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

Jean-Marie Dufour †
McGill University

Alain Trognon ‡
ENSAE, Paris

First version: May 1997
Revised: April 2001, September 2008
This version: September 2008
Compiled: October 10, 2008, 10:41pm

* 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 and Pascale Valéry for several useful comments. Earlier versions of this paper were presented at the Canadian Statistical Society 1997 annual meeting and at CREST (CREST, Paris). This work was supported by the William Dow Chair in Political Economy (McGill University), the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Institut de finance mathématique de Montréal (IFM2), 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 8879; FAX: (1) 514 398 4938; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

‡ Director, Groupe des écoles nationales d'économie et statistique (ENSAE, ENSAI, CREST, CEPE). Mailing address: GENES, 3 avenue Pierre Larousse, 92245 Malakoff Cedex, France. TEL.: 33 1 41175131; FAX: 33 1 41176480. email: trognon@ensae.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; Reparamerization; 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 have no incidence on 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*c*, 1984*b*), 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), 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 7. Proofs appear in appendix.

2. Framework

We consider an inference problem about a parameter of interest $\theta \in \Omega \subseteq \mathbb{R}^p$. This parameter appears in a model which is not fully specified. In order to identify θ, we assume there exists 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

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

$D_\infty(\cdot\,; \theta_0)$ is an application from Ω onto $\mathbb{R}^p$ 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 that

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

and

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

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

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

where W_n is a symmetric positive definite matrix. The method of estimating equations [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

$$\Sigma(W_0) = \left(J_0' W_0 J_0\right)^{-1} J_0' W_0 I_0 W_0 J_0 \left(J_0' W_0 J_0\right)^{-1}, \tag{2.7}$$

$$J_0 = J(\theta_0), \quad I_0 = I(\theta_0), \quad W_0 = \underset{n\to\infty}{\text{plim}}\, W_n, \quad \det(W_0) \neq 0; \tag{2.8}$$

see Gouriéroux and Monfort (1995, Chapter 9).

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

$$D_n(\hat{\theta}_n; Z_n) = 0. \tag{2.9}$$

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

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

This sequence is asymptotically normal with zero mean and asymptotic variance

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

Obviously, condition (2.9) 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 (2.9) also covers most classical asymptotic methods of estimation. 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

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

b) *Generalized method of moments (GMM) method.* θ is identified through a $m \times 1$ vector of conditions of the form:

$$E\left[h_t\left(\theta; z_t\right)\right] = 0\ , \quad t = 1, \ldots, n\ . \tag{2.12}$$

Then one considers the sample analogue of the above mean,

$$\overline{h}_n\left(\theta\right) = \frac{1}{n}\sum_{t=1}^{n} h_t\left(\theta; z_t\right)\ , \tag{2.13}$$

and the quadratic form

$$M_n\left(\theta\right) = \overline{h}_n\left(\theta\right)' W_n \overline{h}_n\left(\theta\right) \tag{2.14}$$

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

$$D_n\left(\theta; Z_n\right) = 2\frac{\partial \overline{h}_n\left(\theta\right)'}{\partial \theta} W_n \overline{h}_n\left(\theta\right)\ . \tag{2.15}$$

c) *M-estimator.* $\hat{\theta}_n$ is defined through an objective function Q_n of the form :

$$Q_n\left(\theta; Z_n\right) = \frac{1}{n}\sum_{t=1}^{n} \xi\left(\theta; z_t\right). \tag{2.16}$$

The score function has the following form:

$$D_n\left(\theta; Z_n\right) = \frac{\partial Q_n}{\partial \theta}\left(\theta; Z_n\right) = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial}{\partial \theta}\xi\left(\theta; z_t\right). \tag{2.17}$$

3. Test statistics

Consider now the problem of testing

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

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

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

has full row rank (at least in an open neighborhood of θ_0). Let $\hat{\theta}_n$ be the unrestricted estimator obtained by minimizing $M_n\left(\theta\right)$, 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 $I\left(\theta_0\right)$ and $J\left(\theta_0\right)$ 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 :

$$\hat{I}(\theta) = \frac{1}{n}\sum_{t=1}^{n} h_t(\theta; z_t)\, h_t(\theta; z_t)'\ \ , \tag{3.4}$$

$$\hat{J}(\theta) = \frac{\partial D_n}{\partial \theta'}(\theta) = H_n(\theta; Z_n)\ \ , \tag{3.5}$$

where θ can be replaced by an appropriate estimator.

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

$$\begin{aligned} \hat{I}(\theta) &= \sum_{s=1}^{n}\sum_{t=1}^{n} w_{st}(n)\, h_s(\theta; z_s)\, h_t(\theta; z_t)' \\ &= h(\theta;\ Z_n)\, W_I(n) h(\theta;\ Z_n)' \end{aligned} \tag{3.6}$$

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

$$h(\theta;\ Z_n) = \left[h_1(\theta; z_1)\,,\ h_2(\theta; z_2)\,,\,,\ \ldots\ ,\ h_n(\theta; z_n)\,\right]. \tag{3.7}$$

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

$$\hat{I}(\theta) = \sum_{t=1}^{n}\left[h_t(\theta; z_t) - \overline{h}(\theta)\right]\left[h_t(\theta; z_t) - \overline{h}(\theta)\right]' \tag{3.8}$$

where $\overline{h}(\theta) = \frac{1}{n}\sum_{t=1}^{n} h_t(\theta; z_t)$. Similarly, so-called "heteroskedastic, autocorrelation consistent (HAC)" covariance matrix estimators can usually be rewritten in the form (3.6). In most cases, such estimators are defined by a formula of the type:

$$\hat{I}(\theta) = \sum_{j=-n+1}^{n-1} \kappa(j/S_n)\ \hat{\Gamma}(j,\ \theta) \tag{3.9}$$

where $\kappa(\cdot)$ is a kernel function, S_n is a bandwidth parameter (which depends on the sample size and, possibly, on the data), and

$$\hat{\Gamma}(j,\,\theta) = \begin{cases} \frac{1}{n}\sum\limits_{t=-j+1}^{n} h_t\,(\theta; z_t)\,h_{t-j}\,(\theta; z_{t-j})'\,, & \text{if } j \geq 0\,, \\ \frac{1}{n}\sum\limits_{t=-j+1}^{n} h_{t+j}\,(\theta; z_{t+j})\,h_t\,(\theta; z_t)'\,, & \text{if } j < 0\,. \end{cases} \tag{3.10}$$

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 the assumption that the referenced inverse matrices do exist, these test criteria can be defined as follows:

(a) *Wald-type statistic,*

$$W\,(\psi) = n\,\psi(\hat{\theta}_n)'\big[\hat{P}\big(\hat{J}'\hat{I}^{-1}\hat{J}\big)^{-1}\hat{P}'\,\big]^{-1}\psi(\hat{\theta}_n) \tag{3.11}$$

where $\hat{P} = P(\hat{\theta}_n)$, $\hat{I} = \hat{I}(\hat{\theta}_n)$ and $\hat{J} = \hat{J}(\hat{\theta}_n)$;

(b) *score-type statistic,*

$$S\,(\psi) = n\,D_n(\hat{\theta}_n^0; Z_n)'\hat{I}_0^{-1}\hat{J}_0\big(\hat{J}_0'\hat{I}_0^{-1}\hat{J}_0\big)^{-1}\hat{J}_0'\hat{I}_0^{-1}D_n(\hat{\theta}_n^0; Z_n) \tag{3.12}$$

where $\hat{I}_0 = \hat{I}(\hat{\theta}_n^0)$ and $\hat{J}_0 = \hat{J}(\hat{\theta}_n^0)$;

(c) *Lagrange-multiplier (LM) type statistic*,

$$LM\,(\psi) = n\,\hat{\lambda}_n'\hat{P}_0\big(\hat{J}_0'\hat{I}_0^{-1}\hat{J}_0\big)^{-1}\hat{P}_0'\hat{\lambda}_n \tag{3.13}$$

where $\hat{P}_0 = P(\hat{\theta}_n^0)$;

(d) $C\,(\alpha)$*-type statistic,*

$$PC(\tilde{\theta}_n^0; \psi) = n\,D_n\big(\tilde{\theta}_n^0; Z_n\big)'\tilde{W}_0 D_n\big(\tilde{\theta}_n^0; Z_n\big) \tag{3.14}$$

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

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

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 :

$$LR(\psi) = S_n(\hat{\theta}_n; Z_n) - S_n(\hat{\theta}_n^0; Z_n) . \tag{3.15}$$

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):

$$D_{NW} = n\left[M_n(\hat{\theta}_n^0, \tilde{I}_0) - M_n(\hat{\theta}_n, \tilde{I}_0)\right] \tag{3.16}$$

where

$$M_n(\theta, \tilde{I}_0) = D_n(\theta; Z_n)' \tilde{I}_0^{-1} D_n(\theta; Z_n) \ , \tag{3.17}$$

$\tilde{I}_0$ 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,

$$H_n(\hat{\theta}_n^0; Z_n)' \tilde{I}_0^{-1} D_n(\hat{\theta}_n^0; Z_n) = P(\hat{\theta}_n^0)' \hat{\lambda}_n \,, \tag{3.18}$$

we see that

$$S(\psi) = LM(\psi) \ , \tag{3.19}$$

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 :

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

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,

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

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

$$\underset{n\to\infty}{\text{plim}}\, W_n = W_0 \ , \quad \det(W_0) \neq 0.$$

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

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

where

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

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

$$PC\big(\tilde{\theta}_n^0; \psi, W_n\big) = n\, D_n\big(\tilde{\theta}_n^0; Z_n\big)' \tilde{Q}[W_n]' \left\{\tilde{Q}[W_n]\, \tilde{I}_0 \tilde{Q}[W_n]'\right\}^{-1} \tilde{Q}[W_n]\, D_n\big(\tilde{\theta}_n^0; Z_n\big) \tag{3.20}$$

where $\tilde{I}_0 = \hat{I}(\tilde{\theta}_n^0)$. Under general regularity conditions, the asymptotic distribution of $PC\big(\tilde{\theta}_n^0; \psi, W_n\big)$ is $\chi^2(p_1)$ under H_0. Indeed, in the following proposition, we provide the asymptotic distribution of an even more general statistic of the form:

$$\overline{PC}(\tilde{\theta}_n^0; \psi, W_n) = n\, D_n\big(\tilde{\theta}_n^0; Z_n\big)' \tilde{Q}[W_n]' \left\{\tilde{Q}[W_n]\, \tilde{I}_0 \tilde{Q}[W_n]'\right\}^{-} \tilde{Q}[W_n]\, D_n\big(\tilde{\theta}_n^0; Z_n\big)\ . \tag{3.21}$$

where the matrices $\tilde{Q}[W_n]$ and $P(\theta_0)$ may not have full row rank and the derivative of $D_n(\theta; Z_n)$ need not be continuous.

Proposition 3.1 ASYMPTOTIC DISTRIBUTION OF GENERALIZED $C(\alpha)$ STATISTIC. *Under the assumptions* **A.1** *to* **A.13** *of Appendix A, let* $\tilde{Q}_n \equiv \tilde{Q}[W_n] = \tilde{P}_n[\tilde{J}_n' W_n \tilde{J}_n]^{-} \tilde{J}_n' W_n$ *where* $\tilde{J}_n = H_n(\tilde{\theta}_n^0)$, $\tilde{P}_n = P(\tilde{\theta}_n^0)$ *and* $[\,\cdot\,]^{-}$ *refers to any generalized inverse. Then*

$$\sqrt{n}\, \tilde{Q}_n D_n(\tilde{\theta}_0; Z_n) \xrightarrow[n\to\infty]{L} N\left[0,\, Q(\theta_0) I(\theta_0) Q(\theta_0)'\right] \tag{3.22}$$

where $Q(\theta_0) = P(\theta_0)\left[J(\theta_0)' W_0 J(\theta_0)\right]^{-1} J(\theta_0)' W_0$ *, and*

$$\overline{PC}(\tilde{\theta}_n^0; \psi, W_n) = n D_n\big(\tilde{\theta}_n^0; Z_n\big)' \tilde{Q}_n' \left[\tilde{Q}_n \tilde{I}_{0n} \tilde{Q}_n'\right]^{-} \tilde{Q}_n D_n\big(\tilde{\theta}_n^0; Z_n\big) \xrightarrow[n\to\infty]{L} \chi^2(r_1) \tag{3.23}$$

where $Q(\theta_0) = P(\theta_0)\left[J(\theta_0)' W_0 J(\theta_0)\right]^{-1} J(\theta_0)' W_0,\ r_1 = \mathrm{rank}[P(\theta_0)]$. *Further, the matrix* $\tilde{Q}_n' \left[\tilde{Q}_n \tilde{I}_{0n} \tilde{Q}_n'\right]^{-} \tilde{Q}_n$ *is invariant to the choice of the generalized inverse* $\left[\tilde{Q}_n \tilde{I}_{0n} \tilde{Q}_n'\right]^{-}$.

The proof of the latter proposition appears in Appendix A. Note that the numerical value of the statistic $\overline{PC}(\tilde{\theta}_n^0; \psi, W_n)$ is invariant to the choice of the generalized inverse $\left(\tilde{Q}[W_n]\, \tilde{I}_0 \tilde{Q}[W_n]'\right)^{-}$. It is clear $\overline{PC}(\tilde{\theta}_n^0; \psi, W_n)$ includes as special cases various other $C(\alpha)$-type statistics proposed in the statistical and econometric literatures.[1] In the sequel, we shall concentrate our discussion on nonsingular cases, where rank $[P(\theta_0)] = p_1$.

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 (3.14). When the number of equations equals the number of parameters $(m = p)$,

[1]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)

we have $\tilde{Q}\left[W_n\right] = \tilde{P}_0 \tilde{J}_0^{-1}$ and $PC(\tilde{\theta}_n^0; \psi, W_n)$ does not depend on the choice of W_n :

$$
\begin{aligned}
PC(\tilde{\theta}_n^0; \psi, W_n) &= PC\big(\tilde{\theta}_n^0; \psi\big) \\
&= D_n\big(\tilde{\theta}_n^0; Z_n\big)'(\tilde{J}_0^{-1})'\tilde{P}_0'\big[\tilde{P}_0\big(\tilde{J}_0'\tilde{I}_0^{-1}\tilde{J}_0\big)^{-1}\tilde{P}_0'\big]^{-1}\tilde{P}_0\tilde{J}_0^{-1}D_n\big(\tilde{\theta}_n^0; Z_n\big)\,. \quad (3.24)
\end{aligned}
$$

In particular, this will be the case if $D_n(\theta; Z_n)$ is the derivative vector of a (pseudo) log-likelihood function. Finally, when $m \geq p$, but $\tilde{\theta}_n^0$ 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

$$\Theta_0 = \{\theta \in \Omega \mid \psi(\theta) = 0\} \quad (4.1)$$

and Ψ the set of differentiable function $\bar{\psi} : \Omega \to \mathbb{R}^m$ such that

$$\big\{\theta \in \Omega \mid \bar{\psi}(\theta) = 0\big\} = \Theta_0\,. \quad (4.2)$$

A test statistic is invariant with respect to Ψ if it is the same for all $\psi \in \Psi$. It is obvious the LR-type statistics $LR(\psi)$ and D_{NW} (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 $\psi \in \Psi$ is used to obtain it. However, $\hat{P}$, $\hat{\lambda}_n$ and $\psi\big(\hat{\theta}_n\big)$ change. Following Dagenais and Dufour (1991), if $\bar{\psi} \in \Psi$, we have:

$$
\begin{aligned}
\bar{P}(\theta) &= \frac{\partial \bar{\psi}}{\partial \theta'} = \bar{P}_1(\theta)\, G(\theta)\ , \quad (4.3) \\
P(\theta) &= \frac{\partial \psi}{\partial \theta'} = P_1(\theta)\, G(\theta)\ , \quad (4.4)
\end{aligned}
$$

where $\bar{P}_1$ and P_1 are two squared 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*

$$\psi(\theta) = 0 \Longleftrightarrow \bar{\psi}(\theta) = 0, \forall \psi, \ \bar{\psi} \in \Psi . \tag{4.5}$$

Then,

$$T(\psi) = T\left(\bar{\psi}\right) \tag{4.6}$$

where T stands for any one of the test statistics $S(\psi)$, $LM(\psi)$, $PC(\hat{\theta}_n^0; \psi)$, $LR(\psi)$, $D_{NW}(\psi)$ and $PC(\hat{\theta}_n^0; \psi, W_n)$ defined in (3.12) to (3.20).

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 $\Omega \subseteq \mathbb{R}^p$ onto $\Omega_* \subseteq \mathbb{R}^p : \theta_* = \bar{g}(\theta)$. $\bar{g}$ 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 $\bar{g}$:

$$k(\theta_*) = \bar{g}^{-1}(\theta_*) = \theta . \tag{4.7}$$

Set

$$G(\theta) = \frac{\partial \bar{g}'}{\partial \theta} \text{ and } K(\theta_*) = \frac{\partial k}{\partial \theta_*'} . \tag{4.8}$$

Since $k[\bar{g}(\theta)] = \theta$ and $\bar{g}[k(\theta_*)] = \theta_*$, we have:

$$K[\bar{g}(\theta)]\, G(\theta) = I_p \text{ and } G[k(\theta_*)]\, K(\theta_*) = I_p \ , \ \forall \theta_* \in \Omega_* \ , \ \forall \theta \in \Omega . \tag{4.9}$$

Let

$$\psi^*(\theta_*) = \psi\left[\bar{g}^{-1}(\theta_*)\right] . \tag{4.10}$$

Clearly,

$$\psi^*(\theta_*) = 0 \Leftrightarrow \psi(\theta) = 0 , \tag{4.11}$$

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 as $\psi^*(\theta_*) = 0$, we will say it is *canonically invariant*.

By the invariance property of Proposition 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

$$P_*(\theta_*) \equiv \frac{\partial \psi^*}{\partial \theta'_*} = \frac{\partial \psi}{\partial \theta'} \frac{\partial \theta}{\partial \theta'_*} = P[k(\theta_*)] K(\theta_*) = P(\theta) K[\bar{g}(\theta)] \ . \tag{4.12}$$

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 $D_n(\theta; Z_n) = \sum_{t=1}^{n} h(\theta; z_t)/n$ as in (3.3) where the values of the scores are unaffected by the reparameterization, but are simply reexpressed in term of θ_* and z_{t*} (*invariant scores*):

$$h_t(\theta_*; z_{t*}) = h_t(\theta; z_t) \ , \quad t = 1, \ldots, n \ , \tag{4.13}$$

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 (4.13), we see easily that

$$H_{n*}(\theta_*; Z_{n*}) = \frac{\partial D_{n*}(\theta_*; Z_{n*})}{\partial \theta'_*} = H_n(\theta; Z_n) K(\theta_*) = H_n(\theta; Z_n) K[\bar{g}(\theta)] \ . \tag{4.14}$$

Further the functions $\hat{I}(\theta)$ and $\hat{J}(\theta)$ in (3.4) - (3.5) are then transformed in the following way :

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

If $\hat{I}(\theta)$ and $\hat{J}(\theta)$ are defined as (3.4) - (3.5), if $W_{n*} = W_n$ and if $\tilde{\theta}_n^0$ 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 (3.20) 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 (3.20)*, and $\tilde{K} = K(\tilde{\theta}_{n*}^0)$ is invertible. Then*

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

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

$\partial\psi^* / \partial\theta'_*$.

It is clear the estimators $\hat{\theta}_n$ and $\hat{\theta}_n^0$ satisfy the equivariance condition, i.e., $\hat{\theta}_{n*} = \bar{g}(\hat{\theta}_n)$ and $\hat{\theta}_{n*}^0 = \bar{g}(\hat{\theta}_n^0)$. 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 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 GERERAL HYPOTHESIS REFORMULATIONS: INVARIANT SCORE CASE. *Let $\psi_* : \Omega_* \to \Omega$ be any continuously differentiable function of $\theta_* \in \Omega_*$ 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

$$T(\psi) = T_*(\psi_*) \tag{4.15}$$

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

$$PC_*(\tilde{\theta}_{n*}^0; \psi_*) = PC(\tilde{\theta}_n^0; \psi) \ . \tag{4.16}$$

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

Cases where (4.14) 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 :

$$D_n(\theta; Z_n) = \frac{1}{n}\frac{\partial S_n(\theta; Z_n)}{\partial\theta} \ . \tag{4.17}$$

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

$$D_{n*}(\theta_*; Z_{n*}) = K(\theta_*)' D_n(\theta; Z_n) = K[\bar{g}(\theta)]' D_n(\theta; Z_n) \ . \tag{4.18}$$

From (2.3) and (4.18), it then follows that

$$\sqrt{n}\, D_{n*}(\theta_{0*}; Z_{n*}) \xrightarrow[n\to\infty]{D} N[0, I_*(\theta_{0*})] \tag{4.19}$$

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

$$I_*(\theta_*) = K(\theta_*)' I[k(\theta_*)] K(\theta_*) = K[\bar{g}(\theta)]' I(\theta) K[\bar{g}(\theta)] \ . \tag{4.20}$$

Further,

$$H_{n*}(\theta_*; Z_n) = K[\bar{g}(\theta)]' H_n(\theta; Z_n) K[\bar{g}(\theta)] + \sum_{i=1}^{p} D_{ni}(\theta; Z_n) K_{i\cdot}^{(1)}[\bar{g}(\theta)] \tag{4.21}$$

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

$$K_{i\cdot}^{(1)}(\theta_*) = \frac{\partial^2 \theta_i}{\partial \theta_* \partial \theta_*'}(\theta_*) = \frac{\partial^2 k_i}{\partial \theta_* \partial \theta_*'}(\theta_*) \ . \tag{4.22}$$

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

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

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

where $\tilde{I}_0$, $\tilde{J}_0$ and W_n are defined as in (3.20)*, 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 $\psi_* : \Omega_* \to \Omega$ be any continuously differentiable function of $\theta_* \in \Omega_*$ 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) $\hat{J}_*[\bar{g}(\theta)] = K[\bar{g}(\theta)]' \hat{J}(\theta) K[\bar{g}(\theta)]$, *where* $K(\theta_*) = \partial \bar{g}^{-1}(\theta) / \partial \theta_*'$ *. Then, provided the relevant matrices are invertible, we have*

$$T(\psi) = T_*(\psi_*) \tag{4.23}$$

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

$$PC_*\big(\tilde{\theta}_{n*}^{0}; \psi_*\big) = PC\big(\tilde{\theta}_{n}^{0}; \psi\big) , \tag{4.24}$$

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

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

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

$$h_{t*}\big(\bar{g}(\theta); z_{t*}\big) = K[\bar{g}(\theta)]' \, h_t(\theta; z_t) \ , \ t = 1, , \ \ldots \ , n, \tag{4.25}$$

where $D_{n*}\big(\bar{g}(\theta); Z_{n*}\big) = \frac{1}{n}\sum_{t=1}^{n} h_{t*}\big(\bar{g}(\theta); z_{t*}\big)$. Consequently, in such a case, any estimator $\hat{I}(\theta)$ of the general form (3.6) 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 (3.9) as soon as the bandwidth parameter S_n only depends on the sample size n. However, this may not hold if S_n is data-dependent [as considered in Andrews and Monahan (1992)].

5. Invariant test criteria

Despite the apparent "positive nature" of the invariance results presented in the previous section, the main conclusion is that none of the proposed test statistics is invariant to general reparameterizations, especially when the score-type function considered is derived from an objective function.[2] In particular, this problem will occur when the score-type function is derived from a (pseudo) likelihood function or, more generally, from the objective function minimized by an M-estimator.

In this section, we propose two ways of doing this. 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.

5.1. Modified Newey–West LR-type statistic

Consider the LR-type statistic

$$D_{NW} = n\big[M_n\big(\hat{\theta}_{n}^{0}, \tilde{I}_0\big) - M_n\big(\hat{\theta}_n, \tilde{I}_0\big)\big]$$

[2]The reader may note that further insight may be gained on the invariance properties of test staistics by using differential geometry arguments; for some applications to statistical problems, see Critchley, Marriott and Salmon (1996), Kass and Voss (1997) and Marriott and Salmon (2000). However, this would go beyond the scope of the present paper.

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 $\theta : \tilde{I}_0 = \hat{I}\left(\bar{\theta}_n\right)$. 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 $M_n\big(\theta, \hat{I}(\theta)\big)$. For example, such an estimation methods 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 $D_{NW} = nM_n(\hat{\theta}_n^0, \tilde{I}_0)$. When $m > p$, this will typically not be the case.

Suppose now the following conditions hold :

$$\begin{aligned} D_{n*}(\bar{g}(\theta), Z_{n*}) &= K[\bar{g}(\theta)]' D_n(\theta; Z_n) \ , && (5.1)\\ \hat{I}_*(\bar{g}(\theta)) &= K[\bar{g}(\theta)]' \hat{I}(\theta) K[\bar{g}(\theta)] \ . && (5.2) \end{aligned}$$

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

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

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

$$\bar{D} = n\big[M_n\big(\tilde{\theta}_n^0; \hat{I}(\tilde{\theta}_n^0)\big) - M_n\big(\tilde{\theta}_n; \hat{I}(\tilde{\theta}_n)\big)\big] \tag{5.4}$$

is invariant to reparameterizations of the type considered in (4.18) - (4.20). Under standard regularity conditions 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. For completeness, we state this result in the following proposition. The proof is provided in Appendix B.

Proposition 5.1 ASYMPTOTIC DISTRIBUTION OF MODIFIED NEWEY-WEST STATISTIC. *Under the assumptions* **A.8**, **A.9** *and* **B.1** *to* **B.7** [*stated in Appendices A and B*] *with* $r_1 = \mathrm{rank}[P(\theta_0)] = p_1$, *the statistic* $D_{NW} = n\big[M_n(\hat{\theta}_n^0) - M_n(\hat{\theta}_n)\big]$ *converges in distribution to a* $\chi^2(p_1)$ *when* $\psi(\theta_0) = 0$.

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 to an exogenous random vector x_t :

$$E\left(y_t \mid x_t\right) = f\left(x_t; \theta\right) \equiv f_t\left(\theta\right)\ ,\ V\left(y_t \mid x_t\right) = \Omega_0(x_t) \tag{5.5}$$

where $f_t(\theta)$ is a known function and θ is the parameter of interest. (5.5) 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 technique (PML) which consists in maximizing a chosen likelihood as it where the true undefined likelihood; see Gouriéroux, Monfort and Trognon (1984*c*).[3] 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:

$$l(y;\mu) = \exp\left[A(\mu) + B(y) + C(\mu)y\right] \tag{5.6}$$

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

$$\prod_{t=1}^{n} \exp\left\{A(f_t(\theta)) + B(y_t) + C[f_t\left(\theta\right)]y_t\right\} \tag{5.7}$$

or equivalently through the following equivalent programme:

$$\underset{\theta}{\text{Max}}\ \sum_{t=1}^{n}\left\{A[f_t\left(\theta\right)] + C[f_t\left(\theta\right)]y_t\right\}\ \text{with}\ \frac{\partial A}{\partial \mu} + \frac{\partial C}{\partial \mu}\,\mu = 0\ . \tag{5.8}$$

The class of linear exponential distributions contains most of the classical statistical models, such as the Gaussian model the Poisson model, the Binomial model, the Gamma model, the negative Binomial model, etc. The constraint in the programme (5.8) ensures that the expectation of the linear exponential pseudo-distribution is μ. The pseudo-likelihood equations have an orthogonal

[3]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).

.

condition form:

$$D_n(\theta) = \sum_{t=1}^{n} \frac{\partial f_t'}{\partial \theta} \frac{\partial C}{\partial \mu} (f_t(\theta))(y_t - f_t(\theta)) = 0 \, . \tag{5.9}$$

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:

$$J(\theta) \;=\; E_x \left\{ \left(\frac{\partial f_t'}{\partial \theta} \right) \left[\frac{\partial C}{\partial \mu} \big(f_t(\theta)\big) \right] \left(\frac{\partial f_t'}{\partial \theta} \right)' \right\} \, , \tag{5.10}$$

$$I(\theta) \;=\; E_x \left\{ \left(\frac{\partial f_t'}{\partial \theta} \right) \left[\left(\frac{\partial C}{\partial \mu} \big(f_t(\theta)\big) \right) \Omega_0 \left(\frac{\partial C}{\partial \mu} \big(f_t(\theta)\big) \right) \right] \left(\frac{\partial f_t'}{\partial \theta} \right)' \right\} \, . \tag{5.11}$$

These matrices can be estimated by :

$$\hat{J} \;=\; \frac{1}{n} \sum_{t=1}^{n} \left(\frac{\partial f_t'}{\partial \theta} (\hat{\theta}) \right) \left[\frac{\partial C}{\partial \mu} \big(f_t(\hat{\theta})\big) \right] \left(\frac{\partial f_t'}{\partial \theta} (\hat{\theta}) \right)' , \tag{5.12}$$

$$\hat{I} \;=\; \frac{1}{n} \sum_{t=1}^{n} S_t\big(\hat{\theta}\big) S_t\big(\hat{\theta}\big)' , \tag{5.13}$$

where

$$S_t\big(\hat{\theta}\big) = \left(\frac{\partial f_t'}{\partial \theta} (\hat{\theta}) \right) \left[\frac{\partial C}{\partial \mu} \big(f_t(\hat{\theta})\big) \right] \big(y - f_t(\hat{\theta})\big) \, . \tag{5.14}$$

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,

$$f_t^*(\theta_*) = f_t^*[\bar{g}(\theta)] = f(\theta) \, , \quad \frac{\partial f_t^*}{\partial \theta_*'} = \left(\frac{\partial f_t^*}{\partial \theta'} \right) \left(\frac{\partial \theta}{\partial \theta_*'} \right) = \left(\frac{\partial f_t}{\partial \theta'} \right) K[\bar{g}(\theta)] \tag{5.15}$$

and

$$\hat{I}_* = K\big[\bar{g}(\hat{\theta})\big]' \hat{I} K\big[\bar{g}(\hat{\theta})\big] \, , \quad \hat{J}_* = K\big[\bar{g}(\hat{\theta})\big]' \hat{J} K\big[\bar{g}(\hat{\theta})\big] . \tag{5.16}$$

The Lagrange, 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:

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

$$-\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 f_t}{\partial\theta\partial\theta'}(\hat{\theta})\frac{\partial C}{\partial\mu}(f_t(\hat{\theta}))(y_t - f_t(\hat{\theta}))$$

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 leads to the same non-invariance problems bears the same defaults as those presented in (3.5) and (4.21). The two last terms of $\tilde{J}$ 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. 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:

$$V(y_t|x_t) = \Omega_0(x_t) = h(x_t, \alpha_0) = h_t(\alpha_0)\,.$$

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:

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

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 the problem

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

The QGPML estimator $\hat{\theta}$ of θ is strongly consistent and asymptotically normal: $\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{L} N[0,\,K]$ with

$$K = \left\{E_x\frac{\partial f_t'}{\partial\theta}g_t(\alpha_0)^{-1}\frac{\partial f_t}{\partial\theta'}\right\}^{-1},$$

$$I_0 = J_0 = E_x \left[\frac{\partial f_t'}{\partial \theta}(\theta_0) g_t(\alpha_0)^{-1} \frac{\partial f_t}{\partial \theta'}(\theta_0) \right] .$$

I_0 and J_0 can be consistently estimated by:

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

where

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

Since $\frac{\partial C}{\partial \mu}(f_t(\hat{\theta}), \Psi(f_t(\tilde{\theta}), g_t(\tilde{\alpha})))$, 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:

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

and the Wald, Lagrange, 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 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:

$$y_t = \gamma + \beta_1 x_{1t}^{(\lambda)} + \beta_2 x_{2t}^{(\lambda)} + u_t \, , \tag{6.1}$$

$$u_t \overset{i.i.d.}{\sim} N[0, \, \sigma^2] \, , \; t = 1, \, \dots \, , \, T, \tag{6.2}$$

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

$$H_0 : \lambda = 1 \, . \tag{6.3}$$

The log-likelihood associated with this model is:

$$l = \sum_{t=1}^{T} l[y_t;\, \gamma,\, \beta_1, \beta_2,\, \lambda,\, \sigma^2]\,, \tag{6.4}$$

$$l[y_t;\, \gamma,\, \beta_1, \beta_2,\, \lambda,\, \sigma^2] = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}u_t^2\,,\ t = 1,\, \ldots\,,\, T. \tag{6.5}$$

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

$$x_{1t*} = kx_{1t}, \quad x_{2t*} = kx_{2t}, \tag{6.6}$$

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

$$_t = \gamma_* + \beta_{1*}x_{1t*}^{(\lambda)} + \beta_{2*}x_{2t*}^{(\lambda)} + u_t\,, \tag{6.7}$$

where the power parameter λ remains the same and

$$\gamma_* = \gamma - k^{(\lambda)}k^{-\lambda}\sum_{1=2}^{2}\beta_i\,, \quad \beta_{i*} = \beta_i k^{-\lambda},\ i = 1,\, 2. \tag{6.8}$$

On interpreting model (6.1) - (6.2) as a pseudo-model and (6.4) 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:

$$\mathsf{E}\left[\sum_{t=1}^{T} u_t\right] = 0\,, \tag{6.9}$$

$$\mathsf{E}\left[\sum_{t=1}^{T} u_t x_{1t}^{(\lambda)}\right] = 0\,, \tag{6.10}$$

$$\mathsf{E}\left[\sum_{t=1}^{T} u_t x_{2t}^{(\lambda)}\right] = 0\,, \tag{6.11}$$

$$\mathsf{E}\left[\sum_{t=1}^{T} \frac{u_t}{\lambda}\left(\sum_{i=1}^{2}\beta_i x_{it}^{\lambda}\ln x_{it} - x_{it}^{(\lambda)}\right)\right] = 0\,, \tag{6.12}$$

$$\mathsf{E}\left[\sum_{t=1}^{T}\left(u_t^2 - \sigma^2\right)\right] = 0\,. \tag{6.13}$$

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

$$\mathsf{E}\left[\sum_{t=1}^{T} u_t x_{1t} x_{2t}\right] = 0\,. \tag{6.14}$$

To get data, we considered the sample size $T = 200$ and generated y_t according to equation (6.1) - (6.2) 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).[4]

Numerical values of the GMM-based test statistics for a number of rescalings are reported in Table 1 for the 5 moment system (6.9) - (6.13) and in Table 2 for the 6 moment system (6.9) - (6.14). 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 $M_n\,(\theta,\,W_n)$ in (2.5) with no weighting $(W_n = I_m)$, 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 [D_{NW}, 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 D_{NW} 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.

7. 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 ob-

[4]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.

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

	Two-step GMM				CUP-GMM				Pseudo ML	
k	D_{NW}	Wald	Score	$C(\alpha)$	D_{NW}	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

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

	Two-step GMM				CUP-GMM			
k	D_{NW}	Wald	Score	$C(\alpha)$	D_{NW}	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

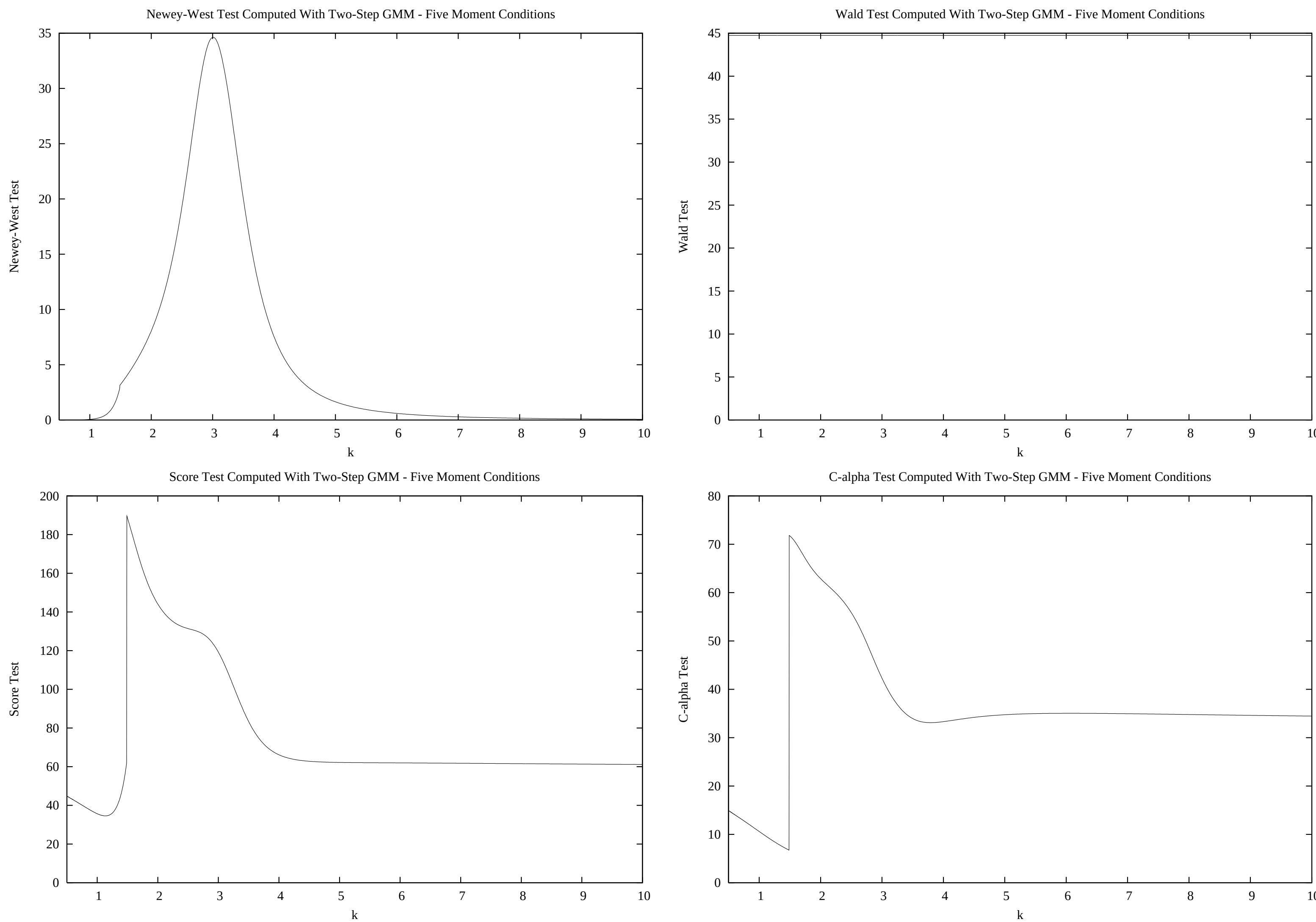

Figure 1. Two-step GMM tests based on 5 moment conditions

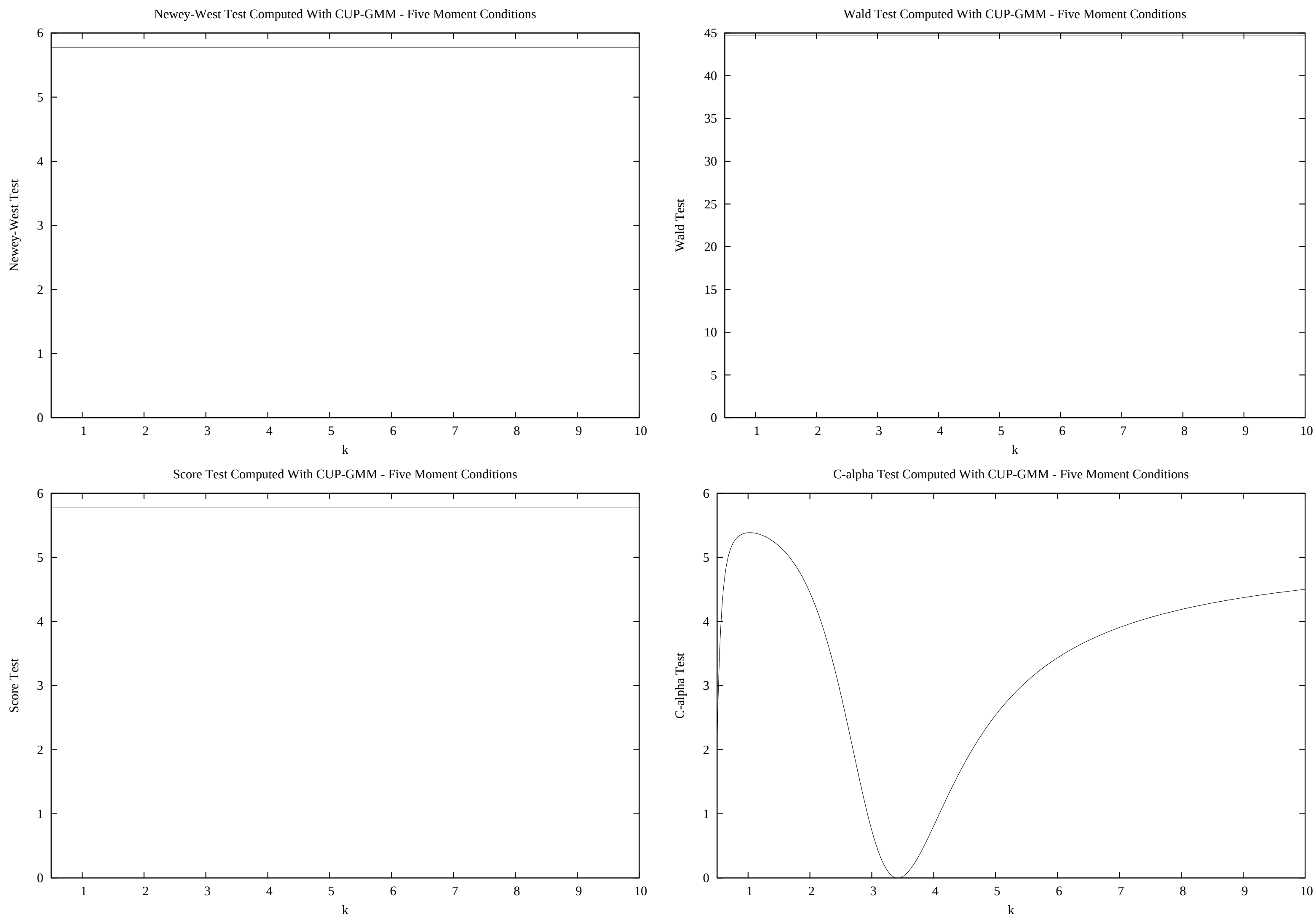

Figure 2. CUP GMM tests based on 5 moment conditions

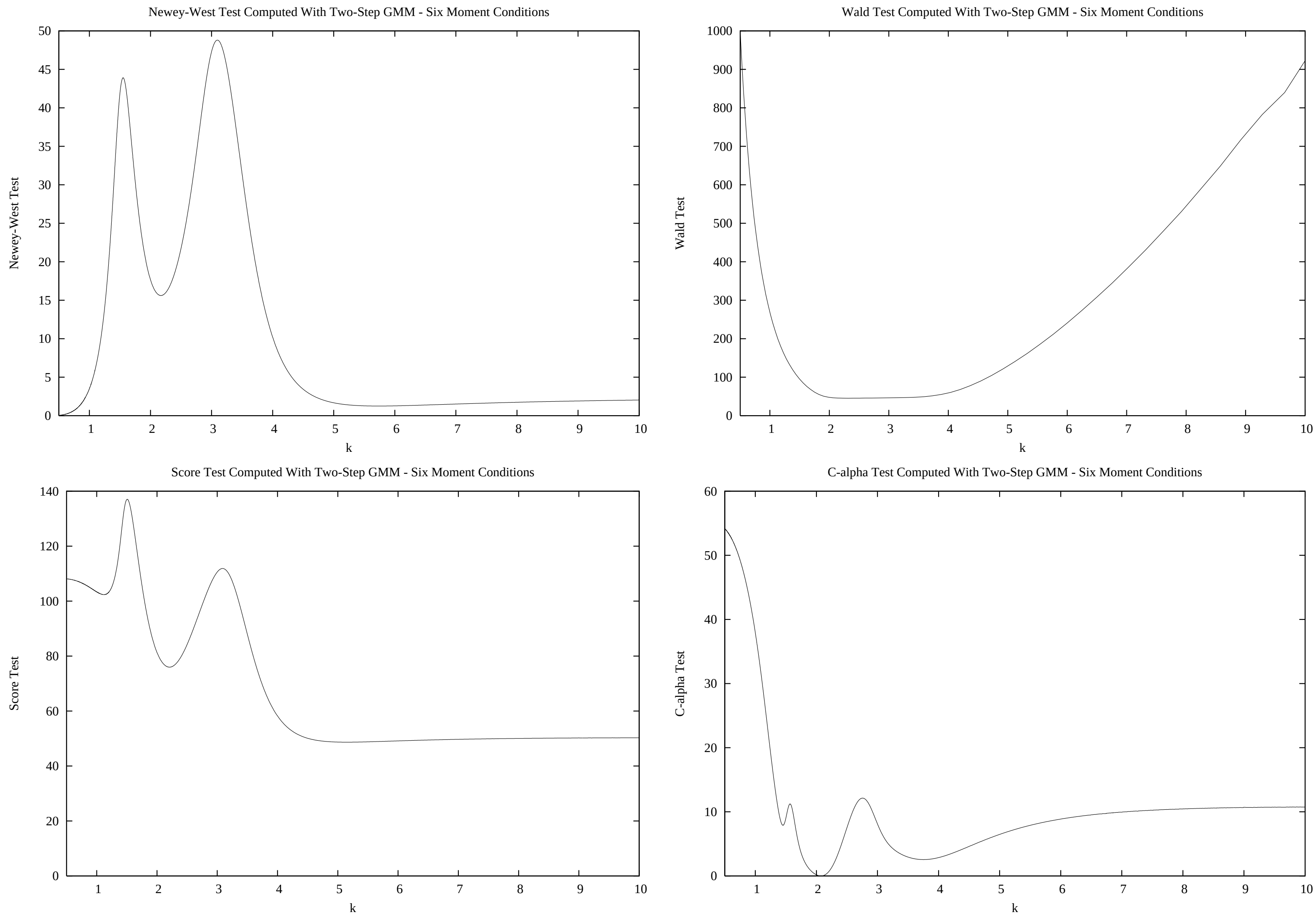

Figure 3. Two-step GMM tests based on 6 moment conditions

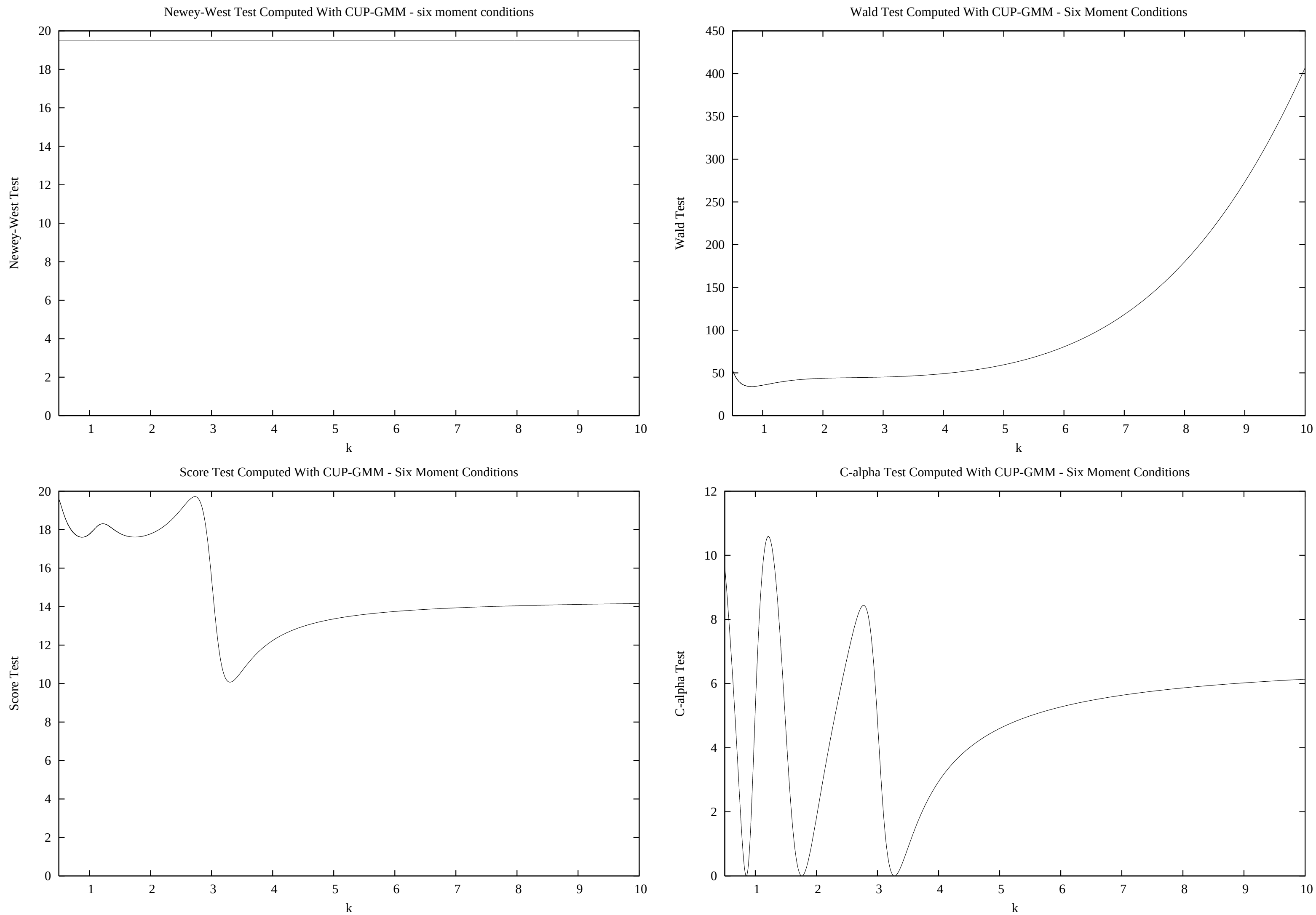

Figure 4. CUP GMM tests based on 6 moment conditions

tain 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: Distribution of the generalized $C(\alpha)$ statistic

In this appendix we derive the asymptotic distribution of the generalized $C(\alpha)$ statistic defined in (3.20) 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.

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

is a sequence of $m \times 1$ *random vectors, defined on a common probability space* $(\mathcal{Z},\, \mathcal{A}_{\mathcal{Z}},\, \mathsf{P})$, *which are functions of a* $p \times 1$ *parameter vector* θ, *where* $\theta \in \Omega \subseteq \mathbb{R}^p$ *and* Ω *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* $\theta_0 \in \Omega$ *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 SCORE ASYMPTOTIC NON-SINGULARITY. $I(\theta_0)$ *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* $\omega \in \mathcal{D}_H$ *and* $\mathcal{D}_H$ *is an event with probability one* (i.e., $\mathsf{P}[\omega \in \mathcal{D}_H] = 1$).

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* $\delta > 0$,

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

where

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

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

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 $\omega \in \mathcal{D}_H$, where the residue $R_n(\theta,\, \theta_0,\, \omega)$ 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

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

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

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 and* $\mathrm{rank}[P(\theta)] = r_1$ *for all $\theta \in N_5$, where $P(\theta) \equiv \dfrac{\partial\psi}{\partial\theta'}$ and $0 \le r_1 \le p_1$.*

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

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

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

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

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

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

Assumption A.12 CONSISTENT ESTIMATOR OF SCORE COVARIANCE MATRIX. *$\widetilde{I}_{0n}$ is a weakly consistent estimator of $I(\theta_0)$,* i.e., $\operatorname*{plim}_{n\to\infty} \widetilde{I}_{0n} = I(\theta_0)\,.$

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

PROOF OF PROPOSITION **3.1** To simplify notation, we shall assume all along that $\omega \in \mathcal{D}_H$ (an event with probability 1) and drop the symbol ω from the random variables considered. In order to obtain the asymptotic null distribution of the generalized $C(\alpha)$ statistic defined in (3.23), we first need to show that $P(\tilde{\theta}_n^0)$ and $H_n(\tilde{\theta}_n^0)$ converge to $P(\theta_0)$ and $J(\theta_0)$ respectively. The consistency of $P(\tilde{\theta}_n^0)$, *i.e.*

$$\operatorname*{plim}_{n\to\infty} [P(\tilde{\theta}_n^0) - P(\theta_0)] = 0\,, \tag{A.1}$$

follows simply from the consistency of $\tilde{\theta}_n^0$ [Assumption **A.10**] and the continuity of $P(\theta)$ at θ_0 [Assumption **A.9**]. Further, since since $P(\theta)$ is continuous in open neighborhood of θ_0, we also have

$$\text{rank}\,[\tilde{P}_n] \xrightarrow[n\to\infty]{\mathsf{p}} \text{rank}\,[P(\theta_0)] = r_1\,. \tag{A.2}$$

Consider now $H_n(\tilde{\theta}_n^0)$. By the assumptions **A.4 - A.5** and **A.10**, for any $\varepsilon > 0$ and $\varepsilon_1 > 0$, we can choose $\delta_1 \equiv \delta(\varepsilon_1,\, \varepsilon) > 0$ and a positive integer $n_1(\varepsilon,\, \delta_1)$ such that: (i) $U_H(\delta_1,\, \varepsilon,\, \theta_0) \leq \varepsilon_1/2$, and (ii) $n > n_1(\varepsilon,\, \delta_1)$ entails

$$\mathsf{P}\,[\Delta_n(\theta_0,\, \delta) > \varepsilon] \equiv \mathsf{P}\big[\{\omega : \Delta_n(\theta_0,\, \delta,\, \omega) > \varepsilon\}\big] \leq U_H(\delta_1,\, \varepsilon,\, \theta_0) \leq \varepsilon_1/2\,.$$

Further, by the consistency of $\tilde{\theta}_n^0$ [Assumption **A.10**], we can choose $n_2(\varepsilon,\, \delta_1)$ such that $n > n_2(\varepsilon,\, \delta_1)$ entails $\mathsf{P}\big[\|\tilde{\theta}_n^0 - \theta_0\| \leq \delta_1\big] \geq 1 - (\varepsilon_1/2)$. Then, for $n > \max\{n_1(\varepsilon,\, \delta_1),\, n_2(\varepsilon,\, \delta_1)\}$, we have, using the Boole-Bonferroni inequality,

$$\begin{aligned}
\mathsf{P}\big[\|H_n(\tilde{\theta}_n^0) - J(\theta_0)\| \leq \varepsilon\big] \quad &\geq \mathsf{P}\big[\|\tilde{\theta}_n^0 - \theta_0\| \leq \delta_1 \text{ and } \|H_n(\tilde{\theta}_n^0) - J(\theta_0)\| \leq \varepsilon\big] \\
&\geq \mathsf{P}\big[\|\tilde{\theta}_n^0 - \theta_0\| \leq \delta_1 \text{ and } \Delta_n\,(\theta_0,\, \delta_1) \leq \varepsilon\big] \\
&\geq 1 - \mathsf{P}\big[\|\tilde{\theta}_n^0 - \theta_0\| > \delta_1\big] - \mathsf{P}\big[\Delta_n\,(\theta_0,\, \delta_1) > \varepsilon\big] \\
&\geq 1 - (\varepsilon_1/2) - (\varepsilon_1/2) = 1 - \varepsilon_1\,.
\end{aligned}$$

Thus,

$$\liminf_{n\to\infty} \mathsf{P}\big[\|H_n(\tilde{\theta}_n^0) - J(\theta_0)\| \leq \varepsilon\big] \geq 1 - \varepsilon_1\,, \text{ for all } \varepsilon > 0,\ \varepsilon_1 > 0\,,$$

hence

$$\lim_{n\to\infty} \mathsf{P}\big[\|H_n(\tilde{\theta}_n^0) - J(\theta_0)\| \leq \varepsilon\big] = 1\,, \text{ for all } \varepsilon > 0\,,$$

or, equivalently,

$$\operatorname*{plim}_{n\to\infty} \big[H_n(\tilde{\theta}_n^0) - J(\theta_0)\big] = 0\,. \tag{A.3}$$

By Assumption **A.6**, we can write [setting $0/0 = 0$] :

$$\begin{aligned}
\|\sqrt{n}\,\big[D_n(\tilde{\theta}_n^0) - D_n(\theta_0)\big] - J(\theta_0)\sqrt{n}\,(\tilde{\theta}_n^0 - \theta_0)\| \quad &= \quad \sqrt{n}\,\|R_n(\tilde{\theta}_n^0,\, \theta_0)\| \\
&= \quad \frac{\|R_n(\tilde{\theta}_n^0,\, \theta_0)\|}{\|\tilde{\theta}_n^0 - \theta_0\|}\sqrt{n}\,\|\tilde{\theta}_n^0 - \theta_0\|
\end{aligned}$$

where

$$\frac{\|R_n(\tilde{\theta}_n^0,\,\theta_0)\|}{\|\tilde{\theta}_n^0-\theta_0\|}\le r_n(\delta,\,\theta_0)\text{ when }\tilde{\theta}_n^0\in N_3\text{ and }\|\tilde{\theta}_n^0-\theta_0\|\le\delta$$

and $\limsup_{n\to\infty}\mathsf{P}\,[r_n(\delta,\,\theta_0)>\varepsilon]<U_D(\delta,\,\varepsilon,\,\theta_0)$. Thus, for any $\varepsilon>0$ and $\delta>0$, we have:

$$\begin{aligned}\mathsf{P}\left[\frac{\|R_n(\tilde{\theta}_n^0,\,\theta_0)\|}{\|\tilde{\theta}_n^0-\theta_0\|}\le\varepsilon\right] &\ge \mathsf{P}\big[r_n(\delta,\,\theta_0)\le\varepsilon\,,\tilde{\theta}_n^0\in N_3\text{ and }\|\tilde{\theta}_n^0-\theta_0\|\le\delta\big]\\ &\ge 1-\mathsf{P}\big[r_n(\delta,\,\theta_0)>\varepsilon\,\big]-\mathsf{P}\big[\,\tilde{\theta}_n^0\notin N_3\text{ or }\|\tilde{\theta}_n^0-\theta_0\|>\delta\big]\end{aligned}$$

hence, using the consistency of $\tilde{\theta}_n^0$,

$$\begin{aligned}\liminf_{n\to\infty}\,\mathsf{P}\left[\|R_n(\tilde{\theta}_n^0,\,\theta_0)\|/\|\tilde{\theta}_n^0-\theta_0\|\le\varepsilon\right] &\ge 1-\,\limsup_{n\to\infty}\mathsf{P}\,[r_n(\delta,\,\theta_0)>\varepsilon\,]\\ &\quad-\limsup_{n\to\infty}\mathsf{P}\big[\,\tilde{\theta}_n^0\notin N_3\text{ or }\|\tilde{\theta}_n^0-\theta_0\|>\delta\big]\\ &\ge 1-U_D(\delta,\,\varepsilon,\,\theta_0)\,.\end{aligned}$$

Since $\lim_{\delta\downarrow 0}U_D(\delta,\,\varepsilon,\,\theta_0)\,=\,0$, it follows that $\lim_{n\to\infty}\mathsf{P}\big[\|R_n(\tilde{\theta}_n^0,\,\theta_0)\|/\|\tilde{\theta}_n^0-\theta_0\|\,\le\,\varepsilon\big]\,=\,1$ for any $\varepsilon>0$, or equivalently,

$$\|R_n(\tilde{\theta}_n^0,\,\theta_0)\|/\|\tilde{\theta}_n^0-\theta_0\|\;\xrightarrow[n\to\infty]{\mathsf{p}}\;0\,.$$

Since $\sqrt{n}\,(\tilde{\theta}_n^0-\theta_0)$ is asymptotically bounded in probability (by Assumption **A.10**), this entails:

$$\sqrt{n}\,\|R_n(\tilde{\theta}_n^0,\,\theta_0)\|=\frac{\|R_n(\tilde{\theta}_n^0,\,\theta_0)\|}{\|\tilde{\theta}_n^0-\theta_0\|}\sqrt{n}\,\|\tilde{\theta}_n^0-\theta_0\|\;\xrightarrow[n\to\infty]{\mathsf{p}}\;0$$

and

$$\|\sqrt{n}\,[D_n(\tilde{\theta}_n^0)-D_n(\theta_0)]-J(\theta_0)\sqrt{n}\,(\tilde{\theta}_n^0-\theta_0)\|\;\xrightarrow[n\to\infty]{\mathsf{p}}\;0\,. \tag{A.4}$$

By Taylor's theorem and assumptions **A.8 - A.9**, we also have the expansion:

$$\psi(\theta)=\psi(\theta_0)+P(\theta_0)(\theta-\theta_0)+R_2(\theta,\,\theta_0)\ , \tag{A.5}$$

for $\theta\in N_5$, where

$$\lim_{\theta\to\theta_0}\|R_2(\theta,\,\theta_0)\,\|/\|\,\theta-\,\theta_0\|=0\ ,$$

i.e., $R_2(\theta,\,\theta_0)=o(\|\theta-\,\theta_0\|)$, so that, using Assumption **A.11**,

$$\sqrt{n}\,P(\theta_0)(\tilde{\theta}_n^0-\theta_0)=\sqrt{n}\,[\psi(\tilde{\theta}_n^0)-\psi(\theta_0)]-\sqrt{n}\,R_2(\tilde{\theta}_n^0,\,\theta_0)=-\sqrt{n}\,R_2(\tilde{\theta}_n^0,\,\theta_0) \tag{A.6}$$

for $\tilde{\theta}_n^0 \in N_5$, and

$$\|\sqrt{n}\,P(\theta_0)(\tilde{\theta}_n^0 - \theta_0)\| = \frac{\|R_2(\tilde{\theta}_n^0 - \theta_0)\|}{\|\tilde{\theta}_n^0 - \theta_0\|}\sqrt{n}\,\|\tilde{\theta}_n^0 - \theta_0\| \xrightarrow[n\to\infty]{\mathsf{p}} 0\,. \tag{A.7}$$

By (A.1), (A.2) and (A.3) jointly with the assumptions **A.7**, **A.8**, **A.9**, **A.12** and **A.13**, we .have:

$$\text{rank}\,[\tilde{P}_n] \xrightarrow[n\to\infty]{\mathsf{p}} r_1\,, \quad \text{rank}\,[\tilde{J}_n] \xrightarrow[n\to\infty]{\mathsf{p}} p\,, \quad \text{rank}\,[\widetilde{I}_{0n}] \xrightarrow[n\to\infty]{\mathsf{p}} m, \quad \text{rank}\,[W_n] \xrightarrow[n\to\infty]{\mathsf{p}} p\,, \tag{A.8}$$

so that the probability that the matrices $\tilde{P}_n,\ \tilde{J}_n,\ \widetilde{I}_{0n}$, and W_n all have full row rank converges to one as $n \to \infty$; in particular, for $\tilde{J}_n,\ \widetilde{I}_{0n}$, and W_n, this follows on observing that $\tilde{J}_n,\ \widetilde{I}_{0n}$, and W_n each converge to matrices with determinants strictly greater than zero and from the continuity of determinants with respect to the elements of the corresponding matrices. Since $\operatorname*{plim}_{n\to\infty} \tilde{P}_n = P(\theta_0)$ and $\operatorname*{plim}_{n\to\infty} \tilde{J}_n = J(\theta_0)$, we can then write:

$$\begin{aligned} \operatorname*{plim}_{n\to\infty} \left[\tilde{J}_n W_n \tilde{J}_n'\right]^- &= \left[J(\theta_0)' W_0 J(\theta_0)\right]^{-1},\ \operatorname*{plim}_{n\to\infty} \tilde{Q}_n = Q(\theta_0)\,, \\ \operatorname*{plim}_{n\to\infty} \tilde{Q}_n \tilde{J}_n &= \operatorname*{plim}_{n\to\infty} \tilde{Q}_n J(\theta_0) = Q(\theta_0) J(\theta_0) = P(\theta_0)\,, \end{aligned}$$

where $\tilde{Q}_n \equiv \tilde{Q}[W_n] = \tilde{P}_n[\tilde{J}_n' W_n \tilde{J}_n]^- \tilde{J}_n' W_n$. Then, using (A.7) and (A.4), it follows that:

$$\begin{aligned} &\operatorname*{plim}_{n\to\infty} \left\{\sqrt{n}\,\tilde{Q}_n D_n(\tilde{\theta}_n^0) - Q(\theta_0)\sqrt{n}\,D_n(\theta_0)\right\} \\ &= \operatorname*{plim}_{n\to\infty} \left\{\sqrt{n}\,\tilde{Q}_n D_n(\tilde{\theta}_n^0) - Q(\theta_0)\sqrt{n}\,D_n(\theta_0) - P(\theta_0)\sqrt{n}\,(\tilde{\theta}_n^0 - \theta_0)\right\} \\ &= \operatorname*{plim}_{n\to\infty} \left\{\tilde{Q}_n\left[\sqrt{n}\,[D_n(\tilde{\theta}_n^0) - D_n(\theta_0)] - J(\theta_0)\sqrt{n}\,(\tilde{\theta}_n^0 - \theta_0)\right]\right\} \\ &\quad + \operatorname*{plim}_{n\to\infty} \left\{\left[\tilde{Q}_n - Q(\theta_0)\right]\sqrt{n}\,D_n(\theta_0) + \left[\tilde{Q}_n J(\theta_0) - P(\theta_0)\right]\sqrt{n}\,(\tilde{\theta}_n^0 - \theta_0)\right\} \\ &= \operatorname*{plim}_{n\to\infty} \left\{\tilde{Q}_n\left[\sqrt{n}\,[D_n(\tilde{\theta}_n^0) - D_n(\theta_0)] - J(\theta_0)\sqrt{n}\,(\tilde{\theta}_n^0 - \theta_0)\right]\right\} = 0\,. \end{aligned}$$

We conclude that the asymptotic distribution of $\sqrt{n}\,\tilde{Q}_n D_n(\tilde{\theta}_0)$ is the same as the one of $Q(\theta_0)\sqrt{n}\,D_n(\theta_0)$, namely by Assumption **A.2**, a $N\,[0, V_\psi(\theta_0)]$ distribution where

$$V_\psi(\theta) = Q(\theta) I(\theta) Q(\theta)'$$

has rank r_1 =rank$[Q(\theta_0)]$ =rank$[P(\theta_0)]$ in an open neighborhood of θ_0. Consequently, the estimator

$$\widetilde{V}_\psi(\tilde{\theta}_n^0) = \tilde{Q}_n \widetilde{I}_{0n} \tilde{Q}_n' \tag{A.9}$$

converges to $V_\psi(\theta_0)$ in probability and, by (A.8),

$$\text{rank}\,\left[\widetilde{V}_\psi(\tilde{\theta}_n^0)\right] \xrightarrow[n\to\infty]{\mathsf{p}} r_1 \text{ and } \mathsf{P}\Big[\text{rank}\,\left[\widetilde{V}_\psi(\tilde{\theta}_n^0)\right] = r_1\Big] \xrightarrow[n\to\infty]{} 1\,. \tag{A.10}$$

Further, when rank $[\widetilde{I}_{0n}]$ = rank $[W_n] = m,$ rank $[\tilde{J}_n] = p$ and rank $[\tilde{P}_n] = r_1$ (an event whose probability converges to one as $n \to \infty$), the matrix $\tilde{Q}_n'[\widetilde{V}_\psi(\tilde{\theta}_n^0)]^- \tilde{Q}_n = \tilde{Q}_n'[\tilde{Q}_n \widetilde{I}_{0n} \tilde{Q}_n']^- \tilde{Q}_n$ is invariant to the choice of the generalized inverse $[\widetilde{V}_\psi(\tilde{\theta}_n^0)]^-$; see Harville (1997, Section 9.4, p. 119). Thus the test criterion

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

is (with probability converging to one) also invariant to the choice of the generalized inverse $\{\tilde{Q}\,[W_n]\,\widetilde{I}_{0n}\tilde{Q}\,[W_n]'\}^-$. Finally, by Theorems 1 and 2 of Andrews (1987), it follows that the asymptotic distribution of $\overline{PC}(\tilde{\theta}_n^0; \psi, W_n)$ is $\chi^2(r_1)$. □

It is of interest to note here that the assumptions **A.5** and **A.6** do not require that the derivative matrix of $D_n(\theta_0)$ [*i.e.*, $H_n(\theta, \omega)$] be continuous with respect to θ, even in an open neighborhood of θ_0. More usual assumptions would consist in assuming that $H_n(\theta, \omega)$ has a.s. a limit $J(\theta)$, at least at every point be continuous in a neighborhood of θ_0, and that both $H_n(\theta, \omega)$ and $J(\theta)$ are continuous. We will now show that the assumptions made for establishing Proposition **3.1** include the standard assumptions as special cases. The latter may be stated in the following form.

Assumption A.14 SCORE DERIVATIVE UNIFORM 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:*

(*a*) *$H_n(\theta, \omega)$ is continuous with respect to θ for all $\theta \in N_2, \omega \in \mathcal{D}_H$ and $n \geq 1$;*

(*b*) $\sup_{\theta \in N_2} \|H_n(\theta, \omega) - J(\theta)\| \xrightarrow[n\to\infty]{\mathsf{p}} 0\,.$

Proposition A.15 SUFFICIENCY OF SCORE JACOBIAN CONTINUITY AND UNIFORM CONVERGENCE. *Suppose the assumptions* **A.1** *to* **A.4** *hold. Then Assumption* **A.14** *entails that:*

(*a*) *$J(\theta)$ is continuous at $\theta = \theta_0$;*

(*b*) *both the assumptions* **A.5** *and* **A.6** *also hold.*

PROOF. Consider the (nonempty) open neighborhood $N_0 = N_1 \cap N_2$ of θ_0. For any $\theta \in N_0$ and $\omega \in \mathcal{Z}$, we can write

$$\begin{aligned}\|J(\theta) - J(\theta_0)\| &\leq \|H_n(\theta, \omega) - J(\theta)\| + \|H_n(\theta_0, \omega) - J(\theta_0)\| \\ &\quad + \|H_n(\theta, \omega) - H_n(\theta_0, \omega)\| \\ &\leq 2 \sup_{\theta \in N_0} \|H_n(\theta, \omega) - J(\theta)\| + \|H_n(\theta, \omega) - H_n(\theta_0, \omega)\|\end{aligned}$$

By Assumption **A.14**(b), we have

$$\operatorname*{plim}_{n\to\infty}\Big(\sup_{\theta\in N_0}\|H_n(\theta, \omega) - J(\theta)\|\Big) \leq \operatorname*{plim}_{n\to\infty}\Big(\sup_{\theta\in N_2}\|H_n(\theta, \omega) - J(\theta)\|\Big) = 0$$

and we can find a subsequence of $\{H_{n_t}(\theta,\omega) : t = 1, 2, \ldots\}$ of $\{H_n(\theta,\omega) : n = 1, 2, \ldots\}$ such that

$$\sup_{\theta\in N_0} \{\|H_{n_t}(\theta,\,\omega) - J(\theta)\|\} \underset{t\to\infty}{\longrightarrow} 0 \quad a.s.$$

Let

$$CS = \left\{\omega \in S : \lim_{t\to\infty}\Big(\sup_{\theta\in N_0}\|H_{n_t}(\theta,\,\omega) - J(\theta)\|\Big) = 0\right\}$$

and $\varepsilon > 0$. By definition, $\mathsf{P}[\omega \in CS] = 1$. For $\theta \in CS$, we can choose $t_0(\varepsilon,\,\omega)$ such that

$$t \geq t_0(\varepsilon,\omega) \Rightarrow 2\sup_{\theta\in N_0}\{\|H_{n_t}(\theta,\,\omega) - J(\theta)\|\} < \varepsilon/2\;.$$

Further, since $H_n(\theta,\,\omega)$ is continuous at θ_0, we can find $\delta(\omega) > 0$ such that

$$\|\theta - \theta_0\| < \delta(n,\omega) \Rightarrow \|H_n(\theta,\,\omega) - H_n(\theta_0,\,\omega)\| < \varepsilon/2\;.$$

Thus, taking $n = n_{t_0}$, we find that $\|\theta - \theta_0\| < \delta(n_{t_0},\omega)$ implies

$$\|J(\theta) - J(\theta_0)\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\;.$$

In other words, for any $\varepsilon > 0$, we can choose $\delta = \delta(n_{t_0},\varepsilon) > 0$ such that

$$\|\theta - \theta_0\| < \delta \Rightarrow \|J(\theta) - J(\theta_0)\| < \varepsilon\;,$$

and the function $J(\theta)$ must be continuous at θ_0. Part (a) of the Proposition is established.

Set $\overline{\varDelta}_n(N_2,\,\omega) \equiv \sup\{\|H_n(\theta,\,\omega) - J(\theta)\| : \theta \in N_2\}$. To get **A.5**, we note that

$$\begin{aligned}\varDelta_n(\theta_0,\,\delta,\,\omega) \;&\equiv\; \sup\{\|H_n(\theta,\,\omega) - J(\theta_0)\| : \theta \in N_2 \text{ and } 0 \leq \|\theta -\ \theta_0\| \leq \delta\}\\ &\leq\; \overline{\varDelta}_n(N_2,\,\omega)\end{aligned}$$

for any $\delta > 0$, hence, by Assumption **A.14**(b),

$$\begin{aligned}\limsup_{n\to\infty}\mathsf{P}\big[\{\omega : \varDelta_n(\theta_0,\,\delta,\,\omega) > \varepsilon\}\big] \;&\leq\; \limsup_{n\to\infty}\mathsf{P}\left[\left\{\omega : \overline{\varDelta}_n(N_2,\,\omega) > \varepsilon\right\}\right] = 0\\ &\leq\; U_H(\delta,\,\varepsilon,\,\theta_0)\end{aligned}$$

for any function $U_H(\delta,\,\varepsilon,\,\theta_0)$ that satisfies the conditions of Assumption **A.5**. The latter thus holds.

To obtain **A.6**, we note that **A.14**(b) entails $D_n(\theta,\,\omega)$ is continuously differentiable in an open neighborhood of θ_0 for all $\omega \in \mathcal{D}_H$, so that we can apply Taylor's formula for a function of several variables [see Edwards (1973, Section II.7)] to each component of $D_n(\theta,\,\omega)$: for all θ in an open neighborhood U of θ_0 (we can take $U \subseteq N_0$), we can write

$$\begin{aligned}D_{in}(\theta,\,\omega) \;&=\; D_{in}(\theta_0,\,\omega) + H_n\Big(\overline{\theta}^{\,i}_n(\omega),\,\omega\Big)'_{i\cdot}(\theta -\ \theta_0)\\ &=\; D_{in}(\theta_0,\,\omega) + J(\theta_0)'_{i\cdot}(\theta -\ \theta_0) + R_{in}\big(\overline{\theta}^{\,i}_n(\omega),\,\theta_0,\,\omega\big)\,,\; i = 1,\,\ldots\,,\,m,\end{aligned}$$

where $H_n(\theta,\,\omega)'_{i\cdot}$ and $J(\theta)'_{i\cdot}$ are the i-th rows of $H_n(\theta,\,\omega)$ and $J(\theta)$ respectively,

$$R_{in}\big(\overline{\theta}_n^i(\omega),\,\theta_0,\,\omega\big) = \big[H_n(\overline{\theta}_n^i(\omega),\,\omega)_{i\cdot} - J(\theta_0)_{i\cdot}\big]'(\theta-\,\theta_0)$$

and $\overline{\theta}_n^i(\omega)$ belongs to the line joining θ and θ_0. Further, for $\theta\in U$,

$$\begin{aligned}\big|R_{in}\big(\overline{\theta}_n^i(\omega),\,\theta_0,\,\omega\big)\big| &\leq \big\|H_n(\overline{\theta}_n^i(\omega),\,\omega)_{i\cdot} - J(\theta_0)_{i\cdot}\big\|\;\|\theta-\,\theta_0\| \\ &\leq \big\|H_n(\overline{\theta}_n^i(\omega),\,\omega) - J(\theta_0)\big\|\;\|\theta-\,\theta_0\| \\ &\leq \|\theta-\,\theta_0\|\;\sup\left\{\|H_n(\theta,\,\omega)-J(\theta)\| : \theta\in N_2\right\}\;,\; i=1,\,\ldots\,,\,m,\end{aligned}$$

hence, on defining $N_3=U$,

$$R_n(\theta,\,\theta_0,\,\omega) = \big[R_{1n}\big(\overline{\theta}_n^1(\omega),\,\theta_0,\,\omega\big),\,\ldots\,,\,R_{mn}\big(\overline{\theta}_n^m(\omega),\,\theta_0,\,\omega\big)\big]'\,,$$

we see that

$$\begin{aligned}\|R_n(\theta,\,\theta_0,\,\omega)\| &\leq \sum_{1=1}^{m}\big|R_{in}\big(\overline{\theta}_n^i(\omega),\,\theta_0,\,\omega\big)\big| \\ &\leq m\;\|\theta-\,\theta_0\|\;\sup_{\theta\in N_2}\left\{\|H_n(\theta,\,\omega)-J(\theta)\|\right\}\end{aligned}$$

and

$$\begin{aligned}r_n(\delta,\,\theta_0,\,\omega) &\equiv \sup\left\{\frac{\|R_n(\theta,\,\theta_0,\,\omega)\|}{\|\theta-\,\theta_0\|} : \theta\in N_3 \text{ and } 0<\|\theta-\,\theta_0\|\leq\delta\right\} \\ &\leq m\,\sup\left\{\|H_n(\theta,\,\omega)-J(\theta)\| : \theta\in N_2\right\}\end{aligned}$$

Thus $r_n(\delta,\,\theta_0,\,\omega)\ \underset{n\to\infty}{\overset{\mathsf{p}}{\longrightarrow}}\ 0$ and

$$\limsup_{n\to\infty}\mathsf{P}\big[\{\omega : r_n(\delta,\,\theta_0,\,\omega)>\varepsilon\}\big]\leq U_D(\delta,\,\varepsilon,\,\theta_0) \tag{A.11}$$

must hold for any function that satisfies the conditions of Assumption **A.6**. This completes the proof. □

B. Appendix: Distribution of the modified NW LR-type statistics

The consistency and asymptotic normality of estimators based on the optimization (minimization in our case) of a random criteria, $M_n(\theta)$ say, rests upon various types of assumptions. Pakes and Pollard (1989) gave conditions for criteria of the following types: $D_n\left(\theta;Z_n\right)'D_n\left(\theta;Z_n\right)$ or $D_n\left(\theta;Z_n\right)'I_n\left(\theta\right)^{-1}D_n\left(\theta;Z_n\right)$. Gouriéroux and Monfort (1995) gave general conditions for

$\hat{\theta}_n = \arg\min \, M_n(\theta)$ to be consistent and asymptotically normal. These conditions are listed here.

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 $\Theta \times \Omega$, such that $M_n(\theta;\omega)$ is a continuous function of θ for all $\omega \in \Omega$.*

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

$$\mathsf{P}(\{\omega : \max_{\theta} \mid M_n(\theta;\omega) - \bar{M}(\theta) \mid \to 0\}) = 1.$$

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.

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

Proposition B.8 CONVERGENCE AND ASYMPTOTIC NORMALITY OF CONTINUOUSLY UPDATED GMM ESTIMATOR.

(*a*) *Under the assumptions* **B.1** *to* **B.5**, *there is a sequence $\widehat{\theta}_n = \arg\min M_n(\theta)$ that converges almost surely to θ_0.*

(*b*) *Under assumptions* **B.1** *to* **B.7**,

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \underset{n\to\infty}{\overset{L}{\longrightarrow}} N\big[0,\, G(\theta_0)^{-1} H(\theta_0) G(\theta_0)^{-1}\big] \,. \tag{B.1}$$

If we consider the specific case $M_n(\theta) = D_n(\theta)' I_n(\theta)^{-1} D_n(\theta)$ assumption **B.2** is fulfilled when $D_n(\theta)$ and $I_n(\theta)$ are continuous vectors and matrices. Assumption **B.3** will be verified if $D_n(\theta)$ and $I_n(\theta)$ are strongly consistent for all θ. Since $\lim D_n(\theta_0) = 0$, the minimum of $\bar{M}(\theta)$ is 0 and since $\lim D_n(\theta) \neq 0$ when $\theta \neq \theta_0$, $\bar{M}(\theta) > 0$ when $\theta \neq \theta_0$, if $I_n(\theta)$ is bounded for all

n and θ. Assumption **B.5** requires $D_n(\theta)$ and $I_n(\theta)^{-1}$ are twice continously differentiable. And in the present context $\sqrt{n}D_n(\theta_0) \underset{n\to\infty}{\xrightarrow{L}} N[0,\ I(\theta_0)]$ and $I_n(\theta_0) \underset{n\to\infty}{\xrightarrow{\mathsf{p}}} I(\theta_0)$ ensures assumption **B.7**.

The unconstrained estimator $\hat{\theta}_n$ satisfies the first order conditions

$$\frac{\partial M_n}{\partial \theta}\big(\hat{\theta}_n\big) = 0 \ ,$$

while the constrained estimator $\hat{\theta}_n^0$ solves the system:

$$\begin{aligned} \frac{\partial M_n}{\partial \theta}\big(\hat{\theta}_n^0\big) - \frac{\partial \psi'}{\partial \theta}\big(\hat{\theta}_n^0\big)\hat{\lambda}_n &= 0 \ , \\ \psi\big(\hat{\theta}_n^0\big) &= 0 \ , \end{aligned}$$

where $\psi\left(\theta\right) = 0$ is the null hypothesis tested.

If we add the assumptions **A.8** and **A.9** of Appendix A to **B.1 - B.7**, $\hat{\theta}_n$ and $\hat{\theta}_n^0$ are strongly consistent and asymptotically normal vectors. We can now prove Proposition **5.1**.

PROOF OF PROPOSITION **5.1** The consistency and normal limits of $\hat{\theta}_n$ and $\hat{\theta}_n^0$ yield the classical developments :

$$0 = \frac{\partial M_n}{\partial \theta}\big(\hat{\theta}_n\big) = \frac{\partial M_n}{\partial \theta}\left(\theta_0\right) + \frac{\partial^2 M_n}{\partial \theta \partial \theta'}\left(\theta_0\right)\big(\hat{\theta}_n - \theta_0\big) + e_{1n} \ , \tag{B.2}$$

$$0 = \frac{\partial M_n}{\partial \theta}\left(\theta_0\right) + \frac{\partial^2 M_n}{\partial \theta \partial \theta'}\left(\theta_0\right)\big(\hat{\theta}_n^0 - \theta_0\big) - \frac{\partial \psi'}{\partial \theta}\big(\hat{\theta}_n^0\big)\hat{\lambda}_n + e_{2n} \ , \tag{B.3}$$

$$0 = \psi\big(\hat{\theta}_n^0\big) = \psi\left(\theta_0\right) + \frac{\partial \psi}{\partial \theta'}\left(\theta_0\right)\big(\hat{\theta}_n^0 - \theta_0\big) + e_{3n} \ , \tag{B.4}$$

where $e_{in},\ i = 1,2,3$ are $o_p\left(1/\sqrt{n}\right)$. Taking into account the assumptions of sections 2 and 3 jointly with the convergence of $I_n\left(\theta_0\right)$ to $I\left(\theta_0\right) \equiv I_0$, we see that:

$$\sqrt{n}\,\frac{\partial M_n}{\partial \theta}\left(\theta_0\right) \underset{n\to\infty}{\xrightarrow{L}} N\left(0, U_0\right) \ , \ \frac{\partial^2 M_n}{\partial \theta \partial \theta'}\left(\theta_0\right) \underset{n\to\infty}{\xrightarrow{\mathsf{p}}} V_0 \ ,$$

where

$$U_0 = 4J_0' I_0^{-1} J_0 = 2V_0 \ .$$

The latter identity can be shown as follows. We have:

$$\sqrt{n}\,\frac{\partial M_n}{\partial \theta}\left(\theta_0\right) = 2\frac{\partial D_n'}{\partial \theta}\left(\theta_0\right) I_n\left(\theta_0\right)^{-1} \sqrt{n}D_n\left(\theta_0\right) + \sqrt{n}\,\zeta_n\left(\theta_0\right)$$

where $\zeta_n\left(\theta_0\right)$ is a $p-$dimensional vector with elements

$$tr\frac{\partial I_n^{-1}}{\partial \theta_i}\left(\theta_0\right) D_n\left(\theta_0\right) D_n\left(\theta_0\right)' ,\ i = 1, 2, \ldots, p \ .$$

Since $\sqrt{n}D_n(\theta_0)$ converges in distribution to $N(0, I_0)$ and $\frac{\partial D_n'}{\partial\theta}(\theta_0)$ converges in probability to J_0, $\sqrt{n}\zeta_n(\theta_0)$ is $o_p(1)$. Consequently,

$$\sqrt{n}\,\frac{\partial M_n}{\partial\theta}(\theta_0) \xrightarrow{L} N(0, U_0) \text{ with } U_0 = 4J_0 I_0^{-1} J_0 \ .$$

Using similar arguments, we see that:

$$\frac{\partial^2 M_n}{\partial\theta\partial\theta'}(\theta_0) = 2\frac{\partial D_n'}{\partial\theta}(\theta_0)\, I_n(\theta_0)^{-1}\,\frac{\partial D_n}{\partial\theta'}(\theta_0) + \varSigma_n$$

where $\varSigma_n$ is a $p\times p$ matrix with elements :

$$\begin{aligned}
&\frac{\partial^2 D_n'(\theta_0)}{\partial\theta_i\partial\theta_j} I_n(\theta_0)^{-1} D_n(\theta_0) + \frac{\partial D_n'}{\partial\theta_i}(\theta_0)\,\frac{\partial I_n^{-1}}{\partial\theta_j}(\theta_0)\, D_n(\theta_0)\\
&+ tr\frac{\partial I_n^{-1}}{\partial\theta_i}(\theta_0)\left[\frac{\partial D_n}{\partial\theta_j}(\theta_0)\, D_n'(\theta_0) + D_n(\theta_0)\,\frac{\partial D_n'}{\partial\theta_j}(\theta_0)\right]\\
&+ tr\frac{\partial I_n^{-1}}{\partial\theta_i}(\theta_0)\left[\frac{\partial D_n}{\partial\theta_j}(\theta_0)\, D_n'(\theta_0) + D_n(\theta_0)\,\frac{\partial D_n'}{\partial\theta_j}(\theta_0)\right]\\
&+ tr\frac{\partial^2 I_n^{-1}}{\partial\theta_i\partial\theta_j}(\theta_0)\, D_n(\theta_0)\, D_n'(\theta_0) \ .
\end{aligned}$$

Since $D_n(\theta_0)$ is $o_p(1)$, $\varSigma_n$ is also $o_p(1)$ and

$$\frac{\partial^2 M_n}{\partial\theta\partial\theta'}(\theta_0) \underset{n\to\infty}{\xrightarrow{\mathsf{p}}} V_0 = 2J_0' I_0^{-1} J_0 = \frac{U_0}{2}\,.$$

Under the null hypothesis $\psi(\theta_0) = 0$, equations (B.2), (B.3) and (B.4) can be rewritten:

$$X_n + V_0\big(\hat\theta_n - \theta_0\big) + e_{1n}^* = 0\,, \tag{B.5}$$
$$X_n + V_0\big(\hat\theta_n^0 - \theta_0\big) - P_0'\hat\lambda_n + e_{2n}^* = 0\,, \tag{B.6}$$
$$P_0\big(\hat\theta_n^0 - \theta_0\big) + e_{3n}^* = 0\,, \tag{B.7}$$

where $X_n = \frac{\partial M_n}{\partial\theta}(\theta_0)$, $\sqrt{n}X_n \underset{n\to\infty}{\xrightarrow{L}} N(0, U_0)$, e_{in}^*, $i = 1, 2, 3$ are $o_p(1/\sqrt{n})$. We can then write, after some algebra :

$$\hat\theta_n - \theta_0 = -V_0^{-1} X_n + \varepsilon_{1n}\,, \tag{B.8}$$
$$\hat\theta_n^0 - \theta_0 = -AX_n + \varepsilon_{2n}\,, \tag{B.9}$$
$$\hat\lambda_n = -CX_n + \varepsilon_{3n}\,, \tag{B.10}$$

where ε_{in}, $i = 1, 2, 3$ are $o_p(1/\sqrt{n})$ and

$$A = V_0^{-1} - V_0^{-1} P_0' \left(P_0 V_0^{-1} P_0' \right)^{-1} P_0 V_0^{-1} \, , \; C = - \left(P_0 V_0^{-1} P_0' \right)^{-1} P_0 V_0^{-1} \; .$$

Consider now the modified Newey-West statistic $D_{NW} = n \, \xi_n$, where

$$\xi_n = M_n \big(\hat{\theta}_n^0 \big) - M_n \big(\hat{\theta}_n \big) \; .$$

Developing ξ_n up to order 2, we see that

$$\begin{aligned} \xi_n \;\; &= \;\; \frac{\partial M_n}{\partial \theta'} (\theta_0) \big(\hat{\theta}_n^0 - \theta_0 \big) - \frac{\partial M_n}{\partial \theta'} (\theta_0) \big(\hat{\theta}_n - \theta_0 \big) \\ &\quad + \frac{1}{2} \left\{ \big(\hat{\theta}_n^0 - \theta_0 \big)' \frac{\partial^2 M_n}{\partial \theta \partial \theta'} (\theta_0) \big(\hat{\theta}_n^0 - \theta_0 \big) \right. \\ &\quad \left. - \big(\hat{\theta}_n - \theta_0 \big)' \frac{\partial^2 M_n}{\partial \theta \partial \theta'} (\theta_0) \big(\hat{\theta}_n - \theta_0 \big) \right\} + e_n \end{aligned} \tag{B.11}$$

where e_n is $o_p \left(\frac{1}{n} \right)$. Using (B.8) and (B.9), (B.11) becomes :

$$\xi_n = X_n' \Big(V_0^{-1} - A + \frac{1}{2} A V_0 A - \frac{1}{2} V_0^{-1} \Big) X_n + e_n^* \tag{B.12}$$

with $e_n^* = o_p \left(\frac{1}{n} \right)$. Since $A V_0 A = A$, we have

$$\begin{aligned} \xi_n \;\; &= \;\; \frac{1}{2} X_n' \big(V_0^{-1} - A \big) X_n + e_n^* \\ &= \;\; \frac{1}{2} X_n' V_0^{-1} P_0' \big(P_0 V_0^{-1} P_0' \big)^{-1} P_0 V_0^{-1} X_n + e_n^* \end{aligned}$$

hence

$$D_{NW} = n \xi_n \xrightarrow[n \to \infty]{L} \frac{1}{2} u' U_0^{1/2} V_0^{-1} P_0' \big(P_0 V_0^{-1} P_0' \big)^{-1} P_0 V_0^{-1} U_0^{1/2} u \, , \quad \text{where } u \sim N \left(0, I_p \right) \; .$$

Using $U_0 = 2 V_0$, we see that

$$\begin{aligned} Q \;\; &= \;\; \frac{1}{2} U_0^{1/2} V_0^{-1} P_0' \left(P_0 V_0^{-1} P_0' \right)^{-1} P_0 V_0^{-1} U_0^{1/2} = V_0^{-1/2} P_0' \left(P_0 V_0^{-1} P_0' \right)^{-1} P_0 V_0^{-1/2} \\ Q' \;\; &= \;\; Q \; , \; QQ = Q \, , \end{aligned}$$

so Q is an orthogonal projector with rank $p_1 = \dim(\psi)$. Thus, under the hypothesis $\psi \left(\theta \right) = 0$,

$$D_{NW} \xrightarrow[n \to \infty]{L} \chi^2 \left(p_1 \right) \; .$$

□

References

Andrews, D. W. K. (1987), 'Asymptotic results for generalized Wald tests', *Econometric Theory* **3**, 348–358.

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.

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

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.

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

Edwards, C. H. (1973), *Advanced Calculus of Several Variables*, Dover Publications, New York.

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.

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.

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.

Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag, New York.

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

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 Voss, 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.

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.

Pakes, A. and Pollard, D. (1989), 'Simulation and the asymptotics of optimization estimators', *Econometrica* **57**, 1027–1058.

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.

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.