Economics Letters 38 (1992) 251–257
North-Holland

On the lack of invariance of some asymptotic tests to rescaling *

Marcel G. Dagenais and Jean-Marie Dufour
Université de Montréal, Montreal, Que., Canada

Received 27 August 1991
Accepted 20 December 1991

Fourteen different tests applicable in nonlinear models are shown not to be invariant to irrelevant changes in measurement units of the variables. The tests studied include procedures suggested by Durbin, Sargan–Mehta, Hausman, and White (e.g., the information matrix test).

1. Introduction

Statistical inference should not depend on incidental elements which have no impact on the decision under consideration, such as the selection of measurement units or equivalent ways of expressing a parameter constraint; see Lehmann (1986, chap. 6). In the classical linear regression model, t and F-tests of linear restrictions possess such invariance properties. In nonlinear models or for nonlinear constraints, only asymptotic tests are usually available and several alternative procedures can be used. The invariance properties of these are more complex and less well known.

The non-invariance of asymptotic Wald tests to equivalent formulations of nonlinear null hypotheses is already well documented [see, e.g., Gregory and Veall (1985) and Dagenais and Dufour (1991)]. Furthermore, Wald tests are not invariant to rescaling in regression models which involve Box–Cox transformations [Spitzer (1984), Dagenais and Dufour (1991)]. Dagenais and Dufour (1991) also showed that $C(\alpha)$ tests [(Neyman (1959), Smith (1987)] as well as a frequently used form of the Lagrange multiplier (LM) test, in which the Hessian of the log-likelihood function is used to estimate the asymptotic covariance matrix of the score vector, are not invariant to rescaling. Likelihood ratio (LR) tests are clearly invariant to such changes.

Several other test procedures have been suggested by econometricians. In this paper, we show that the non-invariance to measurement unit changes that plague Wald tests and certain variants of the Lagrange multiplier test also affect several other asymptotic tests widely used in econometrics. The tests studied include Durbin's (1970) procedure, Sargan and Mehta's (1983) generalization of

Correspondence to: Jean-Marie Dufour, Département de sciences économiques and C.R.D.E., Université de Montréal, C.P. 6128, Succ. A, Montréal, Québec, H3C 3J7, Canada.
* This work was supported by the Centre de recherche et développement en économique (C.R.D.E.) of the Université de Montréal, the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada and the Fonds F.C.A.R. (Government of Québec). We thank Eric Ghysels and Pierre Perron for helpful discussions.

the latter, Hausman's (1978) specification test as well as five 'robust' tests suggested by White (1982) [pseudo-Wald, pseudo-LM, information matrix, Hausman-type and gradient tests]. Further, for the three first procedures, three alternative ways of estimating the information matrix (Hessian, outer product, exact formula) are considered, so that 14 different test statistics are examined overall. To study the invariance properties of these tests, we consider a simple nonlinear regression model with Box–Cox transformations on the explanatory variables. Of the 14 test statistics computed, none is found to be invariant to rescaling of model variables. In several cases, changing measurement units can lead to dramatic changes in the values of the test criteria and thus also in the conclusions drawn from the tests. Although this is not illustrated here, it is not difficult to demonstrate that the same lack of invariance to changes in measurement units will also affect regression models with Box–Cox transformations on the dependent variable and/or various power transformations on the explanatory variables, such as the CES or VES production functions.

The main test criteria studied in this paper are described in section 2. In section 3, we report the numerical illustrations of non-invariance to measurement unit changes. Section 4 contains a few concluding remarks.

2. Description of the tests

Let us consider a general statistical model with independent observations and log-likelihood function

$$L(\theta;\, Z) = \log[\,p(y \mid X,\, \theta)] = \sum_{t=1}^{n} \log[q(y_t \mid x_t,\, \theta)] = \sum_{t=1}^{n} l_t, \tag{1}$$

where $Z = [y,\, X]$, $y = [y,\, y_2, \dots, y_n]'$, $X = [x_1,\, x_2, \dots, x_n]'$, y_t is a $m \times 1$ random vector ('dependent variables'), x_t is a $k \times 1$ vector of fixed (or strictly exogenous) variables ($t = 1, \dots, n$), θ is a $p \times 1$ vector of fixed parameters in the space Ω, n is the number of observations, $y \in U_0$, $X \in U_1$ and $Z \in \Upsilon = U_0 \times U_1$; U_0 and U_1 are the sets of $n \times m$ and $n \times k$ matrices where y and X can take their values; $p(y \mid X,\, \theta) = \prod_{t=1}^{n} q(y_t \mid x_t,\, \theta)$ is the density function of y given X and θ, $q(y_t \mid x_t,\, \theta)$ is the conditional density function of y_t given x_t and $l_t = \log[q(y_t \mid x_t,\, \theta)]$. [1] We suppose that the probability distributions corresponding to different values of θ are distinct (identification condition). Let also $D_t \equiv D_t(\theta;\, Z) = \partial l_t / \partial\theta$,

$$D(\theta;\, Z) = \partial L(\theta;\, Z)/\partial\theta = \sum_{t=1}^{n} D_t, \tag{2}$$

$$H(\theta;\, Z) = \frac{1}{n}\frac{\partial^2 L}{\partial\theta\, \partial\theta'} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t}{\partial\theta\, \partial\theta'}, \qquad I(\theta) = \mathrm{E}_\theta\left[\frac{1}{n}\sum_{t=1}^{n} D_t D_t'\right] = -\mathrm{E}_\theta[H(\theta;\, Z)], \tag{3}$$

where $D(\theta;\, Z)$ and $H(\theta;\, Z)$ have dimensions $p \times 1$ and $p \times p$. The information matrix corresponding to the log-likelihood function $L(\theta;\, Z)$ is $I(\theta) \equiv I(\theta;\, X) = -\mathrm{E}_\theta[H(\theta;\, Z)]$, where the expected

[1] We could also allow x_t to include lagged dependent variables. However, to simplify the exposition, we will assume that x_t is strictly exogenous.

value $E_\theta(\cdot)$ is taken with respect to the distribution of y (conditional on X) when the true parameter is θ.

Under standard regularity conditions [see, e.g., Burguete, Gallant and Souza (1982) and Lehmann (1983, chap. 6)], a consistent maximum likelihood (ML) estimator $\hat{\theta}$ exists and both $D(\theta;\, Z)$ and $\hat{\theta}$ have asymptotic normal distributions:

$$n^{-1/2}D(\theta;\, Z) \to \mathrm{N}\left[0,\, \bar{I}(\theta)\right], \qquad n^{1/2}(\hat{\theta}-\theta) \to \mathrm{N}\left[0,\, \bar{I}(\theta)^{-1}\right], \tag{4}$$

where $\bar{I}(\theta) = \lim_{n\to\infty} I(\theta)$. Depending on circumstances, three alternative consistent estimators of $\bar{I}(\theta)$ are usually considered:

$$\hat{I}(\tilde{\theta})_1 = -H(\tilde{\theta};\, Z), \qquad \hat{I}(\tilde{\theta})_2 = \frac{1}{n}\sum_{t=1}^{n} D_t(\tilde{\theta};\, Z)D_t(\tilde{\theta};\, Z)', \qquad \hat{I}(\tilde{\theta})_3 = I(\tilde{\theta}), \tag{5}$$

where $\tilde{\theta}$ is a consistent estimator of θ. $\hat{I}(\tilde{\theta})_1$ is the Hessian estimator of $I(\theta)$, $\hat{I}(\tilde{\theta})_2$ is the outer product estimator, and $\hat{I}(\tilde{\theta})_3$ is the exact information matrix (evaluated at $\tilde{\theta}$). Which estimator is the most convenient depends on the context. In particular, $\hat{I}(\tilde{\theta})_3$ requires one to compute the expected value of $H(\theta;\, Z)$, which is difficult in many problems. In the sequel, the symbol $\hat{I}(\tilde{\theta})$, with no subscript, will refer to any of the three estimators in (5).

Let $\theta = (\theta_1',\, \theta_2')'$, where θ_1 and θ_2 are $p_1 \times 1$ and $p_2 \times 1$ subvectors of θ and $p_1 + p_2 = p$. Consider the problem of testing the null hypothesis $\mathrm{H}_0 : \theta_1 = \theta_1^0$. As pointed out above, we will concentrate here on the following criteria: Durbin's (1970) procedure, Sargan and Mehta's (1983) generalization of the latter, Hausman-type tests [Hausman (1978)] and a number of tests suggested by White (1982). Since ambiguities can easily arise, we now define some of these test criteria in a common notation. Unless stated otherwise, the asymptotic null distribution of each of the test statistics described below is $\chi^2(p_1)$.

Given $\hat{\theta}^0 = (\theta_1^{0\prime},\, \hat{\theta}_2^{0\prime})'$ the restricted ML estimator of θ, let $\tilde{\theta}_1$ be the estimator of θ_1 obtained by maximizing $L(\theta_1,\, \hat{\theta}_2^0)$ with respect to θ_1. The test criterion suggested by Durbin is

$$K(\tilde{\theta},\, \hat{\theta}^0) = n\left(\tilde{\theta}_1 - \theta_1^0\right)'\left[\hat{I}_{11}^{-1} - \hat{I}_{11}^{-1}\hat{I}_{12}\hat{I}_{22}^{-1}\hat{I}_{21}\hat{I}_{11}^{-1}\right]^{-1}\left(\tilde{\theta}_1 - \theta_1^0\right), \tag{6}$$

where $\tilde{\theta} = (\tilde{\theta}_1',\, \hat{\theta}_2^{0\prime})'$ and $\hat{I} = \hat{I}(\hat{\theta}^0) = [\hat{I}_{\cdot\cdot}]$ is partitioned conformably with $(\theta_1',\, \theta_2')'$.

Sargan and Mehta's (1983) generalized Durbin method is obtained by partitioning $\theta_2 = (\theta_{21}',\, \theta_{22}')'$ and allowing one of the subvectors to be reestimated in the second step of the procedure. Given $\hat{\theta}^0 = (\theta_1^{0\prime},\, \hat{\theta}_{21}^0,\, \hat{\theta}_{22}^{0\prime})'$ the restricted ML estimate of θ, let $\bar{\theta}_1$ and $\bar{\theta}_{21}$ be the estimates obtained by maximizing $L(\theta_1,\, \theta_{21},\, \hat{\theta}_{22}^0)$ with respect to θ_1 and θ_{21}. For testing $\theta_1 = \theta_1^0$, the criterion suggested by Sargan and Mehta takes the form

$$SK(\bar{\theta},\, \hat{\theta}^0) = \begin{bmatrix} \bar{\theta}_1 - \theta_1^0 \\ \bar{\theta}_{21} - \hat{\theta}_{21}^0 \end{bmatrix}' \begin{bmatrix} \hat{I}_{1,1} & \hat{I}_{1,21} \\ \hat{I}_{21,1} & \hat{I}_{21,21} \end{bmatrix} \begin{bmatrix} \hat{I}^{1,1} & \hat{I}^{1,21} \\ \hat{I}^{21,1} & \hat{I}^{21,21} \end{bmatrix} \begin{bmatrix} \hat{I}_{1,1} & \hat{I}_{1,21} \\ \hat{I}_{21,1} & \hat{I}_{21,21} \end{bmatrix} \begin{bmatrix} \bar{\theta}_1 - \theta_1^0 \\ \bar{\theta}_{21} - \hat{\theta}_{21}^0 \end{bmatrix}, \tag{7}$$

where $\bar{\theta} = (\bar{\theta}_1',\, \bar{\theta}_{21}',\, \hat{\theta}_{22}^{0\prime})'$, and $\hat{I} = \hat{I}(\hat{\theta}^0) = [\hat{I}_{\cdot,\cdot}]$ as well as $\hat{I}^{-1} = [\hat{I}^{\cdot\cdot}]$ are now partitioned into nine subvectors conformably with $(\theta_1',\, \theta_{21}',\, \theta_{22}')'$.

When ML estimators are used, Hausman's (1978) test is based on a statistic of the following form [Hausman and Taylor (1981)]:

$$HA\left(\hat{\theta}_2^0, \hat{\theta}\right) = n\left(\hat{\theta}_2^0 - \hat{\theta}_2\right)'\left[\left(\hat{I}_{22} - \hat{I}_{21}\hat{I}_{11}^{-1}\hat{I}_{12}\right)^{-1} - \hat{I}_{22}^{-1}\right]^{-}\left(\hat{\theta}_2^0 - \hat{\theta}_2\right)$$

$$= n\left(\hat{\theta}_2^0 - \hat{\theta}_2\right)'\left[\hat{I}^{22} - \hat{I}_{22}^{-1}\right]^{-}\left(\hat{\theta}_2^0 - \hat{\theta}_2\right), \tag{8}$$

where $\hat{\theta}_2^0$ and $\hat{\theta}$ are the restricted and unrestricted ML estimators of θ_2, $\hat{I} = \hat{I}(\hat{\theta}) = [\hat{I}_{..}]$ as well as $\hat{I}^{-1} = [\hat{I}^{..}]$ are partitioned conformably with $\theta = (\theta_1', \theta_2')'$, and $[\ ]^-$ refers to a generalized inverse. Under appropriate regularity conditions, the asymptotic distributions of $HA(\hat{\theta}_2, \hat{\theta})$ is $\chi^2(\nu)$, where ν is the rank of the asymptotic covariance matrix of $n^{1/2}(\hat{\theta}_2^0 - \hat{\theta}_2)$.

White's (1982) modified Wald and LM tests (for $\theta_1 = \theta_1^0$) are respectively

$$WW\left(\hat{\theta}, \theta_1^0\right) = nh(\hat{\theta})'\left[\nabla h(\hat{\theta})\bar{C}(\hat{\theta})\ \nabla h(\hat{\theta})'\right]^{-1}h(\hat{\theta}), \tag{9}$$

$$WS(\hat{\theta}^0) = nD(\hat{\theta}^0; Z)'A(\hat{\theta}^0)^{-1}\nabla h(\hat{\theta}^0)'\left[\nabla h(\hat{\theta}^0)\bar{C}(\hat{\theta}^0)\ \nabla h(\theta^0)'\right]^{-1}\nabla h(\hat{\theta}^0)$$

$$\times A(\hat{\theta}^0)^{-1}D(\hat{\theta}^0; Z), \tag{10}$$

where $h(\theta) = \theta_1 - \theta_1^0$, $\nabla h(\theta) = \partial h/\partial\theta'$, $\bar{C}(\theta) = A(\theta)^{-1}B(\theta)A(\theta)^{-1}$, $A(\theta) = \hat{I}(\theta)_1$ and $B(\theta) = \hat{I}(\theta)_2$. Here $\hat{\theta}$ and $\hat{\theta}^0$ are ML estimators based on model (1), which could be misspecified. White (1982) also proposed three tests for misspecification: the information matrix test (Theorem 4.1), a Hausman-type test (Theorem 5.1) and the gradient test (Theorem 5.2). By taking the restricted model (with $\theta_1 = \theta_1^0$) as the model tested for misspecification, each of these procedures may be used to test $H_0: \theta_1 = \theta_1^0$. Because these three tests are rather complex to define, we refer to White (1982) for their definition (degrees of freedom may differ from p_1).

3. Numerical illustrations

To illustrate the lack of invariance of the various tests described in section 2, we will use the following nonlinear regression model with Box–Cox transformations on the explanatory variables:

$$y_t = \gamma + \beta_1 x_{1t}^{(\lambda)} + \beta_2 x_{2t}^{(\lambda)} + u_t, \qquad t = 1, \ldots, n, \tag{11}$$

where $x_{it} > 0$ $(i = 1, 2)$, $x_{it}^{(\lambda)} = (x_{it}^{\lambda} - 1)/\lambda$ when $\lambda \neq 0$, and $x_{it}^{(\lambda)} = \ln(x_{it})$ when $\lambda = 0$. The explanatory variables x_{it} are nonstochastic and the disturbances u_t, $t = 1, \ldots, n$, are i.i.d. normal with mean zero and variance $\sigma^2 > 0$. γ, β_1, β_2, λ and σ^2 are unknown coefficients. We consider the problem of testing $H_0: \beta_2 = 0$.

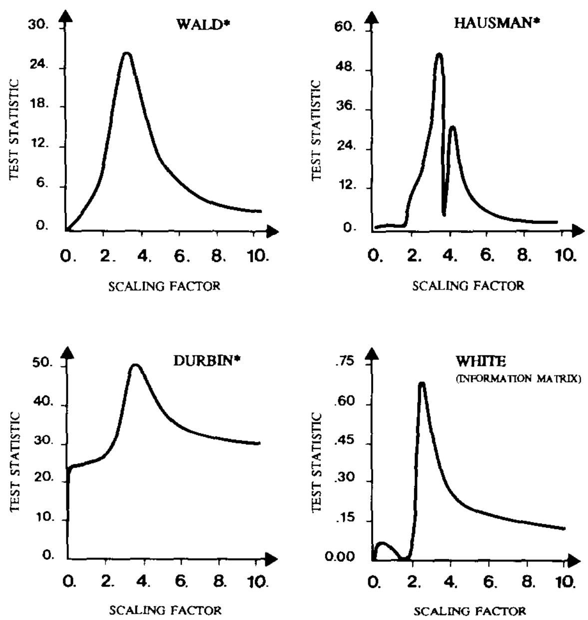
In this model, the choice of the measurement units for y, x_1 and x_2 is a matter of convenience. Given the arbitrariness of the unit choice (provided the model contains an intercept), it is natural to require that the result of a test for $\beta_2 = 0$ be invariant to changes in measurement units.

For the nonlinear model (11), we studied how the test statistics described in section 2 behave when x_1 and x_2 are multiplied by the same scaling factor $k > 0$. [2] The data set used for y, x_1, and

[2] A similar exercise could be performed by multiplying only one variable (y, x_1, or x_2) by k or each variable by a different factor. The results would be similar.

Table 1
Test statistics for $\beta_2 = 0$ in model (11).

Test criterion	Information matrix estimator	Scaling factor $k = 1$	$k = 3$	$k = 10$
1. Hausman	$\hat{I}_1$	1.13347	52.08373	1.70088
	$\hat{I}_2$	0.81449	26.20728	1.39686
	$\hat{I}_3$	1.05921	53.71247	1.55811
2. Durbin	$\hat{I}_1$	−2.42083	146.20932	−1.28795
	$\hat{I}_2$	23.94212	42.99915	28.28281
	$\hat{I}_3$	19.05885	19.05830	19.05683
3. Sargan–Metha	$\hat{I}_1$	−2.42046	146.21033	−1.28956
	$\hat{I}_2$	39.51960	47.95808	2047.79770
	$\hat{I}_3$	19.05842	19.05835	19.05917
4. White				
(a) Wald	–	6.52795	28.18386	2.79509
(b) LM	–	2.22340	24.67461	1.52301
(c) Information matrix	–	0.04093	0.40295	0.10991
(d) Hausman	–	0.29169	4.23203	0.39965
(e) Gradient	–	0.27667	7.33279	0.63055



* Test using $\hat{I}_2$ (outer product matrix).

Fig. 1. Test statistics for $\beta_2 = 0$ in model (11).

x_2 has been produced artificially. [3] The sample size is 50. In table 1, we report the results obtained for the different test statistics considered, when setting $k = 1$ (original units), $k = 3$ and $k = 10$. [4] Graphs illustrating how the Hausman, Durbin and information matrix test statistics change with k appear in fig. 1. For the sake of comparison, we also present a graph of the usual Wald statistic.

To compute Hausman's statistic (8), we used the Moore–Penrose generalized inverse. To implement the Sargan–Mehta test, we took $\theta_1 = \beta_2$, $\theta_{21} = (\gamma, \beta_1)'$ and $\theta_{22} = (\lambda, \sigma^2)'$ in (7). The generalized Wald and LM tests of White (1982) are based on (9) and (10). To get tests of $\beta_2 = 0$ from the three other tests suggested by White (1982), we took the restricted model as the model studied for specification error. The information matrix test was performed on one indicator only, namely the one associated with β_1 ($q = 1$ in Theorem 4.1). For the Hausman-type test and for the gradient test, we took the restricted ML estimator as $\hat{\theta}_n$ (White's notation), the unrestricted ML estimator as $\hat{\gamma}_n$ (White's notation) and $\beta = \beta_1$ in eqs. (5.1) and (5.2) of White (1982).

From the results of table 1 and fig. 1, we see that none of the 14 test statistics considered is invariant to rescaling of the explanatory variables; changes k can produce considerable differences in the values of the test statistics. [5] Scale changes can easily transform a non-significant test statistic (say, at the 5% level) into a significant statistic, or vice versa. Note also that some of the test statistics based on $\hat{I}_1$ can take negative values because there is no general guarantee that $\hat{I}(\tilde{\theta})_1$ be non-negative definite (unless $\tilde{\theta}$ is the *unrestricted* ML estimator).

4. Conclusion

The illustrations of table 1 and fig. 1 strongly suggest that non-invariant asymptotic tests should be avoided or used with great care. On this ground, although they are generally relatively expensive to calculate, likelihood ratio tests or Lagrange multiplier tests using $\hat{I}_2$ or $\hat{I}_3$ should be preferred, when appropriate. Another alternative is to use the modified generalized $C(\alpha)$ tests proposed by Dagenais and Dufour (1991), which can be considerably less costly to calculate than LR of LM tests.

References

Burguete, Jose F., Ronald Gallant and Geraldo Souza, 1982, On unification of the asymptotic theory of nonlinear econometric models, Econometric Reviews 1, 151–211 (with comments).

Dagenais, Marcel G. and Jean-Marie Dufour, 1991, Invariance, nonlinear models and asymptotic tests, Econometrica (forthcoming).

Durbin, James, 1970, Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables, Econometrica 38, 410–421.

Gregory, Allan W. and Michael R. Veall, 1985, Formulating Wald tests of nonlinear restrictions, Econometrica 53, 1465–1468.

Hausman, Jerry A., 1978, Specification tests in econometrics, Econometrica 46, 1251–1272.

Hausman, Jerry A. and William E. Taylor, 1981, A generalized specification test, Economics Letters 8, 239–245.

Lehmann, Eric L., 1983, Theory of point estimation (Wiley, New York).

[3] The data set is available from the authors upon request. Note that the way the data were generated is irrelevant to the issue studied here. These data are identical to those used in Dagenais and Dufour (1991), where additional information on the generation of the data is provided.

[4] Corresponding results for Wald, LR, LM and $C(\alpha)$ tests are available in Dagenais and Dufour (1991).

[5] The main exception is Durbin's test with $\hat{I}_3$, which appears to have very little sensitivity to the value of k for the data set and specification studied here. Highly accurate calculations confirmed, however, that the latter is *not* invariant to changes in k.

Lehmann, Eric L., 1986, Testing statistical hypotheses, second edition (Wiley, New York).

Neyman, Jerzy, 1959, Optimal asymptotic tests of composite statistical hypotheses, in: U. Grenander, ed., Probability and statistics, the Harald Cramer volume (Almqvist and Wiksell, Uppsala) 213–234.

Sargan, J. Dennis and F. Mehta, 1983, A generalization of the Durbin significance test and its application to dynamic specification, Econometrica 51, 1551–1567.

Smith, Richard J., 1987, Alternative asymptotically optimal tests and their application to dynamic specification, Review of Economic Studies 54, 665–680.

Spitzer, John J., 1984, Variance estimates in models with the Box–Cox transformation: implications for estimation and hypothesis testing, Review of Economics and Statistics 66, 645–652.

White, Halbert, 1982, Maximum likelihood estimation of misspecified models, Econometrica 50, 1–25 [Corrigendum, 51 (1983), 513].