Queen's Economics Department Working Paper No. 642

Testing the Specification of Econometric Models in Regression and Non-Regression Directions

Russell Davidson
Queen's University

James G. MacKinnon
Queen's University

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

2-1986

Testing the Specification of Econometric Models in Regression and Non-Regression Directions

Russell Davidson

and

James G. MacKinnon

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

Abstract

The asymptotic power of a statistical test depends on the model being tested, the (implicit) alternative against which the test is constructed, and the process which actually generated the data. The exact way in which it does so is examined for several classes of models and tests. First, we analyze the power of tests of nonlinear regression models in regression directions, that is, tests which are equivalent to tests for omitted variables. Next, we consider the power of heteroskedasticity-robust variants of these tests. Finally, we examine the power of very general tests in the context of a very general class of models.

This research was supported, in part, by grants from the Social Sciences and Humanities Research Council of Canada. This paper was published in *Telecommunications Demand Modelling: An Integrated View*, ed. A. de Fontenay, M. H. Shugard, and D. Sibley, Amsterdam, North-Holland, 1990, 221–240. The references have been updated.

February, 1986

1. Introduction

In any area of applied econometric research, and especially in the usual situation where the experimental design cannot be controlled, it is essential to subject any model to a great many statistical tests before even tentatively accepting it as valid. When that is done routinely, especially early in the model-building process, it is inevitable that most models will fail many tests. The applied econometrician must then attempt to infer from those failures what is wrong with the model. The interpretation of test statistics, both those which reject and those which do not reject the null hypothesis, is thus an important part of the econometrician's job. There is, however, surprisingly little literature on the subject, a recent exception being Davidson and MacKinnon (1985a).

The vast majority of the models that econometricians estimate are regression models of some sort, linear or nonlinear, univariate or multivariate. The tests to which such models may be subjected fall into three broad categories, according to a scheme discussed in Davidson and MacKinnon (1985a). First of all, there are tests in "regression directions", in which the (possibly implicit) alternative model is also a regression model, at least locally. Secondly, there are tests in "higher moment directions", which are only concerned with the properties of the error terms; these might include tests for heteroskedasticity, skewness, and kurtosis. Finally, there are tests in "mixed directions", which combine both regression and higher moment components. This paper will deal primarily with tests in regression directions, although there will be some discussion of tests in mixed directions.

Tests in regression directions form the lion's share of the tests that are commonly employed by econometricians when testing regression models. The asymptotic analysis of such tests is reasonably easy, and it turns out to be remarkably similar to the asymptotic analysis of much more general families of tests; see Davidson and MacKinnon (1987) and Section 5 below. In contrast to Davidson and MacKinnon (1985a), which deals only with linear regression models, we focus on the case of univariate, nonlinear regression models. Because we are dealing with nonlinear models, all of the analysis is asymptotic; this is probably an advantage rather than otherwise, because, even in the linear case, many inessential complexities can be eliminated by focusing on what happens as the sample size tends to infinity.

In Section 2, we discuss nonlinear regression models and tests in regression directions. In Section 3, we analyze the asymptotic power of these tests when the data generating process, or DGP, is not the alternative against which the test is constructed. In Section 4, we discuss heteroskedasticity-robust variants of tests in regression directions. Finally, in Section 5, we discuss tests in mixed directions.

2. Nonlinear Regression Models and Tests in Regression Directions

The model of interest is the nonlinear regression model

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\beta}, \boldsymbol{\gamma}) + \boldsymbol{u}, \quad \mathrm{E}(\boldsymbol{u}) = \mathbf{0}, \quad \mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \sigma^2 \mathbf{I}, \tag{1}$$

where $\boldsymbol{y}$ and $\boldsymbol{u}$ are n–vectors, and $\boldsymbol{f}(\cdot)$ denotes a vector of twice continuously differentiable functions $f_t(\boldsymbol{\beta},\boldsymbol{\gamma})$ which depend on $\boldsymbol{\beta}$, a k–vector, and $\boldsymbol{\gamma}$, an r–vector, of (generally unknown) parameters. The matrices of derivatives of the elements of $\boldsymbol{f}$ with respect to the elements of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ will be denoted $\boldsymbol{F}_{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\gamma})$ and $\boldsymbol{F}_{\boldsymbol{\gamma}}(\boldsymbol{\beta},\boldsymbol{\gamma})$. These matrices are $n \times k$ and $n \times r$, respectively. The quantities $\boldsymbol{f}(\boldsymbol{\beta},\boldsymbol{\gamma})$, $\boldsymbol{F}_{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\gamma})$, and $\boldsymbol{F}_{\boldsymbol{\gamma}}(\boldsymbol{\beta},\boldsymbol{\gamma})$ are all functions of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, as are several other quantities to be introduced below. When one of these, $\boldsymbol{f}(\boldsymbol{\beta},\boldsymbol{\gamma})$ for example, is evaluated at the true values, say $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$, it will simply be denoted $\boldsymbol{f}$.

The functions f_t may depend on past values of y_t, but not on current or future values, since otherwise (1) would not be a regression model and least squares would no longer be an appropriate estimating technique. In order to ensure that all estimators and test statistics behave sensibly as $n \to \infty$, it is assumed that $\boldsymbol{F}_{\boldsymbol{\beta}}^{\top}\boldsymbol{F}_{\boldsymbol{\beta}}/n$, $\boldsymbol{F}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{F}_{\boldsymbol{\gamma}}/n$, and $\boldsymbol{F}_{\boldsymbol{\beta}}^{\top}\boldsymbol{F}_{\boldsymbol{\gamma}}/n$ all tend to finite limiting matrices with ranks k, r, and $\min(k,r)$, respectively, as $n \to \infty$, while the matrix $[\boldsymbol{F}_{\boldsymbol{\beta}} \;\; \boldsymbol{F}_{\boldsymbol{\gamma}}]$ always has rank $k+r$ for large n.

It is remarkably easy to test hypotheses about $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ by means of artificial linear regressions without ever estimating the unrestricted model; see, among others, Engle (1982) and Davidson and MacKinnon (1984). Suppose that we wish to test the null hypothesis $\boldsymbol{\gamma} = \mathbf{0}$. If we estimate (1) by least squares imposing this restriction, we will obtain restricted estimates $\tilde{\boldsymbol{\beta}}$ and $\tilde{\sigma}^2$. Evaluating $\boldsymbol{f}(\boldsymbol{\beta},\boldsymbol{\gamma})$, $\boldsymbol{F}_{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\gamma})$, and $\boldsymbol{F}_{\boldsymbol{\gamma}}(\boldsymbol{\beta},\boldsymbol{\gamma})$ at $\tilde{\boldsymbol{\beta}}$ and $\mathbf{0}$, we obtain the vector $\tilde{\boldsymbol{f}}$ and the matrices $\tilde{\boldsymbol{F}}_{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}}$. Using these, we can construct the artificial linear regression

$$(\boldsymbol{y} - \tilde{\boldsymbol{f}})/\tilde{\sigma} = \tilde{\boldsymbol{F}}_{\boldsymbol{\beta}}\boldsymbol{b} + \tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}}\boldsymbol{c} + \text{errors}. \tag{2}$$

The explained sum of squares from this regression, which is also n times the uncentered R^2, is

$$(\boldsymbol{y} - \tilde{\boldsymbol{f}})^{\top}\tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}}(\tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}}^{\top}\tilde{\boldsymbol{M}}_{\boldsymbol{\beta}}\tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}})^{-1}\tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}}^{\top}(\boldsymbol{y} - \tilde{\boldsymbol{f}})/\tilde{\sigma}^2, \tag{3}$$

where $\tilde{\boldsymbol{M}}_{\boldsymbol{\beta}} \equiv \mathbf{I} - \tilde{\boldsymbol{F}}_{\boldsymbol{\beta}}(\tilde{\boldsymbol{F}}_{\boldsymbol{\beta}}^{\top}\tilde{\boldsymbol{F}}_{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{F}}_{\boldsymbol{\beta}}^{\top}$. It is straightforward to show that this test statistic is asymptotically distributed as chi-squared with r degrees of freedom under the null hypothesis. The proof makes use of the facts that, asymptotically, $\boldsymbol{y} - \tilde{\boldsymbol{f}} = \boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{u}$ and that a central limit theorem can be applied to the r–vector $\boldsymbol{F}_{\boldsymbol{\gamma}}\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{u}$. Note that the numerator of (3) is also the numerator of the ordinary F statistic for a test of $\boldsymbol{c} = \mathbf{0}$ in (2), so that (3) is asymptotically equivalent to an F test.

The test statistic (3) may be thought of as a generic test in what Davidson and MacKinnon (1985a) refers to as “regression directions”. It is generic because the fact that it is based solely on restricted estimates does not prevent it from having the same asymptotic properties, under the null and under local alternatives, as other standard tests. If we had assumed that the u_t were normally distributed, (3) could have been derived as a Lagrange Multiplier test, and standard results on the equivalence of LM, Wald, and Likelihood Ratio tests could have been invoked; see Engle (1984). It is obvious that this equivalence does not depend on normality.

By a regression direction is meant any direction from the null hypothesis, in the space of likelihood functions, which corresponds, at least locally, to a regression model. It is clear that (3) is testing in regression directions because (1) is a regression model whether or not $\boldsymbol{\gamma} = \mathbf{0}$. But a test in regression directions need not be explicitly derived from an alternative hypothesis which is a regression model, although of course the null must be such a model. If we were simply to replace the matrix $\tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}}$ in (2) and (3) by an arbitrary matrix $\boldsymbol{Z}$, asymptotically uncorrelated with $\boldsymbol{u}$ under the null hypothesis, we would obtain the asymptotically valid test statistic

$$(\boldsymbol{y} - \tilde{\boldsymbol{f}})^\top \boldsymbol{Z}(\boldsymbol{Z}^\top \tilde{\boldsymbol{M}}_{\boldsymbol{\beta}} \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top (\boldsymbol{y} - \tilde{\boldsymbol{f}})/\tilde{\sigma}^2. \tag{4}$$

Using this device, specification tests in the sense of Hausman (1978) and Holly (1982) can be computed in the same way as other tests in regression directions whenever the null is a regression model; see Davidson and MacKinnon (1985b). So can non-nested hypothesis tests, encompassing tests, and differencing tests; see Davidson and MacKinnon (1981), Mizon and Richard (1986), and Davidson, Godfrey, and MacKinnon (1985), respectively. In the next section, we shall ignore where the matrix $\boldsymbol{Z}$ and associated test statistic (4) came from and analyze what determines the power of all tests in regression directions.

3. The Local Power of Tests in Regression Directions

In order to say anything about the power of a test statistic, one must specify how the data are actually generated. Since we are concerned with tests in regression directions, we shall restrict our attention to DGPs which differ from the null hypothesis only in such directions. This restriction is of course by no means innocuous, as will be made clear in Section 5 below. The null hypothesis will be (1) with $\boldsymbol{\gamma} = \mathbf{0}$. Since the alternative that $\boldsymbol{\gamma} \neq \mathbf{0}$ is only one of many alternatives against which we may wish to test the null, we shall usually suppress $\boldsymbol{\gamma}$ and write $\boldsymbol{f}(\boldsymbol{\beta}, \mathbf{0})$ as $\boldsymbol{f}(\boldsymbol{\beta})$.

Suppose the data are generated by the sequence of local DGPs

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\beta}_0) + \alpha n^{-1/2} \boldsymbol{a} + \boldsymbol{u}, \quad \mathrm{E}(\boldsymbol{u}) = \mathbf{0}, \quad \mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \sigma_0^2 \mathbf{I}, \tag{5}$$

where $\boldsymbol{\beta}_0$ and σ_0 denote particular values of $\boldsymbol{\beta}$ and σ, $\boldsymbol{a}$ is an n–vector which may depend on exogenous variables, the parameter vector $\boldsymbol{\beta}_0$, and past values of the y_t, and α is a parameter which determines how far the DGP is from the *simple* null hypothesis

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\beta}_0) + \boldsymbol{u}, \quad \mathrm{E}(\boldsymbol{u}) = \mathbf{0}, \quad \mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \sigma_0^2 \mathbf{I}. \tag{6}$$

We assume that $\boldsymbol{a}^\top \boldsymbol{a}/n$, $\boldsymbol{a}^\top \boldsymbol{F}_{\boldsymbol{\beta}}$, and $\boldsymbol{a}^\top \boldsymbol{Z}$ all tend to finite limiting matrices as $n \to \infty$.

The notion of a sequence of local DGPs requires some discussion. Following Pitman (1949) and many subsequent authors, we adopt it because it seems the most reasonable way to deal with power in the context of asymptotic theory. The sequence (5) approaches the simple null (6) at a rate of $n^{-1/2}$. This rate is chosen so that the test statistic (4), and all asymptotically equivalent test statistics, will be of order unity as $n \to \infty$. If, on the contrary, the DGP were held fixed as the sample size was increased, the test statistic would normally tend to blow up, and it would be impossible to talk about its asymptotic distribution.

Sequences like (5) have not been widely used in econometrics. Most authors who investigate the asymptotic power of test statistics have been content to conduct their analysis on the assumption that the sequence of local DGPs actually lies within the compound alternative against which the test is constructed; see, for example, Engle (1984). But this makes it impossible to study how the power of a test depends on the relations among the null, the alternative, and the DGP, so that the ensuing analysis can shed no light on the difficult question of how to interpret significant test statistics. One paper which uses a sequence similar to (5) is Davidson and MacKinnon (1982), which studies the power of various non-nested hypothesis tests. They conclude that several of the tests are asymptotically equivalent under all sequences of local DGPs, a stronger conclusion than would be possible using the conventional assumption.

The sequence (5) provides a perfectly general *local* representation of any regression model which is sufficiently close to the simple null (6). For example, suppose that we wish to see how a certain test performs when the data are generated by an alternative like (1), with $\boldsymbol{\gamma} \neq \mathbf{0}$. We could simply specify the sequence of local DGPs as

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\beta}_0, \alpha n^{-1/2}\boldsymbol{\gamma}_0) + \boldsymbol{u} \tag{7}$$

with the usual assumptions about the vector $\boldsymbol{u}$, where $\boldsymbol{\gamma}_0$ is fixed, and α determines how far (7) is from the simple null hypothesis (6). Because (7) approaches (6) as $n \to \infty$, a first-order Taylor approximation to (7) around $\alpha = 0$ must yield exactly the same results, in an asymptotic analysis, as (7) itself. This approximation is

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\beta}_0) + \alpha n^{-1/2}\boldsymbol{F}_{\boldsymbol{\gamma}}(\boldsymbol{\beta}_0)\boldsymbol{\gamma}_0 + \boldsymbol{u}, \tag{8}$$

where $\boldsymbol{F}_{\boldsymbol{\gamma}}(\boldsymbol{\beta}_0) \equiv \boldsymbol{F}_{\boldsymbol{\gamma}}(\boldsymbol{\beta}_0, \mathbf{0})$. We can see that (8) is simply a particular case of (5), with the vector $\boldsymbol{F}_{\boldsymbol{\gamma}}(\boldsymbol{\beta}_0, \mathbf{0})\boldsymbol{\gamma}_0$ playing the role of $\boldsymbol{a}$.

We now wish to find the asymptotic distribution of the test statistic (4) under the sequence of local DGPs (5). We first rewrite (4) so that all factors are O(1):

$$\big(n^{-1/2}(\boldsymbol{y} - \tilde{\boldsymbol{f}})^{\top}\boldsymbol{Z}\big)\big(n^{-1}\boldsymbol{Z}^{\top}\tilde{\boldsymbol{M}}_{\boldsymbol{\beta}}\boldsymbol{Z}\big)^{-1}\big(n^{-1/2}\boldsymbol{Z}^{\top}(\boldsymbol{y} - \tilde{\boldsymbol{f}})\big)/\tilde{\sigma}^2. \tag{9}$$

The factor of $n^{-1/2}$ in (5) ensures that $\tilde{\sigma}^2 \to \sigma_0^2$ as $n \to \infty$, and it is obvious that

$$\big(n^{-1}\boldsymbol{Z}^{\top}\tilde{\boldsymbol{M}}_{\boldsymbol{\beta}}\boldsymbol{Z}\big)^{-1} \to \Big(\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{Z}^{\top}\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{Z}\Big)^{-1}. \tag{10}$$

Now recall that

$$\boldsymbol{f}(\tilde{\boldsymbol{\beta}}) \overset{a}{=} \boldsymbol{f}(\boldsymbol{\beta}_0) + \boldsymbol{F}_{\boldsymbol{\beta}}(\boldsymbol{F}_{\boldsymbol{\beta}}^{\top}\boldsymbol{F}_{\boldsymbol{\beta}})^{-1}\boldsymbol{F}_{\boldsymbol{\beta}}^{\top}\big(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\beta}_0)\big), \tag{11}$$

where $\stackrel{a}{=}$ means "is asymptotically equal to". It follows from (11) that

$$\begin{aligned} \boldsymbol{y} - \boldsymbol{f}(\tilde{\boldsymbol{\beta}}) &\stackrel{a}{=} \alpha n^{-1/2}\boldsymbol{a} + \boldsymbol{u} - \boldsymbol{F}_{\boldsymbol{\beta}}(\boldsymbol{F}_{\boldsymbol{\beta}}^{\top}\boldsymbol{F}_{\boldsymbol{\beta}})^{-1}\boldsymbol{F}_{\boldsymbol{\beta}}^{\top}(\alpha n^{-1/2}\boldsymbol{a} + \boldsymbol{u}) \\ &= \boldsymbol{M}_{\boldsymbol{\beta}}(\alpha n^{-1/2}\boldsymbol{a} + \boldsymbol{u}). \end{aligned} \tag{12}$$

The test statistic (4) is thus asymptotically equal to

$$\big(\alpha n^{-1}\boldsymbol{a} + n^{-1/2}\boldsymbol{u}\big)^{\top}\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{Z}\Big(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{Z}^{\top}\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{Z}\Big)^{-1}\boldsymbol{Z}^{\top}\boldsymbol{M}_{\boldsymbol{\beta}}\big(\alpha n^{-1}\boldsymbol{a} + n^{-1/2}\boldsymbol{u}\big)/\sigma_0^2. \tag{13}$$

It is easy to find the asymptotic distribution of (13). First, define $\boldsymbol{P}_0$ as an $r \times r$ triangular matrix such that

$$\boldsymbol{P}_0\boldsymbol{P}_0^{\top} = \Big(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{Z}^{\top}\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{Z}\Big)^{-1}, \tag{14}$$

and then define the r–vector $\boldsymbol{\eta}$ as

$$\boldsymbol{\eta} \equiv \boldsymbol{P}_0^{\top}\boldsymbol{Z}^{\top}\boldsymbol{M}_{\boldsymbol{\beta}}\big(\alpha n^{-1}\boldsymbol{a} + n^{-1/2}\boldsymbol{u}\big)/\sigma_0. \tag{15}$$

The test statistic (13) now takes the very simple form $\boldsymbol{\eta}^{\top}\boldsymbol{\eta}$; it is just the sum of r squared random variables which are the elements of the vector $\boldsymbol{\eta}$. It is clear that, asymptotically, the mean of $\boldsymbol{\eta}$ is the vector

$$\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\boldsymbol{P}_0^{\top}\boldsymbol{Z}^{\top}\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{a}\Big) \tag{16}$$

and its variance-covariance matrix is

$$\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\sigma_0^2\boldsymbol{P}_0^{\top}\boldsymbol{Z}^{\top}\boldsymbol{M}_{\boldsymbol{\beta}}\mathrm{E}(\boldsymbol{u}\boldsymbol{u}^{\top})\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{Z}\boldsymbol{P}_0\Big) \tag{17}$$

$$= \boldsymbol{P}_0^{\top}\Big(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{Z}^{\top}\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{Z}\Big)\boldsymbol{P}_0 = \mathbf{I}_r. \tag{18}$$

Since $\boldsymbol{\eta}$ is equal to n times a weighted sum of random variables with mean zero and finite variance, and since our assumptions keep those weights bounded from above and below, a central limit theorem can be applied to it. The test statistic (9) is thus asymptotically equal to a sum of r independent squared normal random variates, each with variance unity, and with means given by the vector (16). Such a sum has the non-central chi-squared distribution with r degrees of freedom and non-centrality parameter, or NCP, given by the squared norm of the mean vector, which in this case is equal to

$$\frac{\alpha^2}{\sigma_0^2}\Big(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{a}^{\top}\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{Z}\Big)\Big(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{Z}^{\top}\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{Z}\Big)^{-1}\Big(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{Z}^{\top}\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{a}\Big). \tag{19}$$

The NCP (19) can be rewritten in a more illuminating way. Consider the vector $\alpha n^{-1/2}\boldsymbol{M_\beta a}$, the length of which, asymptotically, is

$$\alpha^2 \operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\boldsymbol{a}^\top\boldsymbol{M_\beta a}\Big) \tag{20}$$

The quantity (20) is a measure of the distance between the DGP (5) and a linear approximation to the null hypothesis around the simple null (6); in a sense, it tells us how "wrong" the model being tested is. Now consider the artificial regression

$$\alpha n^{-1/2}\boldsymbol{M_\beta a}/\sigma_0 = \boldsymbol{M_\beta Z d} + \text{errors}. \tag{21}$$

The total sum of squares for this regression, asymptotically, is expression (20) divided by σ_0^2. The explained sum of squares, asymptotically, is the NCP (19). Thus the asymptotic uncentered R^2 from regression (21) is

$$\frac{\operatorname{plim}(n^{-1}\boldsymbol{a}^\top\boldsymbol{M_\beta Z})\big(\operatorname{plim} n^{-1}\boldsymbol{Z}^\top\boldsymbol{M_\beta Z}\big)^{-1}\operatorname{plim}(n^{-1}\boldsymbol{Z}^\top\boldsymbol{M_\beta a})}{\operatorname{plim}(n^{-1}\boldsymbol{a}^\top\boldsymbol{M_\beta a})}. \tag{22}$$

Expression (22) has an alternative interpretation. Consider the asymptotic projection of $\alpha n^{-1/2}\boldsymbol{M_\beta a}$ onto the space spanned by $\boldsymbol{F_\beta}$ and $\boldsymbol{Z}$ jointly. This projection is

$$\alpha n^{-1/2}\boldsymbol{M_\beta Z}\Big(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{Z}^\top\boldsymbol{M_\beta Z}\Big)^{-1}\Big(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{Z}^\top\boldsymbol{M_\beta a}\Big). \tag{23}$$

Now let ϕ be the angle between the vector $\alpha n^{-1/2}\boldsymbol{M_\beta a}$ and the projection (23). By the definition of a cosine, it is easily seen that $\cos^2\phi$ is equal to the R^2 (22). Thus we may rewrite the NCP (19) as

$$\frac{\alpha^2}{\sigma_0^2}\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\boldsymbol{a}^\top\boldsymbol{M_\beta a}\Big)\cos^2\phi \tag{24}$$

or simply as

$$\alpha^2\sigma_0^{-2}\cos^2\phi \tag{25}$$

if we normalize the vector $\boldsymbol{a}$ so that $\operatorname{plim}(n^{-1}\boldsymbol{a}^\top\boldsymbol{M_\beta a}) = 1$ and rescale α appropriately.

If a test statistic which has the non-central chi-squared distribution with r degrees of freedom is compared with critical values from the corresponding central chi-squared distribution, the power of the test increases monotonically with the NCP. Expression (25) writes this NCP as the product of three factors. The first factor, α^2, measures the distance between the DGP and the closest point on a linear approximation to the null hypothesis. Note that this distance in no way depends on $\boldsymbol{Z}$; the greater α^2, the more powerful *any* test will be. Like the other two factors, this first factor is independent of n. For the first factor, that comes about because the DGP approaches the null hypothesis at a rate of $n^{-1/2}$. If, on the contrary, the DGP were fixed as n increased, this factor would have to be proportional to n. Of course, the asymptotic analysis we have done would not be valid if the DGP were fixed, although it would provide a useful approximation in most cases.

The second factor in expression (25) is σ_0^{-2}. This tells us that the NCP is inversely proportional to the variance of the DGP, which makes sense because, as the DGP becomes noisier, it should become harder to reject any null hypothesis. What affects the NCP is α^2/σ_0^2, the ratio of the systematic discrepancy between the null and the DGP to the noise in the latter. Note that this ratio does not depend on $\boldsymbol{Z}$. It will be the same for all tests in regression directions of any given null hypothesis with any given data set.

The most interesting factor in expression (25) is the third one, $\cos^2\phi$. It is only through this factor that the choice of $\boldsymbol{Z}$ affects the NCP. A test will have maximal power, for a given number of degrees of freedom, when $\cos^2\phi = 1$, that is, when the artificial regression (21) has an R^2 of one. This will be the case whenever the vector $\boldsymbol{a}$ is a linear combination of the vectors in $\boldsymbol{F_\beta}$ and $\boldsymbol{Z}$, which will occur whenever the DGP lies within the alternative against which the test is constructed. For example, if the null hypothesis were $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\beta}, \boldsymbol{0})$, the alternative were $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\beta}, \boldsymbol{\gamma})$, and the DGP were (8), then $\boldsymbol{Z}$ would be $\boldsymbol{F_\gamma}$, and $\boldsymbol{a}$ would be a linear combination of the columns of $\boldsymbol{F_\gamma}$; see (8). Thus the power of a test is maximized when we test against the truth.

On the other hand, a test will have no power at all when $\cos^2\phi = 0$. This would occur if $\boldsymbol{M_\beta a}$ were asymptotically orthogonal to $\boldsymbol{M_\beta Z}$, something which in general would seem to be highly unlikely. However, special features of a model, or of the sample design, may make such a situation less uncommon than one might think. Nevertheless, it is probably not very misleading to assert that, when a null hypothesis is false in a regression direction, almost any test in regression directions can be expected to have *some* power, although perhaps not very much.

These results make it clear that there is always a tradeoff when we choose what regression directions to test against. By increasing the number of columns in $\boldsymbol{Z}$, we can always increase $\cos^2\phi$, or at worst leave it unchanged, which by itself will increase the power of the test. But doing so also increases r, the number of degrees of freedom, which by itself reduces the power of the test. This tradeoff is at the heart of a number of controversies in the literature on hypothesis testing. These include the debate over non-nested hypothesis tests versus encompassing tests and the literature on specification tests versus classical tests. For the former, see Dastoor (1983) and Mizon and Richard (1986). For the latter, see Hausman (1978) and Holly (1982).

The tradeoff between $\cos^2\phi$ and degrees of freedom is affected by the sample size. As n increases, the NCP can be expected to increase, because in reality the DGP is not approaching the null as $n \to \infty$, so that a given change in $\cos^2\phi$ will have a larger effect on power the larger is n. On the other hand, the effect of r on the critical value is independent of sample size. Thus, when the sample size is small, it is particularly important to use tests with few degrees of freedom. In contrast, when the sample size is large, it becomes feasible to look in many directions at once so as to maximize $\cos^2\phi$.

If we were confident that the null could only be false in a single direction (that is, if we knew exactly what the vector $\boldsymbol{a}$ might be), the optimal procedure would be to have only one column in $\boldsymbol{Z}$, that column being proportional to $\boldsymbol{a}$. In practice, we are rarely in that happy position. There are normally a number of things which we suspect might be wrong with our model, and hence a large number of regression directions in which to test. Faced with this situation, there are at least two ways to proceed.

One approach is to test against each type of potential misspecification separately, with each test having only one or a few degrees of freedom. If the model is in fact wrong in one or a few of the regression directions in which these tests are carried out, such a procedure is as likely as any to inform us of that fact. However, the investigator must be careful to control the overall size of the test, since when one does, say, ten different tests each at the .05 level, the overall size could be as high as .40. Thus one should avoid jumping to the conclusion that the model is wrong in a particular way just because a certain test statistic is significant. Remember that $\cos^2\phi$ will often be well above zero for *many* tests, even if only one thing is wrong with the model.

Alternatively, it is possible to test for a great many types of misspecification at once by putting all the regression directions we want to test against into one big $\boldsymbol{Z}$ matrix. This maximizes $\cos^2\phi$, and hence it maximizes the chance that the test is consistent. It also makes it easy to control the size of the test. But such a test will have many degrees of freedom, so that power may be poor when the sample size is small. Moreover, if such a test rejects the null, that gives us very little information as to what may be wrong with the model.

It may be possible to make some tentative inferences about the true model by looking at the values of several test statistics. Suppose that we test a model against several sets of regression directions, represented by regressor matrices $\boldsymbol{Z}_1$, $\boldsymbol{Z}_2$, and so on, and thus generate test statistics T_1, T_2, and so on. Each of the test statistics T_i can be used to estimate the corresponding NCP, say NCP_i. Since the mean of a non-central chi-squared random variable with r degrees of freedom is r plus the NCP, the obvious estimate of NCP_i is just $T_i - r_i$. It is far from certain that the $\boldsymbol{Z}_i$ with the highest estimated NCP_i, say $\boldsymbol{Z}_i^*$, actually represents truly omitted directions. Nevertheless, modifying the model in the directions represented by $\boldsymbol{Z}_i^*$ would seem to be a reasonable thing to do in many cases, especially when the number of columns in $\boldsymbol{Z}_i^*$ is small. It might be useful to perform a test in all the interesting regression directions one can think of, thus obtaining a test statistic with the largest NCP obtainable. If that test statistic is not much larger than T_i^*, then one might feel reasonably confident that the directions represented by $\boldsymbol{Z}_i^*$ adequately capture the discrepancy between the null and the DGP.

4. Tests that are Robust to Heteroskedasticity of Unknown Form

The distinguishing feature of regression models is that the error term is simply added to a regression function which determines the mean of the dependent variable. This greatly simplifies the analysis of such models. In particular, it means that test statistics such as (4) are asymptotically valid regardless of how the error terms u_t are distributed, provided only that, for all t, $\mathrm{E}(u_t) = 0$, $E(u_t u_s) = 0$ for all $s \neq t$, and $\mathrm{E}(u_t^2) = \sigma^2$. Without normality, least squares will not be asymptotically efficient, and tests based on least squares will not be most powerful, but least squares estimates will be consistent, and tests based on them will be asymptotically valid.

When the error terms u_t display heteroskedasticity of unknown form, least squares estimates remain consistent, but test statistics such as (4) are no longer asymptotically valid. However, the results of White (1980) make it clear that asymptotically valid test statistics can be constructed in this case. Davidson and MacKinnon (1985b) shows how to compute such tests by means of artificial linear regressions and provides some results on their finite-sample properties. These heteroskedasticity-robust tests are likely to be very useful when analyzing cross-section data. In this section, we consider the power properties of such tests.

Under the null hypothesis, the test statistic (4) tends to the random variable

$$\bigl(n^{-1/2}\boldsymbol{u}^\top \boldsymbol{M_\beta Z}\bigr)\Bigl(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{Z}^\top\boldsymbol{M_\beta Z}\Bigr)^{-1}\bigl(n^{-1/2}\boldsymbol{Z}^\top\boldsymbol{M_\beta u}\bigr) \tag{26}$$

as $n \to \infty$. Thus it is evident that (4) is really testing the hypothesis that

$$\lim_{n\to\infty}\mathrm{E}\bigl(n^{-1/2}\boldsymbol{u}^\top\boldsymbol{M_\beta Z}\bigr) = \mathbf{0}. \tag{27}$$

Now suppose that $\mathrm{E}(\boldsymbol{uu}^\top) = \boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is an $n \times n$ diagonal matrix with diagonal elements σ_t^2 bounded from above. The asymptotic variance-covariance matrix of $n^{-1/2}\boldsymbol{u}^\top\boldsymbol{M_\beta Z}$ is

$$\operatorname*{plim}_{n\to\infty}\Bigl(\frac{1}{n}\boldsymbol{Z}^\top\boldsymbol{M_\beta\Omega M_\beta Z}\Bigr). \tag{28}$$

Even though $\boldsymbol{\Omega}$ is unknown and has as many unknown elements as there are observations, the matrix (28) may be estimated consistently in a number of different ways. One of the simplest is to replace $\boldsymbol{\Omega}$ by $\tilde{\boldsymbol{\Omega}}$, a diagonal matrix with diagonal elements $\tilde{u}_t^2$, the $\tilde{u}_t$ being the residuals from least squares estimation of the null hypothesis. Of course, $\boldsymbol{M_\beta}$ must also be replaced by $\tilde{\boldsymbol{M}}_{\boldsymbol{\beta}}$. Other estimators, which may have better finite-sample properties, are discussed in MacKinnon and White (1985).

It is now straightforward to derive a heteroskedasticity-robust test statistic. Written in the same form as (9), so that all factors are O(1), it is

$$\bigl(n^{-1/2}(\boldsymbol{y}-\tilde{\boldsymbol{f}})^\top\boldsymbol{Z}\bigr)\bigl(n^{-1}\boldsymbol{Z}^\top\tilde{\boldsymbol{M}}_{\boldsymbol{\beta}}\tilde{\boldsymbol{\Omega}}\tilde{\boldsymbol{M}}_{\boldsymbol{\beta}}\boldsymbol{Z}\bigr)^{-1}\bigl(n^{-1/2}\boldsymbol{Z}^\top(\boldsymbol{y}-\tilde{\boldsymbol{f}})\bigr). \tag{29}$$

This test statistic is simply n minus the sum of squared residuals from the artificial regression

$$\boldsymbol{\iota} = \tilde{\boldsymbol{U}}\tilde{\boldsymbol{M}}_{\boldsymbol{\beta}}\boldsymbol{Zc} + \text{errors}, \tag{30}$$

where $\boldsymbol{\iota}$ is an n–vector of ones and $\tilde{\boldsymbol{U}} \equiv \mathrm{diag}(\tilde{u}_t)$. That (29) can be calculated in this simple way follows from the facts that

$$\boldsymbol{\iota}^\top\tilde{\boldsymbol{U}}\tilde{\boldsymbol{M}}_{\boldsymbol{\beta}}\boldsymbol{Z} = (\boldsymbol{y}-\tilde{\boldsymbol{f}})^\top\tilde{\boldsymbol{M}}_{\boldsymbol{\beta}}\boldsymbol{Z} = (\boldsymbol{y}-\tilde{\boldsymbol{f}})^\top\boldsymbol{Z} \tag{31}$$

and

$$\boldsymbol{Z}^\top\tilde{\boldsymbol{M}}_{\boldsymbol{\beta}}\tilde{\boldsymbol{U}}^\top\tilde{\boldsymbol{U}}\tilde{\boldsymbol{M}}_{\boldsymbol{\beta}}\boldsymbol{Z} = \boldsymbol{Z}^\top\tilde{\boldsymbol{M}}_{\boldsymbol{\beta}}\tilde{\boldsymbol{\Omega}}\tilde{\boldsymbol{M}}_{\boldsymbol{\beta}}\boldsymbol{Z}. \tag{32}$$

Finding the asymptotic distribution of the heteroskedasticity-robust test statistic (29) is very similar to finding the asymptotic distribution of the ordinary test statistic (9). Under a suitable sequence of local DGPs, it is easy to show that (29) is asymptotically equal to

$$(\alpha n^{-1}\boldsymbol{a} + n^{-1/2}\boldsymbol{u})^\top\Big(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{Z}^\top\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{\Omega}_0\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{Z}\Big)^{-1}(\alpha n^{-1}\boldsymbol{a} + n^{-1/2}\boldsymbol{u}). \tag{33}$$

Here $\boldsymbol{\Omega}_0$ is the $n \times n$ diagonal covariance matrix of the error terms in a sequence of local DGPs similar to (5) except for the heteroskedasticity. An argument very similar to that used earlier can then be employed to show that, asymptotically, (33) has the non-central chi-squared distribution with r degrees of freedom and NCP

$$\alpha^2\Big(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{a}^\top\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{Z}\Big)\Big(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{Z}^\top\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{\Omega}_0\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{Z}\Big)^{-1}\Big(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{Z}^\top\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{a}\Big). \tag{34}$$

Like the earlier NCP (19), expression (34) can also be interpreted as the explained sum of squares from a certain artificial regression. In this case, the t^{th} element of the regressand is $\alpha n^{-1/2}(\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{a})_t\sigma_t$, and the t^{th} row of the regressor matrix is $\sigma_t(\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{Z})_t$. The NCP may then be written as

$$\alpha^2\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\boldsymbol{a}^\top\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{\Omega}_0\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{a}\Big)\psi, \tag{35}$$

where ψ denotes the uncentered, asymptotic R^2 from the artificial regression.

Expression (35) resembles expression (24), but it differs from it in two important respects. First of all, the factor which measures the distance between the DGP and the null hypothesis relative to the noisiness of the DGP is now

$$\alpha^2\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\boldsymbol{a}^\top\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{\Omega}_0\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{a}\Big)$$

rather than

$$\frac{\alpha^2}{\sigma_0^2}\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\boldsymbol{a}^\top\boldsymbol{M}_{\boldsymbol{\beta}}\boldsymbol{a}\Big).$$

Secondly, although ψ plays the same role as $\cos^2\phi$ and is the only factor which is affected by the choice of $\boldsymbol{Z}$, ψ does not have quite the same properties as $\cos^2\phi$. It is possible to make ψ zero by choosing $\boldsymbol{Z}$ appropriately, but it is usually not possible to make ψ unity even by choosing $\boldsymbol{Z}$ so that $\boldsymbol{a}$ lies in the span of the columns of $\boldsymbol{Z}$. That would of course be possible if $\boldsymbol{\Omega}_0$ were proportional to the identity matrix, in which case (35) and (24) would be identical. Thus, when there is no heteroskedasticity, the heteroskedasticity-robust test statistic (29) is asymptotically equivalent, under all sequences of local DGPs, to the ordinary test statistic (9). There will, however, be some loss of power in finite samples; see Davidson and MacKinnon (1985b).

It is clear from expression (35) that when there is in fact heteroskedasticity, the pattern of the error variances will affect the power of the test. Multiplying all elements of $\boldsymbol{\Omega}_0$ by a factor λ will of course reduce the NCP by a factor $1/\lambda$, as in the homoskedastic case. But changes in the pattern of heteroskedasticity, even if they do not affect the average value of σ_t, may well affect $\operatorname{plim}(n^{-1}\boldsymbol{Z}^\top\boldsymbol{M_\beta}\boldsymbol{\Omega}_0\boldsymbol{M_\beta}\boldsymbol{Z})$ and hence affect the power of the test.

As a result of this, the interpretation of heteroskedasticity-robust tests is even harder than the interpretation of ordinary tests in regression directions. As discussed in Section 4, we know in the ordinary case that the NCP will be highest when we test against the truth, and so looking at several test statistics can provide some guidance as to where the truth lies. In the heteroskedasticity-robust case, however, things are not so simple. It is entirely possible that $\boldsymbol{a}$ may lie in the span of the columns of $\boldsymbol{F_\beta}$ and $\boldsymbol{Z}_1$, so that a test against the directions represented by $\boldsymbol{Z}_1$ is in effect a test against the truth, and yet the NCP may be substantially higher when testing against some quite different set of directions represented by $\boldsymbol{Z}_2$. Thus, in the common situation in which several different tests reject the null hypothesis, it may be far from obvious how the model should be modified.

5. Tests in Mixed Non-Regression Directions

The popularity of regression models is easy to understand. They have evolved naturally from the classical problem of estimating a mean; they are easy to write down and interpret; and they are usually quite easy to estimate. The regression specification is, however, very restrictive. By forcing the error term to be additive, it greatly limits the way in which random events outside the model can affect the dependent variable. In order to see whether this is in fact a severe restriction, careful applied workers will usually wish to test their models in non-regression as well as regression directions.

As we saw above, regression directions are those which correspond, at least locally, to a more general regression model in which the null is nested. Non-regression directions, then, are those which correspond either to a regression model with a different error structure (as in tests for heteroskedasticity, skewness, and kurtosis), or to a more general non-regression model. We shall refer to the latter as "mixed" directions, since they typically affect both the mean and the higher moments of the dependent variable.

One non-regression model which has been widely used in applied econometrics is the Box-Cox regression model; see, among others, Zarembka (1974) and Savin and White (1978). This model can be written as

$$y_t(\lambda) = \sum_{i=1}^{k} \beta_i X_{ti}^1(\lambda) + \sum_{j=1}^{m} \gamma_j X_{tj}^2 + u_t, \tag{36}$$

where $x(\lambda)$ denotes the Box-Cox transformation:

$$x(\lambda) \equiv \frac{x^\lambda - 1}{\lambda} \text{ if } \lambda \neq 0; \quad x(\lambda) \equiv \log x \text{ if } \lambda = 0. \tag{37}$$

Here the X_{ti}^1 are regressors which may sensibly be subjected to the Box-Cox transformation, while the X_{tj}^2 are ones which cannot sensibly be thus transformed, such as the constant term, dummy variables, and variables which can take on non-positive values.

Conditional on λ, the Box-Cox model (36) is a regression model. When $\lambda = 1$ and there is a constant term, it is a linear model, and when $\lambda = 0$, it is a loglinear one. Either of these null hypotheses may be tested against the Box-Cox alternative (36), and numerous procedures exist for doing so. These are examples of tests in mixed directions, because (36) is not a regression model when λ is a parameter to be estimated. The reason is that the regression function in (36) determines the mean of $y_t(\lambda)$ rather than the mean of the actual dependent variable. If we were to rewrite (36) so that y_t was on the left-hand side, it would be clear that y_t is actually a nonlinear function of the X_{ti}^1, the X_{tj}^2, and u_t.

The obvious way to test both linear and loglinear null hypotheses against the Box-Cox alternative (36) is to use some form of the Lagrange Multiplier test. Several such tests have been proposed. In particular, Godfrey and Wickens (1981) suggest using the "outer product of the gradient", or OPG, variant of the LM test, while Davidson and MacKinnon (1985c) suggest using the "double-length regression", or DLR, variant. Each of these variants can be computed by means of a single artificial linear regression, and each can handle a wide variety of tests in mixed non-regression directions.

In the remainder of this section, we shall discuss the OPG variant of the LM test. We shall not discuss the DLR variant, which was originally proposed by Davidson and MacKinnon (1984), even though it appears to have substantially better finite-sample properties than the OPG variant. Discussion of the asymptotic power properties of the DLR variant may be found in Davidson and MacKinnon (1985a, 1985c), and there would be no point in repeating that discussion here. Moreover, the DLR variant is applicable to a narrower class of models than the OPG variant, and it is somewhat more complicated to analyze. Note that our results do not apply merely to the OPG form of the LM test; they are equally valid for any form of the LM test, and for asymptotically equivalent Wald and Likelihood Ratio tests as well.

The OPG variant of the LM test is applicable to any model for which the loglikelihood function may be written as a sum of the contributions from all the observations. Thus, if $\ell_t(y_t, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ denotes the contribution from the t^{th} observation (possibly conditional on previous observations), the loglikelihood function is

$$\mathcal{L}(\boldsymbol{y}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \sum_{t=1}^{n} \ell_t(y_t, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2). \tag{38}$$

The derivatives of $\ell_t(y_t, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ with respect to the i^{th} element of $\boldsymbol{\theta}_1$ and the j^{th} element of $\boldsymbol{\theta}_2$ will be denoted $G_{ti}^1(\cdot)$ and $G_{tj}^2(\cdot)$, respectively. These derivatives may then be formed into the $n \times k$ and $n \times r$ matrices $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$. The null hypothesis is that $\boldsymbol{\theta}_2 = \mathbf{0}$. For the computation of the LM test, all quantities will be evaluated at the restricted ML estimates $\tilde{\boldsymbol{\theta}}_1$ and $\mathbf{0}$.

In the particular case of testing the null hypothesis of loglinearity against the Box-Cox alternative (36), $\boldsymbol{\theta}_1$ would be a vector of the β_i, the γ_j, and σ (or σ^2), and $\boldsymbol{\theta}_2$ would be the scalar λ. Assuming normality, it is easy to write down the loglikelihood function, and we see that

$$\ell_t = -\frac{1}{2}\log 2\pi - \log\sigma - \frac{1}{2}u_t^2/\sigma^2 + (\lambda - 1)\log y_t, \tag{39}$$

where u_t is implicitly defined by (36). It is now easy to calculate the elements of the matrix $\boldsymbol{G}$ and to evaluate them under the null hypothesis that $\lambda = 0$, which simply requires that one estimate the loglinear null by OLS and obtain estimates $\tilde{\beta}_i$, for $i = 1, \ldots, k$, $\tilde{\gamma}_j$, for $j = 1, \ldots, m$, and $\tilde{\sigma}$; see Godfrey and Wickens (1981).

The OPG form of the LM test is remarkably easy to calculate. It is simply the explained sum of squares (or n minus the sum of squared residuals) from the artificial regression

$$\boldsymbol{\iota} = \tilde{\boldsymbol{G}}_1\boldsymbol{b} + \tilde{\boldsymbol{G}}_2\boldsymbol{c} + \text{errors}, \tag{40}$$

where $\boldsymbol{\iota}$ is an n–vector of ones, $\tilde{\boldsymbol{G}}_1 = \boldsymbol{G}_1(\tilde{\boldsymbol{\theta}}_1, \mathbf{0})$, and $\tilde{\boldsymbol{G}}_2 = \boldsymbol{G}_2(\tilde{\boldsymbol{\theta}}_1, \mathbf{0})$. The explained sum of squares from regression (40) is

$$\boldsymbol{\iota}^\top\tilde{\boldsymbol{G}}(\tilde{\boldsymbol{G}}^\top\tilde{\boldsymbol{G}})^{-1}\tilde{\boldsymbol{G}}^\top\boldsymbol{\iota} = \boldsymbol{\iota}^\top\tilde{\boldsymbol{G}}_2(\tilde{\boldsymbol{G}}_2{}^\top\tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{G}}_2)^{-1}\tilde{\boldsymbol{G}}_2{}^\top\boldsymbol{\iota}, \tag{41}$$

where $\tilde{\boldsymbol{M}}_1 = \mathbf{I} - \tilde{\boldsymbol{G}}_1(\tilde{\boldsymbol{G}}_1{}^\top\tilde{\boldsymbol{G}}_1)^{-1}\tilde{\boldsymbol{G}}_1{}^\top$. The equality in (41) follows from the facts that the sum of squared residuals from regression (40) is identical to that from the regression

$$\tilde{\boldsymbol{M}}_1\boldsymbol{\iota} = \tilde{\boldsymbol{M}}_1\tilde{\boldsymbol{G}}_2\boldsymbol{c} + \text{errors} \tag{42}$$

and that $\boldsymbol{\iota}^\top\tilde{\boldsymbol{G}}_1 = \mathbf{0}$ by the first-order conditions for $\tilde{\boldsymbol{\theta}}_1$.

Just as a test in regression directions may look in any such directions, not merely those which are suggested by an explicit alternative hypothesis, so may a test in non-regression directions. It is clearly valid to replace $\tilde{\boldsymbol{G}}_2$ in regression (40) by any $n \times r$ matrix $\boldsymbol{Z}$, provided that the elements of $\boldsymbol{Z}$ are $O(1)$ and that the asymptotic expectation of the mean of every column of $\boldsymbol{Z}$ is zero under the null hypothesis. Newey (1985) and Tauchen (1985) exploit this fact to propose families of specification tests, while Lancaster (1984) uses it to provide a simple way to compute the information matrix test of White (1982).

What we are interested in, then, is the asymptotic distribution of the test statistic

$$\big(n^{-1/2}\boldsymbol{\iota}^\top\boldsymbol{Z}\big)\Big(\frac{1}{n}\boldsymbol{Z}^\top\tilde{\boldsymbol{M}}_1\boldsymbol{Z}\Big)^{-1}\big(n^{-1/2}\boldsymbol{Z}^\top\boldsymbol{\iota}\big), \tag{43}$$

which is the explained sum of squares from the artificial regression

$$\boldsymbol{\iota} = \tilde{\boldsymbol{G}}_1\boldsymbol{b} + \boldsymbol{Z}\boldsymbol{c} + \text{errors}, \tag{44}$$

written in such a way that all factors are O(1). Following Davidson and MacKinnon (1987), we shall suppose that the data are generated by a process which can be described by the loglikelihood

$$\mathcal{L}' = \sum_{t=1}^{n} \bigl(\ell_t(y_t, \boldsymbol{\theta}_1^0, \mathbf{0}) + \alpha n^{-1/2} a_t(y_t)\bigr). \tag{45}$$

Here $\boldsymbol{\theta}_1^0$ is a vector of fixed parameters which determines the simple null hypothesis to which the sequence of DGPs (45) tends as $n \to \infty$, α is a parameter which determines how far (45) is from that simple null, and the $a_t(y_t)$ are random variables which are O(1) and have mean zero under the null. The sequence of local DGPs (45) plays the same role here as the sequence (5) did in our earlier analysis. The major difference between the two is that (45) is written in terms of loglikelihoods, so that the DGP may differ from the null hypothesis in *any* direction in likelihood space, while (5) was written in terms of regression functions, so that the DGP could only differ from the null in regression directions.

Using results of Davidson and MacKinnon (1987), it is possible to show that, under the sequence of local DGPs (45), the statistic (43) is asymptotically distributed as non-central chi-squared with r degrees of freedom and non-centrality parameter

$$\alpha^2 \Bigl(\operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{a}^\top \boldsymbol{M}_1 \boldsymbol{Z}\Bigr) \Bigl(\operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{Z}^\top \boldsymbol{M}_1 \boldsymbol{Z}\Bigr)^{-1} \Bigl(\operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{Z}^\top \boldsymbol{M}_1 \boldsymbol{a}\Bigr), \tag{46}$$

where $\boldsymbol{M}_1 \equiv \mathbf{I} - \boldsymbol{G}_1(\boldsymbol{G}_1^\top \boldsymbol{G}_1)^{-1} \boldsymbol{G}_1^\top$ and $\boldsymbol{G}_1 \equiv \boldsymbol{G}(\boldsymbol{\theta}_1^0, \mathbf{0})$. The similarity between expressions (46) and (19) is striking and by no means coincidental. Note that $\boldsymbol{M}_1$ plays exactly the same role here that $\boldsymbol{M}_{\boldsymbol{\beta}}$ did previously, that $\boldsymbol{a}$ and $\boldsymbol{Z}$ play the same roles as before, although of course their interpretation is different, and that σ_0^2 has no place in (46), because the variance parameters (if any) are subsumed in $\boldsymbol{\theta}_1$ and/or $\boldsymbol{\theta}_2$.

Consider the vector $\alpha n^{-1/2} \boldsymbol{M}_1 \boldsymbol{a}$. Its asymptotic projection onto the space spanned by $\boldsymbol{G}_1$ and $\boldsymbol{Z}$ jointly is

$$\alpha n^{-1/2} \boldsymbol{M}_1 \boldsymbol{Z} \Bigl(\operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{Z}^\top \boldsymbol{M}_1 \boldsymbol{Z}\Bigr)^{-1} \Bigl(\operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{Z}^\top \boldsymbol{M}_1 \boldsymbol{a}\Bigr) \tag{47}$$

If ϕ denotes the angle between $\alpha n^{-1/2} \boldsymbol{M}_1 \boldsymbol{a}$ and the projection (47), then

$$\cos^2 \phi = \frac{\operatorname{plim}(n^{-1} \boldsymbol{a}^\top \boldsymbol{M}_1 \boldsymbol{Z}) \bigl(\operatorname{plim} n^{-1} \boldsymbol{Z}^\top \boldsymbol{M}_1 \boldsymbol{Z}\bigr)^{-1} \operatorname{plim}(n^{-1} \boldsymbol{Z}^\top \boldsymbol{M}_1 \boldsymbol{a})}{\operatorname{plim}(n^{-1} \boldsymbol{a}^\top \boldsymbol{M}_1 \boldsymbol{a})}, \tag{48}$$

which is the uncentered asymptotic R^2 from the artificial regression

$$\alpha n^{-1/2} \boldsymbol{M}_1 \boldsymbol{a} = \boldsymbol{M}_1 \boldsymbol{Z} \boldsymbol{b} + \text{errors}. \tag{49}$$

Thus the NCP (46) may be rewritten as

$$\alpha^2 \Bigl(\operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{a}^\top \boldsymbol{M}_1 \boldsymbol{a}\Bigr) \cos^2 \phi. \tag{49}$$

The interpretation of expression (50) is almost exactly the same as the interpretation of expression (24). The first two factors measure the distance between the DGP and the closest point on a linear approximation to the null hypothesis. The larger these factors, the greater will be the power of *any* test statistic like (43), or of any asymptotically equivalent test statistic. The choice of $\boldsymbol{Z}$ only affects the NCP through $\cos^2\phi$, and a test will have maximal power, for a given number of degrees of freedom, when $\cos^2\phi = 1$. This will be the case if $\boldsymbol{a}$ is a linear combination of the vectors in $\boldsymbol{G}_1$ and $\boldsymbol{Z}$, which will happen whenever the DGP lies within the alternative against which the test is constructed. Thus, once again, the power of a test is maximized when we test against the truth.

When $\cos^2\phi = 0$, a test will have no power at all asymptotically. This is a situation which is likely to arise quite often when using tests in non-regression directions. For example, it can be shown that $\cos^2\phi = 0$ whenever the DGP is in a higher moment direction and we test in regression directions, or *vice versa*. This is true for essentially the same reason that the information matrix for a regression model is block-diagonal between the parameters which determine the regression function and those which determine the higher moments of the error terms. Notice that when $\boldsymbol{Z}$ has only one column, expression (48) for $\cos^2\phi$ is symmetrical in $\boldsymbol{a}$ and $\boldsymbol{Z}$; thus, if a test against alternative 1 has power when alternative 2 is true, a test against alternative 2 must have power when alternative 1 is true.

The artificial regression (49) may actually be used to compute NCPs, and values of $\cos^2\phi$, for models and tests where it is too difficult to work them out analytically. This requires a computer simulation, in which n is allowed to become large enough so that the probability limits in expression (46) are calculated with reasonable accuracy. Such a procedure may tell us quite a lot about the ability of certain test statistics to pick up various types of misspecification.

As an illustration of this technique, we consider certain tests for functional form. The null hypothesis is the linear regression model which emerges from the Box-Cox model (36) when $\lambda = 1$:

$$y_t = 1 + \sum_{i=1}^{k} \beta_i (X_{ti}^1 - 1) + \sum_{j=1}^{m} \gamma_j X_{tj}^2 + u_t. \tag{51}$$

If one of the X_{tj}^2 is a constant term, the various 1s that appear on the right-hand side of equation (51) can be ignored. The data are assumed to be generated by the Box-Cox model, or, more precisely, by a sequence of local approximations to that model so that the loglikelihood function has the form of expression (45). The artificial regression (49) will be used to compute $\cos^2\phi$ for three tests of the model (51), none of which is a classical test of (51) against (36).

First of all, we consider the LM test of (51) against the model

$$y_t(\lambda) = \sum_{i=1}^{k} \beta_i X_{ti}^1 + \sum_{j=1}^{m} \gamma_j X_{tj}^2 + u_t, \tag{52}$$

in which the Box-Cox transformation is applied only to the dependent variable. This model is often just as plausible as (36), and would seem to be a reasonable alternative to test against in many cases.

Secondly, we consider a test originally proposed by Andrews (1971) and later extended by Godfrey and Wickens (1981). The basic idea of the Andrews test is to replace the non-regression direction in which a classical test of the linear or loglinear null against (36) would look with a regression direction which approximates it. One first takes a first-order Taylor approximation to the Box-Cox model (36) around $\lambda = 0$ or $\lambda = 1$. The term which multiplies λ in the Taylor approximation necessarily involves y_t, which is replaced by the fitted value of y_t from estimation under the null. This yields an OLS regression, which looks exactly the same as the original regression, with the addition of one extra regressor. The t statistic on that regressor is the test statistic, and, under the usual conditions for t tests to be exact, it actually has the Student's t distribution with $n - k - m - 1$ degrees of freedom. As Davidson and MacKinnon (1985c) showed, and as we will see shortly, the test regressor often provides a poor approximation to the true non-regression direction of the Box-Cox model, so that the Andrews test can be seriously lacking in power.

Finally, we will consider a test against a particular form of heteroskedasticity, namely that associated with the model

$$y_t = \boldsymbol{X}_t\boldsymbol{\beta} + (1 + \alpha\boldsymbol{X}_t\boldsymbol{\beta})u_t, \quad u_t \sim \mathrm{NID}(0, \sigma^2). \tag{53}$$

When α is greater than zero, this model has heteroskedastic errors with variance proportional to $(1 + \alpha\boldsymbol{X}_t\boldsymbol{\beta})^2$. It is well-known to users of the Box-Cox transformation that estimates of models like (36) are very sensitive to heteroskedasticity, and so it seems likely that a test of $\alpha = 0$ in (53) will have power when the DGP is actually the Box-Cox model.

The contribution of the t^{th} observation to the loglikelihood function for the Box-Cox model (36) is given by expression (39). In order to approximate this expression around $\lambda = 1$ by a sequence like (45), we must set

$$a_t(y_t) = \log y_t - \frac{u_t}{\sigma^2}\Big(h(y_t) - \sum_{i=1}^{k} h(X_{ti}^1)\Big), \tag{54}$$

where

$$h(x) = x \log x - x + 1. \tag{55}$$

The vector $\boldsymbol{a}$ depends only on the DGP, and it will be the same for any test we wish to analyze. If we were interested in the power of an LM test of (51) against (36), $\boldsymbol{Z}$ would be a vector and would be identical to the vector $\boldsymbol{a}$, so that $\cos^2\phi$ would necessarily be unity. In fact, however, we are interested in testing (51) against the alternative Box-Cox model (52) and the heteroskedastic model (53), and in the Andrews test. For the first, we find that

$$Z_t = \log y_t - \frac{u_t}{\sigma^2} h(y_t). \tag{56}$$

For the second, we find that

$$Z_t = \boldsymbol{X}_t \boldsymbol{\beta} (u_t^2/\sigma^2 - 1), \tag{57}$$

and for the third, we find that

$$Z_t = \frac{u_t}{\sigma^2}\Big(h(\bar{y}_t) - \sum_{i=1}^{k} h(X_{ti}^1)\Big), \tag{58}$$

where $\bar{y}_t$ is the non-stochastic part of y_t. The matrix $\boldsymbol{G}$ will be the same for all the tests and is easily derived.

In order actually to compute $\cos^2\phi$ by regression (49), we must specify the model and DGP more concretely than has been done so far. For simplicity, we examine a model with only one regressor in addition to the constant term. The regressor is constant dollar quarterly GNP for Canada for the period 1955:1 to 1979:4 (100 observations). The constant term is chosen to be 1000, and the coefficient of the other regressor unity. Based on results in Davidson and MacKinnon (1985c), we expect $\cos^2\phi$ to be very sensitive to the choice of σ, so the calculations are performed for a range of values. In order to obtain reasonably accurate approximations to probability limits, n is set to 5000; this involves repeating the actual 100 observations on the regressor fifty times. The results presented in Table 1 are averages over 200 replications, and they are quite accurate.

The results in Table 1 are quite striking. When σ is small, $\cos^2\phi$ is very close to unity for the Andrews test and the alternative Box-Cox test, and it is very close to zero for the heteroskedasticity test. Note that $\sigma = 10$ is very small indeed, since the mean value of the dependent variable is 21394. For the Andrews test, $\cos^2\phi$ then declines monotonically towards zero as σ increases, a result previously noted by Davidson and MacKinnon (1985c). This could have been predicted by looking at expressions (54) and (58) for a_t and Z_t, respectively. The behavior of $\cos^2\phi$ for the other two tests is more interesting. For the alternative Box-Cox test, it initially declines, essentially to zero, but then begins to increase again as σ is increased beyond 500. By examining expressions (54) and (56), we can see the reason for this: When σ is large, the second terms in these expressions become small, and the first terms, which are $\log y_t$ in both cases, become dominant. For the heteroskedasticity test, $\cos^2\phi$ initially rises as σ increases from zero, but it reaches a maximum around $\sigma = 1000$ and then falls somewhat thereafter. The reason for this is not entirely clear; possibly the fact that (57) has no $\log y_t$ term begins to matter as σ gets large.

This example illustrates that, once we leave the realm of regression models, the power of a test may depend in quite a complicated way on the parameters of the DGP, as well as on the structure of the null, the alternative and the DGP. Thus techniques for computing $\cos^2\phi$ may be quite useful in practice. The technique we have used here is very widely applicable and quite easy to use, but it may be computationally inefficient in many cases. When LM tests can be computed by means of double-length regressions

(Davidson and MacKinnon, 1984), a more efficient but basically similar technique is available; see Davidson and MacKinnon (1985a, 1985c).

This example also shows that approximating a mixed non-regression direction by a regression direction may yield a test with adequate power, as in the case of the Andrews test with σ small, but it may also yield a test with very low power, as in the case of the same test with σ large. Despite the possibly large loss of power, there may sometimes be a reason to do this. Tests in regression directions are asymptotically insensitive to misspecification of the error process, such as normality. Moreover, the techniques of Section 4 can be used to make tests in regression directions robust to heteroskedasticity. Thus, by applying the artificial regression (30) to the Andrews test regression, one could obtain a heteroskedasticity-robust test of linear and loglinear models against Box-Cox alternatives. If such a test rejected the null hypothesis, and the sample was reasonably large, one could be quite confident that rejection was justified.

6. Conclusion

Any test of an econometric model can be thought of as a test in certain directions in likelihood space. If the null is a regression model, these may be regression directions, higher moment directions, or mixed non-regression directions. The power of a test will depend on the model being tested, the process that generated the data, and the directions in which the test is looking. Section 3 provided a detailed analysis of what determines power when the null hypothesis is a univariate nonlinear regression model, the DGP is also a regression model, and we are testing in regression directions. Section 4 extended this analysis to the case of heteroskedasticity-robust tests and obtained the surprising result that a test may not have highest power when looking in the direction of the truth. Section 5 then considered a much more general case, in which the null and the DGP are merely described by loglikelihood functions, and tests may look in any direction. The results are remarkably similar to those for the regression case, and they are concrete enough to allow one to compute the power of test statistics in a variety of cases.

References

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

Dastoor, N. K. (1983). “Some aspects of testing non-nested hypotheses,” *Journal of Econometrics*, 21, 213–228.

Davidson, R., and J. G. MacKinnon (1981). “Several tests for model specification in the presence of alternative hypotheses,” *Econometrica*, 49, 781–793.

Davidson, R., and J. G. MacKinnon (1982). “Some non-nested hypothesis tests and the relations among them,” *Review of Economic Studies*, 49, 551–565.

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

Davidson, R., and J. G. MacKinnon (1985a). “The interpretation of test statistics,” *Canadian Journal of Economics*, 18, 38–57.

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

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

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

Davidson, R., L. G. Godfrey, and J. G. MacKinnon (1985). “A simplified version of the differencing test,” *International Economic Review*, 26, 639–647.

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

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

Holly, A. (1982). “A remark on Hausman’s specification test. *Econometrica*, 50, 749–759.

Lancaster, T. (1984). “The covariance matrix of the information matrix test,” *Econometrica*, 52, 1051–1053.

MacKinnon, J. G., and H. White (1985). “Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties,” *Journal of Econometrics*, 29, 305–325.

Mizon, G. E., and J.-F. Richard (1986). “The encompassing principle and its application to testing non-nested hypotheses,” *Econometrica*, 54, 657–678.

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

Pitman, E. J. G. (1949). “Notes on non-parametric statistical inference,” Columbia University, mimeographed.

Savin, N. E., and K. J. White (1978). “Estimation and testing for functional form and autocorrelation: A simultaneous approach,” *Journal of Econometrics*, 8, 1–12.

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.

White, H. (1982). “Maximum likelihood estimation of misspecified models,” *Econometrica*, 50, 1–25.

Zarembka, P. (1974). “Transformation of variables in econometrics,” in P. Zarembka (ed.), *Frontiers in Econometrics*, New York, Academic Press.

Table 1. Calculations of $\cos^2\phi$

σ	Alternative Box-Cox Test	Andrews Test	Heteroskedasticity Test
10	0.9784	0.9913	0.0095
20	0.9508	0.9661	0.0349
50	0.7907	0.8198	0.1767
100	0.4767	0.5318	0.4574
200	0.1439	0.2207	0.7621
500	0.0025	0.0431	0.9342
1000	0.1220	0.0113	0.9530
1500	0.2958	0.0057	0.9377
2000	0.4596	0.0033	0.9115
2500	0.5932	0.0025	0.8765

All figures were calculated numerically using $n = 5000$ and 200 replications. Standard errors never exceed 0.0022 and are usually much smaller.