Queen's Economics Department Working Paper No. 691

Double-Length Artificial Regressions

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

1987

Double-Length Artificial Regressions

Russell Davidson

and

James G. MacKinnon

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

Abstract

Artificial linear regressions often provide a convenient way to calculate test statistics and estimated covariance matrices. This paper discusses one family of these regressions, called "double-length" because the number of "observations" in the artificial regression is twice the actual number of observations. These double-length regressions can be useful in a wide variety of situations. They are quite easy to calculate, and, in contrast to the more widely applicable OPG regression, seem to have good properties when applied to samples of modest size. We first discuss how they are related to the familiar Gauss-Newton and squared-residuals regressions for nonlinear regression models, then show how they may be used to test for functional form, and finally discuss several other ways in which they may be useful in applied econometric work.

This research was supported, in part, by grants from the Social Sciences and Humanities Research Council of Canada. We are grateful to David Hendry for comments on earlier versions. This paper appeared in *Oxford Bulletin of Economics and Statistics*, Volume 50, 1988, pp. 203-217.

May, 1987

1. Introduction

In recent years, applied econometricians have become familiar with the idea that "artificial" regressions may provide a convenient way to compute many test statistics. Such regressions are run *after* the model of interest has been estimated, using constructed variables which depend on parameter estimates and on the hypotheses to be tested. Test statistics may then be computed in several ways, perhaps as the explained sum of squares from the artificial regression or as n (the sample size) times the uncentred R^2, both of which forms are often derived as variants of the LM statistic, or perhaps as ordinary t or F statistics calculated from the artificial regression.

Three families of artificial regressions are widely used in applied work. The best-known is the Gauss-Newton family, which can be used to test the parameters of a univariate or multivariate nonlinear regression function. In the univariate case, this family simply involves regressing the residuals from the restricted model (that is, the model in which some of the parameters are estimated subject to the restrictions to be tested) on the derivatives of the regression function with respect to all of the parameters of the unrestricted model; see Section 3. An early and important application in econometrics of tests based on the Gauss-Newton family was to testing linear regression models with lagged dependent variables for serial correlation (Durbin, 1970; Godfrey, 1978), but there have subsequently been a great many other applications; see Pagan (1984). For a general discussion of these tests, see Engle (1982a).

A second widely-used family of artificial regressions is also applicable only to regression models, and is useful when one wants to test for some form of heteroskedasticity. In the univariate case, this family simply involves regressing squared residuals on certain regressors with which they should be asymptotically uncorrelated under the null hypothesis of homoskedasticity. Tests for various forms of heteroskedasticity which utilize this family of artificial regressions include those suggested by Breusch and Pagan (1979), White (1980), Koenker (1981) and Engle (1982b).

The third well-known family of artificial regressions is the "outer product of the gradient" or OPG family, in which a vector of ones is regressed on a matrix of the derivatives of the contributions from the individual observations to the loglikelihood for the unrestricted model. This was used to compute a variant of the Lagrange Multiplier test by Godfrey and Wickens (1981), and has subsequently been used for several other applications, notably the calculation of conditional moment tests (Newey, 1985) and information matrix tests (Chesher, 1983; Lancaster, 1984). It can easily be applied to any situation in which maximum likelihood estimation is employed and the loglikelihood function can conveniently be written as a sum of the contributions made by each of the observations.

When one is dealing solely with regression models, the utility of the Gauss-Newton and squared-residuals regressions is unquestioned. In the context of more general models, however, these artificial regressions are not applicable, and the obvious thing to do is to use the OPG regression. Unfortunately, Monte Carlo evidence suggests that tests based on this artificial regression tend to reject much too often in finite samples.

The most dramatic case reported so far is found in Davidson and MacKinnon (1987), where the OPG variant of the LM test incorrectly rejects the null more than 98% of the time at the nominal 5% level for samples of size 100, and more than 53% of the time for samples of size 1600! In this case the test had nine degrees of freedom; a similar test with two degrees of freedom rejected the null 40% of the time for samples of size 100 and 11% of the time for samples of size 1600. These results suggest that the performance of tests based on the OPG regression may deteriorate markedly as dimensionality increases; earlier results (cited below and in Section 4) dealt only with cases having one or a very few degrees of freedom.

Available evidence thus suggests that using tests based on the OPG regression can be extremely misleading, unless the sample size is very large relative to the number of degrees of freedom. The basic reason for this is that the OPG regression provides an estimate of the information matrix which is in general inefficient and can be severely biased downwards in certain cases; see Section 2. As a result, variants of the LM test calculated using the OPG regression are sometimes much too prone to reject the null hypothesis.

Because the OPG regression is so widely applicable, there is no single procedure which can always replace it. But one procedure which can often be used is the "double-length" artificial regression, or DLR, proposed in Davidson and MacKinnon (1984a), a paper which is hereafter referred to as *DM*. This regression can be thought of as a generalization of *both* the Gauss-Newton regression and the squared-residuals regression. It can deal with a wide variety of models, providing they involve only *continuous* random variables. When the finite-sample properties of tests based on the DLR have been studied, they have generally been quite good, and always far better than those of tests based on the OPG regression; see Davidson and MacKinnon (1983, 1985), Bera and McKenzie (1986), and Godfrey, McAleer, and McKenzie (1986).[1] Nevertheless, the DLR procedure does not seem to be well-known to applied econometricians.[2] The purpose of this note is to remedy this situation. Technical details, which may be found in *DM*, are omitted.

2. Double-Length Regressions

The class of models for which these artificial regressions can be defined may be characterized by

$$f_t(y_t, \boldsymbol{\theta}) = \varepsilon_t, \quad t = 1, \ldots, n, \quad \varepsilon_t \sim \text{NID}(0, 1), \tag{1}$$

where each f_t is a (suitably smooth) function which depends on the random variable y_t, a k–vector of parameters $\boldsymbol{\theta}$, and (implicitly) exogenous and/or predetermined variables. This may seem at first sight to be a rather restrictive class of models, but it

[1] In addition, Davidson and MacKinnon (1984b, 1987) provide evidence on the performance of the OPG regression in cases where the DLR is not applicable.

[2] One exception to this generalization is a very recent paper by Larson (1987), which uses the DLR to test functional form in models using pooled time-series and cross-section data.

is actually quite general. For example, any nonlinear regression model with normal errors,

$$y_t = x_t(\boldsymbol{\beta}) + u_t, \quad u_t \sim \text{NID}(0, \sigma^2) \tag{2}$$

can be written in the form of (1) by making the definitions

$$f_t(y_t, \boldsymbol{\theta}) \equiv \big(y_t - x_t(\boldsymbol{\beta})\big)/\sigma, \quad \boldsymbol{\theta}^\top \equiv [\boldsymbol{\beta}^\top \ \sigma].$$

Moreover, the class of models defined by (1) includes many *non-regression* models, namely, models in which the dependent variable cannot simply be written as the sum of a regression function and an error term. Models with non-normal continuously distributed errors can be transformed to the form (1) by transforming the distribution of the errors to the normal distribution. Even many multivariate models can be handled, although, since the algebra is a little complicated, we will not deal with such models in this paper; see *DM*.

For a model of the form (1), the contribution to the loglikelihood $l(\boldsymbol{\theta})$ made by observation t is

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

where

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

is a Jacobian term. Now let us make the definitions

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

and

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

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

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \equiv \boldsymbol{g}(\boldsymbol{\theta}) = -\boldsymbol{F}^\top(\boldsymbol{\theta})\boldsymbol{f}(\boldsymbol{\theta}) + \boldsymbol{K}^\top(\boldsymbol{\theta})\boldsymbol{\iota}, \tag{3}$$

where $\boldsymbol{\iota}$ is an n–vector of ones. The fundamental result of *DM* is that, for this class of models, the information matrix $\mathcal{I}(\boldsymbol{\theta})$ satisfies

$$\operatorname*{plim}_{n\to\infty}\left(\tfrac{1}{n}\big(\boldsymbol{F}^\top(\boldsymbol{\theta})\boldsymbol{F}(\boldsymbol{\theta}) + \boldsymbol{K}^\top(\boldsymbol{\theta})\boldsymbol{K}(\boldsymbol{\theta})\big)\right) = \mathcal{I}(\boldsymbol{\theta}), \tag{4}$$

and so can be consistently estimated by

$$\tfrac{1}{n}\big(\boldsymbol{F}^\top(\ddot{\boldsymbol{\theta}})\boldsymbol{F}(\ddot{\boldsymbol{\theta}}) + \boldsymbol{K}^\top(\ddot{\boldsymbol{\theta}})\boldsymbol{K}(\ddot{\boldsymbol{\theta}})\big) \tag{5}$$

where $\ddot{\boldsymbol{\theta}}$ is any consistent estimate of $\boldsymbol{\theta}$. Note that we have defined $\mathcal{I}$ so that it is of order unity in the sample size; what some authors refer to as the information matrix is in our notation n times $\mathcal{I}(\boldsymbol{\theta})$.

Strictly speaking, the LM statistic in its score form is

$$\tilde{\boldsymbol{g}}^\top (n\mathcal{I})^{-1} \tilde{\boldsymbol{g}}$$

where $\tilde{\boldsymbol{g}} \equiv \boldsymbol{g}(\tilde{\boldsymbol{\theta}})$ is the gradient evaluated at the restricted estimates $\tilde{\boldsymbol{\theta}}$. In practice, however, we rarely know $\mathcal{I}$, and so we must replace it with something which estimates it consistently under the null hypothesis. Doing so does not affect the asymptotic distribution of the resulting test statistic, but different ways of estimating $\mathcal{I}$ can yield test statistics with dramatically different finite-sample properties. In the case of the DLR, if we let $\tilde{\boldsymbol{f}}$, $\tilde{\boldsymbol{F}}$, and $\tilde{\boldsymbol{K}}$ denote, respectively, $f(\tilde{\boldsymbol{\theta}})$, $F(\tilde{\boldsymbol{\theta}})$, and $K(\tilde{\boldsymbol{\theta}})$, (3) and (4) imply that one valid form of the LM statistic for testing hypotheses about (1) is

$$(-\tilde{\boldsymbol{f}}^\top \tilde{\boldsymbol{F}} + \boldsymbol{\iota}^\top \tilde{\boldsymbol{K}})(\tilde{\boldsymbol{F}}^\top \tilde{\boldsymbol{F}} + \tilde{\boldsymbol{K}}^\top \tilde{\boldsymbol{K}})^{-1}(-\tilde{\boldsymbol{F}}^\top \tilde{\boldsymbol{f}} + \tilde{\boldsymbol{K}}^\top \boldsymbol{\iota}). \tag{6}$$

This is the DLR variant of the LM statistic.

The DLR variant (6) is evidently just the explained sum of squares from the double-length artificial regression

$$\begin{bmatrix} \tilde{\boldsymbol{f}} \\ \boldsymbol{\iota} \end{bmatrix} = \begin{bmatrix} -\tilde{\boldsymbol{F}} \\ \tilde{\boldsymbol{K}} \end{bmatrix} \boldsymbol{b} + \text{residuals}. \tag{7}$$

This artificial regression has $2n$ "observations". The regressand is 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" $t+n$, $\tilde{\boldsymbol{F}}_t$ and $\boldsymbol{K}_t$ denoting, respectively, the t^{th} rows of $\tilde{\boldsymbol{F}}$ and $\tilde{\boldsymbol{K}}$. Intuitively, the reason we need a double-length regression here is that each observation makes two contributions to the loglikelihood function: a sum-of-squares term $-\frac{1}{2}f_t^2$ and a Jacobian term k_t. As a result, the gradient and the information matrix each involve two parts as well, and the only way to take both of these into account is to incorporate two "observations" into the artificial regression for each genuine observation. Notice that ordinary t and F statistics for the hypothesis that those elements of $\boldsymbol{b}$ which correspond to restricted parameters are zero, are just as valid, asymptotically, as the DLR variant of the LM statistic, (6).

In contrast to the DLR regression, the OPG regression is based on the general result, utilized by Berndt, Hall, Hall, and Hausman (1974), that

$$\operatorname*{plim}_{n\to\infty} \left(\tfrac{1}{n} \big(\boldsymbol{G}^\top(\boldsymbol{\theta}) \boldsymbol{G}(\boldsymbol{\theta}) \big) \right)$$

where $\boldsymbol{G}(\boldsymbol{\theta})$ is an $n \times k$ matrix with typical element $G_{ti}(\boldsymbol{\theta}) = \partial g_t(\boldsymbol{\theta})/\partial \theta_i$. Using this result, it is evident that the information matrix can be consistently estimated by

$$\tfrac{1}{n} \big(\boldsymbol{G}^\top(\ddot{\boldsymbol{\theta}}) \boldsymbol{G}(\ddot{\boldsymbol{\theta}}) \big) \tag{8}$$

The difference between (5) and (8) is that the former makes use of details about the structure of the model (1) which the latter, because it is applicable to a much wider range of models, ignores. As a result, (5) tends to estimate the information matrix more accurately than does (8).

The OPG variant of the LM statistic is

$$\boldsymbol{\iota}^\top \tilde{\boldsymbol{G}} (\tilde{\boldsymbol{G}}^\top \tilde{\boldsymbol{G}})^{-1} \tilde{\boldsymbol{G}}^\top \boldsymbol{\iota}, \tag{9}$$

which from (8) and the fact that $\tilde{\boldsymbol{g}} = \tilde{\boldsymbol{G}}^\top \boldsymbol{\iota}$ is seen to be equal to the explained sum of squares from the artificial regression

$$\boldsymbol{\iota} = \tilde{\boldsymbol{G}} \boldsymbol{c} + \text{residuals}. \tag{10}$$

For models of the form (1), $\tilde{\boldsymbol{g}} = \tilde{\boldsymbol{G}}^\top \boldsymbol{\iota} = -\tilde{\boldsymbol{F}}^\top \tilde{\boldsymbol{f}} + \tilde{\boldsymbol{K}}^\top \boldsymbol{\iota}$, so that the only difference between the two variants of the LM test, (6) and (9), is that they use different estimates of the information matrix. But this difference can be crucial, since if the estimated information matrix is too small, its inverse, and hence the corresponding LM statistic, will be too large. [3] The evidence cited in Section 1 suggests that this happens very often with the OPG variant of the test.

3. Applications to the Nonlinear Regression Model

The general expression for a DLR, (7), is deceptively simple, since we have said nothing about the structure of the matrices $\tilde{\boldsymbol{F}}$ and $\tilde{\boldsymbol{K}}$. It is therefore useful to consider some special cases. In order to show how the DLR is related to the well-known Gauss-Newton and squared-residuals regressions, we begin by considering a case where the DLR is often unnecessary, namely, nonlinear regression models with (possibly) heteroskedastic errors. When restrictions apply to the regression function alone, or to the skedastic function (which determines the variance) alone, there is no need to use double-length regressions. However, when restrictions affect *both* the regression function and the skedastic function, they become very useful.

Consider a univariate nonlinear regression model like (2), extended to allow for heteroskedasticity of known form by letting the standard deviation of u_t be given by the skedastic function $h(\alpha + \boldsymbol{Z}_t \boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ is a p–vector. When written in the form of (1), this model becomes

$$f_t(y_t, \boldsymbol{\theta}) = \frac{y_t - x_t(\boldsymbol{\beta})}{h(\alpha + \boldsymbol{Z}_t \boldsymbol{\gamma})}, \tag{11}$$

where $\boldsymbol{\beta}$ is a k–vector (so that $\boldsymbol{\gamma}$ is now a $(k + p + 1)$–vector). The null hypothesis of homoskedasticity is evidently equivalent to the hypothesis that $\boldsymbol{\gamma} = \boldsymbol{0}$, since the standard deviation is then $h(\alpha) \equiv \sigma$. We shall initially assume that this hypothesis is

[3] Evidently, the phrase "too small" in connection with an estimate of a matrix is a very loose one, but it probably conveys the flavour of what is happening better than a more rigorous discussion would do.

true, so as to consider tests of restrictions on $\boldsymbol{\beta}$ alone, and will then separately consider tests of the homoskedasticity assumption. The nature and number of the restrictions on $\boldsymbol{\beta}$ is irrelevant; for simplicity, one can think of them as $r \leq k$ zero restrictions. As before, quantities denoted by "$\sim$" are evaluated at ML estimates subject to those restrictions.

Calculating $\tilde{\boldsymbol{f}}$, $\tilde{\boldsymbol{F}}$, and $\tilde{\boldsymbol{K}}$ for (11) with $\boldsymbol{\gamma} = \boldsymbol{0}$ and substituting them into (7) yields the DLR

$$\begin{bmatrix} \tilde{\boldsymbol{\varepsilon}} \\ \boldsymbol{\iota} \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{X}}/\tilde{\sigma} & \tilde{\boldsymbol{u}}/\tilde{\sigma}^2 \\ \boldsymbol{0} & -\boldsymbol{\iota}/\tilde{\sigma} \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ a \end{bmatrix} + \text{residuals}, \tag{12}$$

where $\tilde{\boldsymbol{\varepsilon}} \equiv \tilde{\boldsymbol{u}}/\tilde{\sigma}$ is an n–vector of normalized residuals, and $\tilde{\boldsymbol{X}}$ is an $n \times k$ matrix with typical element $\partial x_t/\partial \beta_i$ evaluated at $\tilde{\boldsymbol{\beta}}$; if the regression model were linear, $\tilde{\boldsymbol{X}}$ would be the usual $\boldsymbol{X}$ matrix. The first k regressors here correspond to $\boldsymbol{\beta}$, while the last one corresponds to α (i.e. σ). Because we are interested only in the explained sum of squares from this regression, we may transform the regressors in any way that does not affect the explanatory power of the regression. In particular, we may multiply all the regressors by $\tilde{\sigma}$ to yield a DLR equivalent to (12):

$$\begin{bmatrix} \tilde{\boldsymbol{\varepsilon}} \\ \boldsymbol{\iota} \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{X}} & \tilde{\boldsymbol{\varepsilon}} \\ \boldsymbol{0} & -\boldsymbol{\iota} \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ a \end{bmatrix} + \text{residuals}. \tag{13}$$

It is now evident that the last regressor is orthogonal to the regressand. It is also orthogonal to all those other regressors which correspond to elements of $\boldsymbol{\beta}$ that were estimated without restriction (by the first-order conditions), and under the null hypothesis it should be uncorrelated with the remaining regressors as well. Thus it must be valid simply to drop this last regressor. But when it is dropped, the second half of the DLR becomes irrelevant, since all remaining regressors are zero, and we are left with the artificial regression

$$\tilde{\boldsymbol{\varepsilon}} = \tilde{\boldsymbol{X}}\boldsymbol{b} + \text{residuals}, \tag{14}$$

which is a variant of the Gauss-Newton regression. It can be shown that the explained sums of squares from (13) and (14) are in fact both functions of the same random variable. They will *not* however be numerically identical, and the exact relationship between them is

$$\text{ESS}_{\text{DLR}} = \frac{\text{ESS}_{\text{GN}}}{1 - \text{ESS}_{\text{GN}}/2n},$$

where ESS_{DLR} and ESS_{GN} are, respectively, the explained sums of squares (i.e. the test statistics) from (13) and (14). Clearly the former will always be larger than the latter, so that the DLR will always be somewhat more prone to reject the null hypothesis than the Gauss-Newton regression. The difference between them will usually be small when the null hypothesis is correct, unless n is very small or ESS_{GN} is very large.[4]

[4] If, instead of the explained sum of squares, t or F statistics are used, it can be shown that the DLR and Gauss-Newton regressions yield numerically identical results, except for slightly different corrections for degrees of freedom.

Now suppose that we impose no restrictions on $\boldsymbol{\beta}$, but wish to test the hypothesis that $\boldsymbol{\gamma} = \mathbf{0}$, i.e. that there is no heteroskedasticity. Calculating $\tilde{\boldsymbol{f}}$, $\tilde{\boldsymbol{F}}$, and $\tilde{\boldsymbol{K}}$ and substituting them into (7) yields the DLR

$$\begin{bmatrix} \tilde{\boldsymbol{\varepsilon}} \\ \boldsymbol{\iota} \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{X}}/\tilde{\sigma} & \tilde{\boldsymbol{u}}/\tilde{\sigma}^2 & \big(h'(\tilde{\alpha})/\tilde{\sigma}^2\big)\tilde{\boldsymbol{u}} * \boldsymbol{Z} \\ \mathbf{0} & -\boldsymbol{\iota}/\tilde{\sigma} & -\big(h'(\tilde{\alpha})/\tilde{\sigma}\big)\boldsymbol{Z} \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ a \\ \boldsymbol{c} \end{bmatrix} + \text{residuals}, \tag{15}$$

where "$*$" denotes the Schur or direct product, so that a typical element of $\tilde{\boldsymbol{u}} * \boldsymbol{Z}$ is $\tilde{u}_t Z_{ti}$. After we multiply all regressors by $\tilde{\sigma}$ and the regressors which correspond to $\boldsymbol{\gamma}$ by $h'(\tilde{\alpha})$, (15) simplifies radically to

$$\begin{bmatrix} \tilde{\boldsymbol{\varepsilon}} \\ \boldsymbol{\iota} \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{X}} & \tilde{\boldsymbol{\varepsilon}} & \tilde{\boldsymbol{\varepsilon}} * \boldsymbol{Z} \\ \mathbf{0} & -\boldsymbol{\iota} & -\boldsymbol{Z} \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ a \\ \boldsymbol{c} \end{bmatrix} + \text{residuals}. \tag{16}$$

Notice that, as with all LM tests for heteroskedasticity, the actual functional form of the skedastic function $h(\cdot)$ is irrelevant.

A more familiar test for heteroskedasticity is Breusch and Pagan's (1979) LM test, which is one-half the explained sum of squares from a regression of the squared normalized residuals, centred about their mean of one, on a constant and the matrix $\boldsymbol{Z}$:

$$\tilde{\boldsymbol{\varepsilon}} * \tilde{\boldsymbol{\varepsilon}} - \boldsymbol{\iota} = a\boldsymbol{\iota} + \boldsymbol{Z}\boldsymbol{c} + \text{residuals}. \tag{17}$$

It is not immediately obvious that the ESS from (16) is equivalent to this. The first set of regressors in (16), those which correspond to $\boldsymbol{\beta}$, are orthogonal both to the regressand and to the regressor which corresponds to α, and are asymptotically independent of the other regressors, so that they can be dropped from the DLR. This leaves

$$\begin{bmatrix} \tilde{\boldsymbol{\varepsilon}} \\ \boldsymbol{\iota} \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{\varepsilon}} & \tilde{\boldsymbol{\varepsilon}} * \boldsymbol{Z} \\ -\boldsymbol{\iota} & -\boldsymbol{Z} \end{bmatrix} \begin{bmatrix} a \\ \boldsymbol{c} \end{bmatrix} + \text{residuals}. \tag{18}$$

The $\boldsymbol{X}^\top\boldsymbol{X}$ matrix from (17) is

$$\begin{bmatrix} \boldsymbol{\iota}^\top\boldsymbol{\iota} & \boldsymbol{\iota}^\top\boldsymbol{Z} \\ \boldsymbol{Z}^\top\boldsymbol{\iota} & \boldsymbol{Z}^\top\boldsymbol{Z} \end{bmatrix}, \tag{19}$$

while the $\boldsymbol{X}^\top\boldsymbol{X}$ matrix from (18) is

$$\begin{bmatrix} \tilde{\boldsymbol{\varepsilon}}^\top\tilde{\boldsymbol{\varepsilon}} + \boldsymbol{\iota}^\top\boldsymbol{\iota} & (\tilde{\boldsymbol{\varepsilon}} * \tilde{\boldsymbol{\varepsilon}})^\top\boldsymbol{Z} + \boldsymbol{\iota}^\top\boldsymbol{Z} \\ \boldsymbol{Z}^\top(\tilde{\boldsymbol{\varepsilon}} * \tilde{\boldsymbol{\varepsilon}}) + \boldsymbol{Z}^\top\boldsymbol{\iota} & (\tilde{\boldsymbol{\varepsilon}} * \tilde{\boldsymbol{\varepsilon}})^\top(\boldsymbol{Z} * \boldsymbol{Z}) + \boldsymbol{Z}^\top\boldsymbol{Z} \end{bmatrix} \tag{20}$$

It can be seen that every element of (20) is equal either to twice the corresponding element of (19), or to a random variable which will on average be twice the corresponding element of (19), under the null hypothesis. This follows from the facts that $\tilde{\boldsymbol{\varepsilon}}^\top\tilde{\boldsymbol{\varepsilon}} = \boldsymbol{\iota}^\top\boldsymbol{\iota} = n$ and that each element of $\tilde{\boldsymbol{\varepsilon}}^\top\tilde{\boldsymbol{\varepsilon}}$ is a random variable which is on average equal to one.

Since (19) is on average equal to one-half of (20), and the inner products of the regressands of (17) and (18) with the regressors are clearly identical, it is easily seen that one-half the explained sum of squares from (17) will be asymptotically equivalent to the explained sum of squares from (18). Hence the LM statistic based on the DLR (16) will be equivalent to the more familiar statistics based on the squared-residuals regression (17), or on asymptotically equivalent regressions in which one simply regresses the ordinary (non-normalized) squared residuals on $\boldsymbol{\iota}$ and $\boldsymbol{Z}$ and then calculates n times the ordinary centred R^2 (Koenker, 1981).

When one simply wants to test a hypothesis about $\boldsymbol{\beta}$, so that the Gauss-Newton regression (14) is applicable, or a hypothesis about α and $\boldsymbol{\gamma}$, so that the squared-residuals regression (17) is applicable, there is no reason to use a DLR. But notice that as soon as one has to deal with a hypothesis which imposes restrictions on $\boldsymbol{\beta}$ *jointly* with α and/or $\boldsymbol{\gamma}$, neither of the conventional procedures can be used any longer. The DLR, in contrast, remains both valid and reasonably easy to use.

For example, consider the following model, in which the variance depends on one of the parameters of the regression function:

$$y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + u_t, \quad u_t \sim \mathrm{N}\big(0, (\alpha + \beta_2 X_{t2})^2\big). \tag{21}$$

Under the null hypothesis that $\beta_2 = 0$, the model (21) is homoskedastic, but under the alternative it is heteroskedastic. A DLR to test this null hypothesis is

$$\begin{bmatrix} \tilde{\boldsymbol{\varepsilon}} \\ \boldsymbol{\iota} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\iota} & \boldsymbol{X}_1 & \tilde{\boldsymbol{\varepsilon}} & \boldsymbol{X}_2 + \tilde{\boldsymbol{\varepsilon}} * \boldsymbol{X}_2 \\ \boldsymbol{0} & \boldsymbol{0} & -\boldsymbol{\iota} & -\boldsymbol{X}_2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ a \\ b_2 \end{bmatrix} + \text{residuals}. \tag{22}$$

The last regressor here is the one that corresponds to β_2, the parameter which is being tested. By comparing (22) with (13) and (16), we see that it is simply the sum of two regressors, one which would test for $\beta_2 \neq 0$ in the regression function $\beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2}$ and one which would test for $\beta_2 \neq 0$ in the skedastic function $\alpha + \beta_2 X_{t2}$.

It may be noted that, for all of the cases examined in this section, the explained sum of squares from the DLR (which is the LM statistic we wish to compute) is equal to $2n$ minus the sum of squared residuals. This will be true whenever the null hypothesis is a regression model, so that the regressand of the DLR consists of n normalized residuals and n ones. This equality makes the LM test statistic particularly easy to calculate when using a regression package which does not print the explained sum of squares around zero.

4. Tests for Functional Form

The DLR form of the LM test becomes very useful when one wants to test the functional form of a model. Consider the class of models

$$\zeta(y_t, \lambda) = x_t(\boldsymbol{\beta}) + u_t, \quad u_t \sim \mathrm{N}(0, \sigma^2), \tag{23}$$

where $\boldsymbol{\beta}$ is an unknown parameter vector and λ is (for simplicity) a scalar. These models are clearly a special case of (1), since we can write

$$f_t(y_t, \boldsymbol{\theta}) = \big(\zeta(y_t, \lambda) - x_t(\boldsymbol{\beta})\big)/\sigma.$$

We have assumed for simplicity that $\boldsymbol{\beta}$ and λ are not functionally related, but this assumption will be relaxed below. Suppose the object is to test the hypothesis that λ takes on a particular value, say λ^*. This is a very common situation in practice. For example, if

$$\zeta(y_t, \lambda) = \frac{y_t^\lambda - 1}{\lambda}, \tag{24}$$

so that $\zeta(\cdot)$ is the familiar Box-Cox transformation, we often wish to test the hypothesis that $\lambda = 0$ (i.e. that the regressand is the logarithm of y_t) or the hypothesis that $\lambda = 1$ (i.e. that the regressand is simply y_t, provided the right-hand side effectively contains a constant term).

The DLR for testing the hypothesis that $\lambda = \lambda^*$ is

$$\begin{bmatrix} \tilde{\boldsymbol{\varepsilon}} \\ \boldsymbol{\iota} \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{X}} & \tilde{\boldsymbol{\varepsilon}} & -\zeta_\lambda(y_t, \lambda^*)/\tilde{\sigma} \\ \boldsymbol{0} & -\boldsymbol{\iota} & \zeta_{y\lambda}(y_t, \lambda^*)/\zeta_y(y_t, \lambda^*) \end{bmatrix} \tag{25}$$

where the scalars in the last column of the regressor matrix, which should be treated as typical elements of n–vectors, are

$$\zeta_y(y_t, \lambda) \equiv \frac{\partial \zeta(y_t, \lambda)}{\partial y_t}, \tag{26}$$

$$\zeta_\lambda(y_t, \lambda) \equiv \frac{\partial \zeta(y_t, \lambda)}{\partial \lambda}, \tag{27}$$

and

$$\zeta_{y\lambda}(y_t, \lambda) \equiv \frac{\partial^2 \zeta(y_t, \lambda)}{\partial y_t \partial \lambda}. \tag{28}$$

Regression (25) will usually be quite easy to implement, provided that the derivatives (26) through (28) are not too hard to evaluate.

It is often desired to test for functional form in situations where theory suggests that both linear and loglinear specifications are plausible. A well-known model that conveniently allows for both as special cases is the Box-Cox regression model

$$y_t(\lambda) = \sum_i \beta_i X_{ti}(\lambda) + \sum_j \gamma_j Z_{tj} + u_t, \tag{29}$$

where $y_t(\lambda)$ and $X_{it}(\lambda)$ denote the Box-Cox transformation (24) applied to the variables y_t and X_{ti}, respectively; the variables Z_{tj} are those, such as the constant term and dummy variables, which cannot sensibly be transformed. This model differs slightly

from (23) because λ appears on the right-hand side of the equals sign as well as the left, but this creates no difficulties.

Godfrey and Wickens (1981) first used the OPG regression to test linear and loglinear null hypotheses against the general alternative (29). Davidson and MacKinnon (1985) showed that it is just as easy, and far more reliable, to use the DLR. For both the null of $\lambda = 1$ and the null of $\lambda = 0$, the regressand of the DLR has t^{th} element $\tilde{\varepsilon}_t$ and $(n+t)^{\text{th}}$ element 1 (where $\tilde{\varepsilon}_t$ denotes the t^{th} normalized residual from the linear or loglinear model as appropriate). The t^{th} and $(n+t)^{\text{th}}$ elements of the regressors are then (first for the null of $\lambda = 1$, then for the null of $\lambda = 0$):

$$
\begin{aligned}
&\text{for } \beta_i: \quad X_{ti} - 1 \text{ and } 0; \quad \log(X_{ti}) \text{ and } 0 \\
&\text{for } \gamma_j: \quad Z_{ti} \text{ and } 0; \quad Z_{ti} \text{ and } 0 \\
&\text{for } \sigma: \quad \tilde{\varepsilon}_t \text{ and } -1; \quad \tilde{\varepsilon}_t \text{ and } -1 \\
&\text{for } \lambda: \quad -\Bigl(y_t \log(y_t) - y_t + 1 - \sum_i \bigl(X_{ti} \log(X_{ti}) - X_{ti} + 1\bigr)\Bigr) \text{ and } \tilde{\sigma} \log(y_t); \\
&\qquad\qquad -\Bigl(\tfrac{1}{2} \log(y_t)^2 - \sum_j \bigl(\tfrac{1}{2} \log(X_{ti})^2\bigr)\Bigr) \text{ and } \tilde{\sigma} \log(y_t).
\end{aligned}
$$

There is a good deal of Monte Carlo evidence on the properties of the LM test based on this DLR in comparison with alternative tests of linear and loglinear regressions; see Davidson and MacKinnon (1983, 1985) and Godfrey, McAleer, and McKenzie (1986). For the most part, the DLR test seems to perform very well even in samples of modest size, while the Godfrey-Wickens OPG test tends to over-reject quite severely. None of the studies finds any serious divergences between the actual and nominal size of the DLR test for samples of 50 or more, although Godfrey *et al.* do find some noticeable discrepancies for samples of size 20.

There is no reason to restrict attention to the Box-Cox transformation, since other transformations are available which may be more widely applicable or more appropriate for certain models. In particular, MacKinnon and Magee (1987) consider transformations of the dependent variable of the form $\rho(\mu y)/\mu$, where the function $\rho(\cdot)$ is assumed to be monotonically increasing in its argument and to possess the following properties:

$$\rho(0) = 0; \quad \rho'(0) = 1; \quad \rho''(0) \neq 0. \tag{30}$$

Many functions $\rho(\cdot)$ possess the properties (30), of course. One of the simplest is the function $y + y^2$, so that the transformation would be

$$\rho(\mu y)/\mu = y + \mu y^2 \tag{31}$$

Evidently, (31) will be a convex function of y when μ is positive and a concave function when μ is negative. This makes it clear that a test of $\mu = 0$ can be interpreted as testing against any form of local quadratic nonlinearity.

For this transformation family, the model (23) would become

$$\rho(\mu y)/\mu = x_t(\boldsymbol{\beta}) + u_t, \quad u_t \sim \mathrm{N}(0, \sigma^2). \tag{32}$$

Like the Box-Cox family, this transformation family includes the linear case in the limit as $\mu \to 0$, but unlike the Box-Cox it can be applied to variables of either sign and to zero variables. MacKinnon and Magee show that the DLR to test the null hypothesis that $\mu = 0$ is the following:

$$\begin{bmatrix} \tilde{\boldsymbol{\varepsilon}} \\ \boldsymbol{\iota} \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{X}} \\ \mathbf{0} \end{bmatrix} \boldsymbol{b} + \begin{bmatrix} \tilde{\boldsymbol{\varepsilon}} \\ -\boldsymbol{\iota} \end{bmatrix} a + \begin{bmatrix} -\frac{1}{2}\boldsymbol{y} * \boldsymbol{y}/\tilde{\sigma} \\ \boldsymbol{y} \end{bmatrix} m + \text{residuals}. \tag{33}$$

where $\boldsymbol{y} * \boldsymbol{y}$ is an n–vector with typical element y_t^2. The simplest test statistic for $\mu = 0$ is then simply $2n$ minus the sum of squared residuals from (33).

Regression (33) is clearly very similar to the DLR for testing a nonlinear regression model against a model with a Box-Cox transformation on the dependent variable alone; only the regressor which corresponds to the parameter being tested (μ in this case, λ in the Box-Cox case) is different. Notice that the explicit form of the function $\rho(\cdot)$ does not affect the test regression, because, like all LM tests, this one depends only on local properties. This test is particularly easy to do, and MacKinnon and Magee show that it is sensitive to several common forms of model misspecification, including nonlinearity in the regression function, heteroskedasticity, and skewness. The test is related to, but in many circumstances more powerful than, the well-known RESET test of Ramsey (1969) and Ramsey and Schmidt (1976).

Especially when dealing with cross-section data, it is often almost as important to transform the dependent variable correctly as it is to specify the appropriate set of regressors. By using double-length regressions, it is very easy to test the validity of any particular transformation. In particular, the test based on (33) is easy to do and may reveal serious problems with the specification of the model. Testing against a Box-Cox transformation can also be useful in cases where the dependent variable is always positive.

5. Additional Features of Double-Length Regressions

Double-length regressions can be useful in many situations where one is not interested in computing test statistics. One major application is calculating estimated covariance matrices. Suppose one obtains ML estimates of a model which is a special case of (1), but not a regression model, using a technique which does not provide a valid covariance matrix estimate. For example, one might estimate a model like (23), which involves a nonlinear transformation of the dependent variable, by searching over λ and at each step doing least squares conditional on λ. The covariance matrix printed by the least squares package will be conditional on the estimate $\hat{\lambda}$ and will therefore not be valid.

This is an ideal situation in which to use a double-length regression. By running the appropriate DLR, one can both verify that a point sufficiently close to the maximum of

the likelihood function has been obtained (since the explanatory power of the regression should be exactly zero at the maximum and very close to zero nearby), and generate the desired covariance matrix. The latter can be estimated by anything which is asymptotically equivalent to 1/n times the inverse of (4), and the simplest thing is to use the OLS covariance matrix estimate from the DLR, which will be

$$\left(\frac{2n}{2n-k-2}\right)\left(\boldsymbol{F}^{\top}(\hat{\boldsymbol{\theta}})\boldsymbol{F}(\hat{\boldsymbol{\theta}})+\boldsymbol{K}^{\top}(\hat{\boldsymbol{\theta}})\boldsymbol{K}(\hat{\boldsymbol{\theta}})\right)^{-1}.$$

The factor $2n/(2n-k-2)$ arises because the DLR will have $k+2$ regressors (k of which correspond to the parameters of the regression function, one of which corresponds to the parameter of the transformation of the dependent variable, and one of which corresponds to the variance of the error terms). Note that, in the case of models like (23), the appropriate DLR will not be (25), since we are no longer evaluating derivatives under the null hypothesis that $\lambda = 0$, and since it is no longer valid to transform the regressors in ways which do not affect the explanatory power of the DLR (because such transformations will affect the OLS covariance matrix estimate).

Double-length regressions can also be useful in at least two other contexts. Since the matrix $\left(\boldsymbol{F}^{\top}(\ddot{\boldsymbol{\theta}})\boldsymbol{F}(\ddot{\boldsymbol{\theta}})+\boldsymbol{K}^{\top}(\ddot{\boldsymbol{\theta}})\boldsymbol{K}(\ddot{\boldsymbol{\theta}})\right)$ approximates minus the Hessian matrix for $\ddot{\boldsymbol{\theta}}$ close to the true parameter vector, it can be used in place of the latter as part of a Newton-type hill-climbing procedure. At any arbitrary point on the hill, say $\ddot{\boldsymbol{\theta}}$, the algorithm can calculate the DLR

$$\begin{bmatrix}\ddot{\boldsymbol{f}}\\ \boldsymbol{\iota}\end{bmatrix}=\begin{bmatrix}-\ddot{\boldsymbol{F}}\\ \ddot{\boldsymbol{K}}\end{bmatrix}\boldsymbol{b}+\text{residuals}. \tag{34}$$

and the estimated coefficient vector $\ddot{\boldsymbol{\beta}}$ will approximate the direction of search which would be used by Newton's Method. Thus by combining (34) with a good line-search algorithm, one can easily create an algorithm for maximizing likelihood functions when the model can be written in the form (1).[5] Such an algorithm may be used in place of algorithms based on the OPG regression, such as the one proposed by Berndt, Hall, Hall, and Hausman (1974).

The result (4), which is the fundamental result on which the DLR is based, can itself be of use in the context of econometric theory. One may often want to find analytical expressions for the information matrix, and it is frequently much easier to start from (4) than from expressions for the Hessian or the outer product of the gradient, as the reader may check using as examples (21), (23), and (32).

Finally, a word of warning. If the regressand or any of the regressors in a DLR is constructed incorrectly, it is not only possible but in practice very likely that the regression will yield a computed test statistic which is large and entirely meaningless.[6]

[5] For a readable introduction to the art of numerical maximization, see Press *et al.* (1986), Chapter 10.

[6] This of course is equally true for the OPG regression.

It is a very good idea to check most of one's calculations by first running the regression *without* those regressors which correspond to the parameters being tested; this regression should have no explanatory power at all if everything has been constructed correctly. However, one cannot check the test regressors in this way, and an error in their construction can easily lead to nonsensical results. For example, think of what would happen if one inadvertently added a constant term to a DLR. Since the second half of the regressand is always a vector of ones, the constant is bound to have substantial ability to explain it, and so one would in all likelihood obtain a "significant" test statistic.

6. Conclusion

Double-length regressions are potentially a very useful part of the econometrician's kit of tools. Like the better-known Gauss-Newton and squared-residuals artificial regressions, to which they are related, they generally have good finite-sample properties, but they can be applied to far more situations. They have a wide variety of potential applications as part of maximization algorithms, as convenient ways to calculate covariance matrix estimates, and as procedures to compute LM, and equivalent, test statistics. We would argue that they should be used routinely to test regression equations for functional form, just as Gauss-Newton regressions are routinely used to test for serial correlation and squared-residuals regressions are routinely used to test for heteroskedasticity.

References

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

Berndt, E. R., Hall, B. H., Hall, R. E., and Hausman, J. A. (1974). "Estimation and inference in nonlinear structural models," *Annals of Economic and Social Measurement*, 3, 653–665.

Breusch, T. S. and Pagan, A. R. (1979). "A simple test for heteroskedasticity and random coefficient variation," *Econometrica*, 47, 1287–1294.

Chesher, A. (1983). "The information matrix test: Simplified calculation via a score test interpretation," *Economics Letters*, 13, 45–48.

Davidson, R. and MacKinnon, J. G. (1983). "Small sample properties of alternative forms of the Lagrange Multiplier test," *Economics Letters*, 12, 269–275.

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

Davidson, R. and MacKinnon, J. G. (1987). "Testing for consistency using artificial regressions," Queen's University Economics Department Working Paper No. 687.

Durbin, J. (1970). "Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables," *Econometrica*, 38, 410–421.

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

Engle, R. F. (1982b). "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1007.

Godfrey, L. G. (1978). "Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables," *Econometrica*, 46, 1293–1301.

Godfrey, L. G., McAleer, M., and McKenzie, C.R. (1986). "Variable addition and Lagrange Multiplier tests for linear and logarithmic regression models: Theory and Monte Carlo evidence," Australian National University Working Paper in Economics and Econometrics No. 136.

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

Koenker, R. (1981). "A note on Studentizing a test for heteroskedasticity," *Journal of Econometrics*, 17, 107–112.

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

Larson, A. C. (1987). "Using the Box-Cox transformation with pooled time series and cross-section data," paper presented at the Australasian Meetings of the Econometric Society, University of Canterbury, Christchurch, New Zealand.

MacKinnon, J. G. and Magee, L. (1987). "Transforming the dependent variable in regression models," McMaster University Department of Economics Discussion Paper No. 87-08.

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

Pagan, A. R. (1984). "Model evaluation by variable addition," in K. Wallis and D. F. Hendry, ed., *Quantitative Economics and Econometric Analysis*, Basil Blackwell, Oxford, 103–133.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1986). *Numerical Recipes*, Cambridge University Press, Cambridge.

Ramsey, J. B. (1969). "Tests for specification errors in classical linear least-squares regression analysis," *Journal of the Royal Statistical Society*, Series B, 31, 350–371.

Ramsey, J. B. and P. Schmidt (1976). "Some further results on the use of OLS and BLUS residuals in specification error tests," *Journal of the American Statistical Association*, 71, 389–390.

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