Econometrica, Vol. 59, No. 6 (November, 1991), 1601–1615

INVARIANCE, NONLINEAR MODELS, AND ASYMPTOTIC TESTS

BY MARCEL G. DAGENAIS AND JEAN-MARIE DUFOUR[1]

The invariance properties of some well known asymptotic tests are studied. Three types of invariance are considered: invariance to the representation of the null hypothesis, invariance to one-to-one transformations of the parameter space (reparameterizations), and invariance to one-to-one transformations of the model variables such as changes in measurement units. Tests that are not invariant include the Wald test and generalized versions of it, a widely used variant of the Lagrange multiplier test, Neyman's $C(\alpha)$ test, and a generalized version of the latter. For all these tests, we show that simply changing measurement units can lead to vastly different answers even when equivalent null hypotheses are tested. This problem is illustrated by considering regression models with Box-Cox transformations on the variables. We observe, in particular, that various consistent estimators of the information matrix lead to test procedures with different invariance properties. General sufficient conditions are then established, under which the generalized $C(\alpha)$ test becomes invariant to reformulations of the null hypothesis and/or to one-to-one transformations of the parameter space as well as to transformations of the variables. In many practical cases where Wald-type tests lack invariance, we find that special formulations of the generalized $C(\alpha)$ test are invariant and hardly more costly to compute than Wald tests. This computational simplicity stands in contrast with other invariant tests such as the likelihood ratio test. We conclude that noninvariant asymptotic tests should be avoided or used with great care. Further, in many situations, the suggested implementation of the generalized $C(\alpha)$ test often yields an attractive substitute to the Wald test (which is not invariant) and to other invariant tests (which are more costly to perform).

KEYWORDS: Asymptotic tests, formulation of restrictions, invariance, measurement units, Neyman's $C(\alpha)$ test, nonlinear models.

1. INTRODUCTION

IT IS A WIDELY RECOGNIZED PRINCIPLE that the inferences drawn from a statistical analysis should not depend on arbitrary incidentals like the selection of measurement units or the labelling of i.i.d. observations; see Hotelling (1936), Pitman (1939), Ferguson (1967, Chap. 4), and Lehmann (1983, Chap. 3; 1986, Chap. 6). More generally, when the unrestricted parameter space is transformed in a one-to-one manner or when the representation of a null hypothesis is changed, it is natural to require that the result of testing equivalent null hypotheses be the same. For example, in the context of the classical linear model, standard t and F tests enjoy such properties for linear transformations of the parameter space. On the other hand, in nonlinear models, several asymptotic tests are usually available; see Engle (1983) and Gouriéroux and Monfort (1989). The invariance properties of these tests may differ markedly.

[1] This work was supported by the Centre de Recherche et Développement en Économique (C.R.D.E., Université de Montréal), the Social Sciences and Humanities Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada, the Fondation F.C.A.R. (Government of Québec), CORE (Université Catholique de Louvain), and CEPREMAP (Paris). We are grateful to a co-editor, Gordon Fisher, Eric Ghysels, James MacKinnon, Pierre Mohnen, Forrest Nelson, Whitney Newey, Pierre Perron, and Michael Wickens for several useful comments. We also wish to thank Tran Cong Liem for his superb programming assistance.

The first purpose of this paper is to study some basic asymptotic tests from the point of view of invariance. Three types of invariance are studied: invariance to equivalent representations of the null hypothesis, invariance to reparameterizations, and invariance to one-to-one transformations of model variables. An important special issue examined is invariance to changes of measurement units in nonlinear models. Four main test criteria are considered: Wald, likelihood ratio, Rao's efficient score (or Lagrange multiplier), and Neyman's $C(\alpha)$. While it is easy to see that the likelihood ratio (LR) criterion enjoys strong invariance properties, some important invariance (or noninvariance) characteristics of the other tests are not well known. Among other things, we observe that the invariance of the Lagrange multiplier (LM) test depends on the way the information matrix is estimated. For example, the Hessian matrix of the log-likelihood function does not lead to an invariant test. Further, the other test criteria considered are not generally invariant, even for simple measurement unit changes.

If one wants to conduct an invariant test, this suggests that only the LR or an appropriate variant of the LM test should be employed. However, an important disadvantage of these two methods is the requirement to reestimate the model under each restriction tested. This stands in contrast with the Wald method which requires estimating only the unrestricted model.[2] Given this difficulty, the second purpose of this paper is to look for possible substitutes to the Wald method, that are invariant as well as less costly to implement than LR or LM tests. We suggest that a class of Neyman's (1959) $C(\alpha)$ tests and the generalization proposed by Smith (1983, 1987) can be useful in meeting this objective. In general, $C(\alpha)$ and generalized $C(\alpha)$ tests are not invariant. However, if one restricts the class of constrained estimators appropriately, the procedure becomes invariant. We give general sufficient conditions under which the generalized $C(\alpha)$ test is invariant. Depending on the model and null hypothesis considered, generalized $C(\alpha)$ tests can be much cheaper to apply than LR or LM tests.

The paper is organized as follows. In Section 2, we state the assumptions made and describe the test criteria studied. In Section 3, we study the invariance properties of the test criteria and describe general sufficient conditions under which generalized $C(\alpha)$ tests are invariant to hypothesis reformulations and reparameterizations. In Section 4, we apply the results of Section 3 to rescaling in models with Box-Cox transformations and to reformulations of restrictions.

2. FRAMEWORK AND TEST CRITERIA

We consider a general statistical model with log-likelihood function of the form

$$(2.1)\qquad L(\Theta;Z)=\log[p(y|X,\Theta)]=\sum_{t=1}^{n}\log[q(y_t|x_t,\Theta)]=\sum_{t=1}^{n}l_t(\Theta;Z)$$

[2] In specification tests, the opposite happens: the null hypothesis is fixed while a large number of alternative hypotheses may be considered. In such situations, LM tests are more convenient.

where $Z=[y, X]$, $y=[y_1, y_2, \ldots, y_n]'$, $X=[x_1, x_2, \ldots, x_n]'$, y_t is an $m \times 1$ random vector ("dependent variables"), x_t is a $k \times 1$ vector of fixed (or strictly exogenous) variables ($t=1,\ldots,n$), $\Theta=(\theta_1,\ldots,\theta_p)'$ is a $p \times 1$ vector of fixed parameters in the space $\Omega \subseteq \mathbb{R}^p$, 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. $p(y|X,\Theta)$ is the density function of y given X, $q(y_t|x_t,\Theta)$ the density of y_t given x_t, and $l_t \equiv l_t(\Theta; Z) \equiv \log[q(y_t|x_t,\Theta)]$.[3] We suppose that the probability distributions corresponding to different values of Θ are distinct (identification). Also let

$$(2.2)\qquad D(\Theta;Z)=\partial L(\Theta;Z)/\partial\Theta = \left[L_1(\Theta;Z),\ldots,L_p(\Theta;Z)\right]' = \sum_{t=1}^{n} D_t(\Theta;Z),$$

$$(2.3)\qquad H(\Theta;Z)=\frac{1}{n}\frac{\partial^2 L}{\partial\Theta\,\partial\Theta'}=\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t}{\partial\Theta\,\partial\Theta'},$$

where $D_t(\Theta;Z)\equiv \partial l_t/\partial\Theta$. The information matrix corresponding to the log-likelihood function $L(\Theta;Z)$ is

$$(2.4)\qquad I(\Theta)\equiv I(\Theta;X) = -E_\Theta[H(\Theta;Z)] = E_\Theta\left[\frac{1}{n}\sum_{t=1}^{n} D_t(\Theta;Z)D_t(\Theta;Z)'\right]$$

where the expected value $E_\Theta(\cdot)$ is taken with respect to the density function $p(y|X,\Theta)$.

We suppose that usual regularity conditions are satisfied; for various sets of conditions, see Bartoo and Puri (1967), Burguete, Gallant, and Souza (1982), Gallant (1987), Gallant and Holly (1980), and Lehmann (1983, Chap. 6). Thus a consistent maximum likelihood (ML) estimator $\hat{\Theta}$ exists and both $D(\Theta,Z)$ and $\hat{\Theta}$ follow normal distributions asymptotically: $n^{-1/2}D(\Theta;Z)\to N[0,\bar{I}(\Theta)]$, $n^{1/2}(\hat{\Theta}-\Theta)\to N[0,\bar{I}(\Theta)^{-1}]$, where $\bar{I}(\Theta)=\lim_{n\to\infty} I(\Theta)$. Three alternative consistent estimators of $\bar{I}(\Theta)$ are usually considered:

$$(2.5)\qquad \hat{I}(\tilde{\Theta})_1 = -H(\tilde{\Theta};Z),$$

$$\hat{I}(\tilde{\Theta})_2=\frac{1}{n}\sum_{t=1}^{n} D_t(\tilde{\Theta};Z)D_t(\tilde{\Theta};Z)', \qquad \hat{I}(\tilde{\Theta})_3 = I(\tilde{\Theta}).$$

Provided $\tilde{\Theta}$ is a consistent estimator of Θ, each of these estimators converges to $\bar{I}(\Theta)$. Even though basic test criteria, like Wald's and Rao's criteria, were originally defined in terms of $\hat{I}(\tilde{\Theta})_3$, the two other estimators are often easier to compute. In many situations, using $\hat{I}(\tilde{\Theta})_3$ is not practical because the analytical evaluation of the expected value $I(\Theta)=-E_\Theta[H(\Theta;Z)]$ is too difficult. In the

[3] 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.

sequel, the symbol $\hat{I}(\tilde{\Theta})$, with no subscript, will refer to any of the three estimators in (2.5).

Consider now the problem of testing H_0: $\psi(\Theta)=0$, where $\psi(\Theta)$ is a $p_1\times 1$ continuously differentiable vector function of Θ $(1\leq p_1\leq p)$, and suppose that the $p_1\times p$ matrix $\partial\psi/\partial\Theta'$ has full row rank (at least in a neighborhood of the true parameter vector Θ). Several criteria may be used for this purpose. In this paper, we will concentrate on the following ones: the likelihood ratio test, the Wald test (Wald, 1943) and generalizations of it, Rao's (1948) score test (or the Lagrange multiplier test (Aitchison and Silvey (1958), Silvey (1959)), Neyman's (1959) $C(\alpha)$ test and a generalization of the latter. The statistics of the LR, Wald, Rao, and $C(\alpha)$ tests for H_0: $\psi(\Theta)=0$ are respectively

$$LR(\psi)=2\left[L(\hat{\Theta};Z)-L(\hat{\Theta}^0;Z)\right], \tag{2.6}$$

$$W(\psi)=n\psi(\hat{\Theta})'\left[P(\hat{\Theta})\hat{I}(\hat{\Theta})^{-1}P(\hat{\Theta})'\right]^{-1}\psi(\hat{\Theta}),$$

$$S(\psi)=\frac{1}{n}D(\hat{\Theta}^0;Z)'\hat{I}(\hat{\Theta}^0)^{-1}D(\hat{\Theta}^0;Z), \tag{2.7}$$

$$PC(\tilde{\Theta}^0;\psi)=\frac{1}{n}D(\tilde{\Theta}^0;Z)'\hat{I}(\tilde{\Theta}^0)^{-1}P(\tilde{\Theta}^0)' \times\left[P(\tilde{\Theta}^0)\hat{I}(\tilde{\Theta}^0)^{-1}P(\tilde{\Theta}^0)'\right]^{-1}P(\tilde{\Theta}^0)\hat{I}(\tilde{\Theta}^0)^{-1}D(\tilde{\Theta}^0;Z), \tag{2.8}$$

where $P(\Theta)\equiv\partial\psi/\partial\Theta'$, $\hat{\Theta}^0$ and $\hat{\Theta}$ are the restricted and unrestricted ML estimators of Θ, and $\tilde{\Theta}^0$ is a root-n consistent estimator of Θ (at least under H_0) that satisfies $\psi(\tilde{\Theta}^0)=0$. We suppose that $P(\hat{\Theta})$, $P(\hat{\Theta}^0)$, and $P(\tilde{\Theta}^0)$ have full row rank, and similarly for $\hat{I}(\hat{\Theta})$, $\hat{I}(\hat{\Theta}^0)$, and $\hat{I}(\tilde{\Theta}^0)$. Under H_0, the asymptotic distribution of each of these test statistics is $\chi^2(p_1)$.

Various generalizations of the Wald test are obtained by replacing the ML estimator $\hat{\Theta}$ by another asymptotically normal estimator $\tilde{\Theta}$ of Θ, and $\hat{I}(\hat{\Theta})^{-1}$ by a consistent estimator of the asymptotic covariance matrix of $\tilde{\Theta}$; see Stroud (1971). Neyman's (1959) $C(\alpha)$ criterion was originally suggested to test hypotheses of the form $\Theta_1=\Theta_1^0$ where $\Theta=(\Theta_1',\Theta_2')'$ and Θ_i is a $p_i\times 1$ subvector of Θ. It can be viewed as a generalization of Rao's score test obtained by replacing the restricted ML estimator $\hat{\Theta}^0$ by $\tilde{\Theta}^0=(\Theta_1^{0\prime},\tilde{\Theta}_2')'$, where $\tilde{\Theta}_2$ is a locally root-n consistent estimator of Θ_2 (for definition of the latter notion, see Neyman (1959, p. 217) and Bühler and Puri (1966, p. 73)). It was generalized by Smith (1983, 1987) to deal with nonlinear restrictions such as $\psi(\Theta)=0$. The PC statistic has the important property of allowing one to use any root-n consistent estimator that satisfies the null hypothesis, not only the restricted ML estimator. It also enjoys optimal local power properties (see Smith (1983, 1987)). For the case in which $\psi(\Theta)=\Theta_1-\Theta_1^0$, it is easy to check that $PC(\tilde{\Theta}^0;\psi)$ reduces to Neyman's $C(\alpha)$ statistic. Further, when $\tilde{\Theta}^0$ is the restricted ML estimator of Θ, $PC(\tilde{\Theta}^0;\psi)=S(\psi)$.

3. INVARIANCE OF TEST CRITERIA

In this section, we start studying the invariance of the statistics defined above. For this purpose, we shall distinguish between three types of transformations: (1) reformulation of the null hypothesis, (2) reparameterization of the model space, (3) transformation of the model variables. Correspondingly, three types of invariance can be defined. The two first ones will be discussed in this section, while the third one, which involves the two others, will be considered in the last section.

The first type of invariance relates to the choice of the function ψ used to represent the null hypothesis. Let Ψ be a family of $p_1 \times 1$ continuously differentiable functions $\psi: \mathbb{R}^p \to \mathbb{R}^{p_1}$ such that

$$(3.1) \qquad \psi(\Theta) = 0 \quad \text{iff } \bar{\psi}(\Theta) = 0, \quad \text{for all} \quad \psi, \bar{\psi} \in \Psi \quad \text{and} \quad \Theta \in \Omega,$$

where "iff" stands for "if and only if." All formulations $\psi(\Theta) = 0$, where $\psi \in \Psi$, describe the same hypothesis about Θ. We say that a test is *invariant with respect to the formulations* in Ψ when the result of testing $\psi(\Theta) = 0$ is the same for all $\psi \in \Psi$.

To define the second type of invariance, we consider a one-to-one transformation $\bar{g}$ from Ω onto $\Omega_* \subseteq \mathbb{R}^p$: $\Theta_* = \bar{g}(\Theta)$. In such a case, it is usually necessary to reformulate the null hypothesis. For example, for a given function $\psi(\Theta)$ and transformation $\bar{g}$ on Ω, one could consider the equivalent formulation

$$(3.2) \qquad \psi^*(\Theta_*) \equiv \psi\left[\bar{g}^{-1}(\Theta_*)\right].$$

The two hypotheses H_0: $\psi(\Theta) = 0$ and H_0^*: $\psi^*(\Theta_*) = 0$ are equivalent representations of the same hypothesis in the two parameterizations Θ and Θ_*. We say that a test is *invariant to the reparameterization* $\Theta_* = \bar{g}(\Theta)$ for the equivalent representations (ψ, ψ^*) when the results of testing H_0 and H_0^* (on the basis of the original and reparameterized models) are the same. Of course, one could also use here a different equivalent formulation $\psi_*(\Theta_*) = 0$, such that $\psi_*(\Theta_*) = 0$ iff $\psi(\Theta) = 0$, for all Θ. But this can be viewed as a reformulation of the null hypothesis (transformation of the first type) in the context of the new parameterization Θ_* (since $\psi_*(\Theta_*) = 0$ iff $\psi^*(\Theta_*) = 0$). Unless stated otherwise, $\psi^*(\Theta_*) = 0$ will henceforth refer to the representation given by (3.2), and $\psi_*(\Theta_*) = 0$ will refer to any (continuously differentiable) equivalent representation of $\psi(\Theta) = 0$ in terms of Θ_*.

Let us now examine how LR, Wald, LM, and $C(\alpha)$ tests behave under the two types of transformations just defined.

Concerning invariance to the way the null hypothesis is expressed, several authors have already discussed and illustrated the noninvariance of Wald tests to the formulation of the null hypothesis; 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), and Phillips and Park (1988). By contrast, LR and LM tests are invariant to hypothesis reformulation. For LR tests, this is straightforward to check because the restricted maximum of the likelihood function does not depend on the formulation of the null hypothesis. The LM and $C(\alpha)$ tests are also invariant to the formulation of the null hypothesis. This follows from the following theorem on $C(\alpha)$ tests.[4]

THEOREM 1: *Let Ψ be a family of $p_1 \times 1$ continuously differentiable functions of Θ such that $\partial\psi/\partial\Theta'$ has full row rank when $\psi(\Theta)=0$ $(1 \le p_1 \le p)$, and $\psi(\Theta)=0$ if and only if $\bar{\psi}(\Theta)=0$, $\forall\psi, \bar{\psi}\in\Psi$. Further, let $\tilde{\Theta}$ be an estimate of Θ such that $\psi(\tilde{\Theta})=0$, and suppose that $\hat{I}(\tilde{\Theta})$ is nonsingular, where $\hat{I}(\tilde{\Theta}) = \hat{I}(\tilde{\Theta})_i$, $i=1$, 2 or 3. Then*

$$PC(\tilde{\Theta};\psi) = PC(\tilde{\Theta};\bar{\psi}), \qquad \forall\psi, \bar{\psi}\in\Psi.$$

PROOF: By assumption, the set $\Omega_0 = \{\Theta\colon \psi(\Theta)=0\}$ is the same for all $\psi\in\Psi$. Let ψ and $\bar{\psi}$ be any two elements of Ψ, and $\Theta=(\Theta_1', \Theta_2')'$ where Θ_i is $p_i\times 1$. Since $\partial\psi/\partial\Theta'$ has full row rank for $\Theta\in\Omega_0$, we can assume, without loss of generality, that $|\partial\psi/\partial\Theta_1'|\neq 0$ for $\Theta\in\Omega_0$. By the implicit function theorem (see Rudin (1976, chap. 9)), there is a differentiable function h such that $\Theta_1 = h(\Theta_2)$ for $\Theta\in\Omega_0$. Then, for $\Theta\in\Omega_0$, we also have $|\partial\bar{\psi}/\partial\Theta_1'|\neq 0$ and

$$\frac{\partial\Theta_1}{\partial\Theta_2'} = \frac{\partial h}{\partial\Theta_2'} = -\left(\frac{\partial\psi}{\partial\Theta_1'}\right)^{-1}\frac{\partial\psi}{\partial\Theta_2'} = -\left(\frac{\partial\bar{\psi}}{\partial\Theta_1'}\right)^{-1}\frac{\partial\bar{\psi}}{\partial\Theta_2'}.$$

Further $P(\Theta) = \partial\psi/\partial\Theta' = P_1(\Theta)G(\Theta)$ and $\bar{P}(\Theta) = \partial\bar{\psi}/\partial\Theta' = \bar{P}_1(\Theta)G(\Theta)$, where $P_1(\Theta) = \partial\psi/\partial\Theta_1'$, $\bar{P}_1(\Theta) = \partial\bar{\psi}/\partial\Theta_1'$, and $G(\Theta) = [I_{p_1}, (\partial\psi/\partial\Theta_1')^{-1}(\partial\psi/\partial\Theta_2')]$. Hence

$$\begin{aligned} PC(\tilde{\Theta};\psi) &= \frac{1}{n}\tilde{D}'\tilde{I}^{-1}\tilde{G}'\tilde{P}_1'\left[\tilde{P}_1\tilde{G}\tilde{I}^{-1}\tilde{G}'\tilde{P}_1'\right]^{-1}\tilde{P}_1\tilde{G}\tilde{I}^{-1}\tilde{D} \\ &= \frac{1}{n}\tilde{D}'\tilde{I}^{-1}\tilde{G}'\left[\tilde{G}\tilde{I}^{-1}\tilde{G}'\right]^{-1}\tilde{G}\tilde{I}^{-1}\tilde{D} \\ &= \frac{1}{n}\tilde{D}'\tilde{I}^{-1}\tilde{G}'\tilde{Q}_1'\left[\tilde{Q}_1\tilde{G}\tilde{I}^{-1}\tilde{G}'\tilde{Q}_1'\right]^{-1}\tilde{Q}_1\tilde{G}\tilde{I}^{-1}\tilde{D} = PC(\tilde{\Theta};\bar{\psi}) \end{aligned}$$

where $\tilde{D} = D(\tilde{\Theta};Z)$, $\tilde{I} = \hat{I}(\tilde{\Theta})_i$ $(i=1$, 2 or 3), $\tilde{G} = G(\tilde{\Theta})$, $\tilde{P}_1 = P_1(\tilde{\Theta})$, and $\tilde{Q}_1 = \bar{P}_1(\tilde{\Theta})$. *Q.E.D.*

As pointed out above, the LM test is a special case of Neyman's $C(\alpha)$ test, so that the theorem also applies to the LM test.

Consider now a reparameterization $\Theta_* = \bar{g}(\Theta)$ of model (2.1) and let $L_* \equiv L_*(\Theta_*;Z)$ be the log-likelihood function of the reparameterized model. Again, it is easy to check that the LR criterion is invariant to this type of transformation. For the other test criteria, we need to examine the behavior of the first and second derivatives of the log-likelihood function. We assume here that the log-likelihood function and the transformation $\Theta_* = \bar{g}(\Theta)$ are twice differen-

[4] Detailed proofs of the results given in this section are available in Dagenais and Dufour (1986).

tiable. Since $L_*(\Theta_*;Z)=L(\Theta;Z)$ for all Θ and Z, we have

$$(3.3)\qquad \frac{\partial L_*}{\partial \Theta_*}=J(\Theta_*)'\frac{\partial L}{\partial \Theta},$$

$$H_*(\Theta_*;Z)=J(\Theta_*)'H(\Theta;Z)J(\Theta_*)+\frac{1}{n}\sum_{k=1}^{p}L_k(\Theta;Z)M_k(\Theta_*),$$

$$(3.4)\qquad I_*(\Theta_*)=-E_\Theta[H_*(\Theta_*;Z)] = -J(\Theta_*)'E_\Theta[H(\Theta;Z)]J(\Theta_*)=J(\Theta_*)'I(\Theta)J(\Theta_*),$$

where $J(\Theta_*)\equiv\partial\Theta/\partial\Theta'_*$, $H_*(\Theta_*;Z)\equiv n^{-1}\partial^2L_*/\partial\Theta_*\,\partial\Theta'_*$, $L_k(\Theta;Z)\equiv \partial L/\partial\theta_k$, $M_k(\Theta_*)\equiv\partial^2\theta_k/\partial\Theta_*\,\partial\Theta'_*$, $k=1,\dots,p$, and $E_\Theta[\partial L/\partial\Theta]=0$. From (3.3) and (3.4), one sees easily how the three estimators $\hat{I}(\tilde{\Theta})_i$, $i=1,2,3$, defined in (2.5), change when the transformation $\Theta_*=\bar{g}(\Theta)$ is applied:

$$(3.5)\qquad \hat{I}_*(\tilde{\Theta}_*)_1=J(\tilde{\Theta}_*)'\hat{I}(\tilde{\Theta})_1J(\tilde{\Theta}_*)-\frac{1}{n}\sum_{k=1}^{p}L_k(\tilde{\Theta};Z)M_k(\tilde{\Theta}_*),$$

$$\hat{I}_*(\tilde{\Theta}_*)_i=J(\tilde{\Theta}_*)'\hat{I}(\tilde{\Theta})_iJ(\tilde{\Theta}_*)\qquad (i=2,3),$$

where $\tilde{\Theta}\equiv\bar{g}^{-1}(\tilde{\Theta}_*)$. When $\partial L/\partial\Theta\neq 0$ (e.g, if it is evaluated at a restricted ML estimator), we see that the second derivatives $M_k(\Theta_*)$ of the transformation $\Theta=\bar{g}^{-1}(\Theta_*)$ can play a role in the determination of $\hat{I}_*(\tilde{\Theta}_*)_1$. By contrast, when $\tilde{\Theta}$ is the unrestricted ML estimator $\hat{\Theta}$, we have $L_k(\tilde{\Theta};Z)=0$, $k=1,\dots,p$, so that $\hat{I}_*(\hat{\Theta}_*)_i=J(\hat{\Theta}_*)'\hat{I}(\hat{\Theta})_iJ(\hat{\Theta}_*)$, $i=1,2,3$, where $\hat{\Theta}_*=\bar{g}(\hat{\Theta})$ is the unrestricted ML estimator of Θ_*, based on the reparameterized model. Thus, provided $J(\hat{\Theta}_*)$ is nonsingular, the Wald statistic for testing H_{0*}: $\psi_*(\Theta_*)=0$ is

$$(3.6)\qquad W_*(\psi_*)=n\psi_*(\hat{\Theta}_*)'\left[\hat{P}_*\hat{I}_*(\hat{\Theta}_*)^{-1}\hat{P}'_*\right]^{-1}\psi_*(\hat{\Theta}_*)$$

$$=n\psi_*(\hat{\Theta}_*)'\left[\hat{P}\hat{I}(\hat{\Theta})^{-1}\hat{P}'\right]^{-1}\psi_*(\hat{\Theta}_*)$$

where $\hat{P}_*=P_*(\hat{\Theta}_*)$, $\hat{P}=\bar{P}(\hat{\Theta})$, $\bar{P}(\Theta)\equiv\partial\bar{\psi}/\partial\Theta'$, $\bar{\psi}(\Theta)\equiv\psi_*[\bar{g}(\Theta)]$, and

$$(3.7)\qquad P_*(\Theta_*)\equiv\partial\psi_*/\partial\Theta'_*=\bar{P}(\Theta)J(\Theta_*).$$

For the special case where the restrictions are formulated as in (3.2), i.e. $\psi_*=\psi^*$, we have $W_*(\psi_*)=W(\psi)$, irrespective of the estimator $\hat{I}(\hat{\Theta})_i$ used ($i=1,2,3$), because then $\psi_*(\hat{\Theta}_*)=\psi[\bar{g}^{-1}(\hat{\Theta}_*)]=\psi(\hat{\Theta})$ and $\bar{P}(\Theta)=P(\Theta)$ for all Θ. In general, however, nothing guarantees that $W_*(\psi_*)=W(\psi)$. It is also clear that similar problems will affect various generalizations of the Wald test.

The LM statistic for testing H_{0*} is given by the appropriate counterpart to (2.7), where

$$(3.8)\qquad D_*(\Theta_*;Z)=\partial L_*/\partial\Theta_*=J(\Theta_*)'D(\Theta;Z).$$

By (3.5), we have

$$(3.9)\quad S_*(\psi_*) = \frac{1}{n}D(\hat{\theta}^0;Z)'J(\hat{\theta}^0_*)\left[J(\hat{\theta}^0_*)'\hat{I}(\hat{\theta}^0)_iJ(\hat{\theta}^0_*)\right]^{-1}J(\hat{\theta}^0_*)'D(\hat{\theta}^0;Z) = \frac{1}{n}D(\hat{\theta}^0;Z)'\hat{I}(\hat{\theta}^0)_i^{-1}D(\hat{\theta}^0;Z) = S(\psi)$$

for $i=2$ and 3, where $\hat{\theta}^0_* = \bar{g}(\hat{\theta}^0)$ is the restricted ML estimator of θ_* based on $L_*(\theta_*;Z)$ and where we assume that $J(\hat{\theta}^0_*)$ and $\hat{I}_*(\hat{\theta}^0_*)_i$, $i=1,2,3$, are nonsingular. Thus this test is invariant when $I(\theta)$ is estimated by $\hat{I}(\hat{\theta}^0)_2$ or $\hat{I}(\hat{\theta}^0)_3$. On the other hand, for $i=1$, we have

$$(3.10)\quad S_*(\psi_*) = \frac{1}{n}D(\hat{\theta}^0;Z)'\left[\hat{I}(\hat{\theta}^0)_1 + M(\hat{\theta}^0)\right]^{-1}D(\hat{\theta}^0;Z)$$

where $M(\hat{\theta}^0) = -(1/n)\Sigma_{k=1}^{p}L_k(\hat{\theta}^0;Z)[J(\hat{\theta}^0_*)^{-1}]'M_k(\hat{\theta}^0_*)J(\hat{\theta}^0_*)^{-1}$. Thus, when the Hessian matrix of the log-likelihood function is used to estimate $I(\theta)$, the LM statistic becomes sensitive to the curvature characteristics (second derivatives) of the transformation $\bar{g}$. A sufficient condition for $S_*(\psi_*) = S(\psi)$ is $M_k(\hat{\theta}^0_*) = 0$, $k=1,\ldots,p$ (e.g, linear transformations). However, this condition is not always satisfied.

The generalized $C(\alpha)$ criterion differs from the LM criterion because it allows one to use any restricted root-n consistent estimator of θ. In order to investigate the invariance of the generalized $C(\alpha)$ criterion, we must specify how the estimator $\tilde{\theta}$ is modified when the model is reparameterized. One such condition is $\tilde{\theta}_* = \bar{g}(\tilde{\theta})$ implying that the restricted estimator is transformed in the same manner as the parameter space Ω. An estimator $\tilde{\theta}_*$ that possesses the latter property is said to be *equivariant* with respect to the reparameterization $\theta_* = \bar{g}(\theta)$ (see Lehmann (1983, chap. 3)). Note that restricted maximum likelihood estimation always selects an equivariant estimator whenever the parameter space is transformed.

As with the LM test, the generalized $C(\alpha)$ is invariant to reparameterizations for $i=2$ or 3 so long as we use an equivariant estimator. Let $PC_*(\tilde{\theta}_*;\psi^*)$ be the generalized $C(\alpha)$ statistic for testing $\psi^*(\theta_*)=0$ (see (3.2)), based on the reparameterized log-likelihood $L_*(\theta_*;Z)$, and suppose that $J(\tilde{\theta}_*)$ is nonsingular. It follows from (3.5)–(3.8) that if $\tilde{\theta}_* = \bar{g}(\tilde{\theta})$,

$$(3.11)\quad PC_*(\tilde{\theta}_*;\psi^*) = \frac{1}{n}\tilde{D}'\tilde{J}_*\left(\tilde{J}'_*\tilde{I}\tilde{J}_*\right)^{-1}\tilde{J}'_*\tilde{P}'\left[\tilde{P}\tilde{J}_*\left(\tilde{J}'_*\tilde{I}\tilde{J}_*\right)^{-1}\tilde{J}'_*\tilde{P}'\right]^{-1} \times\tilde{P}\tilde{J}_*\left(\tilde{J}'_*\tilde{I}\tilde{J}_*\right)^{-1}\tilde{J}'_*\tilde{D} = \frac{1}{n}\tilde{D}'\tilde{I}^{-1}\tilde{P}'\left[\tilde{P}\tilde{I}^{-1}\tilde{P}'\right]^{-1}\tilde{P}\tilde{I}^{-1}\tilde{D} = PC(\tilde{\theta};\psi),$$

where $\tilde{D} = D(\tilde{\theta};Z)$, $\tilde{J}_* = J(\tilde{\theta}_*)$, $\tilde{P} = P(\tilde{\theta})$, and $\tilde{I} = \hat{I}(\tilde{\theta})_i$, $i=2$ or 3. Therefore, when $I(\theta)$ is estimated by $\hat{I}(\tilde{\theta})_2$ or $\hat{I}(\tilde{\theta})_3$, the $C(\alpha)$ test is invariant to the reparameterization $\Theta_* = \bar{g}(\Theta)$ for the equivalent representations (ψ, ψ^*), and the following theorem holds.

THEOREM 2: *Let* $\Theta_* = \bar{g}(\Theta)$, *where* $\bar{g}: \Omega \to \Omega_* \subseteq \mathbb{R}^p$ *is a one-to-one transformation such that* $\bar{g}$ *as well as* $\bar{g}^{-1}$ *are differentiable, let* $\psi(\Theta)$ *be a continuously differentiable vector function of* Θ *such that* $\partial\psi/\partial\Theta'$ *has full row rank when* $\psi(\Theta) = 0$, *and let* $\psi^*(\Theta_*) \equiv \psi[\bar{g}^{-1}(\Theta_*)]$. *Let also* $\tilde{\Theta}$ *and* $\tilde{\Theta}_*$ *be two estimates of* Θ *and* Θ_*, *respectively, such that* $\psi(\tilde{\Theta}) = 0$, $\psi^*(\tilde{\Theta}_*) = 0$ *and* $|J(\tilde{\Theta}_*)| \neq 0$, *where* $J(\Theta_*) \equiv \partial\Theta/\partial\Theta'_*$. *If* $\tilde{\Theta}_* = \bar{g}(\tilde{\Theta})$ *and* $\hat{I}(\tilde{\Theta}) = \hat{I}(\tilde{\Theta})_i$, $i = 2$ *or* 3, *with* $\hat{I}(\tilde{\Theta})$ *nonsingular, then*

$$PC_*(\tilde{\Theta}_*; \psi^*) = PC(\tilde{\Theta}; \psi),$$

where PC and PC_* *are the generalized* $C(\alpha)$ *statistics based on the log-likelihood functions* $L[\Theta; Z]$ *and* $L_*[\Theta_*; Z] \equiv L[\bar{g}^{-1}(\Theta_*); Z]$ *respectively.*

Finally, again when $I(\Theta)$ is estimated by $\hat{I}(\tilde{\Theta})_2$ or $\hat{I}(\tilde{\Theta})_3$, it follows from Theorems 1 and 2 that

$$PC_*(\tilde{\Theta}_*; \psi_*) = PC_*(\tilde{\Theta}_*; \psi^*) = PC(\tilde{\Theta}; \psi),$$

for any representation $\psi_*(\Theta_*) = 0$ equivalent to $\psi(\Theta) = 0$, provided $|J(\tilde{\Theta}_*)| \neq 0$ and $\partial\psi_*/\partial\Theta'_*$ has full row rank when $\psi_*(\Theta_*) = 0$. In other words, the $C(\alpha)$ test is invariant to the reparameterization $\Theta_* = \bar{g}(\Theta)$ for all equivalent representations (ψ, ψ_*).

4. ILLUSTRATIONS AND APPLICATIONS

In this section, we illustrate numerically the noninvariance problems discussed in Section 3 and show, with specific examples, how the theoretical results of Section 3 can be used to obtain simple invariant tests. An important case considered here is the one where the measurement units of the variables in a model are changed (i.e. when these variables are multiplied by fixed constants). It is easy to see that the results of Section 3 continue to hold when reparameterizations and hypothesis reformulations occur in conjunction with one-to-one differentiable transformations of model variables (involving no unknown parameter), such as measurement unit changes.[5]

[5] The only difference in this case is that the log-likelihood function takes the form $L_*(\Theta_*; Z_*) = \log[\kappa(Z_*)] + L(\Theta; Z)$ with $Z_* = g(Z) = [y_*, X_*]$, where g is a one-to-one transformation of Υ onto itself such that $y_* = g_1(y)$ and $X_* = g_2(X)$ have dimensions identical to those of y and X, and $\kappa(Z_*)$ is the Jacobian of the transformation of y_* to y. For further discussion of this case, see Dagenais and Dufour (1986). When model variables are not transformed, we have $Z_* = Z$, so that $\kappa(Z_*) = 1$.

A. Illustrations of Invariance Problems

Let us examine the following nonlinear regression model with Box-Cox transformations on the explanatory variables:

$$y_t = \gamma + \beta_1 x_{1t}^{(\lambda)} + \beta_2 x_{2t}^{(\lambda)} + u_t \qquad (t = 1, \dots, n), \tag{4.1}$$

where $x_{it} > 0$, $x_{it}^{(\lambda)} = (x_{it}^{\lambda} - 1)/\lambda$ when $\lambda \neq 0$, and $x_{it}^{(\lambda)} = \log(x_{it})$ when $\lambda = 0$. The explanatory variables x_{it} are fixed and the disturbances u_t, $t = 1, \dots, n$, 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 whether x_2 can be excluded from the model (H_0: $\beta_2 = 0$). Note that the Box-Cox transformations of x_{1t} and x_{2t} in (4.1) use the same value of λ; provided $\beta_1 \neq 0$, this typically ensures the identification of λ under H_0. In this model, the choice of the measurement units for y, x_1, and x_2 is a matter of convenience and does not alter the form of the model provided the latter contains an intercept (Schlesselman (1971)). Indeed, if the explanatory variables are rescaled, the model may be reexpressed as:

$$y_t = \gamma_* + \beta_{1*} x_{1t*}^{(\lambda_*)} + \beta_{2*} x_{2t*}^{(\lambda_*)} + u_t \qquad (t = 1, \dots, n), \tag{4.2}$$

where $x_{it*} = kx_{it}$, $i = 1, 2$, and

$$\lambda_* = \lambda, \qquad \gamma_* = \gamma - k^{(\lambda)} k^{-\lambda} \sum_{i=1}^{2} \beta_i, \qquad \beta_{i*} = \beta_i k^{-\lambda} \qquad (i = 1, 2). \tag{4.3}$$

Given the arbitrary nature of the unit choice, it is natural to require that the result of testing H_0 be invariant to changes of measurement units. Further, given that $\beta_{i*} = 0$ iff $\beta_i = 0$ $(i = 1, 2)$, H_{0*}: $\beta_{2*} = 0$ is the natural representation of H_0: $\beta_2 = 0$ in the model with rescaled variables. It is readily seen that changing the units of measurement leads spontaneously to a reparameterization of the model and to a reformulation of the null hypothesis of the form $\psi_*(\theta_*) = 0$, but not of the form $\psi^*(\theta_*) = 0$ as defined in (3.2).

For the nonlinear model (4.1) with a sample of 50 observations, we studied how the Wald, Rao, and Neyman test statistics described in Section 2 behave when x_1 and x_2 are multiplied by the same scaling factor $k > 0$.[6] Table I (lines 1–4) reports the values of these test statistics and also the value of the likelihood ratio test for different values of the scaling factor k. To get Neyman's $C(\alpha)$ statistic, we used the unrestricted estimator for all of the parameters

[6] The basic data used for y, x_1, and x_2 are available from the authors; see also Dagenais and Dufour (1986). These data are artificial: x_1 and x_2 were generated so that $x_1^{(-1)}$ and $x_2^{(-1)}$ followed uniform distributions on $(-2.5, 0)$ and $(-3.5, -0.5)$ respectively with a correlation of 0.7 (independent draws between observations); y was generated by setting $\gamma = 10$, $\beta_1 = \beta_2 = 1$, $\lambda = -1$, and by letting the u_t's be independent $N(0, \sigma^2)$ with $\sigma = 0.85$. The noninvariance of the Wald test to rescaling of the dependent variable in regression model with Box-Cox transformation on the dependent variable was pointed out by Spitzer (1984). For further discussion of this case, see Dagenais and Dufour (1986).

TABLE I

TEST STATISTICS FOR $\beta_2 = 0$ IN MODEL (4.1)

Test criterion	Information matrix, estimator	Scaling factor $k = 1$	$k = 3$	$k = 10$
1. Likelihood ratio	*—	23.6308	23.6308	23.6308
2. Wald	$\hat{I}_1$	4.41109	29.02218	2.81834
	$\hat{I}_2$	2.78514	26.86896	2.43995
	$\hat{I}_3$	3.95607	29.55576	2.71635
3. Rao (LM)	$\hat{I}_1$	−2.42076	146.20943	−1.28805
	*$\hat{I}_2$	25.97854	25.97854	25.97854
	*$\hat{I}_3$	19.05830	19.05830	19.05830
4. Neyman's $C(\alpha)$	$\hat{I}_1$	60.82865	61.34041	5091.23103
	$\hat{I}_2$	0.00164	5.84459	0.00877
	*$\hat{I}_3$	30.33716	30.33716	30.33716
5. PC with $\tilde{\theta}_{(1)}$	*$\hat{I}_2$	25.78550	25.78550	25.78550
6. PC with $\tilde{\theta}_{(2)}$	*$\hat{I}_2$	12.31671	12.31671	12.31671
7. PC with $\tilde{\theta}_{(3)}$	*$\hat{I}_2$	5.04102	5.04102	5.04102

* Indicates an invariant test.

except β_2. To guarantee that the restrictions be satisfied, the estimator of β_2 was constrained to be zero. Graphs illustrating how Rao's statistic changes with k when $\hat{I}_1$ is used and how Neyman's $C(\alpha)$ statistic changes when $\hat{I}_2$ is used appear in Figures 1A and 1B. The results confirm the theoretical analysis of Section 3.[7] The problem with the $C(\alpha)$ statistic as performed here is that the estimator $\hat{\gamma}_*$ of the constant term is not related to $\hat{\gamma}$ in the same way that γ_* is related to γ in (4.3): the equivariance condition $\tilde{\theta}_* = \bar{g}(\tilde{\theta})$ is not satisfied. In other words, the *equivariant* of $\hat{\gamma}$ in the transformed parameter space, under H_0, is $\hat{\gamma} - k^{(\hat{\lambda})}k^{-\hat{\lambda}}\hat{\beta}_1$, whereas the estimator used in constructing Figure 1 is $\hat{\gamma} - k^{(\hat{\lambda})}k^{-\hat{\lambda}}(\hat{\beta}_1 + \hat{\beta}_2)$. Note also that in the case of Rao's statistic, $\hat{I}(\hat{\theta}^0)_1$ is not necessarily positive definite because $\hat{\theta}^0$ yields a constrained maximum of the likelihood function.

B. *Applications of the Modified $C(\alpha)$ Test*

Under the transformations considered in Table I, model (4.1) keeps the same form but takes different parameter values, as shown in equations (4.2) and (4.3). From Theorem 2, the PC statistic is invariant provided constrained estimators corresponding to different measurement units are related by the equivariance condition $\tilde{\theta}_* = \bar{g}(\tilde{\theta})$. Further, given the ML estimator for the unconstrained

[7] The only puzzle is the behavior of the $C(\alpha)$ test with $\hat{I}_3$, which did not appear to be sensitive to the value of k. This may be due to the particular form of the model. However, we did not find an explanation of this phenomenon.

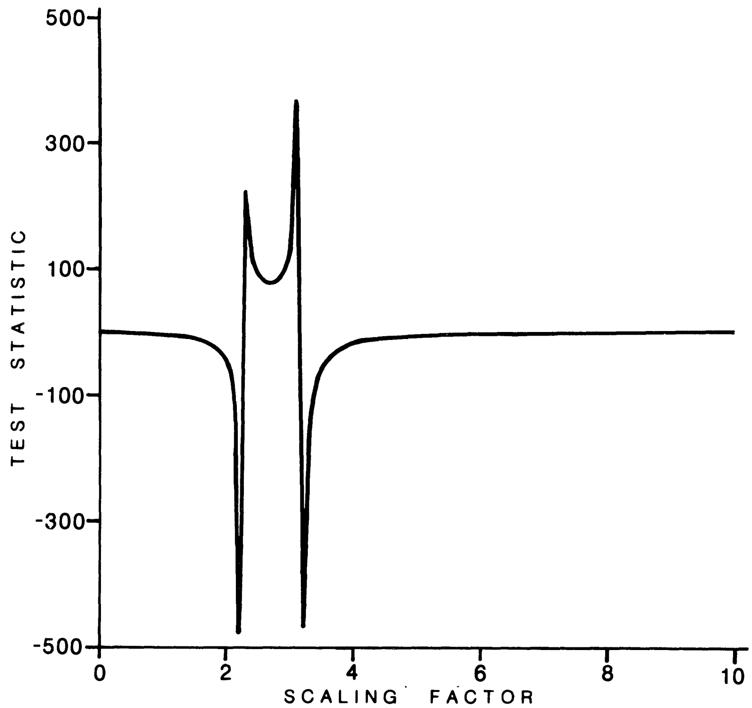

FIGURE 1A.—Rao's test statistics for $\beta_2 = 0$ in Model (4.1)—test using $\hat{I}_1$ (Hessian matrix).

model, we would like to have constrained estimators that are relatively easy to compute. This can be done in several ways.

Let $\beta = (\beta_1, \beta_2)'$, $\theta = (\sigma^2, \gamma, \beta', \lambda)'$, and let $L(\theta; Z)$ be the log-likelihood function of the model. The unconstrained ML estimators $\hat{\theta}$ and $\hat{\theta}_*$ for the original and rescaled data are related by the same equations as θ and θ_*. In particular, $\hat{\lambda} = \hat{\lambda}_*$. Given the latter observation, it is easy to find a constrained equivariant estimator of θ that is simple to compute. Consider the estimator $\tilde{\theta}_{(1)} = (\tilde{\sigma}^2, \tilde{\gamma}, \tilde{\beta}', \hat{\lambda})'$ obtained by fixing $\lambda = \hat{\lambda}$ and estimating the *linear* regression (4.1) with $\lambda = \hat{\lambda}$ and $\beta_2 = 0$. The restricted estimator $\tilde{\theta}_{(1)*} = (\tilde{\sigma}^2_*, \tilde{\gamma}_*, \tilde{\beta}'_*, \hat{\lambda})'$ based on the rescaled model (4.2) is related to $\tilde{\theta}_{(1)}$ by equations (4.3) and the equivariance condition is satisfied. Further, $\tilde{\theta}_{(1)}$ is root-n consistent under H_0: $\beta_2 = 0$ (see Durbin (1970) or Gong and Samaniego (1981)). One can get an even simpler appropriate restricted estimator $\tilde{\theta}_{(2)}$ by fixing $\lambda = \hat{\lambda}$, $\beta_1 = \hat{\beta}_1$, and reestimating γ and σ^2 by least squares. A third estimator $\tilde{\theta}_{(3)}$ can be obtained by reestimating only the intercept γ. In all cases, the intercept needs to be adjusted because the coefficient fixed by the null hypothesis (β_2) is involved in the transformation for γ (see (4.3)). In Table I (lines 5–7), we report the values of the invariant PC statistic based on the three restricted estimators $\tilde{\theta}_{(1)}$, $\tilde{\theta}_{(2)}$, and $\tilde{\theta}_{(3)}$. We can observe that the statistic is invariant to changes in measure-

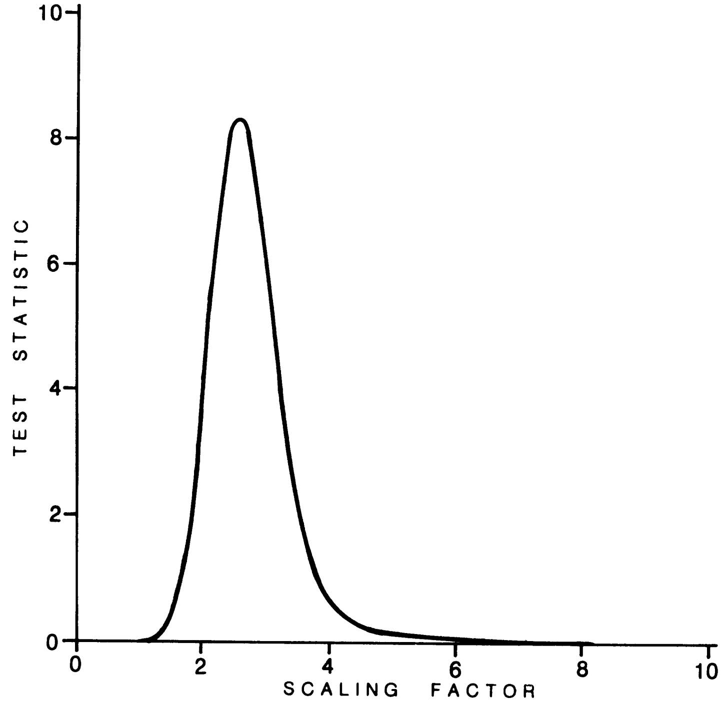

FIGURE 1B.—Neyman's test statistics for $\beta_2 = 0$ in Model (4.1)—test using $\hat{I}_2$ (outer product matrix).

ment units. It is interesting to note that the value of the invariant PC statistic based on $\tilde{\Theta}_{(1)}$ is very similar to the LM and LR statistics. On the other hand, the PC statistics based on $\tilde{\Theta}_{(2)}$ and $\tilde{\Theta}_{(3)}$ differ more from the latter. Clearly, the $C(\alpha)$ procedure is sensitive to the choice of the estimator $\tilde{\Theta}$. Studying which one is preferable according to size and power is an interesting question to be investigated.

From Theorem 1, we also know that the generalized $C(\alpha)$ statistic $PC(\tilde{\Theta}; \psi)$ is invariant to the formulation of the restriction $\psi(\Theta) = 0$, provided the restricted estimator $\tilde{\Theta}$ is itself invariant to the formulation of the restriction. To show how these results can be applied, consider again model (4.1). Clearly, the three "hypotheses" $\beta_2 - 1 = 0$, $\beta_2^3 - 1 = 0$, and $\beta_2^5 - 1 = 0$ are equivalent. Similarly, the two hypotheses $\beta_2 = 0$ and $\exp(2\beta_2) - 1 = 0$ are equivalent. For these two examples, we report in Table II the results of the invariant PC tests. We computed two different restricted estimators $\tilde{\Theta}_{(1)}$ and $\tilde{\Theta}_{(2)}$. In both cases, we set β_2 at the value given by the null hypothesis ($\beta_2 = 1$ or $\beta_2 = 0$). To get $\tilde{\Theta}_{(1)}$, λ is set at its unrestricted ML estimate ($\lambda = \hat{\lambda}$) and all the other coefficients $(\gamma, \beta, \sigma^2)$ are reestimated as if λ were known; to get $\tilde{\Theta}_{(2)}$, both λ and β_1 are set at their unrestricted ML estimates ($\lambda = \hat{\lambda}$, $\beta_1 = \hat{\beta}_1$) and only γ and σ^2 are

TABLE II

INVARIANCE TO FORMULATION OF RESTRICTIONS IN MODEL (4.1)

Null hypothesis	Wald tests ($\hat{I}_2$)	Restricted estimator	PC tests ($\hat{I}_2$)
$\beta_2 - 1 = 0$	0.40015	$\tilde{\theta}_{(1)}$	2.76157
		$\tilde{\theta}_{(2)}$	5.16401
$\beta_2^3 - 1 = 0$	0.17901	$\tilde{\theta}_{(1)}$	2.76157
		$\tilde{\theta}_{(2)}$	5.16401
$\beta_2^5 - 1 = 0$	0.09178	$\tilde{\theta}_{(1)}$	2.76157
		$\tilde{\theta}_{(2)}$	5.16401
$\beta_2 = 0$	2.78514	$\tilde{\theta}_{(1)}$	25.78550
		$\tilde{\theta}_{(2)}$	12.31671
$c^{2\beta_2} - 1 = 0$	0.24747	$\tilde{\theta}_{(1)}$	25.78550
		$\tilde{\theta}_{(2)}$	12.31671

reestimated. Results show that contrary to the Wald test, the PC tests are not influenced by the formulation of the restrictions.

C.R.D.E., Université de Montréal, C.P. 6128, succursale A, Montréal, Québec, H3C 3J7 Canada

Manuscript received April, 1987; final revision received April, 1991.

REFERENCES

AITCHISON, J., AND S. D. SILVEY (1958): "Maximum-Likelihood Estimation of Parameters Subject to Restraints," *Annals of Mathematical Statistics*, 29, 813–828.

BARTOO, J. B., AND P. S. PURI (1967): "On Optimal Asymptotic Tests of Composite Statistical Hypotheses," *Annals of Mathematical Statistics*, 38, 1845–1852.

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).

BÜHLER, W. J., AND P. S. PURI (1966): "On Optimal Asymptotic Tests of Composite Hypotheses with Several Restraints," *Z. Wahrscheinlichkeitstheorie*, 5, 71–88.

COX, D. R., AND D. V. HINKLEY (1974): *Theoretical Statistics*. London: Chapman and Hall.

DAGENAIS, M., AND J.-M. DUFOUR (1986): "Invariance, Nonlinear Models and Asymptotic Tests," Working Paper, C.R.D.E. and Département de Sciences Économiques, Université de Montréal.

DURBIN, J. (1970): "Testing for Serial Correlation in Least Squares Regression when Some of the Regressors are Lagged Dependent Variables," *Econometrica*, 38, 410–421.

ENGLE, R. F. (1983): "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics," in *Handbook of Econometrics*, Volume 2, ed. by Z. Griliches and M.D. Intrilligator. Amsterdam: North-Holland, pp. 775–826.

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 and Sons.

GALLANT, A. R., AND A. HOLLY (1980): "Statistical Inference in an Implicit Nonlinear, Simultaneous Equation Model in the Context of Maximum Likelihood Estimation," *Econometrica*, 48, 697–720.

GONG, G., AND F. J. SAMANIEGO (1981): "Pseudo Maximum Likelihood Estimation: Theory and Applications," *The Annals of Statistics*, 9, 861–869.
GOURIÉROUX, C., AND A. MONFORT (1989): "A General Framework for Testing a Null Hypothesis in a 'Mixed' Form," *Econometric Theory*, 5, 63–82.
GREGORY, A. W., AND M. R. VEALL (1985): "Formulating Wald Tests of Nonlinear Restrictions," *Econometrica*, 53, 1465–1468.
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 and Sons.
——— (1986): *Testing Statistical Hypotheses*, Second Edition. New York: John Wiley and Sons.
NEYMAN, J. (1959): "Optimal Asymptotic Tests of Composite Statistical Hypotheses," in *Probability and Statistics, the Harald Cramer Volume*, ed. by U. Grenander. Uppsala: Almqvist and Wiksell, pp. 213–234.
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.
RAO, C. R. (1948): "Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications to Problems of Estimation," *Proceedings of the Cambridge Philosophical Society*, 44, 50–57.
RUDIN, W. (1976): *Principles of Mathematical Analysis*. New York: McGraw-Hill.
SCHLESSELMAN, J. (1971): "Power Families: A Note on the Box and Cox Transformation," *Journal of the Royal Statistical Society B*, 3, 307–311.
SILVEY, S. D. (1959): "The Lagrangian Multiplier Test," *Annals of Mathematical Statistics*, 30, 389–407.
SMITH, R. (1983): "Alternative Asymptotically Optimal Tests in Econometrics," Discussion Paper No. 544, Institute for Economic Research, Queen's University.
——— (1987): "Alternative Asymptotically Optimal Tests and their Application to Dynamic Specification," *Review of Economic Studies*, 54, 665–680.
SPITZER, J. 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.
STROUD, T. W. F. (1971): "On Obtaining Large-Sample Tests for Asymptotically Normal Estimators," *Annals of Mathematical Statistics*, 42, 1412–1424.
WALD, A. (1943): "Tests of Statistical Hypotheses Concerning Several Parameters when the Number of Observations is Large," *Transactions of the American Mathematical Society*, 54, 426–482.