appears in the mean of an endogenous $G \times 1$ random vector y_t conditional on an exogenous random vector x_t:

$$\mathsf{E}(y_t \mid x_t) = f(x_t; \theta) \equiv f_t(\theta) \;,\; V(y_t \mid x_t) = \Omega_0(x_t) \tag{5.6}$$

where $f_t(\theta)$ is a known function and θ is the parameter of interest. (5.6) provides a non-linear generalized regression model with unspecified variance. Even if a likelihood function with a finite number of parameters is not available for such a semi-parametric model, θ can be estimated through a pseudo-maximum likelihood (PML) technique which consists in maximizing a chosen likelihood as if it were the true undefined likelihood; see Gouriéroux, Monfort and Trognon (1984*c*).[5] In particular, it is shown in the latter reference that this pseudo-likelihood must belong to the specific class of linear exponential distributions adapted for the mean. These distributions have the following general form:

$$l(y; \mu) = \exp\left[A(\mu) + B(y) + C(\mu)y\right] \tag{5.7}$$

where $\mu \in \mathbb{R}^G$ and $C(\mu)$ is a row vector of size G. The vector μ is the mean of y if

$$\frac{\partial A}{\partial \mu} + \frac{\partial C}{\partial \mu}\mu = 0\,.$$

Irrespective of the true data generating process, a consistent and asymptotically normal estimator of θ can be obtained by maximizing

$$\prod_{t=1}^{n} \exp\left\{A\big(f_t(\theta)\big) + B(y_t) + C\big(f_t(\theta)\big)y_t\right\} \tag{5.8}$$

or equivalently through the following equivalent programme:

$$\max_{\theta} \sum_{t=1}^{n} \left\{A\big(f_t(\theta)\big) + C\big(f_t(\theta)\big)y_t\right\} \text{ with } \frac{\partial A}{\partial \mu} + \frac{\partial C}{\partial \mu}\mu = 0\,. \tag{5.9}$$

The class of linear exponential distributions contains most of the classical statistical models, such

[5]For further discussion of such methods, the reader may consult: Gong and Samaniego (1981), Gouriéroux, Monfort and Trognon (1984*a*), Trognon (1984), Bourlange and Doz (1988), Trognon and Gouriéroux (1988), Gouriéroux and Monfort (1993), Crépon and Duguet (1997) and Jorgensen (1997).

.

as the Gaussian model, the Poisson model, the Binomial model, the Gamma model, the negative Binomial model, etc. The constraint in the programme (5.9) ensures that the expectation of the linear exponential pseudo-distribution is μ. The pseudo-likelihood equations have an orthogonal condition form:

$$D_n(\theta) = \sum_{t=1}^{n} \frac{\partial f_t'}{\partial \theta} \left[\frac{\partial C}{\partial \mu}\big(f_t(\theta)\big)\right] \big(y_t - f_t(\theta)\big) = 0\,. \tag{5.10}$$

The PML estimator solution of these first order conditions is consistent and asymptotically normal $N\big[0,\, (J'I^{-1}J)^{-1}\big]$, and we can write:

$$J(\theta) = \mathsf{E}_x\left\{\left(\frac{\partial f_t'}{\partial \theta}\right)\left[\frac{\partial C}{\partial \mu}\big(f_t(\theta)\big)\right]\left(\frac{\partial f_t'}{\partial \theta}\right)'\right\}\,, \tag{5.11}$$

$$I(\theta) = \mathsf{E}_x\left\{\left(\frac{\partial f_t'}{\partial \theta}\right)\left[\left(\frac{\partial C}{\partial \mu}\big(f_t(\theta)\big)\right)\Omega_0\left(\frac{\partial C}{\partial \mu}\big(f_t(\theta)\big)\right)'\right]\left(\frac{\partial f_t'}{\partial \theta}\right)'\right\}\,. \tag{5.12}$$

These matrices can be estimated by:

$$\hat{J} = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial f_t'}{\partial \theta}(\hat{\theta})\right)\left[\frac{\partial C}{\partial \mu}\big(f_t(\hat{\theta})\big)\right]\left(\frac{\partial f_t'}{\partial \theta}(\hat{\theta})\right)'\,, \tag{5.13}$$

$$\hat{I} = \frac{1}{n}\sum_{t=1}^{n} S_t\big(\hat{\theta}\big)\, S_t\big(\hat{\theta}\big)'\,, \tag{5.14}$$

where

$$S_t\big(\hat{\theta}\big) = \left(\frac{\partial f_t'}{\partial \theta}(\hat{\theta})\right)\left[\frac{\partial C}{\partial \mu}\big(f_t(\hat{\theta})\big)\right]\big(y - f_t(\hat{\theta})\big)\,. \tag{5.15}$$

Since $\frac{\partial C}{\partial \mu}\big(f_t(\hat{\theta})\big)$ and $y_t - f_t(\hat{\theta})$ are invariant to reparameterizations, $\hat{I}$ and $\hat{J}$ are modified only through $\dfrac{\partial f_t'}{\partial \theta}$. Further,

$$f_t^*(\theta_*) = f_t^*\left[\bar{g}(\theta)\right] = f_t(\theta)\,,\ \frac{\partial f_t^*}{\partial \theta_*'} = \left(\frac{\partial f_t^*}{\partial \theta'}\right)\left(\frac{\partial \theta}{\partial \theta_*'}\right) = \left(\frac{\partial f_t}{\partial \theta'}\right) K\left[\bar{g}(\theta)\right] \tag{5.16}$$

and

$$\hat{I}_* = K\big[\bar{g}(\hat{\theta})\big]'\hat{I}K\big[\bar{g}(\hat{\theta})\big]\,,\quad \hat{J}_* = K\big[\bar{g}(\hat{\theta})\big]'\hat{J}K\big[\bar{g}(\hat{\theta})\big]\,. \tag{5.17}$$

The Lagrange multiplier, score and $C(\alpha)$-type pseudo-asymptotic tests are then invariant to a reparameterization, though of course Wald tests will not be generally invariant to hypothesis reformulations. Consequently, this provides a strong argument for using pseudo true densities in the linear exponential family (instead of other types of densities) as a basis for estimating parameters of conditional means when the error distribution has unknown type.

The estimation of the J matrix could be obtained through direct second derivative calculus of the objective function. For example, when y_t is univariate $(G=1)$, we have:

$$\begin{aligned}\tilde{J} \;=\;& \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial f_t}{\partial\theta}(\hat{\theta})\right)\left[\frac{\partial C}{\partial\mu}(f_t(\hat{\theta}))\right]\left(\frac{\partial f_t}{\partial\theta}(\hat{\theta})\right)' - \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial f_t}{\partial\theta}(\hat{\theta})\right)\left[\frac{\partial^2 C}{\partial\mu^2}(f_t(\hat{\theta}))\right]\left(\frac{\partial f_t}{\partial\theta}(\hat{\theta})\right)'(y_t - f_t(\theta)) \\ & -\frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial^2 f_t}{\partial\theta\partial\theta'}(\hat{\theta})\right]\left(\frac{\partial C}{\partial\mu}(f_t(\hat{\theta}))\right)(y_t - f_t(\hat{\theta}))\end{aligned}$$

The first two terms of this estimator behave after reparameterization as $\hat{J}$, but the last term is based on second derivatives of $f_t(\theta)$ and so leads to non-invariance problems [see (3.4) and (4.20)]. The two last terms of J vanish asymptotically, they can be dropped as in the estimation method proposed by Gouriéroux et al. (1984*c*). For the invariance purpose, to discard the last term is the correct way to proceed.

5.2.2. Quasi generalized PML (QGPML) methods

Gouriéroux et al. (1984*c*) pointed out that some lower efficiency bound can be achieved by a two-step estimation procedure, when the functional form of the true conditional second order moment of y_t given x_t is known:

$$V(y_t|x_t) = \Omega_0(x_t) = g(x_t,\alpha_0) = g_t(\alpha_0)\,.$$

The method is based on various classical exponential families (negative-binomial, gamma, normal) which depend on an additional parameter η linked with the second order moment of the pseudo-distribution. If μ and Σ are the expectation and the variance-covariance matrix of this pseudo-distribution: $\eta = \Psi(\mu,\Sigma)$, where Ψ defines for any μ, a one to one relationship between η and Σ.

The class of linear exponential distributions depending upon the extra parameter η is of the following form:

$$l^*(y,\mu,\eta) = \exp\{A(\mu,\eta) + B(\eta,y) + C(\mu,\eta)y\}\,.$$

If we consider the negative binomial pseudo distribution $A(\mu,\eta) = -\eta\,\ln\left(1+\frac{\mu}{\eta}\right)$ and $C(\mu,\eta) = \ln\left(\mu/(\eta+\mu)\right)$; if otherwise we use the Gamma pseudo distribution: $A(\mu,\eta) = -\eta\ln(\mu)$ and $C(\mu,\eta) = -\frac{\eta}{\mu}$. In the former case: $\eta = \Psi(\mu,\sigma^2) = \mu\sigma^2/(1-\sigma^2)$ and in the latter $\eta = \Psi(\mu,\sigma^2) = \mu^2\sigma^2$.

With preliminary consistent estimators $\tilde{\alpha}$, $\tilde{\theta}$ of α, θ where $\tilde{\theta}$ and $\tilde{\alpha}$ are equivariant with respect to $\bar{g}$, computed for example as in Trognon (1984), the QGPML estimator of θ is obtained by solving a problem of the type

$$\max_{\theta}\sum_{t=1}^{n} l^*\Big(y_t,\, f_t(\theta), \Psi\big(f_t(\tilde{\theta}),\, g_t(\tilde{\alpha})\big)\Big)\,.$$

The QGPML estimator $\hat{\theta}$ of θ is strongly consistent and asymptotically normal: $\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{L}$

$N\left[0,\, \Sigma_Q\right]$ with

$$\Sigma_Q = \left\{ \mathsf{E}_x \left[\left(\frac{\partial f_t'}{\partial \theta} \right) g_t(\alpha_0)^{-1} \left(\frac{\partial f_t}{\partial \theta'} \right) \right] \right\}^{-1}, \quad I_0 = J_0 = \mathsf{E}_x \left[\left(\frac{\partial f_t'}{\partial \theta}(\theta_0) \right) g_t(\alpha_0)^{-1} \left(\frac{\partial f_t}{\partial \theta'}(\theta_0) \right) \right].$$

I_0 and J_0 can be consistently estimated by:

$$\hat{I} = \frac{1}{n} \sum_{t=1}^{n} S_t(\hat{\theta},\, \tilde{\alpha},\, \tilde{\theta})\, S_t(\hat{\theta},\, \tilde{\alpha},\, \tilde{\theta})', \quad \hat{J} = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{\partial f_t'}{\partial \theta}(\hat{\theta}) \right) \left[\frac{\partial C}{\partial \mu} \Big(f_t(\hat{\theta}),\, \Psi\big(f_t(\tilde{\theta}),\, g_t(\tilde{\alpha})\big) \Big) \right] \left(\frac{\partial f_t}{\partial \theta'}(\hat{\theta}) \right),$$

where

$$S_t(\hat{\theta}, \tilde{\alpha}, \tilde{\theta}) = \frac{\partial f_t'}{\partial \theta} \left[\frac{\partial C}{\partial \mu} \Big(f_t(\hat{\theta}),\, \Psi\big(f_t(\tilde{\theta}),\, g_t(\tilde{\alpha})\big) \Big) \right] (y_t - f_t(\hat{\theta})).$$

Since $\frac{\partial C}{\partial \mu} \Big(f_t(\hat{\theta}),\, \Psi\big(f_t(\tilde{\theta}),\, g_t(\tilde{\alpha})\big) \Big)$ and $y_t - f_t(\hat{\theta})$ are invariant to reparameterizations if $\tilde{\theta}$ and $\tilde{\alpha}$ are equivariant, we face the same favorable case as before:

$$\hat{I}_* = K[\bar{g}(\hat{\theta})]'\, \hat{I}\, K[\bar{g}(\hat{\theta})], \quad \hat{J}_* = K[\bar{g}(\hat{\theta})]'\, \hat{J}\, K[\bar{g}(\hat{\theta})],$$

and the Wald, Lagrange multiplier, score pseudo-asymptotic tests are invariant to a reparameterization. These quasi-generalized pseudo-asymptotic tests are locally more powerful than the corresponding pure pseudo-asymptotic tests under local alternatives [see Trognon (1984)].

Furthermore the quasi-generalized LR statistic (QGLR) is invariant provided, the first-step estimators $\tilde{\theta}$ and $\tilde{\alpha}$ are equivariant under reparameterization. And as shown in Trognon (1984), the QGLR statistic is asymptotically equivalent to the other pseudo-asymptotic statistic under the null and under local alternatives.

6. Numerical results

In order to illustrate numerically the (non-)invariance problems discussed above, we consider the model derived from the following equations:

$$y_t = \gamma + \beta_1 x_{1t}^{(\lambda)} + \beta_2 x_{2t}^{(\lambda)} + u_t\,, \quad u_t \overset{i.i.d.}{\sim} N[0,\, \sigma^2]\,,\; t = 1, \ldots, n, \tag{6.1}$$

where $x_{it}^{(\lambda)} = (x_{it}^{\lambda} - 1)/\lambda$, $i = 1, 2$, $x_{it} > 0$ with $x_{it}^{(\lambda)} = \log(x_{it})$ for $\lambda = 0$, and the explanatory variables x_{1t} and x_{2t} are fixed. The null hypothesis to be tested is:

$$H_0 : \lambda = 1\,. \tag{6.2}$$

The log-likelihood associated with this model is:

$$l = \sum_{t=1}^{n} l[y_t;\, \gamma,\, \beta_1, \beta_2,\, \lambda,\, \sigma^2]\,, \tag{6.3}$$

Table 1. Test statistics for $H_0 : \lambda = 1$ for different measurement units
5 moment models

	Two-step GMM				CUP-GMM				Pseudo ML	
k	$\bar{D}$	Wald	Score	$C(\alpha)$	$\bar{D}$	Wald	Score	$C(\alpha)$	LR	Mod. score
0.2	0.001	44.750	84.810	33.972	5.771	44.750	5.771	5.066	66.408	31.060
0.4	0.000	44.746	47.692	16.726	5.771	44.746	5.771	0.922	66.408	31.060
0.6	0.001	44.745	42.983	14.106	5.771	44.745	5.771	4.482	66.408	31.060
0.8	0.010	44.744	39.161	12.369	5.771	44.744	5.771	5.282	66.408	31.060
1.0	0.056	44.743	35.676	10.593	5.771	44.743	5.771	5.3838	66.408	31.060
3.0	34.629	44.743	118.876	42.124	5.771	44.743	5.771	0.6720	66.408	31.060
5.0	1.641	44.743	62.195	34.746	5.771	44.743	5.771	2.5545	66.408	31.060
7.0	0.282	44.742	61.766	34.953	5.771	44.742	5.771	3.9336	66.408	31.060
10.0	0.068	44.739	61.147	34.465	5.771	44.739	5.771	4.5010	66.408	31.060

$$l[y_t; \gamma, \beta_1, \beta_2, \lambda, \sigma^2] = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}u_t^2\,,\; t = 1, \dots, n. \tag{6.4}$$

It is easy to see that changing the measurement units on x_{1t} and x_{2t} leaves the form of model (6.1) and the null hypothesis invariant. For example, if both x_{1t} and x_{2t} are multiplied by a positive constant k, *i.e.*

$$x_{1t*} = kx_{1t}, \quad x_{2t*} = kx_{2t}, \tag{6.5}$$

(6.1) can be reexpressed in terms of the scaled variables x_{1t*} and x_{2t*} as

$$y_t = \gamma_* + \beta_{1*}x_{1t*}^{(\lambda)} + \beta_{2*}x_{2t*}^{(\lambda)} + u_t\,, \tag{6.6}$$

where the power parameter λ remains the same and

$$\gamma_* = \gamma - k^{(\lambda)}k^{-\lambda}\sum_{i=1}^{2}\beta_i\,, \quad \beta_{i*} = \beta_i k^{-\lambda},\; i = 1, 2. \tag{6.7}$$

On interpreting model (6.1) as a pseudo-model and (6.3) as a pseudo-likelihood, we will examine the effect of rescaling on GMM-based and pseudo-likelihood tests. Moment equations can be derived from the above model by differentiating the log-likelihood with respect to model parameters and equating the expectation to zero. This yields following five moment conditions:

$$\mathsf{E}\left[u_t\right] = 0\,,\; \mathsf{E}\left[u_t x_{1t}^{(\lambda)}\right] = 0\,,\; \mathsf{E}\left[u_t x_{2t}^{(\lambda)}\right] = 0\,, \tag{6.8}$$

$$\mathsf{E}\left[\frac{u_t}{\lambda}\left(\sum_{i=1}^{2}\beta_i\left[x_{it}^{\lambda}\ln x_{it} - x_{it}^{(\lambda)}\right]\right)\right] = 0\,,\; \mathsf{E}\left[u_t^2 - \sigma^2\right] = 0\,,\; t = 1, \dots, n. \tag{6.9}$$

Table 2. Test statistics for $H_0 : \lambda = 1$ for different measurement units
6 moment models

	Two-step GMM				CUP-GMM			
k	$\bar{D}$	Wald	Score	$C(\alpha)$	$\bar{D}$	Wald	Score	$C(\alpha)$
0.2	0.016	416.546	106.734	54.462	19.480	359.380	11.107	3.189
0.4	0.036	221.829	108.142	54.852	19.480	83.743	16.296	7.318
0.6	0.248	213.918	107.764	52.818	19.480	40.481	18.637	7.063
0.8	1.068	178.757	106.053	47.539	19.480	34.101	17.678	0.661
1.0	3.562	139.364	103.364	37.915	19.480	35.580	17.769	5.215
3.0	47.490	46.214	110.751	7.960	19.480	45.146	15.250	4.650
5.0	1.651	129.698	48.704	6.518	19.480	59.667	13.367	4.611
7.0	1.511	384.944	49.719	9.978	19.480	118.911	13.937	5.639
10.0	2.031	905.870	50.264	10.747	19.480	406.974	14.162	6.136

These equations provide an exactly identified system of equations. To get a system with 6 moment equations (hence overidentified), we add the equation:

$$\mathsf{E}\left[u_t x_{1t} x_{2t}\right] = 0\,. \tag{6.10}$$

To get data, we considered the sample size $n = 200$ and generated y_t according to equation (6.1) with the parameter values $\gamma = 10$, $\beta_1 = 1.0$, $\beta_2 = 1.0$, $\lambda = -1.0$, $\sigma^2 = 0.85$. The values of the regressors x_{1t} and x_{2t} were selected by transforming the values used in Dagenais and Dufour (1991).[6]

Numerical values of the GMM-based test statistics for a number of rescalings are reported in Table 1 for the 5 moment system (6.8) - (6.9) and in Table 2 for the 6 moment system (6.8) - (6.10). Results for the pseudo-likelihood tests appear in Table 1. Graphs of the non-invariant test statistics are also presented in figures 1 - 4. In these calculations, the first-step estimator of the two-step GMM tests is obtained by minimizing $M_n(\theta)$ in (2.5) with $W_n = I_m$ (equal weights), while the second step uses the weight matrix defined in (3.4). No correction for serial correlation is applied (although this could also be studied).

These results confirm the theoretical expectations of the theory presented in the previous sections. Namely, the GMM-based test statistics [$\bar{D}(\psi)$, Wald, score, $C(\alpha)$] are not invariant to measurement unit changes and, indeed, can change substantially (even if both the null and the alternative hypotheses remain the same under the rescaling considered here). Noninvariance is especially strong for the overidentified system (6 equations). In contrast, the $\bar{D}(\psi)$ and score tests based on the continuously updated GMM criterion are invariant. The same holds for the LR and adjusted score criteria based on linear exponential pseudo likelihoods.

[6]The numerical values of x_{1t}, x_{1t} and y_t used are available from the authors upon request. It is important to note that this is *not a simulation exercise* aimed at studying the statistical properties of the tests, but only an illustration of the *numerical properties* of the test statistics considered.

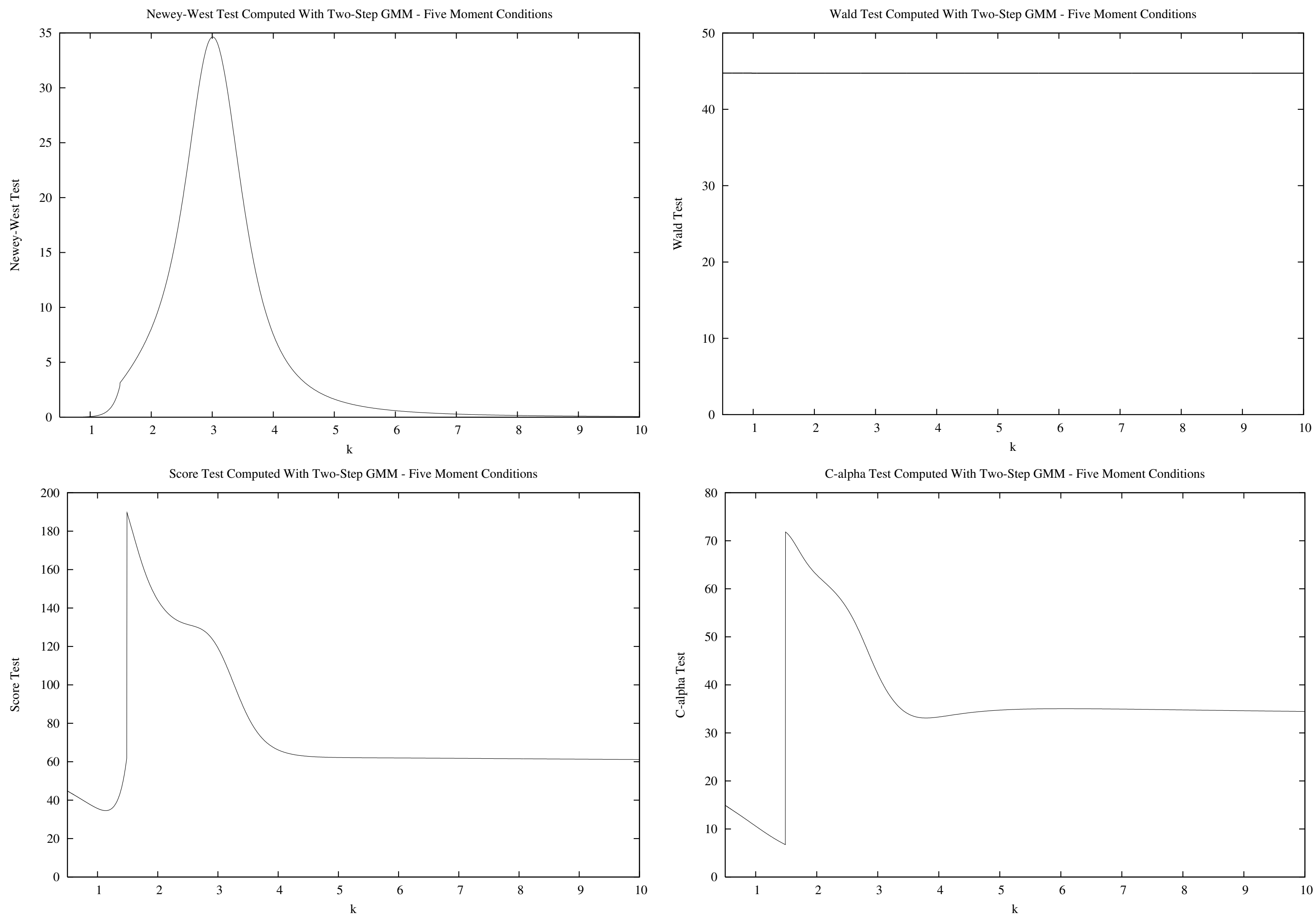

Figure 1. Two-step GMM tests based on 5 moment conditions

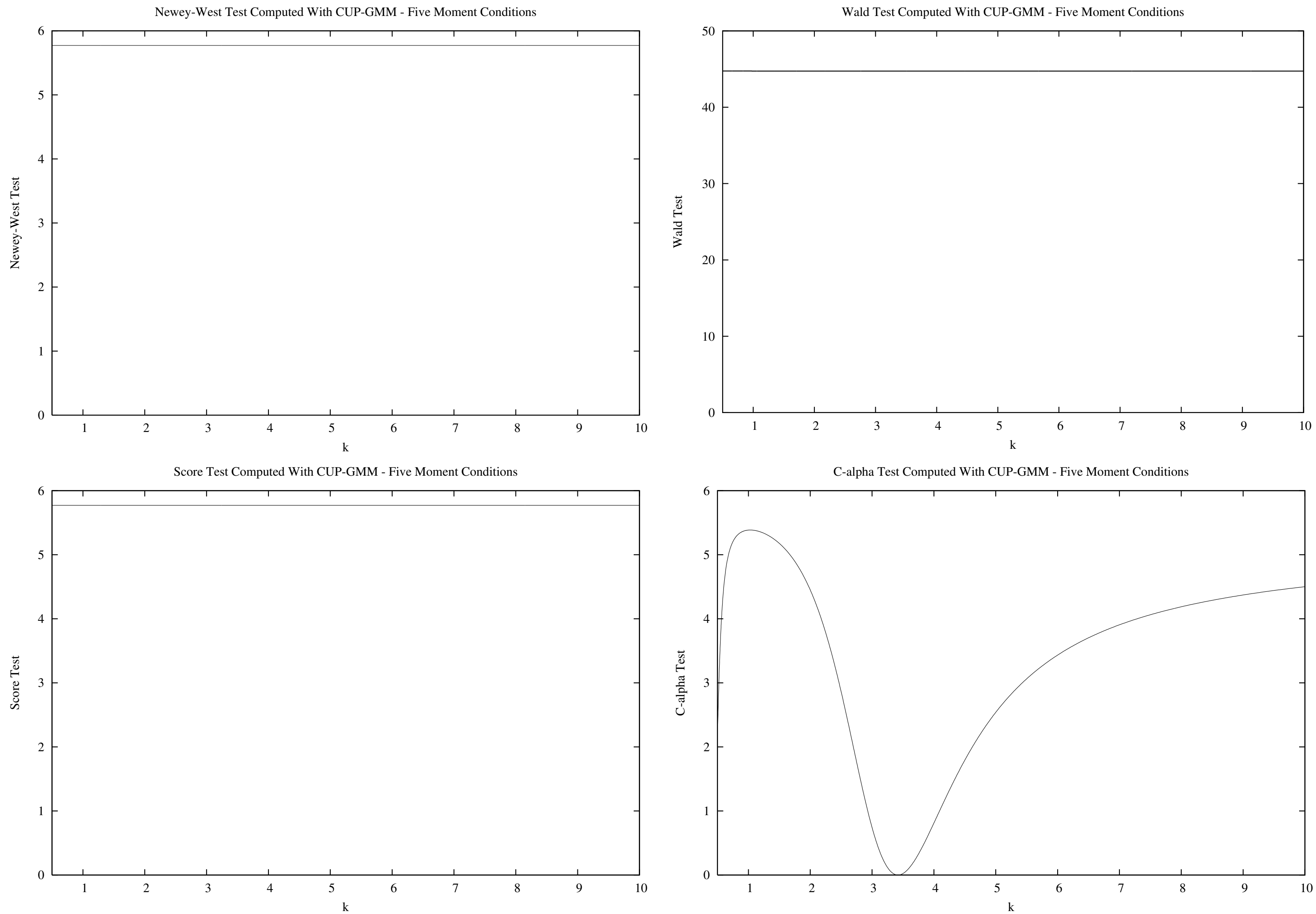

Figure 2. CUP GMM tests based on 5 moment conditions

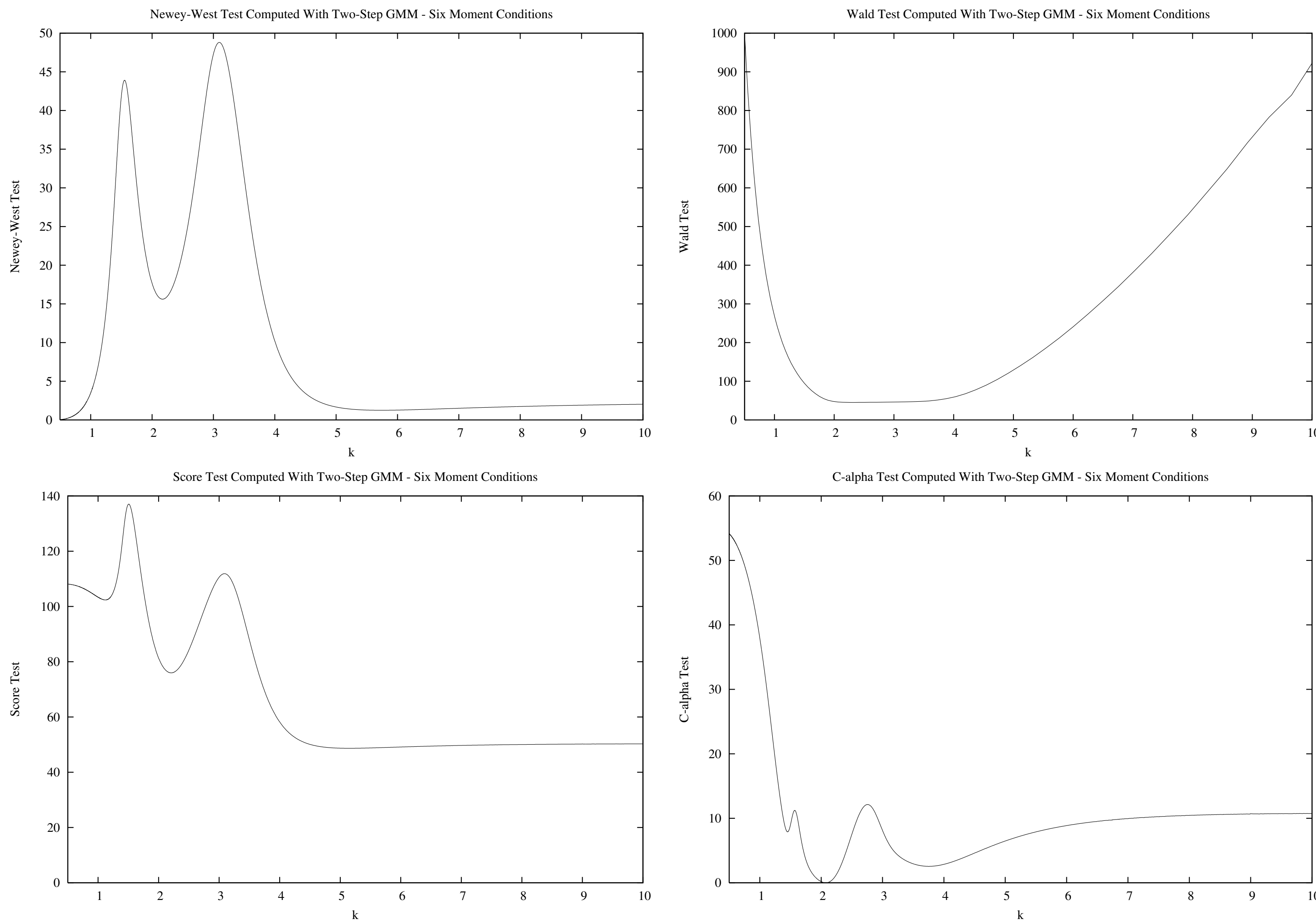

Figure 3. Two-step GMM tests based on 6 moment conditions

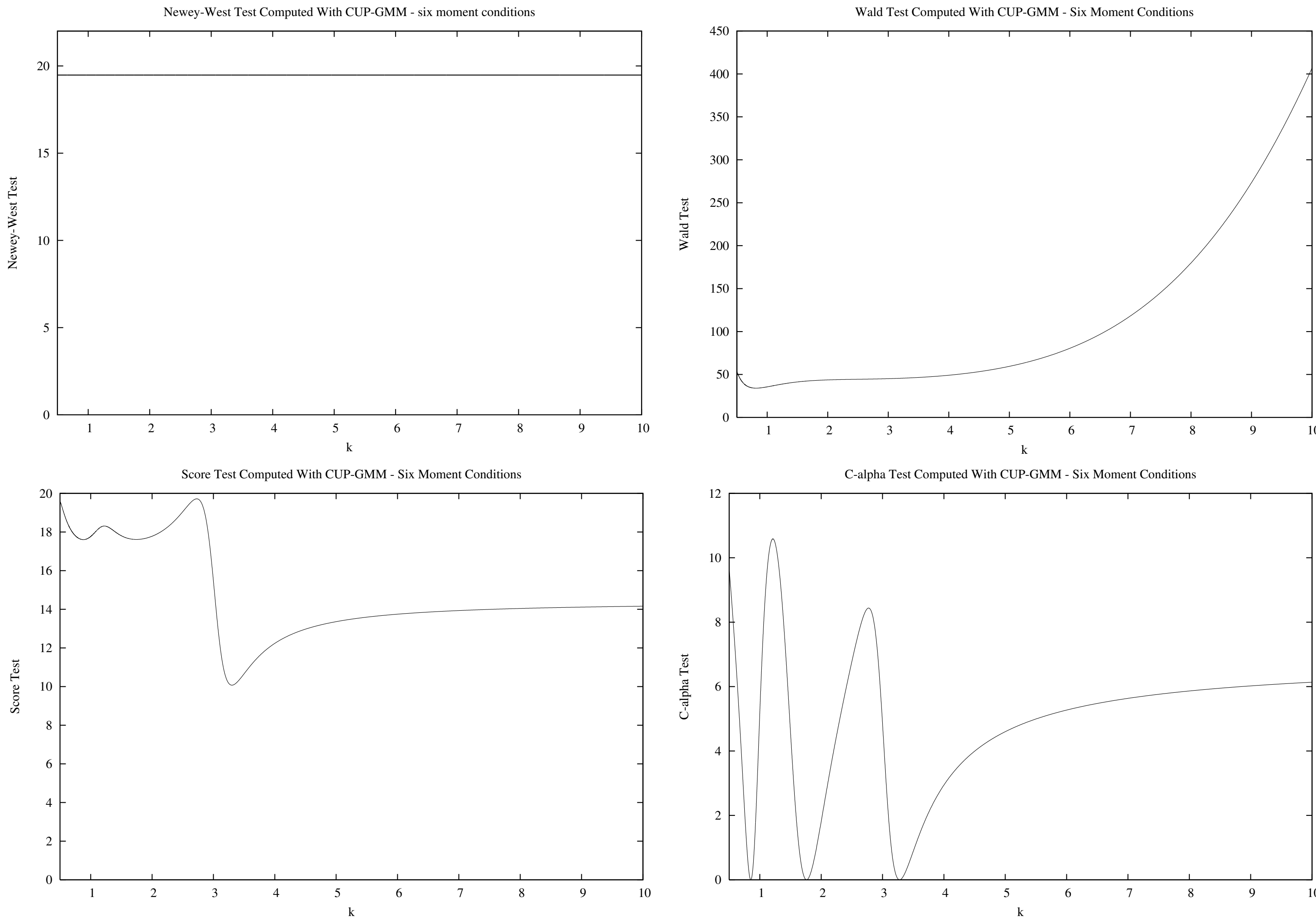

Figure 4. CUP GMM tests based on 6 moment conditions

7. Empirical illustration: linear stochastic discount factor models

In the context of linear stochastic discount factor model, it is shown that procedures based on non-invariant test statistics could lead to drastically different results depending on the form of identifying restrictions imposed. While an in-depth analysis of this problem is provided by Burnside (2010) from the perspective of model misspecification and identification, we aim to shed light on this issue from invariance considerations. The linear stochastic discount factor model is described by the following two equations:

$$\mathsf{E}[m_t R_t^e] = 0, \tag{7.1}$$

$$m_t = a - f_t' b, \tag{7.2}$$

where m_t is the stochastic discount factor (SDF); f_t is a $k \times 1$ vector of factors; R_t^e is the excess return (the difference between the gross asset return and the risk free rate); a and b are scalar and $p \times 1$ vector of unknown parameters, respectively; $\mathsf{E}[\cdot]$ is an expectation operator conditional on information up to time $t-1$. The equations (7.1) and (7.2) can equivalently be written as

$$\mathsf{E}\left[(a - f_t' b) R_t^e\right] = 0. \tag{7.3}$$

Since the unknowns a and b are not identified individually, we consider the following two normalizations [see Burnside (2010), Cochrane (2005)]:

$$\begin{aligned} \text{Normalization 1 :} \quad & \mathsf{E}\left[\frac{m_t}{a} R_t^e\right] = 0 \\ \text{Normalization 2 :} \quad & \mathsf{E}\left[\frac{m_t}{\mathsf{E}[m_t]} R_t^e\right] = 0. \end{aligned}$$

By applying the normalizations to (7.3), we have

$$\mathsf{E}\left[(1 - f_t'\theta) R_t^e\right] = 0\ ,\ \mathsf{E}\left[(1 - (f_t - \mu_f)'\theta_*) R_t^e\right] = 0, \tag{7.4}$$

where $\mu_f = \mathsf{E}[f_t]$, $\theta = b/a$ and $\theta_* = b/\mathsf{E}[m_t]$. The implied two sets of sample moments are:

$$D_n(\theta; Z_n) = \begin{pmatrix} \frac{1}{n}\sum_{t=1}^n (R_t^e - R_t^e f_t'\theta) \\ \frac{1}{n}\sum_{t=1}^n f_t - \mu_f \end{pmatrix}\ ,\ D_{n*}(\theta_*, Z_{n*}) = \begin{pmatrix} \frac{1}{n}\sum_{t=1}^n (R_t^e - R_t^e (f_t - \mu_f)'\theta_*) \\ \frac{1}{n}\sum_{t=1}^n f_t - \mu_f \end{pmatrix}.$$

It is clear that the sample moments satisfy

$$D_{n*}(\bar{g}(\theta), Z_{n*}) = K[\bar{g}(\theta)]' D_n(\theta; Z_n)$$

with $K[\bar{g}(\theta)] = \mathrm{diag}\{a/\mathsf{E}[m_t], 1\}$; one set of moments can be derived from the other by affine transformation of f_t. Let $\hat{I}_*(\theta_*)$ be the HAC estimator of $I(\theta_*)$ with Bartlett kernel and $\hat{I}(\theta)$ be

Table 3. J statistic for the validity of (7.4) under different identifying restrictions (p-values in parentheses)

Two-step GMM				CUP-GMM			
Normalization 1		Normalization 2		Normalization 1		Normalization 2	
QS	Bartlett	QS	Bartlett	QS	Bartlett	QS	Bartlett
0.218 (0.897)	0.106 (0.949)	17.214 (0.000)	5.670 (0.059)	1.235 (0.539)	1.774 (0.412)	1.759 (0.415)	1.866 (0.393)

defined similarly. Then we have

$$\hat{I}_*(\bar{g}(\theta)) = K[\bar{g}(\theta)]'\hat{I}(\theta)K[\bar{g}(\theta)].$$

Therefore, by virtue of equation (5.3), the continuously updated GMM (CUP-GMM) objective function and the statistic $\bar{D}$ are invariant to affine transformation of f_t i.e., they are not affected by the form of normalization employed. The model is estimated using the observed returns on 5 stocks [Weis Markets (WMK), Unisys Corporation (UIS), Orbital Sciences Corporation (ORB), Mattel (MAT) and Abaxis (ABAX)], and the three factors Rm-Rf [return of the market portfolio (Rm) minus the risk-free return (Rf)], SMB [Small (market capitalization) Minus Big] and HMB [High (book-to-market ratio) Minus Low] from the Fama-French data set over the period from January 5th, 1993 - March 16th, 1993. All calculations were carried out in R Version 3.0.2 (R Development Core Team (2013)) using the package **gmm** developed by Pierre Chaussé [Chaussé (2010)]. The data we use are readily available in the `Finance` data set contained in **gmm**. The estimation methods are two-step GMM and CUP-GMM with covariance matrix estimated with Bartlett and Quadratic Spectral (QS) kernels. Table 3 reports the values of J statistic for testing the validity of the restrictions (7.4). For the two-step GMM, it is clear that the values of test statistics differ greatly across the normalizations, and are sensitive to the choice of kernels. Furthermore, the test rejects the null of correct specification under Normalization 2 with QS kernel, but the conclusion is reversed under Normalization 1. In the case of CUP-GMM with Bartlett kernel, though there is a small incongruity in the values of test statistics (possibly due to an optimization error), the model is not rejected under both normalizations. The difference between test statistics under the CUP-GMM with QS kernel may be attributed to the non-invariance of the objective function with QS kernel. The main message of this exercise is that procedures based on non-invariant test statistics can be quite sensitive to the identifying restrictions employed and may result in conflicting conclusions. For a thorough discussion on the effect of normalizations on estimation and inferences, we refer the reader to Hamilton, Waggoner and Zha (2007).

8. Conclusion

In this paper, we have studied the invariance properties of hypothesis tests applicable in the context of incompletely specified models, such as models formulated in terms of estimating functions and moment conditions, which are usually estimated by GMM procedures, or models estimated by pseudo-likelihood and *M*-estimation methods. The test statistics examined include Wald-type, LR-type, LM-type, score-type, and $C(\alpha)$-type criteria. We found that all these procedures are *not generally invariant* to (possibly nonlinear) hypothesis reformulations and reparameterizations, such as those induced by measurement unit changes. This means that testing two equivalent hypotheses in the context of equivalent models may lead to completely different inferences. For example, this may occur after an apparently innocuous rescaling of some model variables.

In view of avoiding such undesirable properties, we studied restrictions that can be imposed on the objective functions used for pseudo-likelihood (or M-estimation) as well as the structure of the test criteria used with estimating functions and GMM procedures to obtain invariant tests. In particular, we showed that using linear exponential pseudo-likelihood functions allows one to obtain invariant score-type and $C(\alpha)-$type test criteria, while in the context of estimating function (or GMM) procedures it is possible to modify a LR-type statistic proposed by Newey and West (1987) to obtain a test statistic that is invariant to general reparameterizations. The invariance associated with linear exponential pseudo-likelihood functions is interpreted as a strong argument for using such pseudo-likelihood functions in empirical work. Furthermore, the LR-type statistic is the one associated with using continuously updated GMM estimators based on appropriately restricted weight matrices. Of course, this provides an extra argument for such GMM estimators.

A. Appendix: Assumptions for the generalized $C(\alpha)$ statistic

The asymptotic distribution of the generalized $C(\alpha)$ statistic defined in (3.24) is derived under the following set of assumptions. Note $\|\cdot\|$ refers to the Euclidean distance, applied to either vectors or matrices.

Assumption A.1 EXISTENCE OF SCORE-TYPE FUNCTIONS.

$$D_n(\theta,\,\omega) = \big(D_{1n}(\theta,\,\omega),\,\dots,\,D_{mn}(\theta,\,\omega)\big)',\ \omega \in \mathscr{Z},\ n = 1,\,2,\,\dots$$

is a sequence of $m\times 1$ random vectors, defined on a common probability space $(\mathscr{Z},\,\mathscr{A}_{\mathscr{Z}},\,\mathsf{P})$, which are functions of a $p\times 1$ parameter vector θ, where $\theta \in \Theta \subseteq \mathbb{R}^p$ and Θ is a non-empty open subset of $\mathbb{R}^p$. All the random variables considered here as well in the following assumptions are functions of ω, so the symbol ω may be dropped to simplify notations [e.g., $D_n(\theta) \equiv D_n(\theta,\omega)$].

Assumption A.2 SCORE ASYMPTOTIC NORMALITY. *There is a value $\theta_0 \in \Theta$ such that*

$$\sqrt{n}D_n(\theta_0) \xrightarrow[n\to\infty]{\mathsf{p}} D_\infty(\theta_0) \text{ where } D_\infty(\theta_0) \sim N[0,I(\theta_0)]\,.$$

Assumption A.3 NON-SINGULARITY OF THE SCORE VARIANCE. *$I(\theta_0)$ is nonsingular.*

Assumption A.4 SCORE DIFFERENTIABILITY. *$D_n(\theta,\omega)$ is almost surely (a.s.) differentiable with respect to θ, for all n, in a non-empty open neighborhood N_1 of θ_0. The derivative matrix of $D_n(\theta,\omega)$ is denoted*

$$H_n(\theta,\,\omega) = \frac{\partial D_n(\theta,\,\omega)}{\partial \theta'}$$

where the sequence of matrices $H_n(\theta,\,\omega)$, $n \geq 1$, is well-defined for $\omega \in \mathscr{D}_H$ and $\mathscr{D}_H$ is an event with probability one (i.e., $\mathsf{P}[\omega \in \mathscr{D}_H] = 1$).

Assumption A.5 SCORE DERIVATIVE CONVERGENCE. *There is an $m\times p$ (nonrandom) matrix function $J(\theta)$ and a non-empty open neighborhood N_2 of θ_0 such that, for all $\varepsilon > 0$ and $\delta > 0$,*

$$\limsup_{n\to\infty} \mathsf{P}\big[\{\omega : \Delta_n(\theta_0,\,\delta,\,\omega) > \varepsilon\}\big] \leq U_H(\delta,\,\varepsilon,\,\theta_0)$$

where

$$\Delta_n(\theta_0,\,\delta,\,\omega) \equiv \sup\{\|H_n(\theta,\,\omega) - J(\theta_0)\| : \theta \in N_2 \text{ and } 0 \leq \|\theta - \theta_0\| \leq \delta\}\,,$$

$U_H(\delta,\,\varepsilon,\,\theta_0) \geq 0$ *and* $\lim_{\delta\downarrow 0} U_H(\delta,\,\varepsilon,\,\theta_0) = 0\,.$

Assumption A.6 SCORE EXPANSION. *For θ in a non-empty open neighborhood N_3 of θ_0, $D_n(\theta)$ admits an expansion of the form*

$$D_n(\theta,\,\omega) = D_n(\theta_0,\,\omega) + J(\theta_0)(\theta - \theta_0) + R_n(\theta,\,\theta_0,\,\omega)$$

for $\omega \in \mathscr{D}_H$, *where the remainder* $R_n(\theta,\, \theta_0,\, \omega)$ *satisfies the following condition: for any* $\varepsilon > 0$ *and* $\delta > 0$, *we have*

$$\limsup_{n\to\infty} \mathsf{P}\big[\{\omega : r_n(\delta,\, \theta_0,\, \omega) > \varepsilon\}\big] \le U_D(\delta,\, \varepsilon,\, \theta_0)$$

where

$$r_n(\delta,\, \theta_0,\, \omega) = \sup\left\{ \frac{\|R_n(\theta,\, \theta_0,\, \omega)\|}{\|\theta - \theta_0\|} : \theta \in N_3 \text{ and } 0 < \|\theta - \theta_0\| \le \delta \right\},$$

$U_D(\delta,\, \varepsilon,\, \theta_0) \ge 0$ *and* $\lim\limits_{\delta\downarrow 0} U_D(\delta,\, \varepsilon,\, \theta_0) = 0$.

Assumption A.7 SCORE DERIVATIVE NON-DEGENERACY. $\mathrm{rank}[J(\theta)] = p$, *for all* θ *in a non-empty open neighborhood* N_4 *of* θ_0.

Assumption A.8 RESTRICTION DIFFERENTIABILITY. $\psi(\theta)$ *is a* $p_1 \times 1$ *differentiable vector function of* θ.

Assumption A.9 RESTRICTION RANK. *There is a non-empty open neighborhood* N_5 *of* θ_0 *such that* $\psi(\theta)$ *is continuously differentiable with derivative* $P(\theta) \equiv \dfrac{\partial \psi}{\partial \theta'}$ *and such that*

$$\psi(\theta) = 0 \text{ and } \theta \in N_5 \Rightarrow \mathrm{rank}\,[P(\theta)] = p_1$$

where $0 \le p_1 \le p$.

Assumption A.10 ESTIMATOR $\sqrt{n}$ CONVERGENCE. $\tilde{\theta}^0_n \equiv \tilde{\theta}^0_n(\omega)$ *is a consistent estimator of* θ_0, i.e.,

$$\operatorname*{plim}_{n\to\infty}\big(\tilde{\theta}^0_n - \theta_0\big) = 0,$$

such that $\sqrt{n}\big(\tilde{\theta}^0_n - \theta_0\big)$ *is asymptotically bounded in probability,* i.e.,

$$\limsup_{n\to\infty}\ \mathsf{P}\big[\{\omega : \sqrt{n}\big\|\tilde{\theta}^0_n - \theta_0\big\| \ge y\}\big] \le U(y;\, \theta_0)\ , \forall y > 0,$$

where $U(y;\theta_0)$ *is a function such that* $\lim\limits_{y\to\infty} U(y;\theta_0) = 0$.

Assumption A.11 RESTRICTED ESTIMATOR. $\psi(\tilde{\theta}^0_n) = \psi(\theta_0) = 0$ *with probability* 1.

Assumption A.12 CONSISTENT ESTIMATOR OF SCORE COVARIANCE MATRIX. $\widetilde{I}_0$ *is a weakly consistent estimator of* $I(\theta_0)$, i.e., $\operatorname*{plim}\limits_{n\to\infty} \widetilde{I}_0 = I(\theta_0)$.

Assumption A.13 WEIGHT MATRIX CONSISTENCY. $W_n,\ n \ge 1$, *is a sequence of* $m \times m$ *matrices such that* $\operatorname*{plim}\limits_{n\to\infty} W_n = W_0$ *where* W_0 *is nonsingular.*

B. Appendix: Assumptions for the modified NW LR-type statistics

This appendix presents the regularity conditions under which the modified Newey-West LR type has asymptotically chi-square distributed.

Assumption B.1 COMPACT PARAMETER SPACE. *$\theta \in \Theta$, where Θ is a compact set of $\mathbb{R}^p$.*

Assumption B.2 OBJECTIVE FUNCTION CONTINUITY. *$M_n(\theta) = M_n(\theta, \omega)$ is a real function on $\Theta \times \mathscr{Z}$, such that $M_n(\theta, \omega)$ is a continuous function of θ for all $\omega \in \mathscr{Z}$.*

Assumption B.3 OBJECTIVE FUNCTION UNIFORM CONVERGENCE. *There is a fixed (non-random) function $\bar{M}(\theta)$ such that*

$$\mathsf{P}(\{\omega : \max_{\theta} \mid M_n(\theta, \omega) - \bar{M}(\theta) \mid \underset{n \to \infty}{\longrightarrow} 0\}) = 1.$$

Assumption B.4 ASYMPTOTIC IDENTIFICATION. *$\bar{M}(\theta)$ has a unique minimum at $\theta = \theta_0$ in the interior of Θ.*

Assumption B.5 UNIFORM CONVERGENCE OF SECOND DERIVATIVES. *$M_n(\theta, \omega)$ is a twice continuously differentiable function in θ and there is a fixed (non-random) function $G(\theta)$ such that*

$$\mathsf{P}\Big[\{\omega : \sup_{\theta} \Big\| \frac{\partial^2 M_n}{\partial\theta\partial\theta'}(\theta, \omega) - G(\theta) \Big\| \underset{n \to \infty}{\longrightarrow} 0\}\Big] = 1.$$

Assumption B.6 OBJECTIVE FUNCTION ASYMPTOTIC REGULARITY. *For all $\theta \in \Theta$, $G(\theta)$ is a nonsingular matrix.*

Assumption B.7 OBJECTIVE FUNCTION ASYMPTOTIC NORMALITY.

$$\sqrt{n}\,\frac{\partial M_n}{\partial \theta}(\theta_0; \omega) \underset{n \to \infty}{\overset{L}{\longrightarrow}} N[0, H(\theta_0)]\,.$$

Assumption B.8 RESTRICTION DIFFERENTIABILITY. *$\psi(\theta)$ is a $p_1 \times 1$ differentiable vector function of θ.*

Assumption B.9 RESTRICTION RANK. *There is a non-empty open neighborhood N_5 of θ_0 such that $\psi(\theta)$ is continuously differentiable with derivative $P(\theta) \equiv \dfrac{\partial \psi}{\partial \theta'}$ and such that*

$$\psi(\theta) = 0 \text{ and } \theta \in N_5 \Rightarrow \operatorname{rank}[P(\theta)] = r_1$$

where $0 \le r_1 \le p_1$.

References

Amari, S.-I. (1990), *Diffrential-Geometrical Methods in Statistics*, Vol. 28 of *Lecture Notes in Statistics*, Springer-Verlag, Berlin.

Andrews, D. W. K. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* **59**, 817–858.

Andrews, D. W. K. and Monahan, J. C. (1992), 'An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator', *Econometrica* **60**, 953–966.

Basawa, I. V. (1985), Neyman-Le Cam tests based on estimating functions, *in* L. Le Cam and R. A. Olshen, eds, 'Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer', Wadsworth, Belmont, California, pp. 811–825.

Basawa, I. V., Godambe, V. P. and Taylor, R. L., eds (1997), *Selected Proceedings of the Symposium on Estimating Functions*, Vol. 32 of *IMS Lecture Notes Monograph Series*, Institute of Mathematical Statistics, Hayward, California.

Bates, D. M. and Watts, D. G. (1980), 'Relative curvature measures of nonlinearity', *Journal of the Royal Statistical Society, Series B* **42**, 1–25.

Bates, D. M. and Watts, D. G. (1981), 'Parameter transformations for improved approximate confidence regions in nonlinear least squares', *The Annals of Statistics* **9**, 1152–1167.

Berger, A. and Wallenstein, S. (1989), 'On the theory of C_α-tests', *Statistics and Probability Letters* **7**, 419–424.

Bourlange, D. and Doz, C. (1988), 'Pseudo-maximum de vraisemblance: expériences de simulations dans le cadre de modèle de Poisson', *Annales d'Économie et de Statistique* **10**, 139–178.

Breusch, T. S. and Schmidt, P. (1988), 'Alternative forms of the Wald test: How long is a piece of string?', *Communications in Statistics, Theory and Methods* **17**, 2789–2795.

Burguete, W. J., Gallant, A. R. and Souza, G. (1982), 'On unification of the asymptotic theory of nonlinear econometric models', *Econometric Reviews* **1**, 151–211 (with comments).

Burnside, C. (2010), Identification and inference in linear stochastic discount factor models, Technical report. NBER WP 16634.

Burnside, C. and Eichenbaum, M. S. (1996), 'Small-sample properties of GMM-based Wald tests', *Journal of Business and Economic Statistics* **14**, 294–308.

Chaussé, P. (2010), 'Computing generalized method of moments and generalized empirical likelihood with r', *Journal of Statistical Software* **34**(11), 1–35.

Cochrane, J. H. (2005), *Asset Pricing*, Princeton University Press, Princeton, New Jersey.

Cox, D. R. and Hinkley, D. V. (1974), *Theoretical Statistics*, Chapman & Hall, London.

Crépon, B. and Duguet, E. (1997), 'Research and development, competition and innovation: Pseudo maximum likelihood and simulated maximum likelihood applied to count data models with heterogeneity', *Journal of Econometrics* **79**, 355–378.

Critchley, F., Marriott, P. and Salmon, M. (1996), 'On the differential geometry of the Wald test with nonlinear restrictions', *Econometrica* **64**, 1213–1222.

Cushing, M. J. and McGarvey, M. G. (1999), Covariance matrix estimation, *in* (Mátyás 1999), chapter 3, pp. 63–95.

Dagenais, M. G. and Dufour, J.-M. (1991), 'Invariance, nonlinear models and asymptotic tests', *Econometrica* **59**, 1601–1615.

Dagenais, M. G. and Dufour, J.-M. (1992), 'On the lack of invariance of some asymptotic tests to rescaling', *Economics Letters* **38**, 251–257.

Davidson, R. and MacKinnon, J. G. (1991), 'Artificial regressions and $C(\alpha)$ tests', *Economics Letters* **35**, 149–153.

Davidson, R. and MacKinnon, J. G. (1993), *Estimation and Inference in Econometrics*, Oxford University Press, New York.

Dufour, J.-M. (1997), 'Some impossibility theorems in econometrics, with applications to structural and dynamic models', *Econometrica* **65**, 1365–1389.

Dufour, J.-M. and Dagenais, M. G. (1992), 'Nonlinear models, rescaling and test invariance', *Journal of Statistical Planning and Inference* **32**, 111–135.

Dufour, J.-M. and Jasiak, J. (2001), 'Finite sample limited information inference methods for structural equations and models with generated regressors', *International Economic Review* **42**, 815–843.

Dufour, J.-M., Trognon, A. and Tuvaandorj, P. (2013*a*), Invariant tests based on *M*-estimators, estimating functions, and the generalized method of moments, Technical report, Department of Economics, McGill University. http://www.jeanmariedufour.com.

Dufour, J.-M., Trognon, A. and Tuvaandorj, P. (2013*b*), A note on asymptotics of GMM distance test, Technical report, McGill University and CREST-ENSAE Paris.

Dufour, J.-M., Trognon, A. and Tuvaandorj, P. (2015), Generalized $C(\alpha)$ tests for estimating functions with serial dependence, Technical report, McGill University and CREST-ENSAE Paris.

Durbin, J. (1960), 'Estimation of parameters in time series regression models', *Journal of the Royal Statistical Society, Series A* **22**, 139–153.

Ferguson, T. S. (1967), *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, New York.

Gallant, A. R. (1987), *Nonlinear Statistical Models*, John Wiley & Sons, New York.

Gallant, A. R. and White, H. (1988), *Estimation and Inference for Nonlinear Dynamic Models*, Blackwell, New York .

Godambe, V. P. (1960), 'An optimum property of regular maximum likelihood estimation', *The Annals of Mathematical Statistics* **31**, 1208–1212. Ackowledgement 32 (1960), 1343.

Godambe, V. P., ed. (1991), *Estimating Functions*, Clarendon Press, Oxford, U.K.

Gong, G. and Samaniego, F. J. (1981), 'Pseudo maximum likelihood estimation: Theory and applications', *The Annals of Statistics* **13**, 861–869.

Gouriéroux, C. and Monfort, A. (1989), 'A general framework for testing a null hypothesis in a 'mixed' form', *Econometric Theory* **5**, 63–82.

Gouriéroux, C. and Monfort, A. (1993), Pseudo-likelihood methods, *in* G. S. Maddala, C. R. Rao and H. D. Vinod, eds, 'Handbook of Statistics 11: Econometrics', North-Holland, Amsterdam, chapter 12, pp. 335–362.

Gouriéroux, C. and Monfort, A. (1995), *Statistics and Econometric Models, Volumes One and Two*, Cambridge University Press, Cambridge, U.K. Translated by Quang Vuong.

Gouriéroux, C., Monfort, A. and Trognon, A. (1984*a*), Estimation and test in probit models with serial correlation, *in* J. P. Florens, M. Mouchart, J. P. Roualt and L. Simar, eds, 'Alternative Approaches to Time Series Analysis', Facultes Universitaires Saint-Louis, Bruxelles, Belgium, pp. 169–209.

Gouriéroux, C., Monfort, A. and Trognon, A. (1984*b*), 'Pseudo maximum likelihood methods: Applications to Poisson models', *Econometrica* **52**, 701–720.

Gouriéroux, C., Monfort, A. and Trognon, A. (1984*c*), 'Pseudo maximum likelihood methods: Theory', *Econometrica* **52**, 681–700.

Gregory, A. and Veall, M. (1985), 'Formulating Wald tests of nonlinear restrictions', *Econometrica* **53**, 1465–1468.

Hall, A. R. (1999), Hypothesis testing in models estimated by GMM, *in* (Mátyás 1999), chapter 4, pp. 96–127.

Hall, A. R. (2004), *Generalized Method of Moments*, Advanced Texts in Econometrics, Oxford University Press, Oxford, U.K.

Hamilton, J. D., Waggoner, D. F. and Zha, T. (2007), 'Normalization in econometrics', *Econometric Reviews* **26**, 221–252.

Hansen, B. E. (1992), 'Consistent covariance matrix estimation for dependent heterogeneous processes', *Econometrica* **60**, 967–972.

Hansen, L. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica* **50**, 1029–1054.

Hansen, L. P., Heaton, J. and Yaron, A. (1996), 'Finite-sample properties of some alternative GMM estimators', *Journal of Business and Economic Statistics* **14**, 262–280.

Hotelling, H. (1936), 'Relations between two sets of variables', *Biometrika* **28**, 321–377.

Hougaard, P. (1982), 'Parameterisations of nonlinear models', *Journal of the Royal Statistical Society, Series B* **44**, 244–252.

Huber, P. J. (1981), *Robust Statistics*, John Wiley & Sons, New York.

Jorgensen, B. (1997), *The Theory of Dispersion Models*, Chapman & Hall, London, U.K.

Kass, R. E. and Vos, P. W. (1997), *Geometrical Foundations of Asymptotic Inference*, John Wiley & Sons, New York.

Kocherlakota, S. and Kocherlakota, K. (1991), 'Neyman's $C(\alpha)$ test and Rao's efficient score test for composite hypotheses', *Statistics and Probability Letters* **11**, 491–493.

Lafontaine, F. and White, K. J. (1986), 'Obtaining any Wald statistic you want', *Economics Letters* **21**, 35–40.

Larsen, P. V. and Jupp, P. E. (2003), 'Parametrization-invariant Wald tests', *Bernoulli* **9**(1), 167–182.

Le Cam, L. (1990), 'On the standard asymptotic confidence ellipsoids of Wald', *International Statistical Review* **58**, 129–152.

Lehmann, E. L. (1983), *Theory of Point Estimation*, John Wiley & Sons, New York.

Lehmann, E. L. (1986), *Testing Statistical Hypotheses*, 2nd edn, John Wiley & Sons, New York.

Marriott, P. and Salmon, M., eds (2000), *Applications of Differential Geometry to Econometrics*, Cambridge University Press, Cambridge, U.K.

Mátyás, L., ed. (1999), *Generalized Method of Moments Estimation*, Cambridge University Press, Cambridge, U.K.

Newey, W. K. (1985), 'GMM specification testing', *Jornal of Econometrics* **29**, 229–256.

Newey, W. K. and McFadden, D. (1994), Large sample estimation and hypothesis testing, *in* R. F. Engle and D. L. McFadden, eds, 'Handbook of Econometrics, Volume 4', North-Holland, Amsterdam, chapter 36, pp. 2111–2245.

Newey, W. K. and West, K. D. (1987), ‘Hypothesis testing with efficient method of moments estimators’, *International Economic Review* **28**, 777–787.

Phillips, P. C. B. and Park, J. Y. (1988), ‘On the formulation of Wald tests of nonlinear restrictions’, *Econometrica* **56**, 1065–1083.

Pitman, E. J. G. (1939), ‘Tests of hypotheses concerning location and scale parameters’, *Biometrika* **31**, 200–215.

Podivinsky, J. M. (1999), Finite sample properties of GMM estimators and tests, *in* (Mátyás 1999), chapter 5, pp. 128–148.

R Development Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. *http://www.R-project.org

Ronchetti, E. (1987), ‘Robust $C(\alpha)$-type tests for linear models’, *Sankhyā Series A* **49**, 1–16.

Smith, R. J. (1987), ‘Alternative asymptotically optimal tests and their application to dynamic specification’, *Review of Economic Studies* **LIV**, 665–680.

Trognon, A. (1984), ‘Généralisation des tests asymptotiques au cas où le modèle est incomplètement spécifié’, *Cahiers du Séminaire d’Économétrie* **26**, 93–109.

Trognon, A. and Gouriéroux, C. (1988), Une note sur l’efficacité des procédures d’estimation en deux étapes, *in* P. Champsaur, M. Deleau, J.-M. Grandmont, C. Henry, J.-J. Laffont, G. Laroque, J. Mairesse, A. Monfort and Y. Younès, eds, ‘Mélanges économiques. Essais en l’honneur de Edmond Malinvaud’, Economica, Paris, pp. 1056–1074.

Vaeth, M. (1985), ‘On the use of Wald’s test in exponential families’, *International Statistical Review* **53**, 199–214.

Vuong, Q. (1989), ‘Likelihood ratio tests for model selection and non-nested hypotheses’, *Econometrica* **57**, 307–333.

White, H. (1982), ‘Maximum likelihood estimation of misspecified models’, *Econometrica* **50**, 1–25.