INVARIANT TESTS BASED ON M-ESTIMATORS, ESTIMATING FUNCTIONS, AND THE GENERALIZED METHOD OF MOMENTS *

Jean-Marie Dufour † and Alain Trognon
C.R.D.E., Université de Montréal and INSEE, Paris

May 1997
This version: January 31, 2000

* The authors thank Russell Davidson, V. P. Godambe and Pascale Valéry for several useful comments. This work was supported by the Natural Sciences and Humanities Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Government of Québec (Fonds FCAR).

† Centre de recherche et développement en économique, Université de Montréal, C.P. 6128, succursale Centre-Ville, Montréal, Québec H3C3J7, Canada.
Tel.: 1 (514) 343 6557. FAX: 1 (514) 343 5831. e-mail: dufour@crde.umontreal.ca
Web page: http://tornade.ere.umontreal.ca/crde/dufour/dufour.html

ABSTRACT

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, *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.

Keywords: 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

1 INTRODUCTION

It is a widely accepted principle in statistics and econometrics that inferences should not depend on arbitrary incidentals like the labelling of i.i.d. observations or the selection of measurement unit changes, when those elements have no incidence on the interpretation of the null and alternative hypotheses; see Hotelling (1936), Pitman (1939), Lehmann (1983; Chap. 3; 1986, Chap. 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. For example, standard t and F tests in linear regressions are invariant to linear hypothesis reformulations and reparameterizations. In nonlinear models, the situation is however more complex.

It is well known that Wald-type tests are not invariant to equivalent hypothesis reformulations; see Cox and Hinkley (1974, p. 302), Burguete, Gallant and Souza (1982, p. 185), Gregory and Veall (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, both 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. The invariance properties of test procedures applicable in models which are incompletely specified or misspecified.

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 pseudo-likelihood (Gouriéroux, Monfort and Trognon, 1984) and M-estimation methods. For general reviews of inference in such models, the reader may consult Davidson and MacKinnon (1993), Gallant (1987), Godambe (1991), Gouriéroux and Monfort (1995) and Newey and McFadden (1994). A striking feature of inference in such models is the fact that likelihood ratio (LR) tests are difficult to apply because their asymptotic distributions involve unknown nuisance parameters [*e.g.*, see Trognon(1984)]. This is quite unfortunate from the point of view of obtaining invariant tests because LR test statistics enjoy very strong invariance qualities. 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, 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. 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) to obtain a test statistic that is invariant to general reparameterizations. The invariance associated with linear exponential pseudo-likelihood functions can be interpreted as a strong argument for using such pseudo-likelihood functions in empirical work.

In Section 2, we describe the general setup considered and define the test statistics that will be studied. The invariance properties of the available test statistics are studied in Section 3. In Section 4, we make suggestions for obtaining tests that are invariant to general hypothesis reformulations and reparameterizations.

2 FRAMEWORK AND TEST STATISTICS

2.1 Assumptions

We consider an inference problem about a parameter of interest $\theta \in \Omega \subseteq \mathbb{R}^p$. θ 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:

$$D_n(\theta; Z_n) \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} D_\infty(\theta;\theta_o) \ . \tag{1}$$

$D_\infty(.;\theta_o)$ is an application of Ω onto $\mathbb{R}^p$ such that : $D_\infty(\theta;\theta_o) = 0 \Longleftrightarrow \theta = \theta_o$. Furthermore, we assume that

$$\sqrt{n} D_n(\theta_o; Z_n) \underset{n\to\infty}{\overset{D}{\longrightarrow}} N[0, I(\theta_o)] \tag{2}$$

and $H_n(\theta_o; Z_n) = \dfrac{\partial}{\partial\theta'} D_n(\theta_o; Z_n) \underset{n\to\infty}{\overset{P}{\longrightarrow}} J(\theta_o)$, where $I(\theta_o)$ and $J(\theta_o)$ 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

$$M_n(\theta, S_n) = D_n(\theta; Z_n)' S_n D_n(\theta; Z_n) \ , \tag{3}$$

where S_n is a symmetric positive definite matrix. The method of estimating equations [Durbin (1960), Godambe (1960, 1991)], the generalized method of moments [Hansen (1982)], 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:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \underset{n\to\infty}{\overset{D}{\to}} N[0, \Sigma(S_0)]$$

where

$$\Sigma(S_0) = (J_0' S_0 J_0)^{-1} J_0' S_0 I_0 S_0 J_0 (J_0' S_0 J_0)^{-1}$$

where $S_0 = p\lim_{n\to\infty} S_n$, $\det(S_0) \neq 0$, $J_0 = J(\theta_0)$ and $I_0 = I(\theta_0)$; 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

$$D_n(\hat{\theta}_n; Z_n) = 0 \; . \tag{4}$$

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

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

This sequence is asymptotically normal with zero mean and asymptotic variance

$$\Omega(\theta_o) = J(\theta_o)^{-1} I(\theta_o) J(\theta_o)^{-1} \; ; \tag{5}$$

see Gouriéroux and Monfort (1995). Obviously, condition (4) 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 (4) 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 $D_n(\theta; Z_n) = \frac{\partial}{\partial \theta} L_n(\theta; Z_n)$.

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

$$E[h(\theta; z_i)] = 0 \; , \quad i = 1, \ldots, n \; .$$

Then one considers the sample analogue of the above mean,

$$h_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} h(\theta; z_i) \; ,$$

and the quadratic form $M_n(\theta) = h_n(\theta)' S_n h_n(\theta)$ where S_n is a symmetric positive definite matrix. In this case, the score-type function is :

$$D_n(\theta; Z_n) = 2 \frac{\partial h_n(\theta)'}{\partial \theta} S_n h_n(\theta) \; .$$

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

$$Q_n(\theta; Z_n) = \frac{1}{n}\sum_{i=1}^{n} \xi(\theta; z_i).$$

The score function has the following form:

$$D_n(\theta; Z_n) = \frac{\partial Q_n}{\partial \theta}(\theta; Z_n) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \theta}\xi(\theta; z_i).$$

2.2 Test statistics

Consider now the problem of testing $H_o : \psi(\theta) = 0$, where $\psi(\theta)$ is a $p_1 \times 1$ continuously differentiable function of θ, $1 \leq p_1 \leq p$ and suppose the $p_1 \times p$ matrix $P(\theta) = \frac{\partial \psi}{\partial \theta'}$ has full row rank (at most in a neighborhood of θ_o). Let $\hat{\theta}_n$ be the unrestricted estimator obtained by minimizing $M_n(\theta)$, and $\hat{\theta}_n^o$ the corresponding constrained estimator under H_0.

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

$$D_n(\theta; Z_n) = \frac{1}{n}\sum_{i=1}^{n} h(\theta; z_i), \tag{6}$$

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

$$\hat{I}(\theta) = \frac{1}{n}\sum_{i=1}^{n} h(\theta; z_i)\, h(\theta; z_i)' \ , \tag{7}$$

$$\hat{J}(\theta) = \frac{\partial D_n}{\partial \theta'}(\theta) = H_n(\theta; Z_n) \ , \tag{8}$$

where θ can be replaced by appropriate estimators. But other estimators may be considered, *e.g.* in view of taking into account serial dependence [see Newey and West (1987)].

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. These can be defined as follows :

a) *Wald-type statistic,*

$$W(\psi) = \psi(\hat{\theta}_n)' \left[\hat{P}\hat{J}^{-1}\hat{I}\left(\hat{J}^{-1}\right)'\hat{P}' \right]^{-1} \psi(\hat{\theta}_n) \tag{9}$$

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,*

$$S(\psi) = nD_n(\hat{\theta}_n^o; Z_n)'\hat{I}_o^{-1}\hat{J}_o\left(\hat{J}_o'\hat{I}_o^{-1}\hat{J}_o\right)^{-1}\hat{J}_o'\hat{I}_o^{-1}D_n(\hat{\theta}_n^o; Z_n) \tag{10}$$

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

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

$$LM(\psi) = n\hat{\lambda}_n'\hat{P}_o\left(\hat{J}_o'\hat{I}_o^{-1}\hat{J}_o\right)^{-1}\hat{P}_o'\hat{\lambda}_n \tag{11}$$

where $\hat{P}_o = P(\hat{\theta}_n^o)$;

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

$$PC(\tilde{\theta}_n^o; \psi) = nD_n(\tilde{\theta}_n^o; Z_n)'\tilde{W}_oD_n(\tilde{\theta}_n^o; Z_n) \tag{12}$$

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

$$\tilde{W}_o = \tilde{I}_o^{-1}\tilde{J}_o\left(\tilde{J}_o'\tilde{I}_o^{-1}\tilde{J}_o\right)^{-1}\tilde{P}_o'\left[\tilde{P}_o\left(\tilde{J}_o'\tilde{I}_o^{-1}\tilde{J}_o\right)^{-1}\tilde{P}_o'\right]^{-1}\tilde{P}_o\left(\tilde{J}_o'\tilde{I}_o^{-1}\tilde{J}_o\right)^{-1}\tilde{J}_o'\tilde{I}_o^{-1}$$

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

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 (1987, 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:

$$LR(\psi) = S_n(\hat{\theta}_n; Z_n) - S_n(\hat{\theta}_n^o; Z_n) \,. \tag{13}$$

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)]. 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):

$$D_{NW} = n[M_n(\hat{\theta}_n^o, \tilde{I}_o) - M_n(\hat{\theta}_n, \tilde{I}_o)] \tag{14}$$

where

$$M_n(\theta_n, \tilde{I}_o) = D_n(\theta; Z_n)' \, \tilde{I}_o^{-1} D_n(\theta; Z_n) \ ,$$

$\tilde{I}_o$ is a consistent estimator of $I(\theta_o)$, $\hat{\theta}_n$ minimizes $M_n(\theta, \tilde{I}_o)$ without restriction and $\hat{\theta}_n^o$ 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,

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

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

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

where W_n is a positive definite (possibly random) $p \times p$ matrix such that

$$p \lim_{n\to\infty} W_n = W_o, \quad \det(W_o) \neq 0,$$

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

$$\tilde{Q}_o[W_n] = \tilde{P}_o\left(\tilde{J}_o' W_n \tilde{J}_o\right)^{-1} \tilde{J}_o' W_n\,,$$

where $\tilde{P}_o = P(\tilde{\theta}_n^o)$, and $\tilde{J}_o = \hat{J}(\tilde{\theta}_n^o)$. Under standard regularity conditions, it is easy to see that

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

where

$$Q(\theta_o) = p\lim_{n\to\infty} \tilde{Q}_o[W_n] = P(\theta_o)\left[J(\theta_o)' W_o J(\theta_o)\right]^{-1} J(\theta_o)' W_o$$

and $rank\left[Q(\theta_o)\right] = p$. This suggests the following generalized $C(\alpha)$ criterion:

$$PC(\tilde{\theta}_n^o; \psi, W_n) = nD_n(\tilde{\theta}_n^o; Z_n)' \tilde{Q}_o[W_n]' \left\{\tilde{Q}_o[W_n]\, \tilde{I}_o \tilde{Q}_o[W_n]'\right\}^{-1} \tilde{Q}_o[W_n]\, D_n(\tilde{\theta}_n^o; Z_n)\,, \tag{15}$$

where $\tilde{I}_o = \hat{I}(\tilde{\theta}_n^o)$, whose asymptotic distribution is $\chi^2(p_1)$ under H_0.

On taking $W_n = \tilde{I}_o^{-1}$, as suggested by efficiency arguments, $PC(\tilde{\theta}_n^o; \psi, W_n)$ reduces to $PC(\tilde{\theta}_n^o; \psi)$ in (12). When the number of equations equals the number of parameters $(m = p)$, we have $\tilde{Q}_o[W_n] = \tilde{P}_o \tilde{J}_o^{-1}$ and $PC(\tilde{\theta}_n^o; \psi, W_n)$ does not depend on the choice of W_n :

$$\begin{aligned} PC(\tilde{\theta}_n^o; \psi, W_n) &= PC(\tilde{\theta}_n^o; \psi) \\ &= D_n(\tilde{\theta}_n^o; Z_n)' \left(\tilde{J}_o^{-1}\right)' \tilde{P}_o' \left[\tilde{P}_o \tilde{J}_o^{-1} \tilde{I}_o \left(\tilde{J}_o^{-1}\right)' \tilde{P}_o'\right]^{-1} \tilde{P}_o \tilde{J}_o^{-1} D_n(\tilde{\theta}_n^o; Z_n). \end{aligned}$$

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 $\tilde{\theta}_n^o$ is obtained by minimizing $M_n(\theta) =$

$D_n(\theta; Z_n)' \tilde{I}_o^{-1} D_n(\theta; Z_n)$ subject to $\psi(\theta) = 0$, we can write $\tilde{\theta}_n^o \equiv \hat{\theta}_n^o$ and $PC(\tilde{\theta}_n^o; \psi, W_n)$ is identical to the score (or LM) -type statistic suggested by Newey and West (1987). Since the statistic $PC(\tilde{\theta}_n^o; \psi, W_n)$ is quite comprehensive, it will be convenient for establishing general invariance results.

3 INVARIANCE

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

3.1 Hypothesis reformulation

Let $\Theta_o = \{\theta \in \Omega \mid \psi(\theta) = 0\}$ and Ψ the set of differentiable function $\bar{\psi} : \Omega \to \mathbb{R}^m$ such that $\left\{\theta \in \Omega \mid \bar{\psi}(\theta) = 0\right\} = \Theta_0$. 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 $\hat{I}, \hat{J}, \hat{I}_o$ and $\hat{J}_o$. The same holds for $\tilde{I}_o$ and $\tilde{J}_o$ provided the restricted estimator $\tilde{\theta}_n^o$ 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 $\psi\left(\hat{\theta}_n\right)$ change. Following Dagenais and Dufour (1991), if $\bar{\psi} \in \Psi$, we have $\bar{P}(\theta) = \dfrac{\partial \bar{\psi}}{\partial \theta'} = \bar{P}_1(\theta) G(\theta)$ and $P(\theta) = \dfrac{\partial \psi}{\partial \theta'} = P_1(\theta) G(\theta)$ 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 $\bar{P}_1^{o\prime} \bar{\lambda}_n = \hat{P}_1^{o\prime} \hat{\lambda}_n$ where $\bar{P}_1^o = \bar{P}_1(\hat{\theta}_n^o)$, $\hat{P}_1^o = P_1(\hat{\theta}_n^o)$ 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 1 *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 $\psi(\theta) = 0$ if and only if $\bar{\psi}(\theta) = 0$, $\forall \psi, \bar{\psi} \in \Psi$. Then, $T(\psi) = T(\bar{\psi})$, where T stands for any one of the test statistics $S(\psi)$, $LM(\psi)$, $PC(\hat{\theta}_n^o; \psi)$, $LR(\psi)$, $D_{NW}(\psi)$ and $PC(\hat{\theta}_n^o; \psi, W_n)$ defined in (10) to (15).*

3.2 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 of $\bar{g} : k(\theta_*) = \bar{g}^{-1}(\theta_*) = \theta$. Set $G(\theta) = \frac{\partial \bar{g}'}{\partial \theta}(\theta)$ and $K(\theta_*) = \frac{\partial k}{\partial \theta_*'}(\theta_*)$. Since $k[\bar{g}(\theta)] = \theta$ and $\bar{g}[k(\theta_*)] = \theta_*$, $K[\bar{g}(\theta)]\, G(\theta) = I_p$ and $G[k(\theta_*)]\, K(\theta_*) = I_p$, $\forall \theta_* \in \Omega_*$ and $\theta \in \Omega$. Let

$$\psi^*(\theta_*) = \psi\left[\bar{g}^{-1}(\theta_*)\right] \ . \tag{16}$$

Clearly, $\psi^*(\theta_*) = 0 \Leftrightarrow \psi(\theta) = 0$, and $H_0^* : \psi^*(\theta_*) = 0$ is an equivalent reformulation of $H_0 : \psi(\theta) = 0$ in terms of θ_*. Other (possibly more "natural") reformulations are of course possible, but the latter has the convenient property that $\psi^*(\theta_*) = \psi(\theta)$. 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

$$P_*(\theta_*) \equiv \frac{\partial \psi^*}{\partial \theta_*'} = \frac{\partial \psi}{\partial \theta'} \frac{\partial \theta}{\partial \theta_*'} = P[k(\theta_*)]\, K(\theta_*) = P(\theta)\, K[\bar{g}(\theta)] \ . \tag{17}$$

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 of assuming that $D_n(\theta; Z_n) = \sum_{i=1}^{n} h(\theta; z_i)/n$ as in (6) where the values of the scores are unaffected by the reparameterization, but are simply reexpressed in term of θ_* and z_{i*} :

$$h(\theta_*; z_{i*}) = h(\theta; z_i) \ , \quad i = 1, \ldots, n \ , \tag{18}$$

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 (18), we see easily that

$$H_{n*}(\theta_*; Z_{n*}) = \frac{\partial D_{n*}(\theta_*; Z_{n*})}{\partial \theta_*'} = H_n(\theta; Z_n) K(\theta_*) = H_n(\theta; Z_n) K[\bar{g}(\theta)] \ . \tag{19}$$

Further the functions $\hat{I}(\theta)$ and $\hat{J}(\theta)$ in (7) - (8) 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 (7) - (8), if $W_{n*} = W_n$ and if $\tilde{\theta}_n^o$ is equivariant with respect to $\bar{g}$ $\left[\text{i.e., } \tilde{\theta}_{n*}^o = \bar{g}\left(\tilde{\theta}_n^o\right)\right]$, it is easy to check that the generalized $C(\alpha)$ statistic defined in (15) is invariant to the reparameterization $\theta_* = \bar{g}(\theta)$. This suggests the following general sufficient condition for the invariance of $C(\alpha)$ statistics.

Proposition 2 *Let $\psi^*(\theta_*) = \psi[\bar{g}^{-1}(\theta_*)]$, and suppose the following conditions hold :*

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

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

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 (15), and $\tilde{K} = K(\tilde{\theta}_{n*}^o)$ is invertible. Then

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

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

It is clear the estimators $\hat{\theta}_n$ and $\hat{\theta}^{o}_{n}$ satisfy the equivariance condition, *i.e.*, $\hat{\theta}_{n*} = \bar{g}\left(\hat{\theta}_n\right)$ and $\hat{\theta}^{o}_{n*} = \bar{g}\left(\hat{\theta}^{o}_{n}\right)$. 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.

Proposition 3 *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 $K(\theta_{*}) = \dfrac{\partial \bar{g}^{-1}(\theta_{*})}{\partial \theta'_{*}}$. *Then, provided the relevant matrices are invertible, we have* $T(\psi) = T_{*}(\psi_{*})$ *where* T *stands for any one of the test statistics* $S(\psi)$, $LM(\psi)$, $LR(\psi)$ *and* $D_{NW}(\psi)$. *If* $\hat{\theta}^{o}_{n*} = \bar{g}(\hat{\theta}^{o}_{n})$, *we also have* $PC_{*}(\tilde{\theta}^{o}_{n*}; \psi_{*}) = PC(\tilde{\theta}^{o}_{n}; \psi)$. *If* $\psi_{*}(\theta) = \psi\left[\bar{g}^{-1}(\theta)\right]$, *the Wald statistic is invariant :* $W_{*}(\psi_{*}) = W(\psi)$.

Cases where (19) 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 : $D_n(\theta; Z_n) = \dfrac{\partial S_n(\theta; Z_n)}{\partial \theta}$. 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

$$D_{n*}(\theta_*; Z_{n*}) = K(\theta_*)' \ D_n(\theta; Z_n) = K[\bar{g}(\theta)]' \ D_n(\theta; Z_n) \ . \tag{20}$$

From (2) and (20), it then follows that :

$$\sqrt{n} D_{n*}(\theta_{0*}; Z_{n*}) \underset{n\to\infty}{\overset{D}{\longrightarrow}} N[0, I_*(\theta_{0*})] \ , \tag{21}$$

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

$$I_*(\theta_*) = K(\theta_*)' I[k(\theta_*)] K(\theta_*) = K[\bar{g}(\theta)]' I(\theta) K[\bar{g}(\theta)] \ , \tag{22}$$

and

$$H_{n*}(\theta_*; Z_n) = K[\bar{g}(\theta)]' H_n(\theta; Z_n) K[\bar{g}(\theta)] + \sum_{j=1}^{p} D_{nj}(\theta; Z_n) M_j[\bar{g}(\theta)] , \tag{23}$$

where $D_{nj}(\theta; Z_n), j = 1, \ldots, p$, are the coordinates of $D_n(\theta; Z_n)$ and $M_j(\theta_*) = \dfrac{\partial^2 \theta_j}{\partial \theta_* \partial \theta_*'}(\theta_*) = \dfrac{\partial^2 k_j}{\partial \theta_* \partial \theta_*'}(\theta_*)$.

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 are sensitive to a reparameterization, unless some specific estimator of J is used. At this generality level, the following result 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_*.

Proposition 4 *Let $\psi^*(\theta_*) = \psi[\bar{g}^{-1}(\theta_*)]$, and suppose the following conditions hold :*

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

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

c) $\tilde{I}_{0*} = \tilde{K}'\tilde{I}_0\tilde{K}, \tilde{J}_{0*} = \tilde{K}'\tilde{J}_0\tilde{K}$, *d)* $W_{n*} = \tilde{K}^{-1} W_n \left(\tilde{K}^{-1}\right)'$,

where $\tilde{I}_0$, $\tilde{J}_0$ *and* W_n *are defined as in (15), and* $\tilde{K} = K(\tilde{\theta}^o_{n*})$. *Then, provided the relevant matrices are invertible,*

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

Proposition 5 *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 $K(\theta_*) = \dfrac{\partial \bar{g}^{-1}(\theta)}{\partial \theta'_*}$. *Then, provided the relevant matrices are invertible, we have* $T(\psi) = T_*(\psi_*)$ *where* T *stands for any one of the test statistics* $S(\psi)$, $LM(\psi)$, $LR(\psi)$ *and* $D_{NW}(\psi)$. *If* $\tilde{\theta}^o_{n*} = \bar{g}(\tilde{\theta}^o_n)$, *we also have* $PC_*(\tilde{\theta}^o_{n*}; \psi_*) = PC(\tilde{\theta}^o_n; \psi)$. *If* $\psi_*(\theta) = \psi[\bar{g}^{-1}(\theta)]$, $W_*(\psi_*) = W(\psi)$.

4 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. In particular, this problem will occur when the score-type function is derived from a (pseudo) likelihood function or, more generally, 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.

4.1 Modified Newey–West LR-type statistic

Consider the LR-type statistic

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

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

However, there is a simple way of creating the appropriate invariance as soon as the function $\hat{I}\left(\theta\right)$ is a reasonably smooth function of θ. Instead of estimating θ by minimizing $M_n\left(\theta, \tilde{I}_o\right)$, estimate θ by minimizing $M_n\left(\theta, \hat{I}\left(\theta\right)\right)$. Suppose further the following conditions hold :

$$D_{n*}\left(\bar{g}\left(\theta\right), Z_{n*}\right) = K\left[\bar{g}\left(\theta\right)\right]' D_n\left(\theta; Z_n\right) , \quad (24)$$

$$\hat{I}_*\left(\bar{g}\left(\theta\right)\right) = K\left[\bar{g}\left(\theta\right)\right]' \hat{I}\left(\theta\right) K\left[\bar{g}\left(\theta\right)\right] . \quad (25)$$

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

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

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

LR-type statistic

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

is invariant to reparameterizations of the type considered in (20)-(22). 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 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.

4.2 Pseudo-maximum likelihood methods

Gouriéroux, Monfort and Trognon (1984) studied inference on the parameter which appears in the mean of an endogenous random vector conditional to an exogenous random vector x_i :

$$E(y_i \mid x_i) = f(x_i; \theta) \equiv f_i(\theta) \ , V(y_i \mid x_i) = \Omega_o \, .$$

A consistent and asymptotically normal estimator of θ can be obtained through the PML method :

$$\underset{\theta}{\text{Max}} \sum_{i=1}^{n} \left\{ A(f_i(\theta)) + C(f_i(\theta))' y_i \right\} \text{ with } \frac{\partial A}{\partial m} + \frac{\partial C}{\partial m} m = 0 \, .$$

The PML estimator is consistent and asymptotically normal, and we can write:

$$\begin{aligned} J(\theta) &= E_x \frac{\partial f_i}{\partial \theta} \frac{\partial C}{\partial m} (f_i(\theta)) \cdot \frac{\partial f_i}{\partial \theta'} \, , \\ I(\theta) &= E_x \frac{\partial f_i}{\partial \theta} \frac{\partial C}{\partial m} (f_i(\theta)) \, \Omega_o \frac{\partial C}{\partial m'} \frac{\partial f_i}{\partial \theta'} \, . \end{aligned}$$

These matrices can be estimated by :

$$\hat{J} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial f_i'}{\partial \theta} \left(\hat{\theta}\right) \frac{\partial C}{\partial m} \left(f_i\left(\hat{\theta}\right)\right) \frac{\partial f_i}{\partial \theta'} \left(\hat{\theta}\right),$$

$$\hat{I} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial f_i'}{\partial \theta} \left(\hat{\theta}\right) \frac{\partial C}{\partial m} \left(f_i\left(\hat{\theta}\right)\right) \left(y - f_i\left(\hat{\theta}\right)\right) \left(y_i - f_i\left(\hat{\theta}\right)\right)' \frac{\partial C}{\partial m'} \left(f_i\left(\hat{\theta}\right)\right) \cdot \frac{\partial f_i}{\partial \theta'} \left(\hat{\theta}\right).$$

Since $\frac{\partial C'}{\partial m}\left(f_i\left(\hat{\theta}\right)\right)$ and $y_i - f_i\left(\hat{\theta}\right)$ are invariant to reparameterizations, $\hat{I}$ and $\hat{J}$ are modified only through $\frac{\partial f_i}{\partial \theta'}$. Further,

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

and

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

The Wald, Lagrange, score and $C\left(\alpha\right)$-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 paramenters of conditional means when the error distribution has unknown type.

References

Breusch, T. S. and P. Schmidt (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., A. R. Gallant, and G. Souza (1982). On unification of the asymptotic theory of nonlinear econometric models. *Econometric Reviews 1*, 151–211 (with comments).

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

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

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

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

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

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

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

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

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics 31*, 1208–1211.

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

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

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

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

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

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

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

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

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

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics, Volume IV*, pp. 2111–2245. Amsterdam: Elsevier Science.

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

Phillips, P. C. B. and J. Y. Park (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.

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