Queen's Economics Department Working Paper No. 707

Specification Tests Based on Artificial Regressions

Russell Davidson
Queen's University and GREQAM

James G. MacKinnon
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

1-1988

Specification Tests Based on Artificial Regressions

Russell Davidson

GREQAM
Centre de la Vieille Charité
2 rue de la Charité
F-13236 Marseille cedex 02
France

Department of Economics
Queen's University
Kingston, Ontario, Canada
K7L 3N6

russell@ehess.univ-mrs.fr

and

James G. MacKinnon

Department of Economics
Queen's University
Kingston, Ontario, Canada
K7L 3N6

jgm@econ.queensu.ca

Abstract

Many specification tests can be computed by means of artificial linear regressions. These are linear regressions designed to be used as calculating devices to obtain test statistics and other quantities of interest. In this paper, we discuss the general principles which underlie all artificial regressions, and the use of such regressions to compute Lagrange Multiplier and other specification tests based on estimates under the null hypothesis. We demonstrate the generality and power of artificial regressions as a means of computing test statistics, show how Durbin-Wu-Hausman, conditional moment, and other tests which are not explicitly Lagrange Multiplier tests may be computed, and discuss a number of special cases which serve to illustrate the general results and can also be very useful in practice. These include tests of parameter restrictions in nonlinear regression models and tests of binary choice models such as the logit and probit models.

This research was supported, in part, by grants from the Social Sciences and Humanities Research Council of Canada. We are grateful to an anonymous referee for comments on earlier versions. This paper was published in the *Journal of the American Statistical Association*, 85, 1990, 220–227, and the references have been updated.

January, 1988

1. Introduction

In recent years, numerous specification tests have been proposed which can be computed by means of artificial linear regressions. These regressions are artificial in the sense that they are designed to be used solely as calculating devices, the regressand and regressors being constructed so that the desired test statistic is equal to, or can easily be computed from, one of the quantities normally calculated by an OLS regression program. Artificial regressions essentially the same as those used to calculate test statistics can also be used for other purposes, such as calculating consistent estimates of the asymptotic covariance matrix of a vector of parameter estimates and computing one-step efficient estimates from an initial consistent estimate.

The test statistic from an artificial regression is often n (the sample size) times the R^2, sometimes the explained sum of squares, or sometimes an ordinary t test or F test based on the artificial regression. These tests are in some cases derived explicitly as Lagrange Multiplier (LM) tests in their score form, and in other cases are equivalent to such tests. Although many of the tests based on artificial regressions are well known, there has not, to our knowledge, been an exposition of the general principles which underlie them, and which may be used to develop new tests and extend existing ones. The objective of this paper is provide such a general exposition, to demonstrate the generality and power of artificial regressions as a means of computing test statistics, and to discuss a number of special cases which serve to illustrate the general results and can be useful in practice.

2. Some Examples

In this introductory section, we shall present three well-known artificial regressions and indicate how they may be used for the calculation of test statistics. The discussion will be informal at this point. In the next section, we shall discuss matters more formally and in greater generality.

The best-known artificial regression is almost certainly the Gauss-Newton (or G-N) regression, which was originally derived as a way to calculate least squares estimates for nonlinear regression models (Hartley, 1961). Its use for testing restrictions on nonlinear regression models is discussed by Engle (1982, 1984). These papers deal with multivariate as well as univariate models. However, the simplest form of the G-N regression applies to the class of univariate nonlinear regression models:

$$\boldsymbol{y} = \boldsymbol{x}(\boldsymbol{\beta}) + \boldsymbol{u}, \tag{1}$$

where $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^\top \; \boldsymbol{\beta}_2^\top]^\top$ is a k–vector of parameters (with $k = k_1 + k_2$), and $\boldsymbol{x}(\boldsymbol{\beta})$ is an n–vector of nonlinear functions, which would usually depend on exogenous variables and perhaps also on lagged values of the dependent variable. The parameter vector $\boldsymbol{\beta}$ has been partitioned because we wish to consider testing the hypothesis that $\boldsymbol{\beta}_2 = \mathbf{0}$.

The G-N regression may be obtained as a first-order Taylor-series approximation to (1) around some value of $\boldsymbol{\beta}$, say $\boldsymbol{\beta}^*$. Its general form is

$$\boldsymbol{y} - \boldsymbol{x}(\boldsymbol{\beta}^*) = \boldsymbol{X}(\boldsymbol{\beta}^*)\boldsymbol{b} + \text{residuals}. \tag{2}$$

Here and elsewhere, when we write "+ residuals" we mean simply whatever happens to be the difference between the regressand and the rest of the right-hand side of the regression. It is a way of indicating that no statistical meaning is intended: We merely have a linear regression, which when run may yield useful results. The matrix of derivatives $\boldsymbol{X}(\boldsymbol{\beta}) \equiv [\boldsymbol{X}_1(\boldsymbol{\beta}) \;\; \boldsymbol{X}_2(\boldsymbol{\beta})]$ is an $n \times n$ matrix with ti^{th} element the derivative of $x_t(\boldsymbol{\beta})$ with respect to β_i.

If we estimate (1) subject to the restriction that $\boldsymbol{\beta}_2 = \mathbf{0}$, so as to obtain restricted estimates $\tilde{\boldsymbol{\beta}} = [\tilde{\boldsymbol{\beta}}_1^\top \; \mathbf{0}^\top]^\top$, the G-N regression (2) becomes

$$\tilde{\boldsymbol{u}} = \tilde{\boldsymbol{X}}\boldsymbol{b} + \text{residuals} = \tilde{\boldsymbol{X}}_1\boldsymbol{b}_1 + \tilde{\boldsymbol{X}}_2\boldsymbol{b}_2 + \text{residuals}, \tag{3}$$

where $\tilde{\boldsymbol{u}} \equiv \boldsymbol{y} - \boldsymbol{x}(\tilde{\boldsymbol{\beta}})$ and $\tilde{\boldsymbol{X}} \equiv \boldsymbol{X}(\tilde{\boldsymbol{\beta}})$. Here and elsewhere we employ a useful notation whereby functions of parameter vectors (in this case $\boldsymbol{\beta}$) which are evaluated at particular values such as $\tilde{\boldsymbol{\beta}}$ may be written without making the argument explicit. The regressand of (3) is simply the vector of residuals from restricted estimation of (1), and there are k regressors, each of which is a vector of the derivatives of $\boldsymbol{x}(\boldsymbol{\beta})$ with respect to one of the elements of $\boldsymbol{\beta}$, evaluated at $\tilde{\boldsymbol{\beta}}$. If $\boldsymbol{x}(\boldsymbol{\beta})$ were a linear regression function with $\boldsymbol{Z}$ the matrix of independent variables, $\tilde{\boldsymbol{X}}$ would simply be equal to $\boldsymbol{Z}$. When the artificial regression (3) is run, nR^2 is asymptotically equivalent to any asymptotically efficient chi-squared or F test of the restrictions $\boldsymbol{\beta}_2 = \mathbf{0}$. When the restrictions are valid, the statistic will, under appropriate regularity conditions, be distributed as central $\chi^2(k_2)$. An ordinary F test for $\boldsymbol{b}_2 = \mathbf{0}$ would be asymptotically equivalent to the nR^2 test.

A second well-known example of an artificial regression is provided by the so-called outer-product-of-the-gradient regression (or OPG regression, for short). Unlike the G-N regression, which applies only to nonlinear least squares models, the OPG regression is almost universally applicable to models that can be estimated by maximum likelihood. We may suppose that there is a sample of size n which gives rise to a loglikelihood function

$$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \sum_{t=1}^{n} \ell_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2), \tag{4}$$

where $\boldsymbol{\theta}_1$ is a k_1–vector and $\boldsymbol{\theta}_2$ is a k_2–vector of parameters with $k = k_1 + k_2$. We begin by defining a matrix $\boldsymbol{G}(\boldsymbol{\theta})$ with typical element

$$G_{tj}(\boldsymbol{\theta}) = \frac{\ell_t(\boldsymbol{\theta})}{\partial \theta_j}.$$

This is the contribution to the gradient of the loglikelihood function with respect to the j^{th} parameter made by the t^{th} observation. The vector of scores associated with $\ell(\boldsymbol{\theta})$, that is, the gradient of the loglikelihood function (4), will be denoted by $\boldsymbol{g}(\boldsymbol{\theta}) \equiv \boldsymbol{G}^\top(\boldsymbol{\theta})\boldsymbol{\iota}$, where $\boldsymbol{\iota}$ is an n–vector of ones.

As is well-known, the matrix $n^{-1}\boldsymbol{G}^\top(\ddot{\boldsymbol{\theta}})\boldsymbol{G}(\ddot{\boldsymbol{\theta}})$ consistently estimates the expectation of the outer product of the gradient, which is the information matrix $\mathcal{I}(\boldsymbol{\theta})$, whenever

$\ddot{\boldsymbol{\theta}}$ consistently estimates $\boldsymbol{\theta}$. Then, if we let $\tilde{\boldsymbol{\theta}}$ denote the ML estimates obtained by maximizing the loglikelihood function (4) subject to the restrictions $\boldsymbol{\theta}_2 = \mathbf{0}$, the OPG artificial regression used for testing these restrictions is

$$\boldsymbol{\iota} = \boldsymbol{G}(\tilde{\boldsymbol{\theta}})\boldsymbol{c} + \text{residuals}. \tag{5}$$

The quantity nR^2 from regression (5), which in this case is equal to the explained sum of squares from the regression, is a test statistic that is asymptotically distributed as $\chi^2(k_2)$ under the null hypothesis. An early application of this procedure may be found in Godfrey and Wickens (1981).

Since the OPG regression is almost always available, one may ask why artificial regressions are needed at all. They are needed because the finite-sample properties of the OPG regression are often rather poor. The explained sum of squares from the OPG regression (5) with $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$ is

$$\boldsymbol{\iota}^\top \tilde{\boldsymbol{G}}(\tilde{\boldsymbol{G}}^\top \tilde{\boldsymbol{G}})^{-1}\tilde{\boldsymbol{G}}^\top \boldsymbol{\iota} = \tilde{\boldsymbol{g}}^\top (\tilde{\boldsymbol{G}}^\top \tilde{\boldsymbol{G}})^{-1}\tilde{\boldsymbol{g}}. \tag{6}$$

The only difference between this and any other form of the LM statistic is that the matrix $n^{-1}\tilde{\boldsymbol{G}}^\top\tilde{\boldsymbol{G}}$ is used to estimate the information matrix $\mathcal{I}(\boldsymbol{\theta})$. Presumably because $n^{-1}\tilde{\boldsymbol{G}}^\top\tilde{\boldsymbol{G}}$ often provides a poor estimate of $\mathcal{I}(\boldsymbol{\theta})$ when n is not very large, test statistics based on the OPG regression often have finite-sample distributions which are poorly approximated by their asymptotic distributions. Monte Carlo evidence on this point has been provided by Davidson and MacKinnon (1984b, 1985a), Bera and McKenzie (1986) and Godfrey, McAleer, and McKenzie (1988), among others. All these papers found that variants of the LM statistic (6) were much more prone incorrectly to reject the null hypothesis than alternative forms of the LM test, many of which were based on different (and less generally applicable), artificial regressions.

The last artificial regression that we shall consider in this section is associated with the simplest type of binary choice model. The dependent variable y_t may be either zero or one, and it is assumed that $\Pr(y_t = 1) = \Psi(\boldsymbol{X}_t\boldsymbol{\beta})$, where $\Psi(x)$ is a thrice continuously differentiable function which maps from the real line to the 0–1 interval, is weakly increasing in x, and has the properties

$$\Psi(x) \geq 0, \quad \Psi(-\infty) = 0, \quad \Psi(\infty) = 1, \text{ and } \Psi(-x) = 1 - \Psi(x).$$

The most commonly used binary choice models are the probit model, where $\Psi(x)$ is the cumulative standard normal distribution function, and the logit model, where $\Psi(x)$ is the logistic function $\big(1 + \exp(-x)\big)^{-1}$.

For binary choice models of this type, the artificial regression uses as regressand a vector $\boldsymbol{r}(\boldsymbol{\beta})$ with typical element

$$r_t(\boldsymbol{\beta}) = \frac{y_t - \Psi(\boldsymbol{X}_t\boldsymbol{\beta})}{\big(\Psi(\boldsymbol{X}_t\boldsymbol{\beta})\Psi(-\boldsymbol{X}_t\boldsymbol{\beta})\big)^{1/2}}, \tag{7}$$

and as regressors a matrix $\boldsymbol{R}(\boldsymbol{\beta})$ with typical element

$$R_{ti}(\boldsymbol{\beta}) = \frac{\psi(\boldsymbol{X}_t\boldsymbol{\beta})X_{ti}}{\big(\Psi(\boldsymbol{X}_t\boldsymbol{\beta})\Psi(-\boldsymbol{X}_t\boldsymbol{\beta})\big)^{1/2}}, \tag{8}$$

where $\psi(x)$ is the first derivative of $\Psi(x)$. This artificial regression can be derived as a variant of the G-N regression. We can write a binary choice model as

$$y_t = \Psi(\boldsymbol{X}_t\boldsymbol{\beta}) + u_t \tag{9}$$

where u_t equals either $1 - \Psi(\boldsymbol{X}_t\boldsymbol{\beta})$ or $-\Psi(\boldsymbol{X}_t\boldsymbol{\beta})$ and can easily be shown to have variance $\Psi(\boldsymbol{X}_t\boldsymbol{\beta})\Psi(-\boldsymbol{X}_t\boldsymbol{\beta})$. Taylor-expanding (9) as if it were an ordinary nonlinear regression model, and correcting for the heteroskedasticity of the u_t, yields the artificial regression defined by (7) and (8).

If the parameter vector $\boldsymbol{\beta}$ is partitioned as $\boldsymbol{\beta}^\top = [\boldsymbol{\beta}_1^\top \; \boldsymbol{\beta}_2^\top]^\top$ and the binary choice model is estimated by maximum likelihood subject to the restrictions that $\boldsymbol{\beta}_2 = \mathbf{0}$, with restricted ML estimates $\tilde{\boldsymbol{\beta}} = [\tilde{\boldsymbol{\beta}}_1^\top \; \mathbf{0}^\top]^\top$, then running the artificial regression

$$\boldsymbol{r}(\tilde{\boldsymbol{\beta}}) = \boldsymbol{R}(\tilde{\boldsymbol{\beta}})\boldsymbol{b} + \text{residuals}$$

yields both an nR^2 and an explained sum of squares, either of which can serve as a test statistic for the restrictions. The two test statistics are both asymptotically distributed as $\chi^2(k_2)$, but they are not numerically equal, and there is reason to prefer the explained sum of squares in finite samples. For more details, and extensions, see Engle (1984) and Davidson and MacKinnon (1984b).

3. The General Case

All of the artificial regressions discussed in the previous section, along with many others, can be understood in terms of a general framework of artificial regressions that share a set of basic properties and can therefore be used for a wide variety of purposes. We shall deal with the following general case. There is a fully specified, parametrized model characterized by its loglikelihood function, which for a sample of size n can be written as

$$\ell(\boldsymbol{\theta}) = \sum_{t=1}^{n} \ell_t(\boldsymbol{\theta}), \tag{10}$$

where $\boldsymbol{\theta}$ is a k–vector of model parameters. We shall frequently wish to partition $\boldsymbol{\theta}$ as $[\boldsymbol{\theta}_1^\top \; \boldsymbol{\theta}_2^\top]^\top$ in order to consider the restrictions $\boldsymbol{\theta}_2 = \mathbf{0}$. In such cases, $\boldsymbol{\theta}_i$ is a k_i–vector, for $i = 1, 2$, with $k = k_1 + k_2$. As before, $\boldsymbol{g}(\boldsymbol{\theta})$ denotes the k–vector of scores, and $\boldsymbol{G}(\boldsymbol{\theta})$ denotes the $n \times k$ matrix of the derivatives of $\ell_t(\boldsymbol{\theta})$ with respect to the elements of the vector $\boldsymbol{\theta}$.

We assume that the data were generated by a data-generating process, or DGP, characterized by the loglikelihood (10) for some true (but unknown) parameter vector $\boldsymbol{\theta}^0$ such that

$$\boldsymbol{\theta}^0 \equiv \begin{bmatrix} \boldsymbol{\theta}_1^0 \\ \boldsymbol{\theta}_2^0 \end{bmatrix}.$$

Often, we additionally assume that $\boldsymbol{\theta}_2^0 = \mathbf{0}$. The model represented by (10) is assumed to satisfy all the usual conditions for maximum likelihood estimation and inference to be asymptotically valid; see, for example, Amemiya (1985, Chapter 4). In particular, we assume that the true parameter vector $\boldsymbol{\theta}^0$ is interior to a compact parameter space Θ and that the information matrix

$$\mathcal{I}(\boldsymbol{\theta}) \equiv \lim_{n\to\infty} \mathrm{E}\Big(\frac{1}{n}\boldsymbol{g}(\boldsymbol{\theta})\boldsymbol{g}^\top(\boldsymbol{\theta})\Big),$$

which in this case is $k \times k$, is a finite, non-singular matrix for all $\boldsymbol{\theta}$ in Θ.

Various artificial regressions can be associated with the model (10). They always involve two things: a regressand, say $\boldsymbol{r}(\boldsymbol{\theta})$, and a matrix of regressors, say $\boldsymbol{R}(\boldsymbol{\theta})$. The artificial regression can be evaluated at any point $\boldsymbol{\theta} \in \Theta$. It may be written as:

$$\boldsymbol{r}(\boldsymbol{\theta}) = \boldsymbol{R}(\boldsymbol{\theta})\boldsymbol{b} + \text{residuals}. \tag{11}$$

Note that we again use "residuals" as a neutral term to avoid any implication that (11) is a statistical model.

Our theory of artificial regressions depends on assumptions that $\boldsymbol{r}(\boldsymbol{\theta})$ and $\boldsymbol{R}(\boldsymbol{\theta})$ have certain defining properties. These properties are as follows, where all probability limits are calculated with as DGP any process characterized by the loglikelihood (10) for some set of parameters in Θ.

Property 1: Under the DGP characterized by $\boldsymbol{\theta}$,

$$\rho(\boldsymbol{\theta}) \equiv \operatorname*{plim}_{n\to\infty} n^{-1}\boldsymbol{r}^\top(\boldsymbol{\theta})\boldsymbol{r}(\boldsymbol{\theta})$$

exists and is a finite, smooth, real-valued function of $\boldsymbol{\theta}$.

Property 2: $\boldsymbol{R}^\top(\boldsymbol{\theta})\boldsymbol{r}(\boldsymbol{\theta}) = \rho(\boldsymbol{\theta})\boldsymbol{g}(\boldsymbol{\theta})$.

Property 3: If $\ddot{\boldsymbol{\theta}} \to \boldsymbol{\theta}^0$, then $n^{-1}\boldsymbol{R}^\top(\ddot{\boldsymbol{\theta}})\boldsymbol{R}(\ddot{\boldsymbol{\theta}}) \to \rho(\boldsymbol{\theta}^0)\boldsymbol{g}(\boldsymbol{\theta}^0)$.

These properties are not shared by every linear regression used solely to calculate quantities of interest, which we may in general call auxiliary regressions. Nevertheless, in this article, any regression termed "artificial" will satisfy these three properties.

The two crucial features of artificial regressions satisfying properties 1 through 3 are expressed in the following two theorems.

Theorem 1: Suppose that the artificial regression (11) is associated with the fully specified, parametrized model (10), in the sense that Properties 1, 2, and 3 are satisfied for all $\boldsymbol{\theta} \in \operatorname{int}\Theta$, a compact k–dimensional parameter space. Suppose further that the model (10) satisfies the regularity conditions of Amemiya (1985, Chapter 4). Then, if the artificial regression is evaluated at some $\acute{\boldsymbol{\theta}} \in \Theta$ such that $\acute{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 = O_p(n^{-1/2})$, the artificial parameter estimates $\acute{\boldsymbol{b}}$ obtained by OLS have the property that

$$n^{1/2}\acute{\boldsymbol{b}} = n^{1/2}(\hat{\boldsymbol{\theta}} - \acute{\boldsymbol{\theta}}) + o_p(1) \tag{12}$$

as $n \to \infty$, where $\hat{\boldsymbol{\theta}}$ is the (asymptotically efficient) ML estimator of the model (10).

Proof: The proof of Theorem 1 is both simple and illuminating. A Taylor expansion of the gradient $\boldsymbol{g}(\acute{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}^0$ yields

$$\boldsymbol{g}(\acute{\boldsymbol{\theta}}) \cong \boldsymbol{g}(\boldsymbol{\theta}^0) + \boldsymbol{H}(\boldsymbol{\theta}^0)(\acute{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) = \boldsymbol{g}^0 + \boldsymbol{H}^0(\acute{\boldsymbol{\theta}} - \boldsymbol{\theta}^0), \tag{13}$$

where $\boldsymbol{H}^0 \equiv \boldsymbol{H}(\boldsymbol{\theta}^0)$ denotes the Hessian matrix of the loglikelihood function $\ell(\boldsymbol{\theta})$, and "$\cong$" denotes asymptotic equivalence. Multiplying all quantities in (13) by appropriate powers of n, so that they are $O_p(1)$, and using Properties 2 and 3 and the fact that $\acute{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 = O_p(n^{-1/2})$, we see that

$$\begin{aligned} n^{-1/2}\acute{\boldsymbol{R}}^\top\acute{\boldsymbol{r}} &= n^{-1/2}\acute{\rho}\boldsymbol{g}(\acute{\boldsymbol{\theta}}) \\ &\cong n^{-1/2}\rho^0\boldsymbol{g}^0 - (n^{-1}\boldsymbol{R}^{0\top}\boldsymbol{R}^0)n^{1/2}(\acute{\boldsymbol{\theta}} - \boldsymbol{\theta}^0). \end{aligned} \tag{14}$$

If (14) is now evaluated at the ML estimator $\hat{\boldsymbol{\theta}}$ instead of at $\acute{\boldsymbol{\theta}}$, the left-hand side of the equation is zero by the first-order conditions for the maximum of the loglikelihood function, so that

$$\boldsymbol{0} \cong n^{-1/2}\rho^0\boldsymbol{g}^0 - (n^{-1}\boldsymbol{R}^{0\top}\boldsymbol{R}^0)n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0). \tag{15}$$

Subtracting (15) from (14) and rearranging then yields

$$(n^{-1}\boldsymbol{R}^{0\top}\boldsymbol{R}^0)n^{1/2}(\hat{\boldsymbol{\theta}} - \acute{\boldsymbol{\theta}}) \cong n^{-1/2}\acute{\boldsymbol{R}}^\top\acute{\boldsymbol{r}},$$

which implies that

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \acute{\boldsymbol{\theta}}) \cong (n^{-1}\boldsymbol{R}^{0\top}\boldsymbol{R}^0)^{-1}n^{-1/2}\acute{\boldsymbol{R}}^\top\acute{\boldsymbol{r}} \cong \acute{\boldsymbol{b}}.$$

The final step here makes use of the fact that $n^{-1}\boldsymbol{R}^{0\top}\boldsymbol{R}^0 \cong n^{-1}\acute{\boldsymbol{R}}^\top\acute{\boldsymbol{R}}$. The last equation is just a restatement of (12), and so the theorem is proved.

Theorem 2: Under the regularity conditions of Theorem 1, nR^2 from the artificial regression (11), evaluated at any $\acute{\boldsymbol{\theta}} \in \Theta$ such that $\acute{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 = O_p(n^{-1/2})$, is asymptotically equal to

$$\tfrac{1}{n}\acute{\boldsymbol{g}}^\top\mathcal{I}^{-1}(\boldsymbol{\theta}^0)\acute{\boldsymbol{g}}. \tag{16}$$

Proof: The R^2 from (11) is equal to the ratio of the explained sum of squares to the total sum of squares. The total sum of squares, divided by the sample size n, tends to ρ^0 as $n \to \infty$, by Property 1 and the fact that $\acute{\boldsymbol{\theta}} \to \boldsymbol{\theta}^0$. The explained sum of squares is

$$\acute{\boldsymbol{r}}^\top\acute{\boldsymbol{R}}(\acute{\boldsymbol{R}}^\top\acute{\boldsymbol{R}})^{-1}\acute{\boldsymbol{R}}^\top\acute{\boldsymbol{r}}.$$

By Property 2, this becomes

$$n^{-1}\acute{\boldsymbol{r}}^\top\acute{\boldsymbol{R}}(n^{-1}\acute{\boldsymbol{R}}^\top\acute{\boldsymbol{R}})^{-1}\acute{\boldsymbol{R}}^\top\acute{\boldsymbol{r}} = n^{-1}\acute{\rho}^2\acute{\boldsymbol{g}}^\top(n^{-1}\acute{\boldsymbol{R}}^\top\acute{\boldsymbol{R}})^{-1}\acute{\boldsymbol{g}},$$

which, by Property 3, is asymptotically equal to

$$n^{-1}\rho^0 \acute{\boldsymbol{g}}^\top \mathcal{I}^{-1}(\boldsymbol{\theta}^0)\acute{\boldsymbol{g}}.$$

Multiplying this by n and dividing it by the total sum of squares removes the factor of ρ^0. The asymptotic equivalence of (16) and nR^2 from (11) then follows at once.

By itself, Theorem 2 does not say anything about the distribution of nR^2, but it underlies the use of artificial regressions for LM tests. Let $\tilde{\boldsymbol{\theta}}$ denote a vector of ML estimates subject to the restriction that $\boldsymbol{\theta}_2 = \mathbf{0}$. The score form of the LM test statistic for testing the hypothesis $\boldsymbol{\theta}_2 = \mathbf{0}$ against the alternative $\boldsymbol{\theta}_2 \neq \mathbf{0}$ is

$$\frac{1}{n}\tilde{\boldsymbol{g}}^\top \mathcal{I}^{-1}(\boldsymbol{\theta}^0)\tilde{\boldsymbol{g}}. \tag{17}$$

If in fact $\boldsymbol{\theta}_2 = \mathbf{0}$, it is a familiar result — see, for example, Cox and Hinkley (1974, Chapter 9) — that (17) is asymptotically distributed as $\chi^2(k_2)$. Moreover, any test statistic of the form of (17), but with the matrix $\mathcal{I}(\boldsymbol{\theta}^0)$ replaced by any matrix which estimates $\mathcal{I}(\boldsymbol{\theta}^0)$ consistently under the null hypothesis, will be asymptotically equivalent to (17). It follows directly from Theorem 2 that if we are given $\tilde{\boldsymbol{r}} \equiv \boldsymbol{r}(\tilde{\boldsymbol{\theta}})$ and $\tilde{\boldsymbol{R}} \equiv \boldsymbol{R}(\tilde{\boldsymbol{\theta}})$ which satisfy Properties 1–3, then the LM statistic (17), or a test statistic asymptotically equivalent to it, may be computed as nR^2 from the artificial regression

$$\tilde{\boldsymbol{r}} = \tilde{\boldsymbol{R}}\boldsymbol{b} + \text{residuals}. \tag{18}$$

In many cases, $\rho(\boldsymbol{\theta})$ is equal to one, perhaps after rescaling of the artificial variables $\tilde{\boldsymbol{r}}$ and $\tilde{\boldsymbol{R}}$, and in some cases $n^{-1}\tilde{\boldsymbol{r}}^\top\tilde{\boldsymbol{r}}$ is equal to one. In all such cases, the explained sum of squares from (18) is asymptotically equal to the test statistic (17). Other expressions asymptotically equal to (17) can also be found. It is easy to show that k_2 times the F statistic for the (artificial) hypothesis $\boldsymbol{b}_2 = \mathbf{0}$ is such an expression. When $k_2 = 1$, so that only one restriction is being tested, the square of the t statistic for $b_2 = 0$ is also asymptotically equal to (17), and the t test itself is asymptotically valid. Which of these variants of the LM test statistic will in finite samples have a distribution closest to the nominal asymptotic one depends on the details of the model under test.

It is sometimes useful to make explicit the distinction between $\tilde{\boldsymbol{R}}_1$ and $\tilde{\boldsymbol{R}}_2$. Regression (18) can be written as

$$\tilde{\boldsymbol{r}} = \tilde{\boldsymbol{R}}_1\boldsymbol{b}_1 + \tilde{\boldsymbol{R}}_2\boldsymbol{b}_2 + \text{residuals}. \tag{19}$$

The first-order conditions for $\tilde{\boldsymbol{\theta}}$ imply that $\tilde{\boldsymbol{r}}^\top\tilde{\boldsymbol{R}}_1 = \mathbf{0}$, so that the explained and total sums of squares from (19) must be identical to those from the regression

$$\tilde{\boldsymbol{r}} = \tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{R}}_2\boldsymbol{b}_2 + \text{residuals},$$

where

$$\tilde{\boldsymbol{M}}_1 \equiv \mathbf{I} - \tilde{\boldsymbol{R}}_1(\tilde{\boldsymbol{R}}_1^\top\tilde{\boldsymbol{R}}_1)^{-1}\tilde{\boldsymbol{R}}_1^\top$$

is the matrix that projects orthogonally onto the orthogonal complement of the subspace spanned by $\tilde{\boldsymbol{R}}_1$. Thus we see that the numerator of the nR^2 form of the test can also be written as

$$\begin{aligned}\tilde{\boldsymbol{r}}^\top \tilde{\boldsymbol{M}}_1 \tilde{\boldsymbol{R}}_2 (\tilde{\boldsymbol{R}}_2^\top \tilde{\boldsymbol{M}}_1 \tilde{\boldsymbol{R}}_2)^{-1} \tilde{\boldsymbol{R}}_2^\top \tilde{\boldsymbol{M}}_1 \tilde{\boldsymbol{r}} \\ = (n^{-1/2} \tilde{\boldsymbol{r}}^\top \tilde{\boldsymbol{M}}_1 \tilde{\boldsymbol{R}}_2)(n^{-1} \tilde{\boldsymbol{R}}_2^\top \tilde{\boldsymbol{M}}_1 \tilde{\boldsymbol{R}}_2)^{-1}(n^{-1/2} \tilde{\boldsymbol{R}}_2^\top \tilde{\boldsymbol{M}}_1 \tilde{\boldsymbol{r}}).\end{aligned} \tag{20}$$

Writing the test statistic in this form makes it quite clear that it must have k_2 degrees of freedom, since $n^{-1/2} \tilde{\boldsymbol{R}}_2^\top \tilde{\boldsymbol{M}}_1 \tilde{\boldsymbol{r}}$ is a k–vector.

In some cases of practical interest, some of the parameters $\boldsymbol{\theta}_1$ that may vary freely under the null hypothesis can be treated as nuisance parameters and the artificial regressors corresponding to them dropped from the artificial regression. This situation arises if the information matrix is block-diagonal between these nuisance parameters and all of the other parameters of the model, so that those columns of $\boldsymbol{R}(\boldsymbol{\theta})$ which correspond to the nuisance parameters will be asymptotically orthogonal to all the remaining columns of $\boldsymbol{R}(\boldsymbol{\theta})$. This situation arises in the case of the nonlinear regression model with normal errors and allows us to show that the Gauss-Newton regression (3) can be regarded as a special case of the general class of artificial regressions discussed above.

If we assume that the vector $\boldsymbol{u}$ in (1) is normally distributed, the contribution from the t^{th} observation to the loglikelihood function is

$$\ell_t(\boldsymbol{\beta}, \sigma) = -\tfrac{1}{2}\log(2\pi) - \log(\sigma) - \frac{\big(y_t - x_t(\boldsymbol{\beta})\big)^2}{2\sigma^2}.$$

The derivatives of $\ell_t(\boldsymbol{\beta}, \sigma)$ are

$$\frac{\partial \ell_t}{\partial \beta_i} = \frac{1}{\sigma^2} \boldsymbol{X}_{ti}(\boldsymbol{\beta}) \big(y_t - x_t(\boldsymbol{\beta})\big) \tag{21}$$

and

$$\frac{\partial \ell_t}{\partial \sigma} = \frac{-1}{\sigma} + \frac{\big(y_t - x_t(\boldsymbol{\beta})\big)^2}{\sigma^2}. \tag{22}$$

When (21) is multiplied by (22), the expectation of the resulting product is zero, which establishes the familiar result that the information matrix for nonlinear regression models is block-diagonal between $\boldsymbol{\beta}$ and σ. If we are only interested in restrictions on $\boldsymbol{\beta}$, we can construct an artificial regression in which σ is treated as a nuisance parameter. The block-diagonality property implies that the artificial regression need not include a regressor corresponding to σ.

If we make the following definitions corresponding to the artificial variables used in the G-N regression (2), namely,

$$r_t(\boldsymbol{\beta}) = y_t - x_t(\boldsymbol{\beta}) \quad \text{and} \quad R_{ti}(\boldsymbol{\beta}) = X_{ti}(\boldsymbol{\beta}) \quad \text{for } i = 1, \ldots, k,$$

then we see that these artificial variables satisfy the defining properties 1, 2, and 3. In particular, $\rho(\boldsymbol{\beta}) = \sigma^2$, and

$$\boldsymbol{R}^\top(\boldsymbol{\beta})\boldsymbol{r}(\boldsymbol{\beta}) = \boldsymbol{X}^\top(\boldsymbol{\beta})\big(\boldsymbol{y} - \boldsymbol{x}(\boldsymbol{\beta})\big),$$

which is the gradient of the normal loglikelihood with respect to $\boldsymbol{\beta}$, times $\rho(\boldsymbol{\beta})$, and

$$\tfrac{1}{n}\boldsymbol{R}^\top(\boldsymbol{\beta})\boldsymbol{R}(\boldsymbol{\beta}) = \tfrac{1}{n}\boldsymbol{X}^\top(\boldsymbol{\beta})\boldsymbol{X}(\boldsymbol{\beta}),$$

which is $\rho(\boldsymbol{\beta})$ times the $\boldsymbol{\beta} - -\boldsymbol{\beta}$ block of the information matrix.

It is well known that the assumption of normality is quite unnecessary for the asymptotic theory of nonlinear regression models estimated by least squares. The conventional theory of the G-N regression shows that it, too, can validly be used with models in which the errors are non-normal. It may well be possible to extend the general theory of artificial regressions to a semi-parametric context, and then the use of the G-N regression with non-normal errors would be covered by such an extended theory. Such an extension is beyond the scope of this paper, however.

In contrast, the assumption of homoskedasticity is essential if we are to make asymptotic inferences based on the usual estimated covariance matrix $s^2(\hat{\boldsymbol{X}}^\top\hat{\boldsymbol{X}})^{-1}$. But the work of Eicker (1963) and White (1980), among others, has shown that it is possible to make valid inferences asymptotically even in the presence of heteroskedasticity of unknown form. There exist auxiliary regressions which permit the straightforward calculation of test statistics robust to the presence of such heteroskedasticity; see, for example, Davidson and MacKinnon (1985b) and Wooldridge (1991). However, these auxiliary regressions have only k_2 rather than k regressors, and thus they cannot satisfy properties 1 through 3. Although they can be very useful, they are therefore not artificial regressions in the sense that the term is used in this paper.

Artificial regressions can be useful for a variety of purposes besides calculating test statistics. Suppose, for example, that we evaluate $\boldsymbol{r}(\boldsymbol{\theta})$ and $\boldsymbol{R}(\boldsymbol{\theta})$ at the unrestricted estimates $\hat{\boldsymbol{\theta}}$ and run the regression

$$\hat{\boldsymbol{r}} = \hat{\boldsymbol{R}}\boldsymbol{b} + \text{residuals}. \tag{23}$$

It is obvious by the first-order conditions that the OLS estimate $\hat{\boldsymbol{\theta}}$ must be identically zero, so running (23) is an easy way to verify that $\hat{\boldsymbol{\theta}}$ does indeed satisfy the first-order conditions. Moreover, the OLS covariance matrix estimate from (23) will be

$$\frac{\hat{\boldsymbol{r}}^\top\hat{\boldsymbol{r}}}{n-k}(\hat{\boldsymbol{R}}^\top\hat{\boldsymbol{R}})^{-1},$$

and, by properties 1 and 3, this is evidently a valid estimate of the inverse of the information matrix. Whether it is an estimate that we would actually want to use in practice will depend, in part, on whether $n^{-1}\hat{\boldsymbol{r}}^\top\hat{\boldsymbol{r}}$ equals one identically or not, and on whether the degrees of freedom correction is deemed to be appropriate.

Of considerably greater interest to the results on testing which follow is the fact that artificial regressions can be used to calculate one-step efficient estimates. Suppose that we are somehow able to obtain a vector $\acute{\boldsymbol{\theta}}$ of consistent, but asymptotically inefficient, estimates. The manner in which $\acute{\boldsymbol{\theta}}$ is obtained is unimportant; all we require is that $\acute{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 = O_p(n^{-1/2})$. If we then evaluate $\boldsymbol{r}(\boldsymbol{\theta})$ and $\boldsymbol{R}(\boldsymbol{\theta})$ at $\acute{\boldsymbol{\theta}}$ and run the artificial regression

$$\acute{\boldsymbol{r}} = \acute{\boldsymbol{R}}\boldsymbol{b} + \text{residuals},$$

the resulting OLS estimates $\acute{\boldsymbol{b}}$ have the property (12) by Theorem 1. Consequently, the one-step estimator

$$\grave{\boldsymbol{\theta}} \equiv \acute{\boldsymbol{\theta}} + \acute{\boldsymbol{b}} = \acute{\boldsymbol{\theta}} + (\acute{\boldsymbol{R}}^\top \acute{\boldsymbol{R}})^{-1} \acute{\boldsymbol{R}}^\top \acute{\boldsymbol{r}}$$

is asymptotically equivalent to $\hat{\boldsymbol{\theta}}$. In situations where $\hat{\boldsymbol{\theta}}$ is difficult or expensive to obtain but $\acute{\boldsymbol{\theta}}$ is readily obtainable, this can be a very valuable result.

More specifically, Theorem 1 allows us to go in one step from the restricted estimates $\tilde{\boldsymbol{\theta}}$, obtained by ML estimation of the model (10) under the null hypothesis that $\boldsymbol{\theta}_2 = \boldsymbol{0}$, to estimates asymptotically equivalent to the unrestricted ML estimates $\hat{\boldsymbol{\theta}}$ of the same model, where the asymptotic equivalence is good whether or not $\boldsymbol{\theta}_2 = \boldsymbol{0}$, provided only that $\boldsymbol{\theta}_2 = O_p(n^{-1/2})$. Further, one can go in one step in the reverse direction as well, starting from the unrestricted estimates $\hat{\boldsymbol{\theta}}$ and obtaining estimates asymptotically equivalent under the null hypothesis to the restricted estimates $\tilde{\boldsymbol{\theta}}$. The trick is to use in the former case what we may call the artificial regression for the unrestricted model, that is, the regression with the full set of artificial regressors $[\tilde{\boldsymbol{R}}_1 \ \tilde{\boldsymbol{R}}_2]$, and in the latter case the artificial regression for the restricted model, where the regressors $\hat{\boldsymbol{R}}_2$ are absent. Formally, the $\tilde{\boldsymbol{b}}$ obtained by OLS on regression (19), when added to $\tilde{\boldsymbol{\theta}} \equiv [\tilde{\boldsymbol{\theta}}_1^\top \ \boldsymbol{0}^\top]^\top$ give an answer equal through order $n^{-1/2}$ to $\hat{\boldsymbol{\theta}}$, while the $\hat{\boldsymbol{b}}_1$ obtained by OLS on the regression

$$\hat{\boldsymbol{r}} = \hat{\boldsymbol{R}}_1 \boldsymbol{b}_1 + \text{residuals},$$

when added to $\hat{\boldsymbol{\theta}}_1$, give an answer equivalent to $\tilde{\boldsymbol{\theta}}_1$ to the same order of approximation. These very convenient results follow from Theorem 1 and the facts that $\tilde{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\theta}^0$ if $\boldsymbol{\theta}_2 = O_p(n^{-1/2})$ and that $\hat{\boldsymbol{\theta}}_1$ is consistent for $\boldsymbol{\theta}_1^0$ if $\boldsymbol{\theta}_2 = \boldsymbol{0}$.

4. Tests where the Alternative is Implicit

Up to this point, all the tests we have discussed have involved an explicit alternative hypothesis. The null is a special case of the alternative, and the restrictions being tested are (or can be reformulated as) zero restrictions. In this section, we show that artificial regressions can be used to perform tests where there is no explicit alternative hypothesis. Thus we can write down an artificial regression analogous to (19), with $\tilde{\boldsymbol{R}}_2$ replaced by an $n \times l$ matrix $\tilde{\boldsymbol{Z}} \equiv \boldsymbol{Z}(\tilde{\boldsymbol{\theta}})$:

$$\tilde{\boldsymbol{r}} = \tilde{\boldsymbol{R}}_1 \boldsymbol{c}_1 + \tilde{\boldsymbol{Z}} \boldsymbol{c}_2 + \text{ residuals}. \tag{24}$$

Provided that the matrix $\tilde{\boldsymbol{Z}}$ satisfies certain conditions (discussed below), which essentially give it the same properties as $\tilde{\boldsymbol{R}}_2$, and assuming that the matrix $[\tilde{\boldsymbol{R}}_1 \ \tilde{\boldsymbol{Z}}]$ has

full rank, the nR^2 from (24) and other asymptotically equivalent statistics must be asymptotically distributed as $\chi^2(l)$ when the data are actually generated by (10) with $\boldsymbol{\theta}_2 = \mathbf{0}$. Thus (24) provides a way to compute a wide variety of test statistics, which need not necessarily be derived explicitly as LM statistics.

We now briefly indicate how to prove the above proposition. Since the proof is similar to standard proofs for LM tests based on artificial regressions, many details are omitted. As noted above, it is necessary that $\tilde{\boldsymbol{Z}}$ satisfy certain conditions, in order that it should have essentially the same properties as $\tilde{\boldsymbol{R}}_2$. First, we require that

$$\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\tilde{\boldsymbol{r}}^\top\tilde{\boldsymbol{Z}}\Big) = \mathbf{0}$$

under the null hypothesis; if this condition were not satisfied, we obviously could not expect the plim of $\boldsymbol{c}_2$ in (24) to be zero. Second, we require that

$$\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\tilde{\boldsymbol{Z}}^\top\tilde{\boldsymbol{r}}\tilde{\boldsymbol{r}}^\top\tilde{\boldsymbol{Z}}\Big) = \rho^0 \operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\tilde{\boldsymbol{Z}}^\top\tilde{\boldsymbol{Z}}\Big) \tag{25}$$

and

$$\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\tilde{\boldsymbol{Z}}^\top\tilde{\boldsymbol{r}}\tilde{\boldsymbol{r}}^\top\tilde{\boldsymbol{R}}_1\Big) = \rho^0 \operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\tilde{\boldsymbol{Z}}^\top\tilde{\boldsymbol{R}}_1\Big), \tag{26}$$

conditions which are similar to the requirement that

$$\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\tilde{\boldsymbol{R}}_1^\top\tilde{\boldsymbol{r}}\tilde{\boldsymbol{r}}^\top\tilde{\boldsymbol{R}}_1\Big) = \rho^0 \operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\tilde{\boldsymbol{R}}_1^\top\tilde{\boldsymbol{R}}_1\Big). \tag{27}$$

This last requirement does not have to be assumed separately, because it is a consequence of properties 2 and 3, the definition of the information matrix $\mathcal{I}(\boldsymbol{\theta})$, and the consistency of $\tilde{\boldsymbol{\theta}}$. Third, we require that laws of large numbers be applicable to the quantities whose probability limits appear on the right-hand sides of (25), (26), and (27). Finally, we require that a central limit theorem be applicable to the vector

$$n^{-1/2}\tilde{\boldsymbol{Z}}^\top\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{r}}. \tag{28}$$

All of these conditions and assumptions must, of course, be verified in individual cases. Since this paper is concerned with the general properties of artificial regressions, it seems inappropriate to consider any particular case in detail. In some instances, it is not too difficult to find sufficient conditions that will guarantee the needed regularity, while in others, especially in time-series contexts, rather elaborate arguments may be necessary, especially for a central limit theorem to apply to (28). There is no doubt, however, that in numerous cases of interest sufficient regularity is present.

Now consider the vector (28). Asymptotically, it has mean zero under the null hypothesis, and its asymptotic covariance matrix is

$$\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\tilde{\boldsymbol{Z}}^\top\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{r}}\tilde{\boldsymbol{r}}^\top\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{Z}}\Big),$$

which is equal to

$$\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\big(\tilde{\boldsymbol{Z}}^\top\tilde{\boldsymbol{r}}\tilde{\boldsymbol{r}}^\top\tilde{\boldsymbol{Z}} - \tilde{\boldsymbol{Z}}^\top\tilde{\boldsymbol{r}}\tilde{\boldsymbol{r}}^\top\tilde{\boldsymbol{R}}_1(\tilde{\boldsymbol{R}}_1^\top\tilde{\boldsymbol{R}}_1)^{-1}\tilde{\boldsymbol{R}}_1^\top\tilde{\boldsymbol{Z}} - \tilde{\boldsymbol{Z}}^\top\tilde{\boldsymbol{R}}_1(\tilde{\boldsymbol{R}}_1^\top\tilde{\boldsymbol{R}}_1)^{-1}\tilde{\boldsymbol{R}}_1^\top\tilde{\boldsymbol{r}}\tilde{\boldsymbol{r}}^\top\tilde{\boldsymbol{Z}}$$
$$+ \tilde{\boldsymbol{Z}}^\top\tilde{\boldsymbol{R}}_1(\tilde{\boldsymbol{R}}_1^\top\tilde{\boldsymbol{R}}_1)^{-1}\tilde{\boldsymbol{R}}_1^\top\tilde{\boldsymbol{r}}\tilde{\boldsymbol{r}}^\top\tilde{\boldsymbol{R}}_1(\tilde{\boldsymbol{R}}_1^\top\tilde{\boldsymbol{R}}_1)^{-1}\tilde{\boldsymbol{R}}_1^\top\tilde{\boldsymbol{Z}}\big)\Big). \tag{29}$$

Rewriting (29) so that each term is a product of $O_p(1)$ probability limits, using (25), (26), and (27), and simplifying, we find that

$$\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\tilde{\boldsymbol{Z}}^\top\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{r}}\tilde{\boldsymbol{r}}^\top\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{Z}}\Big) = \rho^0 \operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\tilde{\boldsymbol{Z}}^\top\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{Z}}\Big). \tag{30}$$

This plus the asymptotic normality of (28) implies that the expression

$$\big(n^{-1/2}\tilde{\boldsymbol{r}}^\top\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{Z}}\big)\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\tilde{\boldsymbol{Z}}^\top\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{Z}}\Big)\big(n^{-1/2}\tilde{\boldsymbol{Z}}^\top\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{r}}\big)$$

is asymptotically distributed as ρ^0 times $\chi^2(l)$. This expression is what the numerator of nR^2 from regression (24) tends to as $n \to \infty$. By property 1, the denominator of nR^2 tends to ρ^0. Thus we conclude that nR^2 from regression (24) is asymptotically distributed as $\chi^2(l)$ if the DGP satisfies the null hypothesis.

There are numerous examples of tests, not designed against explicit alternatives, which can be based on artificial regressions. One example is the class of tests called Durbin-Wu-Hausman tests, which we consider in the next section. These can be based on any artificial regression. Other examples are provided by several tests based on the OPG regression, which we will consider in the remainder of this section.

Newey (1985) suggested using the OPG regression to calculate what he called "conditional moment tests"; see also Tauchen (1985), who used a related auxiliary regression for quite similar purposes. The basic idea of conditional moment tests is that parametric statistical models are generally based on assumptions which imply that certain moment conditions must hold. For example, suppose that a model depends on underlying error terms u_t which are assumed to be NID$(0, \sigma^2)$. Estimation of the model will generally yield observable counterparts of these, say $\hat{u}_t$, which are functions of the parameter estimates $\hat{\boldsymbol{\theta}}$. Then, in large samples, we would expect that

$$\frac{1}{n}\sum_{t=1}^{n}\hat{u}_t^2 \quad \text{and} \quad \frac{1}{n}\sum_{t=1}^{n}(\hat{u}_t^4 - 3\hat{u}_t^2), \tag{31}$$

the empirical counterparts of conditions on the third and fourth moments of the u_t, should both be approximately equal to zero. Newey demonstrates that moment conditions such as these can be tested by running the artificial regression

$$\boldsymbol{\iota} = \hat{\boldsymbol{G}}\boldsymbol{c}_1 + \hat{\boldsymbol{Z}}\boldsymbol{c}_2 + \text{residuals}, \tag{32}$$

where $\hat{\boldsymbol{G}} \equiv \boldsymbol{G}(\hat{\boldsymbol{\theta}})$, and $\hat{\boldsymbol{Z}}$ is chosen so that $\boldsymbol{\iota}^\top \hat{\boldsymbol{Z}}$ generates the moment conditions to be tested. In the case of (31), $\hat{\boldsymbol{Z}}$ would consist of two vectors, with typical elements $\hat{u}_t^3$ and $\hat{u}_t^4 - 3\hat{\sigma}^4$.

Moment conditions may also arise from the theory of the phenomenon being modeled. For example, in the context of models of rational behavior by economic agents, error terms are often supposed to be orthogonal to everything in agents' information sets. Thus, if $\boldsymbol{W}_t$ denotes an l–vector of variables that should be orthogonal to u_t, we could define $\hat{\boldsymbol{Z}}_t$ as the $l \times 1$ vector $\boldsymbol{W}_t \hat{u}_t$ and then test the l orthogonality conditions by regressing $\boldsymbol{\iota}$ on $\hat{\boldsymbol{G}}$ and $\hat{\boldsymbol{Z}}_t$. However, because of the poor finite-sample properties of tests based on the OPG regression, conditional moment tests should be used with great caution when the sample size is not very large.

5. Durbin-Wu-Hausman Tests

Hausman (1978), following Durbin (1954) and Wu (1973), suggested that it may often be useful to test whether there is any significant difference between two sets of estimates, one of which is consistent and efficient under relatively strong conditions, and one of which is consistent under weaker conditions. The original application was comparing estimates obtained by least squares with ones obtained by instrumental variables. It might also be natural to compare the vector of restricted estimates $\tilde{\boldsymbol{\theta}}_1$ with the vector of unrestricted estimates $\hat{\boldsymbol{\theta}}_1$ from model (10). Because of the possibility of a one-step artificial regression, it is not actually necessary to obtain $\hat{\boldsymbol{\theta}}_1$ in order to do so, and this observation provides the easiest way to see how to use an artificial regression to perform a Durbin-Wu-Hausman, or DWH, test.

The estimate of $\boldsymbol{b}_1$ from the artificial regression (19) is

$$\tilde{\boldsymbol{b}}_1 = (\tilde{\boldsymbol{R}}_1^\top \tilde{\boldsymbol{M}}_2 \tilde{\boldsymbol{R}}_1)^{-1} \tilde{\boldsymbol{R}}_1^\top \tilde{\boldsymbol{M}}_2 \tilde{\boldsymbol{r}},$$

where $\tilde{\boldsymbol{M}}_2 \equiv \mathbf{I} - \tilde{\boldsymbol{R}}_2 (\tilde{\boldsymbol{R}}_2^\top \tilde{\boldsymbol{R}}_2)^{-1} \tilde{\boldsymbol{R}}_2^\top$. Adding this quantity to $\tilde{\boldsymbol{\theta}}_1$ yields a one-step estimator which is asymptotically equivalent to $\hat{\boldsymbol{\theta}}_1$ in the sense discussed in Section 4. Hence a test based on a comparison of $\tilde{\boldsymbol{\theta}}_1$ and $\tilde{\boldsymbol{\theta}}_1 + \tilde{\boldsymbol{b}}_1$ is equivalent to one based on a comparison of $\tilde{\boldsymbol{\theta}}_1$ and $hat\boldsymbol{\theta}_1$.

There is no need to restrict attention to regression (19); regression (24), of which the former is a special case, can equally well be used to obtain one-step estimates of $\boldsymbol{\theta}_1$. The estimate of $\boldsymbol{c}_1$ from (24) is

$$\begin{aligned} \tilde{\boldsymbol{c}}_1 &= (\tilde{\boldsymbol{R}}_1^\top \tilde{\boldsymbol{M}}_{\boldsymbol{Z}} \tilde{\boldsymbol{R}}_1)^{-1} \tilde{\boldsymbol{R}}_1^\top \tilde{\boldsymbol{M}}_{\boldsymbol{Z}} \tilde{\boldsymbol{r}} \\ &= -(\tilde{\boldsymbol{R}}_1^\top \tilde{\boldsymbol{M}}_Z \tilde{\boldsymbol{R}}_1)^{-1} \tilde{\boldsymbol{R}}_1^\top \tilde{\boldsymbol{P}}_{\boldsymbol{Z}} \tilde{\boldsymbol{r}}, \end{aligned} \tag{33}$$

where $\tilde{\boldsymbol{P}}_{\boldsymbol{Z}} \equiv \tilde{\boldsymbol{Z}} (\tilde{\boldsymbol{Z}}^\top \tilde{\boldsymbol{Z}})^{-1} \tilde{\boldsymbol{Z}}^\top$ and $\tilde{\boldsymbol{M}}_{\boldsymbol{Z}} \equiv \mathbf{I} - \tilde{\boldsymbol{P}}_{\boldsymbol{Z}}$. The corresponding one-step estimate of $\hat{\boldsymbol{\theta}}_1$ is simply

$$\tilde{\boldsymbol{\theta}}_1 + \tilde{\boldsymbol{c}}_1 = \tilde{\boldsymbol{\theta}}_1 + (\tilde{\boldsymbol{R}}_1^\top \tilde{\boldsymbol{M}}_{\boldsymbol{Z}} \tilde{\boldsymbol{R}}_1)^{-1} \tilde{\boldsymbol{R}}_1^\top \tilde{\boldsymbol{M}}_{\boldsymbol{Z}} \tilde{\boldsymbol{r}}. \tag{34}$$

These one-step estimates are less efficient than $\tilde{\boldsymbol{\theta}}_1$ if $\boldsymbol{\theta}_2 = \mathbf{0}$, because, as (34) makes clear, they are equal to $\tilde{\boldsymbol{\theta}}_1$ plus something which should be random noise when the

model is correctly specified. If the model were not correctly specified, however, the second term in (34) would not be random noise, and $\tilde{\boldsymbol{c}}_1$ would differ systematically from zero. Thus the DWH test simply asks whether or not the second term in (34) is random noise.

We have seen that the difference between the one-step estimate and the restricted ML estimate $\tilde{\boldsymbol{\theta}}_1$ is $\tilde{\boldsymbol{c}}_1$, which is given in (33). The DWH test is thus concerned with whether the vector

$$n^{-1/2}\tilde{\boldsymbol{R}}_1^\top \tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{r}} = n^{-1/2}\tilde{\boldsymbol{R}}_1^\top \tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{r}} \tag{35}$$

has mean zero asymptotically. Note that the equalities in (33) and (35) both follow from the fact that $\tilde{\boldsymbol{R}}_1^\top \tilde{\boldsymbol{r}} = \boldsymbol{0}$. The vector on the right-hand side of (35) looks just like the vector (28), with $\tilde{\boldsymbol{R}}_1^\top \tilde{\boldsymbol{P}}_{\boldsymbol{Z}}$ playing the role of $\tilde{\boldsymbol{Z}}^\top$. Hence the result (30) implies that

$$\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\tilde{\boldsymbol{R}}_1^\top \tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{r}}\tilde{\boldsymbol{r}}^\top \tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{R}}_1\Big) = \rho^0 \operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\tilde{\boldsymbol{R}}_1^\top \tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{R}}_1\Big).$$

It is now evident that we may test the hypothesis that expression (35) has mean zero asymptotically by using the test statistic

$$\frac{n}{\tilde{\boldsymbol{r}}^\top\tilde{\boldsymbol{r}}}\tilde{\boldsymbol{r}}^\top \tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{R}}_1(\tilde{\boldsymbol{R}}_1^\top \tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{R}}_1)^+\tilde{\boldsymbol{R}}_1^\top \tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{r}}, \tag{36}$$

where $(\cdot)^+$ denotes a generalized inverse. We must use a generalized inverse here because the matrix

$$\tilde{\boldsymbol{R}}_1^\top \tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{R}}_1$$

may not have full rank k_1; in fact, it can have rank at most equal to $\min(k_1, l)$.

The test statistic (36) may, of course, be calculated by means of an artificial regression, namely,

$$\tilde{\boldsymbol{r}} = \tilde{\boldsymbol{R}}_1\boldsymbol{d}_1 + \tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{R}}_1^*\boldsymbol{d}_2 + \text{residuals}, \tag{37}$$

where $\tilde{\boldsymbol{R}}_1^*$ is a matrix which consists of as many columns of $\tilde{\boldsymbol{R}}_1$ as possible, subject to the constraint that the matrix $[\tilde{\boldsymbol{R}}_1 \;\; \tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{R}}_1^*]$ must have full rank. Note that the nR^2 from regression (37) is

$$\frac{n}{\tilde{\boldsymbol{r}}^\top\tilde{\boldsymbol{r}}}\tilde{\boldsymbol{r}}^\top \tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{R}}_1^*(\tilde{\boldsymbol{R}}_1^{*\top} \tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{R}}_1^*)^{-1}\tilde{\boldsymbol{R}}_1^{*\top} \tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{r}},$$

and this is numerically equal to the test statistic (36). Thus we have provided a general procedure for performing a DWH test by means of an artificial regression. DWH tests are potentially of interest when either the less efficient set of estimates is not explicitly obtained by relaxing a set of restrictions — although, as Davidson and MacKinnon (1987) prove, this is always implicitly the case — or when k_1 is substantially smaller than k_2, so that the DWH test involves substantially fewer degrees of freedom than a classical test would.

Ruud (1984) and Newey (1985) have previously shown that tests asymptotically equivalent to DWH tests can be computed as score tests, so that various artificial regressions

can be used to compute these tests. However, the only artificial regression which has been explicitly suggested for this purpose is the OPG regression discussed in Section 2. The above results are very much more general. They show that for any test statistic which can be computed by means of an artificial regression, whether or not it is explicitly an LM test, there is a DWH version which can be computed by a similar artificial regression. One simply replaces the $n \times l$ matrix $\tilde{\boldsymbol{Z}}$, whatever it may be, by the matrix $\tilde{\boldsymbol{P}}_{\boldsymbol{Z}}\tilde{\boldsymbol{R}}_1^*$, which will in regular cases be $n \times \min(k_1, l)$. For more on the interpretation of DWH tests, see Davidson and MacKinnon (1989).

6. Double-length Regressions

In Section 2, we discussed three widely-used artificial regressions. There are many others, most of which, we suspect, have yet to be discovered. In this section, we discuss one useful but not yet widely used class of artificial regressions, the "double-length" artificial regressions proposed by Davidson and MacKinnon (1984a). These apply to any model of the form

$$f_t(y_t, \bar{\boldsymbol{y}}_t, \boldsymbol{\theta}) = \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, 1), \tag{38}$$

where $f_t(\cdot)$ is a nonlinear function which may depend on exogenous variables (hence the t subscript), $y_t$is an observation on the dependent variable, $\bar{\boldsymbol{y}}_t$ is a vector of observations on lagged values of y_t, and $\boldsymbol{\theta}$ is a vector of parameters. Suitable regularity conditions must be assumed, of course; see the cited article for details. Since the nonlinear function $f_t(\cdot)$ may include a transformation to the standard normal, (38) is actually a rather general class of models; univariate and multivariate nonlinear regression models with normal errors are both special cases of it, for example.

For a model of this class, the contribution to the loglikelihood made by the t^{th} observation is

$$\ell_t = -\frac{1}{2}\log(2\pi) - \frac{1}{2}f_t^2 + k_t,$$

where

$$k_t(y_t, \bar{\boldsymbol{y}}_t, \boldsymbol{\theta}) \equiv \log\left|\frac{\partial f_t(y_t, \bar{\boldsymbol{y}}_t, \boldsymbol{\theta})}{\partial \theta_i}\right|$$

is a Jacobian term. Now let us make the definitions

$$F_{ti}(y_t, \bar{\boldsymbol{y}}_t, \boldsymbol{\theta}) \equiv \frac{\partial f_t(y_t, \bar{\boldsymbol{y}}_t, \boldsymbol{\theta})}{\partial \theta_i},$$

and

$$K_{ti}(y_t, \bar{\boldsymbol{y}}_t, \boldsymbol{\theta}) \equiv \frac{\partial k_t(y_t, \bar{\boldsymbol{y}}_t, \boldsymbol{\theta})}{\partial \theta_i}.$$

Further, we define $\boldsymbol{F}(\boldsymbol{\theta})$ and $\boldsymbol{K}(\boldsymbol{\theta})$ as the $n \times k$ matrices with typical elements $F_{ti}(y_t, \bar{\boldsymbol{y}}_t, \boldsymbol{\theta})$ and $K_{ti}(y_t, \bar{\boldsymbol{y}}_t, \boldsymbol{\theta})$, respectively. It is easy to see that the gradient is

$$\boldsymbol{g}(\boldsymbol{\theta}) = -\boldsymbol{F}^\top(\boldsymbol{\theta})\boldsymbol{f}(\boldsymbol{\theta}) + \boldsymbol{K}^\top(\boldsymbol{\theta})\boldsymbol{\iota}. \tag{39}$$

The fundamental result proved in Davidson and MacKinnon (1984a) is that, for this class of models, the information matrix $\mathcal{I}(\boldsymbol{\theta})$ satisfies

$$\mathcal{I}(\boldsymbol{\theta}) = \operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\big(\boldsymbol{F}^\top(\boldsymbol{\theta})\boldsymbol{F}(\boldsymbol{\theta}) + \boldsymbol{K}^\top(\boldsymbol{\theta})\boldsymbol{K}(\boldsymbol{\theta})\big)\Big), \tag{40}$$

and so it can be consistently estimated by the matrix

$$\frac{1}{n}\big(\boldsymbol{F}^\top(\ddot{\boldsymbol{\theta}})\boldsymbol{F}(\ddot{\boldsymbol{\theta}}) + \boldsymbol{K}^\top(\ddot{\boldsymbol{\theta}})\boldsymbol{K}(\ddot{\boldsymbol{\theta}})\big),$$

where $\ddot{\boldsymbol{\theta}}$ is any consistent estimate of $\boldsymbol{\theta}$. Hence one valid form of the LM statistic for testing hypotheses about $\boldsymbol{\theta}$ is

$$(-\tilde{\boldsymbol{f}}^\top\tilde{\boldsymbol{F}} + \boldsymbol{\iota}^\top\tilde{\boldsymbol{K}})(\tilde{\boldsymbol{F}}^\top\tilde{\boldsymbol{F}} + \tilde{\boldsymbol{K}}^\top\tilde{\boldsymbol{K}})^{-1}(-\tilde{\boldsymbol{F}}^\top\tilde{\boldsymbol{f}} + \tilde{\boldsymbol{K}}^\top\boldsymbol{\iota}), \tag{41}$$

where, as usual, quantities with a tilde are evaluated at $\tilde{\boldsymbol{\theta}}$, the vector of ML estimates of $\boldsymbol{\theta}$ subject to the restrictions to be tested. The test statistic (41) is evidently just the explained sum of squares from the double-length artificial regression

$$\begin{bmatrix} \tilde{\boldsymbol{f}} \\ \boldsymbol{\iota} \end{bmatrix} = \begin{bmatrix} -\tilde{\boldsymbol{F}} \\ \tilde{\boldsymbol{K}} \end{bmatrix}\boldsymbol{b} + \text{residuals}. \tag{42}$$

This artificial regression has $2n$ "observations." The regressand is $\tilde{f}_t$ for "observation" t and unity for "observation" $t+n$, and the regressors corresponding to $\boldsymbol{\theta}$ are $-\tilde{\boldsymbol{F}}_t$ for "observation" t and $\tilde{\boldsymbol{K}}_t$ for "observation" $t+n$, with $\tilde{\boldsymbol{F}}_t$ and $\tilde{\boldsymbol{K}}_t$ denoting the t^{th} rows of $\tilde{\boldsymbol{F}}$ and $\tilde{\boldsymbol{K}}$, respectively. It is clear from (39) and (40) that regression (42) satisfies properties 1, 2, and 3, provided that, in property 1, n is replaced by $2n$, corresponding to the double length of the artificial regression. Similarly, all results mentioning nR^2 from the regression should in this context be read as $2nR^2$.

Double-length artificial regressions are particularly useful for testing whether or not the dependent variable in a regression model should be transformed in some way. For example, they may be used to test both the hypothesis that $\lambda = 0$ and the hypothesis that $\lambda = 1$ in the Box-Cox regression model

$$(y_t^\lambda - 1)/\lambda = x_t(\boldsymbol{\beta}) + u_t, \quad u_t \sim \text{NID}(0, \sigma^2); \tag{43}$$

see Box and Cox (1964). Under the null hypothesis that $\lambda = 0$, the regressand of (43) is simply $\log y_t$, whereas under the null hypothesis that $\lambda = 1$, and provided that $x_t(\boldsymbol{\beta})$ includes the equivalent of a constant term, the regressand is effectively just y_t. LM tests for $\lambda = 0$ and $\lambda = 1$ based on double-length regressions were derived by Davidson and MacKinnon (1985a). They found that those test statistics have finite-sample distributions much closer to their asymptotic distributions than similar LM test statistics based on the OPG regression that were proposed by Godfrey and Wickens (1981), a result confirmed by Godfrey, McAleer, and McKenzie (1988). They also found, analytically, that except when σ^2 is quite small, both forms of the LM statistic have much greater power than the well-known test proposed by Andrews (1971). Further applications of double-length regressions are discussed in Davidson and MacKinnon (1988).

7. Conclusion

In this paper, we have discussed several aspects of the general theory of artificial linear regressions. We have shown that, whenever it is possible to construct an artificial regression which satisfies our properties 1, 2, and 3, one can evaluate that regression at restricted estimates and use n times the R^2 from the regression as a test statistic. Moreover, it is possible in many cases to calculate tests based on artificial regressions without ever explicitly specifying an alternative hypothesis. One simply has to construct test regressors so that they satisfy certain properties. This opens the door to a wide range of tests. One-step efficient estimates are readily calculated by means of artificial regressions, and we have used this fact to show that it is always possible to calculate a Durbin-Wu-Hausman variant of any test based on an artificial regression. We have also discussed a number of specific artificial regressions which can be very useful in practice.

References

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, Mass.: Harvard University Press.

Andrews, D. F. (1971). "A note on the selection of data transformations," *Biometrika*, 58, 249–254.

Bera, A. K., and McKenzie, C. R. (1986). "Alternative forms and properties of the score test," *Journal of Applied Statistics*, 13, 13–25.

Box, G. E. P., and Cox, D. R. (1964). "An analysis of transformations," *Journal of the Royal Statistical Society*, Series B, 26, 211–252.

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

Davidson, R., and MacKinnon, J. G. (1984a). "Model specification tests based on artificial linear regressions," *International Economic Review*, 25, 485–502.

Davidson, R., and MacKinnon, J. G. (1984b). "Convenient specification tests for logit and probit models," *Journal of Econometrics*, 25, 241–262.

Davidson, R., and MacKinnon, J. G. (1985a). "Testing linear and loglinear regressions against Box-Cox alternatives," *Canadian Journal of Economics*, 25, 499–517.

Davidson, R., and MacKinnon, J. G. (1985b). "Heteroskedasticity-robust tests in regression directions," *Annales de l'INSEE*, 59/60, 183–218.

Davidson, R., and MacKinnon, J. G. (1987). "Implicit alternatives and the local power of test statistics," *Econometrica*, 55, 1305–1329.

Davidson, R., and MacKinnon, J. G. (1988). "Double-length artificial regressions," *Oxford Bulletin of Economics and Statistics*, 50, 203–217.

Davidson, R., and MacKinnon, J. G. (1989). "Testing for consistency using artificial regressions," *Econometric Theory*, 5, 363–384.

Durbin, J. (1954). "Errors in variables," *Review of the International Statistical Institute*, 22, 23–32.

Eicker, F. (1963). "Asymptotic normality and consistency of the least squares estimators for families of linear regressions," *Annals of Mathematical Statistics*, 34, 447–456.

Engle, R. F. (1982). "A general approach to Lagrange Multiplier model diagnostics," *Journal of Econometrics*, 20, 83–104.

Engle, R. F. (1984). "Wald, likelihood ratio and Lagrange multiplier tests in econometrics," in Z. Griliches and M. Intriligator, ed., *Handbook of Econometrics*. Amsterdam: North Holland.

Godfrey, L. G., McAleer. M., and McKenzie, C. R. (1988). "Variable addition and Lagrange Multiplier tests for linear and logarithmic regression models," *Review of Economics and Statistics*, 70, 492–503.

Godfrey, L. G., and Wickens, M. R. (1981). "Testing linear and log-linear regressions for functional form," *Review of Economic Studies*, 48, 487–496.

Hartley, H. O. (1961). "The modified Gauss-Newton method for the fitting of nonlinear regressions by least squares," *Technometrics*, 3, 269–280.

Hausman, J. A. (1978). "Specification tests in econometrics," *Econometrica*, 46, 1251–1272.

Newey, W. K. (1985). "Maximum likelihood specification testing and conditional moment tests," *Econometrica*, 53, 1047–1070.

Ruud, P. A. (1984). "Tests of specification in econometrics," *Econometric Reviews*, 3, 211–242.

Tauchen, G. E. (1985). "Diagnostic testing and evaluation of maximum likelihood models," *Journal of Econometrics*, 30, 415–443.

White, H. (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica* 48, 817–838.

Wooldridge, J. M. (1991). "Specification testing and quasi-maximum likelihood estimation," *Journal of Econometrics*, 48, 29–55.

Wu, D. (1973). "Alternative tests of independence between stochastic regressors and disturbances," *Econometrica*, 41, 733–750.