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

Jean-Marie Dufour, Alain Trognon & Purevdorj Tuvaandorj

To cite this article: Jean-Marie Dufour, Alain Trognon & Purevdorj Tuvaandorj (2017) Invariant tests based on M-estimators, estimating functions, and the generalized method of moments, Econometric Reviews, 36:1-3, 182-204, DOI: 10.1080/07474938.2015.1114285

To link to this article: http://dx.doi.org/10.1080/07474938.2015.1114285
Invariant tests based on $M$-estimators, estimating functions, and the generalized method of moments

Jean-Marie Dufour, Alain Trognon, and Purevdorj Tuvaandorj

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) 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, 1984b,c) 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), 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 generalized method of moments (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 is interpreted as a strong argument for using such pseudo-likelihood functions in empirical work.

1. Introduction

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

It is well known that Wald-type tests are not invariant to equivalent hypothesis reformulations and reparameterizations; see Cox and Hinkley (1974, p. 302), Burguete et al. (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 article, 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 et al., 1984b, c; Gouriéroux and Monfort, 1993). For general discussions of inference in such models, the reader may consult White (1982), Newey (1985), Gallant (1987), Newey and West (1987), Gallant and White (1988), Gouriéroux and Monfort (1988, 1995), Godambe (1991), Davidson and MacKinnon (1993), Newey and McFadden (1994), Hall (1999), and Mátyás (1999); for studies of the performance of some test procedures based on GMM estimators, see also Burnside and Eichenbaum (1996) and Podivinsky (1999).

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

In view of avoiding such undesirable properties, we study restrictions that can be imposed on the objective functions used for pseudo-likelihood (or $M$-estimation) as well as the structure of the test criteria used with estimating functions and GMM procedures to obtain invariant tests. In particular, we show that using linear exponential pseudo-likelihood functions allows one to obtain invariant score-type and $C(\alpha)$-type test criteria, while in the context of estimating function (or GMM) procedures it is possible to modify a LR-type statistic proposed by Newey and West (1987) to obtain a test statistic that is invariant to general reparameterizations. The invariance associated with linear exponential pseudo-likelihood functions can be viewed as a strong argument for using such pseudo-likelihood functions in empirical work. Of course, the fact that Wald-type tests are not invariant to both hypothesis reformulations and reparameterizations is by itself a strong argument to avoid using this type of procedure (when they are not equivalent to other procedures) and suggest as well that Wald-type tests can be quite unreliable in finite samples; for further arguments going in the same direction, see Burnside and Eichenbaum (1996), Dufour (1997), and Dufour and Jasiak (2001).
In Section 2, we describe the general setup considered, while the test statistics studied are defined in Section 3. The invariance properties of the available test statistics are studied in Section 4. In Section 5, we make suggestions for obtaining tests that are invariant to general hypothesis reformulations and reparameterizations. Numerical illustrations of the invariance (and noninvariance) properties discussed are provided in Section 6. In Section 7, we consider linear stochastic discount factor model as an empirical example and show that noninvariant procedures may yield drastically different outcomes depending on the identifying restrictions imposed. We conclude in Section 8.

2. Framework

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

\[
D_n (\theta; Z_n) \xrightarrow{p} D_\infty (\theta; \theta_0).
\]

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

\[
D_\infty (\theta; \theta_0) = 0 \iff \theta = \theta_0,
\]

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

\[
\sqrt{n} D_n (\theta_0; Z_n) \xrightarrow{L} N [0, I (\theta_0)],
\]

\[
H_n (\theta_0; Z_n) = \frac{\partial}{\partial \theta} D_n (\theta_0; Z_n) \xrightarrow{p} J (\theta_0),
\]

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 \( \theta \) an expression of the form

\[
M_n (\theta; W_n) = D_n (\theta; Z_n)' W_n D_n (\theta; Z_n)
\]

where \( W_n \) is a symmetric positive definite matrix. The method of estimating functions (Durbin, 1960; Godambe, 1960, 1991; Basawa et al., 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

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

where

\[
\Sigma (W_0) = (J_0' W_0 J_0)^{-1} J_0' W_0 I_0 W_0 J_0 (J_0' W_0 J_0)^{-1},
\]

\( J_0 = J (\theta_0) \), \( I_0 = I (\theta_0) \), \( W_0 = \text{plim} \ W_n \), \( \det (W_0) \neq 0 \); see Gouriéroux and Monfort (1995, Ch. 9).

Note also that “asymptotic estimation efficiency” arguments suggest one to use \( W_n = I_n^{-1} \) as weighting matrix, where \( I_n \) is consistent estimator of \( I_0 \).

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

\[
D_n (\hat{\theta}_n; Z_n) = 0.
\]

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

\[
D_n (\theta; Z_n) = \frac{\partial S_n (\theta; Z_n)}{\partial \theta}, \quad H_n (\theta; Z_n) = \frac{\partial^2 S_n (\theta; Z_n)}{\partial \theta \partial \theta'}.
\]  

(2.9)

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

\[
\Sigma_D (\theta_0) = \left[ I (\theta_0) - I (\theta_0) - I (\theta_0) \right]^{-1} = \left( J_0^{-1} J_0 \right)^{-1}.
\]

(2.10)

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

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

\[
D_n (\theta; Z_n) = \frac{1}{n} \frac{\partial}{\partial \theta} L_n (\theta; Z_n).
\]

(2.11)

b) GMM. \( \theta \) is identified through a \( m \times 1 \) vector of conditions of the form: \( E [h_t (\theta; z_t)] = 0, t = 1, \ldots, n \). Then one considers the sample analogue of this mean,

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

(2.12)

and the quadratic form

\[
M_n (\theta) = \bar{h}_n (\theta)' W_n \bar{h}_n (\theta),
\]

(2.13)

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

\[
D_n (\theta; Z_n) = 2 \frac{\partial \bar{h}_n (\theta)'}{\partial \theta} W_n \bar{h}_n (\theta).
\]

(2.14)

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

\[
\tilde{M}_n (\theta; Z_n) = \frac{1}{n} \sum_{t=1}^{n} \xi (\theta; z_t).
\]

(2.15)

The score function has the following form:

\[
D_n (\theta; Z_n) = \frac{\partial \tilde{M}_n (\theta; Z_n)}{\partial \theta} = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} \xi (\theta; z_t).
\]

(2.16)

3. Test statistics

Consider now the problem of testing

\[
H_0 : \psi (\theta) = 0,
\]

(3.1)

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

\[
P (\theta) = \frac{\partial \psi}{\partial \theta'}
\]

(3.2)

has full row rank (at least in an open neighborhood of \( \theta_0 \)). Let \( \hat{\theta}_n \) be the unrestricted estimator obtained by minimizing \( M_n (\theta) \), and \( \hat{\theta}_n^0 \) the corresponding constrained estimator under \( H_0 \).

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

\[
D_n(\theta; Z_n) = \frac{1}{n} \sum_{t=1}^{n} h_t(\theta; z_t),
\]

(3.3)

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

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

(3.4)

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

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

\[
\hat{I}(\theta) = \sum_{s=1}^{n} \sum_{t=1}^{n} w_{st}(n) h_s(\theta; z_s) h_t(\theta; z_t)' = h(\theta; Z_n) W_t(n)h(\theta; Z_n)',
\]

(3.5)

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

\[
h(\theta; Z_n) = [h_1(\theta; z_1), h_2(\theta; z_2), \ldots, h_n(\theta; z_n)].
\]

(3.6)

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

\[
\hat{I}(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left[ h_t(\theta; z_t) - \bar{h}(\theta) \right] \left[ h_t(\theta; z_t) - \bar{h}(\theta) \right]',
\]

(3.7)

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

\[
\hat{I}(\theta) = \sum_{j=-n+1}^{n-1} \tilde{k}(j/B_n) \hat{\Gamma}(j, \theta),
\]

(3.8)

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

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

(3.9)

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

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

(a) The Wald-type statistic,

\[
W(\psi) = n \psi(\hat{\theta}_n)' [\hat{P}(\hat{\Gamma}^{-1})]^{-1} \hat{P}' [\hat{P}(\hat{\Gamma}^{-1})]^{-1} \psi(\hat{\theta}_n),
\]

(3.10)

where \( \hat{P} = P(\hat{\theta}_n), \hat{\Gamma} = \hat{I}(\hat{\theta}_n), \) and \( \hat{J} = \hat{J}(\hat{\theta}_n); \)
(b) The score-type statistic,
\[
S(\psi) = nD_n(\hat{\theta}_n^0; Z_n)\hat{J}_0^{-1}\hat{J}_0^{-1}D_n(\hat{\theta}_n^0; Z_n),
\]
where \(\hat{\theta}_n\) and \(\hat{J}_0\) are the score and LM statistics, respectively.

(c) The LM-type statistic,
\[
LM(\psi) = n\hat{\lambda}_n\hat{P}_0\hat{J}_0^{-1}\hat{J}_0^{-1}\hat{P}_0\hat{\lambda}_n,
\]
where \(\hat{\lambda}_n\) is the LM statistic.

(d) The C(\alpha)-type statistic,
\[
PC(\hat{\theta}_n^0; \psi) = nD_n(\hat{\theta}_n^0; Z_n)\hat{W}_0D_n(\hat{\theta}_n^0; Z_n),
\]
where \(\hat{\theta}_n^0\) is any root-n consistent estimator of \(\theta\) that satisfies \(\psi(\hat{\theta}_n^0) = 0\), and
\[
\hat{W}_0 = \hat{J}_0^{-1}\hat{J}_0^{-1}\hat{P}_0\hat{J}_0^{-1}\hat{J}_0^{-1}\hat{P}_0\hat{J}_0^{-1}\hat{J}_0^{-1}\hat{P}_0\hat{J}_0^{-1}\hat{J}_0^{-1}\hat{P}_0 - I(\hat{\theta}_n^0).
\]

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:
\[
LR(\psi) = S_n(\hat{\theta}_n; Z_n) - S_n(\hat{\theta}_n^0; Z_n).
\]

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 \(\alpha\) and \(\beta\) coefficients depending upon nuisance parameters; see, for example, Trognon (1984) and Vuong (1989).

Using the constrained minimization condition,
\[
H_n(\hat{\theta}_n^0; Z_n)\hat{J}_0^{-1}D_n(\hat{\theta}_n^0; Z_n) = \hat{P}_0(\hat{\theta}_n^0)\hat{\lambda}_n,
\]
we see that
\[
S(\psi) = LM(\psi),
\]
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
\[
s(\hat{\theta}_n^0, W_n) = \sqrt{n} \hat{Q}[W_n] D_n(\hat{\theta}_n^0; Z_n),
\]
where $\tilde{\theta}_n^0$ is consistent restricted estimate of $\theta_0$ such that $\psi(\tilde{\theta}_n^0) = 0$ and $\sqrt{n}(\tilde{\theta}_n^0 - \theta_0)$ is asymptotically bounded in probability,

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

(3.20)

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

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

(3.21)

Under standard regularity conditions, we have

$$
s(\tilde{\theta}_n^0; Z_n) \xrightarrow{L} N\left[0, Q(\theta_0) I(\theta_0) Q(\theta_0)'ight],
$$

(3.22)

where

$$
Q(\theta_0) = \text{plim } \tilde{Q}[W_n] = P(\theta_0) \left[J(\theta_0)' W_0 J(\theta_0)\right]^{-1} J(\theta_0)' W_0
$$

(3.23)

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

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

(3.24)

where $\tilde{J}_0 = J(\tilde{\theta}_n^0)$. Under general regularity conditions, the asymptotic distribution of $PC(\tilde{\theta}_n^0; \psi, W_n)$ is $\chi^2(p_1)$ under $H_0$.\footnote{The regularity conditions and a rigorous proof of the latter assertion appear in the working paper version of this article Dufour et al. (2013); see also Dufour et al. (2015).} $PC(\tilde{\theta}_n^0; \psi, W_n)$ can be viewed as the extension of the classical procedure of Neyman (1959) to general estimating functions and GMM setups, and it includes as special cases various other $C(\alpha)$-type statistics proposed in the statistical and econometric literatures.\footnote{For further discussion of $C(\alpha)$ tests, the reader may consult Bernshtein (1981), Basawa (1985), Ronchetti (1987), Smith (1987), Berger and Wallenstein (1989), Dagenais and Dufour (1991), Davidson and MacKinnon (1991, 1993), Kocharlakota and Kocharlakota (1991), Bera and Bilias (2001), and Dufour and Valéry (2009).} On taking $W_n = \tilde{J}_0^{-1}$, as suggested by asymptotic estimation efficiency arguments, $PC(\tilde{\theta}_n^0; \psi, W_n)$ reduces to $PC(\tilde{\theta}_n^0; \psi)$ in (3.13). When the number of equations equals the number of parameters $(m = p)$, we have $Q[W_n] = \tilde{P}_0 \tilde{J}_0^{-1}$ and $PC(\tilde{\theta}_n^0; \psi, W_n)$ does not depend on the choice of $W_n$:

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

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

4. Invariance

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

4.1. Hypothesis reformulation

Let

$$
\Theta_0 = \{\theta \in \Theta | \psi(\theta) = 0\}
$$

(4.1)
and \( \Psi \) be the set of differentiable functions \( \tilde{\psi} : \Theta \to \mathbb{R}^m \) such that
\[
\{ \theta \in \Theta | \tilde{\psi}(\theta) = 0 \} = \Theta_0.
\] (4.2)

A test statistic is invariant with respect to \( \Psi \) if it is the same for all \( \psi \in \Psi \). It is obvious the LR-type statistics \( LR(\psi) \) and \( D_{NW}(\psi) \) (when applicable) are invariant to such hypothesis reformulations because the optimal values of the objective function (restricted or unrestricted) do not depend on the way the restrictions are written. Now, a reformulation does not affect \( \hat{\theta}_n \) provided the restricted estimator \( \hat{\theta}_n \) is used with \( C(\alpha) \) tests does not depend on which function \( \psi \in \Psi \) is used to obtain it. However, \( \hat{\theta}_n \), \( \hat{\psi}_n \), and \( \hat{\psi}_n(\hat{\theta}_n) \) change. Following Dagenais and Dufour (1991), if \( \psi \in \Psi \), we have
\[
\tilde{P}(\theta) = \frac{\partial \tilde{\psi}}{\partial \theta} = \tilde{P}_1(\theta) G(\theta), \quad P(\theta) = \frac{\partial \psi}{\partial \theta} = P_1(\theta) G(\theta),
\] (4.3)
where \( \tilde{P}_1 \) and \( P_1 \) are two \( p_1 \times p_1 \) invertible functions and \( G(\theta) \) is a \( p_1 \times p \) full row-rank matrix. Since \( \tilde{P}_1^{\psi} \tilde{\psi}_n = \tilde{P}_1^{\psi} \tilde{\psi}_n \) where \( \tilde{P}_1^{\psi} = \tilde{P}_1(\hat{\theta}_n^0) \), \( \tilde{P}_1^{\psi} = P_1(\hat{\theta}_n^0) \) and \( \tilde{\psi}_n \) is the Lagrange multiplier associated with \( \tilde{\psi} \), 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 \( \Psi \) be a family of \( p_1 \times 1 \) continuously differentiable functions of \( \theta \) such that \( \frac{\partial \psi}{\partial \theta} \) has full row rank when \( \psi(\theta) = 0 \) \((1 \leq p_1 \leq p)\), and
\[
\psi(\theta) = 0 \iff \tilde{\psi}(\theta) = 0, \forall \psi, \tilde{\psi} \in \Psi.
\] (4.4)

Then, \( T(\psi) = T(\tilde{\psi}) \) 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 \( PC(\hat{\theta}_n^0, \tilde{\psi}, W_n) \) defined in (3.11)–(3.15) and (3.24).

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

### 4.2. Reparameterization

Let \( \tilde{g} \) be a one-to-one differentiable transformation from \( \Theta \subseteq \mathbb{R}^p \) to \( \Theta_s \subseteq \mathbb{R}^p : \theta_s = \tilde{g}(\theta) \). \( \tilde{g} \) represents a reparameterization of the parameter vector \( \theta \) to a new one, \( \theta_s \). The latter is often determined by a one-to-one transformation of the data \( Z_{ns} = g(Z_n) \), as occurs for example when variables are rescaled (measurement unit changes). But it may also represent a reparameterization without any variable transformation. Let \( k = \tilde{g}^{-1} \) be the inverse function associated with \( \tilde{g} \):
\[
k(\theta_s) = \tilde{g}^{-1}(\theta_s) = \theta.
\] (4.5)

Set
\[
\tilde{G}(\theta) = \frac{\partial \tilde{g}(\theta)}{\partial \theta} \quad \text{and} \quad K(\theta_s) = \frac{\partial k}{\partial \theta_s}. \quad (4.6)
\]

Since \( k[\tilde{g}(\theta)] = \theta \) and \( \tilde{g}[k(\theta_s)] = \theta_s \), we have by the chain rule of differentiation
\[
K[\tilde{g}(\theta)] \tilde{G}(\theta) = I_p \quad \text{and} \quad \tilde{G}[k(\theta_s)] K(\theta_s) = I_p, \quad \forall \theta_s \in \Theta_s, \forall \theta \in \Theta. \quad (4.7)
\]

Let
\[
\psi^*(\theta_s) = \psi[\tilde{g}^{-1}(\theta_s)].
\] (4.8)

Clearly,
\[
\psi^*(\theta_s) = 0 \iff \psi(\theta) = 0, \quad (4.9)
\]
and $H_0^n : \psi^* (\theta_s) = 0$ is an equivalent reformulation of $H_0 : \psi (\theta) = 0$ in terms of $\theta_s$. We shall call $\psi^* (\theta_s) = 0$ the canonical reformulation of $\psi (\theta) = 0$ in terms of $\theta_s$. Other (possibly more “natural”) reformulations are of course possible, but the latter has the convenient property that $\psi^* (\theta_s) = \psi (\theta)$. If a test statistic is invariant to reparameterizations when the null hypothesis is reformulated as $\psi^* (\theta_s) = 0$, we will say it is canonically invariant.

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

$$P_\psi (\theta_s) \equiv \frac{\partial \psi^*}{\partial \theta_s'} = \frac{\partial \psi}{\partial \theta'} \frac{\partial \theta'}{\partial \theta_s'} = P [k (\theta_s)] K (\theta_s) = P (\theta) K [\tilde{g} (\theta)] . \quad (10.40)$$

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 $D_n (\theta; Z_n) = \sum_{i=1}^n h_1 (\theta; z_1) / n$ as in (3.3) where the values of the scores are unaffected by the reparameterization, but are simply reexpressed in terms of $\theta_s$ and $z_{ts}$ (invariant scores)

$$h_1 (\theta_s; z_{ts}) = h_1 (\theta; z_1), \quad t = 1, \ldots, n , \quad (11.41)$$

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

Under condition (4.11), we see easily that

$$H_{ns} (\theta_s; Z_{ns}) = \frac{\partial D_{ns} (\theta_s; Z_{ns})}{\partial \theta_s'} = H_n (\theta; Z_n) K (\theta_s) = H_n (\theta; Z_n) K [\tilde{g} (\theta)] . \quad (12.42)$$

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

$$\tilde{I}_s (\theta_s) = \tilde{I} (\theta) , \quad \tilde{J}_s (\theta_s) = \tilde{J} (\theta) K [\tilde{g} (\theta)] . \quad (13.43)$$

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

**Proposition 4.2 (C(α) Canonical Invariance to Reparameterizations: Invariant Score Case).** Let $\psi^* (\theta_s) = \psi \left[ g^{-1} (\theta_s) \right]$, and suppose the following conditions hold:

(a) $\tilde{\theta}_n^0 = \tilde{g} (\tilde{\theta}_n^0)$;
(b) $D_{ns} (\tilde{\theta}_n^0; Z_{ns}) = D_n (\tilde{\theta}_n^0; Z_n)$;
(c) $\tilde{I}_0 = \tilde{I}_0$ and $\tilde{J}_0 = \tilde{J}_0 K$;
(d) $W_{ns} = W_n$.

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

$$PC_\psi (\tilde{\theta}_n^0; \psi^*, W_{ns}) \equiv n \tilde{D}_{ns} \tilde{Q}_{0*} (\tilde{Q}_{0*} \tilde{J}_0 \tilde{Q}_{0*})^{-1} \tilde{Q}_{0*} \tilde{D}_{ns} = PC (\tilde{\theta}_n^0; \psi, W_n),$$

where $\tilde{D}_{ns} = D_{ns} (\tilde{\theta}_n^0; Z_{ns})$, $\tilde{Q}_{0*} = \tilde{P}_{0*} (\tilde{J}_0 \tilde{W}_n \tilde{J}_0)^{-1} \tilde{J}_0 \tilde{W}_n$, $\tilde{P}_{0*} = P_\psi (\tilde{\theta}_n^0)$, and $P_\psi (\theta_s)$ is the derivative of $\psi^*$ with respect to $\theta_s'$.

It is clear that the estimators $\hat{\theta}_n$ and $\tilde{\theta}_n^0$ satisfy the equivariance condition, i.e., $\hat{\theta}_n = \tilde{g} (\hat{\theta}_n)$ and $\tilde{\theta}_n^0 = \tilde{g} (\tilde{\theta}_n^0)$. Consequently, the above invariance result also applies to score (or LM) statistics. It is also interesting to observe that $W_n (\psi^*) = W (\psi)$. This holds, however, only for the special reformulation $\psi^* (\theta_s) = \psi \left[ g^{-1} (\theta_s) \right] = 0$, not for all equivalent reformulations $\psi^* (\theta_s) = 0$. On applying Proposition 4.1, this type of invariance holds for the other test statistics. These observations are summarized in the following proposition.
Theorem 4.3 (Test Invariance to Reparameterizations and General Hypothesis Reformulations: Invariant Score Case). Let \( \psi_* : \Theta_* \to \Theta \) be any continuously differentiable function of \( \theta_* \in \Theta_* \) such that \( \psi_* (\tilde{g} (\theta)) = 0 \Leftrightarrow \psi (\theta) = 0 \), let \( m = p \), and suppose:

(a) \( D_{n*} (\tilde{g} (\theta); Z_{n*}) = D_{n*} (\theta; Z_n) \);
(b) \( \hat{T}_n [\tilde{g} (\theta)] = \tilde{T} (\theta) \) and \( \hat{T}_n [\hat{g} (\theta)] = \tilde{T} (\theta) K [\tilde{g} (\theta)] \).

Here \( K (\theta) = \partial \tilde{g}^{-1} (\theta) / \partial \theta^\prime_* \). 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 \( \hat{\theta}^0_{n*} = \tilde{g} (\hat{\theta}^0_n) \), we also have

\[
PC_* (\hat{\theta}^0_{n*}, \psi_*) = PC (\hat{\theta}^0_n, \psi).
\]

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

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

\[
D_n (\theta; Z_n) = \frac{\partial S_n (\theta; Z_n)}{\partial \theta}.
\]

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

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

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

\[
\sqrt{n} D_{n*} (\theta_0*; Z_{n*}) \xrightarrow{L} \mathcal{N} [0, I_*(\theta_0*)], \tag{4.18}
\]

where \( \theta_0* = \tilde{g} (\theta_0) \) and

\[
I_*(\theta_0*) = K (\theta_*)' I [k (\theta_0)] K (\theta_*) = K [\tilde{g} (\theta)]' I (\theta) K [\tilde{g} (\theta)]. \tag{4.19}
\]

Further,

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

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

\[
K^{(1)}_i (\theta_*) = \frac{\partial^2 \theta_i}{\partial \theta_* \partial \theta^\prime_*} (\theta_*) = \frac{\partial^2 k_i}{\partial \theta_* \partial \theta^\prime_*} (\theta_*). \tag{4.21}
\]

By a set of arguments analogous to those used in Dagenais and Dufour (1991), it appears that all the statistics (except the LR-type statistic) are based upon \( H_n \) and so they are sensitive to a reparameterization, unless some specific estimator of \( J \) is used. At this level of generality, the following results can be presented using the following notations: \( \hat{I}, \hat{J}, \hat{P} \) are the estimated matrices for a parameterization in \( \theta \) and \( \hat{I}_*, \hat{J}_*, \hat{P}_* \) are the estimated matrices for a parameterization in \( \theta_* \). 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(α) Canonical Invariance to Reparameterizations). Let \( \psi^* (\theta_n) = \psi [\tilde{g}^{-1} (\theta_n)] \), and suppose the following conditions hold:

(a) \( \tilde{\theta}_{n0} = \tilde{g} (\hat{\theta}_{n0}) \);
(b) \( D_{n*} (\tilde{\theta}_{n0}; Z_{n*}) = K \left[ \tilde{g} (\hat{\theta}_{n0}) \right]' D_n (\theta; Z_n) \);
(c) \( \tilde{I}_{n0} = \tilde{K} \tilde{I}_0 \tilde{K} , \tilde{I}_{n0} = \tilde{K}^{-1} \tilde{I}_0 \tilde{K} ; \)
(d) \( W_{n*} = \tilde{K}^{-1} W_n (\tilde{K}^{-1})' \).

Here \( \tilde{I}_{n0} \), \( \tilde{I}_0 \), and \( W_n \) are defined as in (3.24), and \( \tilde{K} = K (\hat{\theta}_{n*}) \). Then, provided the relevant matrices are invertible,

\[
PC_n (\tilde{\theta}_{n*}; \psi^*, W_{n*}) = PC (\hat{\theta}_{n*}; \psi, W_n).
\]

**Theorem 4.5** (Test Invariance to Reparameterizations and General Equivalent Hypothesis Reformulations). Let \( \psi_n : \Theta_n \rightarrow \Theta \) be any continuously differentiable function of \( \theta_n \in \Theta_n \) such that \( \psi_n [\tilde{g} (\theta)] = 0 \iff \psi (\theta) = 0 \), let \( m = p \), and suppose:

(a) \( D_{n*} (\tilde{g} (\theta); Z_{n*}) = K \left[ \tilde{g} (\hat{\theta}_{n0}) \right]' D_n (\theta; Z_n) \);
(b) \( \tilde{T}_n [\tilde{g} (\theta)] = K \left[ \tilde{g} (\hat{\theta}_{n0}) \right]' \tilde{T} (\theta) K \left[ \tilde{g} (\hat{\theta}_{n0}) \right] \);
(c) \( \tilde{f}_n [\tilde{g} (\theta)] = K \left[ \tilde{g} (\hat{\theta}_{n0}) \right]' \tilde{f} (\theta) K \left[ \tilde{g} (\hat{\theta}_{n0}) \right] \).

Here \( K (\theta_n) = \partial \tilde{g}^{-1} (\theta) / \partial \theta_n \). Then, provided the relevant matrices are invertible, we have

\[
T (\psi) = T_n (\psi_n),
\]
where \( T \) stands for any one of the test statistics \( S (\psi) , LM (\psi) , LR (\psi) \), and \( D_{NW} (\psi) \). If \( \hat{\theta}_{n*} = \tilde{g} (\hat{\theta}_{n0}) \), we also have

\[
PC_n (\tilde{\theta}_{n*}; \psi_n) = PC (\hat{\theta}_{n*}; \psi),
\]
and, in the case where \( \psi_n (\theta) = \psi [\tilde{g}^{-1} (\theta)] \),

\[
W_n (\psi_n) = W (\psi).
\]

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

\[
h_{n*} (\tilde{g} (\theta); z_{n*}) = K \left[ \tilde{g} (\hat{\theta}_{n0}) \right] h_t (\theta; z_t), \quad t = 1, \ldots, n,
\]
where \( D_{n*} (\tilde{g} (\theta); Z_{n*}) = \frac{1}{n} \sum_{t=1}^n h_{n*} (\tilde{g} (\theta); z_{n*}) \). Consequently, in such a case, any estimator \( \tilde{T} (\theta) \) of the general form (3.5) will satisfy (b) provided the matrix \( W_f (n) \) remains invariant under reparameterizations. This will be the case, in particular, for most HAC estimators of the form (3.8) as soon as the bandwidth parameter \( B_n \) only depends on the sample size \( n \). However, this may not hold if \( B_n \) is data-dependent (as considered in Andrews and Monahan, 1992).

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

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

5.1. Modified Newey–West LR-type statistic

Consider the LR-type statistic

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

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

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 \(\theta\). Instead of estimating \(\theta\) by minimizing \(M_n(\theta, \hat{I}_0)\), estimate \(\theta\) by minimizing \(M_n(\theta, \hat{I}(\theta))\). For example, such an estimation method was studied by Hansen et al. (1996). When the score vector \(D_n\) and the parameter vector \(\theta\) have the same dimension \((m = p)\), the unrestricted objective function will typically be zero \([D_n(\hat{\theta}_n; Z_n) = 0]\), so the statistic reduces to \(D_{NW}(\psi) = nM_n(\hat{\theta}_n^0, \hat{I}_0)\). When \(m > p\), this will typically not be the case.

Suppose now the following conditions hold:

\[
D_{ns}(\tilde{g}(\theta); Z_{ns}) = K\left[\tilde{g}(\theta)\right]'D_n(\theta; Z_n),
\]

\[
\tilde{I}_s(\tilde{g}(\theta)) = K\left[\tilde{g}(\theta)\right]'\hat{I}(\theta)K\left[\tilde{g}(\theta)\right].
\]

Then, for \(\theta_s = \tilde{g}(\theta)\),

\[
M_{ns}(\theta_s; \tilde{I}_s(\theta_s)) = D_{ns}(\tilde{g}(\theta); Z_{ns})'\tilde{I}_s(\tilde{g}(\theta))^{-1}D_{ns}(\tilde{g}(\theta); Z_{ns}) = D_n(\theta; Z_n)'\hat{I}(\theta)^{(-1)}D_n(\theta; Z_n).
\]

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

\[
J = nM_n(\hat{\theta}_n; \hat{I}(\hat{\theta}_n)),
\]

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

are invariant to reparameterizations of the type considered in (4.17)–(4.19). 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 \(\overline{D}\) and \(D_{NW}\) are asymptotically equivalent (at least under the null hypothesis) and so have the same asymptotic \(\chi^2(p_1)\) distribution.\(^5\)

---

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

\(^5\)The regularity conditions and a proof of the asymptotic distribution are given in our working paper (Dufour et al., 2013).
5.2. Pseudo-maximum likelihood (PML) methods

5.2.1. PML methods

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

$$E(y_t | x_t) = f_t(x_t; \theta), \ V(y_t | x_t) = \Omega_0(x_t), \quad (5.6)$$

where $f_t(\theta)$ is a known function and $\theta$ is the parameter of interest. (5.6) provides a nonlinear generalized regression model with unspecified variance. Even if a likelihood function with a finite number of parameters is not available for such a semiparametric model, $\theta$ can be estimated through a pseudo-maximum likelihood technique (PML) which consists in maximizing a chosen likelihood as if it were the true undefined likelihood; see Gouriéroux et al. (1984c).

In particular, it is shown in the latter reference that this pseudo-likelihood must belong to the specified class of linear exponential distributions adapted for the mean. These distributions have the following general form:

$$l(y; \mu) = \exp \left[ A(\mu) + B(y) + C(\mu)y \right], \quad (5.7)$$

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

$$\frac{\partial A}{\partial \mu} + \frac{\partial C}{\partial \mu} \mu = 0. \quad (5.8)$$

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

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

or equivalently through the following equivalent program:

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

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 program (5.9) ensures that the expectation of the linear exponential pseudo-distribution is $\mu$. The pseudo-likelihood equations have an orthogonal condition form

$$D_n(\theta) = \sum_{t=1}^{n} \frac{\partial f_t^\prime}{\partial \theta} \frac{\partial C}{\partial \mu}(f_t(\theta))(y_t - f_t(\theta)) = 0. \quad (5.10)$$

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

$$J(\theta) = E_x \left\{ \left( \frac{\partial f_t^\prime}{\partial \theta} \right) \left[ \frac{\partial C}{\partial \mu}(f_t(\theta)) \right]^\prime \left( \frac{\partial f_t^\prime}{\partial \theta} \right)^\prime \right\}, \quad (5.11)$$

$$I(\theta) = E_x \left\{ \left( \frac{\partial f_t^\prime}{\partial \theta} \right) \left[ \frac{\partial C}{\partial \mu}(f_t(\theta)) \right] \Omega_0 \left( \frac{\partial C}{\partial \mu}(f_t(\theta)) \right)^\prime \left( \frac{\partial f_t^\prime}{\partial \theta} \right)^\prime \right\}. \quad (5.12)$$

These matrices can be estimated by

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

(5.13)

\[
\hat{\mu} = \frac{1}{n} \sum_{t=1}^{n} S_t(\hat{\theta}) S_t(\hat{\theta})',
\]

(5.14)

where

\[
S_t(\hat{\theta}) = \left( \frac{\partial f_t}{\partial \hat{\theta}} \right) \left( \frac{\partial C}{\partial \hat{\mu}} (f_t(\hat{\theta})) \right) (y - f_t(\hat{\theta})).
\]

(5.15)

Since \( \frac{\partial C}{\partial \hat{\mu}} (f_t(\hat{\theta})) \) and \( y_t - f_t(\hat{\theta}) \) are invariant to reparameterizations, \( \hat{T} \) and \( \hat{\mu} \) are modified only through \( \frac{\partial f_t}{\partial \hat{\theta}} \). Further,

\[
f_t^* (\theta_*) = f_t^* \left[ \hat{g}(\theta) \right] = f_t (\theta), \quad \frac{\partial f_t^*}{\partial \theta_*} = \left( \frac{\partial \hat{g}}{\partial \theta} \right) \left( \frac{\partial \hat{g}}{\partial \theta_*} \right) = \left( \frac{\partial f_t}{\partial \theta} \right) K \left[ \hat{g}(\theta) \right]
\]

(5.16)

and

\[
\hat{T}_* = K \left[ \hat{g}(\hat{\theta}) \right] K \left[ \hat{g}(\hat{\theta}) \right], \quad \hat{\mu}_* = K \left[ \hat{g}(\hat{\theta}) \right] K \left[ \hat{g}(\hat{\theta}) \right].
\]

(5.17)

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

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

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

\[
- \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 f_t}{\partial \hat{\theta} \partial \theta_*} (\hat{\theta}) \frac{\partial C}{\partial \hat{\mu}} (f_t(\hat{\theta})) (y_t - f_t(\hat{\theta})).
\]

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

5.2.2. Quasi generalized PML (QGPML) methods

Gouriéroux et al. (1984c) 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 \( \eta \) linked with the second order moment of the pseudo-distribution. If \( \mu \) and \( \Sigma \) are the expectation and the variance-covariance matrix of this pseudo-distribution: \( \eta = \Psi(\mu, \Sigma) \), where \( \Psi \) defines for any \( \mu \), a one-to-one relationship between \( \eta \) and \( \Sigma \).
The class of linear exponential distributions depending upon the extra parameter \( \eta \) is of the following form:

\[
I^n(y, \mu, \eta) = \exp[A(\mu, \eta) + B(\eta, y) + C(\mu, \eta)y].
\]

If we consider the negative binomial pseudo distribution \( A(\mu, \eta) = -\eta \ln \left( 1 + \frac{\mu}{\eta} \right) \) and \( C(\mu, \eta) = \ln \left( \mu/(\mu + \eta) \right) \); 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 \( \hat{\alpha}, \hat{\theta} \) of \( \alpha, \theta \) where \( \hat{\theta} \) and \( \alpha \) are equivariant with respect to \( \tilde{g} \), computed for example as in Trognon (1984), the QGPML estimator of \( \theta \) is obtained by solving a problem of the type

\[
\max_\theta \sum_{i=1}^n I^n(y_t, f_t(\theta), \Psi(f_t(\tilde{\theta}), g_t(\tilde{\alpha}))).
\]

The QGPML estimator \( \hat{\theta} \) of \( \theta \) is strongly consistent and asymptotically normal: \( \sqrt{n}(\hat{\theta} - \theta_0) \overset{L}{\rightarrow} N[0, \Sigma_Q] \) with

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

\( I_0 \) and \( J_0 \) can be consistently estimated by

\[
\hat{I} = \frac{1}{n} \sum_{i=1}^n S_i(\hat{\theta}, \tilde{\alpha}, \tilde{\theta}) S_i(\hat{\theta}, \tilde{\alpha}, \tilde{\theta})', \quad \hat{J} = \frac{1}{n} \sum_{i=1}^n \frac{\partial f_t}{\partial \theta} (\hat{\theta}) \left[ \frac{\partial C}{\partial \mu} (f_t(\hat{\theta}), f_t(\tilde{\theta}), g_t(\tilde{\alpha})) \right] \frac{\partial f_t}{\partial \theta} (\hat{\theta}),
\]

where

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

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

\[
\hat{I}_s = K[\tilde{g}(\tilde{\theta})]' \hat{I} K[\tilde{g}(\tilde{\theta})], \quad \hat{J}_s = K[\tilde{g}(\tilde{\theta})]' \hat{J} K[\tilde{g}(\tilde{\theta})],
\]

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 as shown in Trognon (1984), the QGLR statistic is asymptotically equivalent to the other pseudo-asymptotic statistic under the null and under local alternatives.

6. Numerical results

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

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

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

\[
H_0 : \lambda = 1.
\]
The log-likelihood associated with this model is

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

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

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

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

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

\[ y_t = y_s + \beta_1 x_{1*}^{(\lambda)} + \beta_2 x_{2*}^{(\lambda)} + u_t, \tag{6.6} \]

where the power parameter \( \lambda \) remains the same and

\[ y_s = \gamma - k^{(\lambda)} k^{-\lambda} \sum_{i=1}^{2} \beta_i, \quad \beta_i* = \beta_i k^{-\lambda}, \quad i = 1, 2. \tag{6.7} \]

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

\[ \mathbb{E}[u_t] = 0, \quad \mathbb{E}[u_t x_{1*}^{(\lambda)}] = 0, \quad \mathbb{E}[u_t x_{2*}^{(\lambda)}] = 0, \tag{6.8} \]

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

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

\[ \mathbb{E}[u_t x_{1t} x_{2t}] = 0. \tag{6.10} \]

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

²The numerical values of \( x_{1t}, x_{1*}, \) 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.
Numerical values of the GMM-based test statistics for a number of rescalings are reported in Table 1 for the five moment system (6.8)–(6.9) and in Table 2 for the six-moment system (6.8)–(6.10). Results for the pseudo-likelihood tests appear in Table 1. Graphs of the noninvariant test statistics are also presented in Figs. 1–3. 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 $W_n = I_m$ (equal weights), while the second step uses the weight matrix defined in (3.4). No correction for serial correlation is applied (although this could also be studied).

These results confirm the theoretical expectations of the theory presented in the previous sections. Namely, the GMM-based test statistics $[\bar{D}(\psi), \text{Wald}, \text{Score}, C(\alpha)]$ 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 (six equations). In contrast, the $\bar{D}(\psi)$ and score tests based on the continuously updated GMM criterion are invariant. The same holds for the LR and adjusted score criteria based on linear exponential pseudo likelihoods.

### 7. Empirical illustration: Linear stochastic discount factor models

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

\[
E\left[m_t R_t^e\right] = 0, \quad (7.1)
\]

\[
m_t = a - f_t^{'b}, \quad (7.2)
\]

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

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

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

**Normalization 1 :** \[ E\left[\frac{m_t}{a} R_t^e\right] = 0 \]

**Normalization 2 :** \[ E\left[\frac{m_t}{E[m_t]} R_t^e\right] = 0. \]

<table>
<thead>
<tr>
<th>$k$</th>
<th>Two-step GMM</th>
<th></th>
<th></th>
<th>CUP-GMM</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\bar{D}$</td>
<td>Wald</td>
<td>Score</td>
<td>$C(\alpha)$</td>
<td>$\bar{D}$</td>
<td>Wald</td>
</tr>
<tr>
<td>0.2</td>
<td>0.016</td>
<td>416.546</td>
<td>106.734</td>
<td>54.462</td>
<td>19.480</td>
<td>359.380</td>
</tr>
<tr>
<td>0.4</td>
<td>0.036</td>
<td>221.829</td>
<td>108.142</td>
<td>54.852</td>
<td>19.480</td>
<td>40.481</td>
</tr>
<tr>
<td>0.6</td>
<td>1.068</td>
<td>213.918</td>
<td>107.764</td>
<td>52.818</td>
<td>19.480</td>
<td>34.101</td>
</tr>
<tr>
<td>0.8</td>
<td>1.068</td>
<td>178.757</td>
<td>106.053</td>
<td>47.539</td>
<td>19.480</td>
<td>34.101</td>
</tr>
<tr>
<td>1.0</td>
<td>3.562</td>
<td>139.364</td>
<td>103.364</td>
<td>37.915</td>
<td>19.480</td>
<td>35.580</td>
</tr>
<tr>
<td>3.0</td>
<td>47.490</td>
<td>46.214</td>
<td>110.751</td>
<td>7.960</td>
<td>19.480</td>
<td>45.146</td>
</tr>
</tbody>
</table>
Figure 1. Two-step GMM tests based on five moment conditions.

Figure 2. Two-step GMM tests based on six moment conditions.
By applying the normalizations to (7.3), we have

$$E \left[ (1 - f_t' \theta) R_t^c \right] = 0, \quad E \left[ (1 - (f_t - \mu f)' \theta_s) R_t^c \right] = 0,$$

(7.4)

where $\mu f = E[f_t], \theta = b/a$, and $\theta_s = b/E[m_1]$. The implied two sets of sample moments are

$$D_n(\theta; Z_n) = \left( \frac{1}{n} \sum_{t=1}^{n} (R_t^c - R_t^{c'} \theta) \right), \quad D_n(\theta_s; Z_{n s}) = \left( \frac{1}{n} \sum_{t=1}^{n} (R_t^c - R_t^c (f_t - \mu f)' \theta_s) \right).$$

It is clear that the sample moments satisfy

$$D_n(\bar{g}(\theta); Z_{n s}) = K \left[ \bar{g}(\theta) \right]' D_n(\theta; Z_n)$$

with $K \left[ \bar{g}(\theta) \right] = \text{diag}[a/E[m_1], 1]$; one set of moments can be derived from the other by affine transformation of $f_t$. Let $\hat{T}_n(\theta_s)$ be the HAC estimator of $I(\theta_s)$ with Bartlett kernel and $\hat{T}(\theta)$ be defined similarly. Then we have

$$\hat{T}_n(\bar{g}(\theta)) = K \left[ \bar{g}(\theta) \right]' \hat{T}(\theta) K \left[ \bar{g}(\theta) \right].$$

Therefore, by virtue of Eq. (5.3), the continuously updated GMM (CUP-GMM) objective function and the statistic $\bar{D}$ are invariant to affine transformation of $f_t$, i.e., they are not affected by the form of normalization employed. The model is estimated using the observed returns on five stocks: Weis Markets (WMK), Unisys Corporation (UIS), Orbital Sciences Corporation (ORB), Mattel (MAT) and Abaxis (ABAX), and the three factors $Rm-Rf$, Small [market capitalization] Minus Big (SMB), and High [book-to-market ratio] Minus Low (HML), from the Fama–French data set over the period from January 5th, 1993–March 16th, 1993. All calculations were carried out in R Version 3.0.2 (R Development Core Team (2013)) using the package gmm developed by Pierre Chaussé (Chaussé,
Table 3. \( J \) statistic for the validity of (7.4) under different identifying restrictions (\( p \)-values in parentheses).

<table>
<thead>
<tr>
<th></th>
<th>Normalization 1</th>
<th>Normalization 2</th>
<th>CUP-GMM</th>
<th>Normalization 1</th>
<th>Normalization 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-step GMM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QS Bartlett</td>
<td>0.218 ( (0.897) )</td>
<td>0.106 ( (0.949) )</td>
<td>17.214 ( (0.000) )</td>
<td>5.670 ( (0.059) )</td>
<td></td>
</tr>
<tr>
<td>QS Bartlett</td>
<td>1.235 ( (0.539) )</td>
<td>1.774 ( (0.412) )</td>
<td>1.759 ( (0.415) )</td>
<td>1.866 ( (0.393) )</td>
<td></td>
</tr>
</tbody>
</table>

The data we use are readily available in the Finance data set contained in gmm. The estimation methods are two-step GMM and CUP-GMM with covariance matrix estimated with Bartlett and Quadratic Spectral (QS) kernels. Table 3 reports the values of \( J \) statistic for testing the validity of the restrictions (7.4). For the two-step GMM, it is clear that the values of test statistics differ greatly across the normalizations, and are sensitive to the choice of kernels. Furthermore, the test rejects the null of correct specification under Normalization 2 with QS kernel, but the conclusion is reversed under Normalization 1. In the case of CUP-GMM with Bartlett kernel, though there is a small incongruity in the values of test statistics (possibly due to an optimization error), the model is not rejected under both normalizations. The difference between test statistics under the CUP-GMM with QS kernel may be attributed to the noninvariance of the objective function with QS kernel. The main message of this exercise is that procedures based on noninvariant test statistics can be quite sensitive to the identifying restrictions employed and may result in conflicting conclusions. For a thorough discussion on the effect of normalizations on estimation and inferences, we refer the reader to Hamilton et al. (2007).

8. Conclusion

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

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

Acknowledgements

The authors thank Marine Carrasco, Jean-Pierre Cotton, Russell Davidson, Abdeljelil Farhat, V. P. Godambe, Christian Gouriéroux, Stéphane Grégoir, Hervé Mignon, Denis Pelletier, Mohamed Taamouti, Pascale Valéry, three anonymous referees, and the Editor Esfandiar Maasoumi for several useful comments.
Funding

This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universitat Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Canadian Network of Centres of Excellence [program on Mathematics of Information Technology and Complex Systems (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

References


