COMMUN. STATIST.—THEORY METH., 28(6), 1435–1446 (1999)

AN OBSERVATION ON REGRESSION-BASED SPECIFICATION TESTS

Mark Kamstra

Department of Economics
Simon Fraser University
Burnaby, British Columbia, Canada V5A 1S6

Key Words: specification; Lagrange Multiplier test; omission of regressors.

ABSTRACT

The use of regression-based specification tests, such as the nR^2 form of the Lagrange Multiplier test, has become quite widespread over the last 20 years. The popularization of the nR^2 form of the Lagrange Multiplier (LM) test, perhaps the most widely used class of regression-based tests, has come about in large part from the ease of its application to many tests of nonlinear restrictions and its asymptotic equivalence to Likelihood Ratio and Wald tests. Properly performed, these regression-based tests invariably include regressors which are orthogonal by construction to the dependent variable of the regression. The purpose of this paper is to motivate the inclusion of such variables by investigating implications for the test size and power if these regressors are erroneously omitted. It is straightforward to show that both the size and power of the test are adversely affected by omitting these regressors.

www.dekker.com

1. INTRODUCTION

The use of regression-based specification or diagnostic tests has become quite widespread over the last 20 years. The popularization of the nR^2 form of the Lagrange Multiplier (LM) test, perhaps the most widely used class of regression-based tests, has come about due to its ease of application to many tests of nonlinear restrictions and its asymptotic equivalence to Likelihood Ratio (LR) and Wald tests. The notation used here has n as the sample size and the R^2 comes from the auxiliary OLS regression of the errors in estimation of the restricted null model on an appropriate set of regressors, pre-supposing normality and additive errors. The equivalence of the LM, LR and Wald tests holds under fairly general conditions.

A device frequently used to derive the appropriate set of regressors for the auxiliary regression is to manipulate mechanically the equations of the first and second derivative matrices of the Gaussian likelihood until the nR^2 just falls out. This exposes the presence of a subset of regressors, call these X, which are uncorrelated *by construction* with the dependent variable of the auxiliary regression, the errors in estimation of the restricted null model. An interesting question is whether these regressors X are required for size considerations, power considerations or both.

Consider the example of the LM test for the inclusion of Z in the following model:

$$y = X\beta + Z\gamma + \epsilon \tag{1.1}$$

$$H_0 : \gamma = 0 \quad H_1 : \gamma \neq 0 \tag{1.2}$$

where $X, n \times k$, and $Z, n \times q$ are independent of $\epsilon \sim N(0, I_n)$, and β and γ are fixed parameters. It is straightforward to show that the LM test can be written as nR^2 from the auxiliary regression of $e = (y - X\hat{\beta})$ on X and Z, where $\hat{\beta} = (X'X)^{-1}X'y$. This follows provided standard regularity conditions obtain. See, for example, Breusch and Pagan (1980).

The need to include Z in the auxiliary regression is perhaps obvious, as correlation between Z and e supports the contention that Z belongs in model (1.1). The derivation of the LM statistic makes clear that X is also essential for the auxiliary regression, but the X are uncorrelated by construction with e, so that the need for X is perhaps less obvious. The presence of X in the auxiliary regression is not typically mentioned in journal articles, such as Breusch and Pagan (1980), or in texts, such as Chow (1983), Judge et. al. (1985) or Lehmann (1986). (Exceptions are MacKinnon (1992) and Davidson and MacKinnon (1993). In the context of Gauss-Newton regressions – that is, artificial regressions designed for specification tests among other uses – MacKinnon (1992) and Davidson and MacKinnon (1993) motivate the inclusion of regressors which are uncorrelated by construction with the dependent variable).

The contribution of this paper is to provide a pedagogically convenient, clear and formal presentation in the style of White's (1984) *Asymptotic Theory for Econometricians* to show that regressors orthogonal to the dependent variable of the auxiliary regression are required for both power and size considerations. The presentation does not require a Maximum Likelihood framework, and is thus applicable to a wide class of regression-based tests. Although the intuition for this result is simple, I believe the formal analysis is quite helpful for driving home both this result and in general the need for care when forming regression-based specification tests. Examples of such specification tests include LM nR^2 tests as well as specification tests based on Gauss-Newton regressions as MacKinnon (1992) presents (which actually include the LM nR^2 tests as a special case).

In Section 2, I motivate the inclusion of X regressors uncorrelated with the dependent variable of a regression-based test, and I provide a demonstration that the inclusion of X is necessary following from power and size considerations. In Section 3, I discuss situations in which it is common to fail to include the necessary regressors. Section 4 concludes.

2. REGRESSION-BASED SPECIFICATION TESTS

Consider again the LM test of model (1.1) above. Notice that Z may be decomposed into orthogonal terms $X\alpha$ and Z_*, with suitable choice of α, without imposing any restriction on Z.

$$Z_* = Z - X\alpha \ \ and \ \ E[X'Z_*] = 0.$$

Special cases include the degenerate ones with either Z_* or α identically zero. It is straightforward to show that the R^2 from a regression of $e = (y - X\hat{\beta})$ on X and Z, where $\hat{\beta} = (X'X)^{-1}X'y$, is equivalent to the R^2 from the regression of e on Z_*. This is not true in general, but follows here because e is orthogonal to X by construction. The proper form of the regression-based LM test controls for correlation between Z and X with the inclusion of X in the auxiliary regression, and this is the only purpose of X in the auxiliary regression. Failing to control for X in the auxiliary regression lowers both power and size. Ruling out degenerate cases, the R^2 is *strictly* downward biased with the omission of X in the auxiliary regression, hence the inclusion of X is necessary for proper power and size. A demonstration of this follows.

2.1 THE MODEL AND ASSUMPTIONS

Suppose

A1) $y_t = X_t\beta + Z_t\gamma + \epsilon_t, \;\; Z_t = Z_{t,*} + X_t\alpha, \;\; t = 1, ..., n;$

A2) $n^{-1}X'\epsilon \xrightarrow{p} 0, \;\; n^{-1}Z_*'\epsilon \xrightarrow{p} 0, \;\; n^{-1}X'Z_* \xrightarrow{p} 0;$

A3) $n^{-1}Z_*'Z_* \xrightarrow{p} M_z$, and $n^{-1}X'X \xrightarrow{p} M_x$, where M_z and M_x are finite and positive definite. These restrictive assumptions are made for notational convenience and expositional clarity. It should be noted that these assumptions could be considerably weakened with no substantive change to the results presented here.

Immediate implications of these assumptions include the following:

$$n^{-1}Z'Z \xrightarrow{p} M_z + \alpha' M_x \alpha, \quad Z'Z_* \xrightarrow{p} M_z.$$

The power of the LM test of the hypothesis (1.2) is investigated in Section 2.2, then the size is investigated in Section 2.3. In each case the implications of (erroneously) omitting X from the auxiliary regression of the LM test is determined by comparing the test statistic resulting from an auxiliary regression *excluding* X to the test statistic resulting from an auxiliary regression *including* X. The nR^2 LM test statistic can be derived under the restriction $\epsilon \sim N(0, \sigma^2 I_n)$, where $\sigma^2 > 0$, finite. Regression-based test statistics can often be derived under much weaker distributional assumptions however, as Koenker (1981) demonstrates. The results to be presented for the power and size of the LM test statistic do not require the assumption of normality.

2.2 THE TEST UNDER H_1: POWER CONSIDERATIONS

Notice that the true data generating process may be rewritten as

$$y = X(\beta + \alpha\gamma) + Z_*\gamma + \epsilon.$$

Defining $\hat{\beta} = (X'X)^{-1}X'y$, and given A1-A3, we have $\hat{\beta} - (\beta + \alpha\gamma) = o_p(1)$. Define

$$\begin{aligned} e &= y - X\hat{\beta} \\ &= Z_*\gamma + \epsilon + X\left(\beta + \alpha\gamma - \hat{\beta}\right) \\ &= Z_*\gamma + \epsilon + o_p(1). \end{aligned}$$

Consider first the regression of e on Z alone. Call this regression (A). Define $\hat{\gamma}_A = (Z'Z)^{-1}Z'e$. The R^2 for regression (A) is

$$\begin{aligned} R_A^2 &= \hat{\gamma}_A'Z'Z\hat{\gamma}_A/e'e \\ &= e'Z(Z'Z)^{-1}Z'e/e'e. \end{aligned}$$

Notice that

$$\begin{aligned} &e'Z(Z'Z)^{-1}Z'e/e'e \\ &= \left(Z_*\gamma + \epsilon + X(\beta + \alpha\gamma - \hat{\beta})\right)' Z(Z'Z)^{-1}Z' \\ &\quad \times \left(Z_*\gamma + \epsilon + X(\beta + \alpha\gamma - \hat{\beta})\right)/e'e \\ &= n^{-1}\left(Z_*\gamma + \epsilon + X(\beta + \alpha\gamma - \hat{\beta})\right)' Z(n^{-1}Z'Z)^{-1}Z' \\ &\quad \times n^{-1}\left(Z_*\gamma + \epsilon + X(\beta + \alpha\gamma - \hat{\beta})\right)/(n^{-1}e'e) \\ &= \gamma'(n^{-1}Z_*'Z)(n^{-1}Z'Z)^{-1}n^{-1}Z'Z_*\gamma/(n^{-1}e'e) + o_p(1) \end{aligned}$$

as $n^{-1}(\beta + \alpha\gamma - \hat{\beta})'X'Z$ and $n^{-1}\epsilon'Z$ are $o_p(1)$. Further,

$$\gamma'(n^{-1}Z_*'Z)(n^{-1}Z'Z)^{-1}(n^{-1}Z'Z_*)\gamma/(n^{-1}e'e) \xrightarrow{p} \frac{\gamma'M_z(M_z + \alpha'M_x\alpha)^{-1}M_z\gamma}{\sigma^2 + \gamma'M_z\gamma}$$

so that we have

$$R_A^2 \xrightarrow{p} \frac{\gamma'M_z(M_z + \alpha'M_x\alpha)^{-1}M_z\gamma}{\sigma^2 + \gamma'M_z\gamma}. \tag{2.1}$$

Now consider the regression of e on both Z and X. First residualize Z with respect to X, and define this residual to be $\hat{Z}_*$.

$$\begin{aligned}\hat{Z}_* &= Z - X(X'X)^{-1}X'Z\\ &= Z_* + X\alpha - X(X'X)^{-1}X'(Z_* + X\alpha)\\ &= Z_* - X(X'X)^{-1}X'Z_*\\ &= Z_* - X(n^{-1}X'X)^{-1}(n^{-1}X'Z_*)\\ &= Z_* + o_p(1)\end{aligned}$$

Note that e is orthogonal to X by construction, so that e regressed on Z and X yields numerically identical residuals to e regressed on $\hat{Z}_*$ alone. Davidson and MacKinnon (1993) refer to this result as the Frisch-Waugh-Lovell theorem and a proof of this result can be found there.

Call the regression of e on $\hat{Z}_*$ regression (B). Define $\hat{\gamma}_B = (\hat{Z}_*'\hat{Z}_*)^{-1}\hat{Z}_*'e$. The R^2 for regression (B) is

$$\begin{aligned}R_B^2 &= \hat{\gamma}_B'\hat{Z}_*'\hat{Z}_*\hat{\gamma}_B/e'e\\ &= e'Z_*(Z_*'Z_*)^{-1}Z_*'e/e'e + o_p(1)\end{aligned}$$

as $\hat{Z}_* = Z_* + o_p(1)$. Notice that

$$\begin{aligned}&e'Z_*(Z_*'Z_*)^{-1}Z_*'e/e'e\\ &= \left(Z_*\gamma + \epsilon + X(\beta + \alpha\gamma - \hat{\beta})\right)' Z_*(Z_*'Z_*)^{-1}Z_*'\\ &\quad \times \left(Z_*\gamma + \epsilon + X(\beta + \alpha\gamma - \hat{\beta})\right)/e'e\\ &= n^{-1}\left(Z_*\gamma + \epsilon + X(\beta + \alpha\gamma - \hat{\beta})\right)' Z_*(n^{-1}Z_*'Z_*)^{-1}Z_*'\\ &\quad \times n^{-1}\left(Z_*\gamma + \epsilon + X(\beta + \alpha\gamma - \hat{\beta})\right)/(n^{-1}e'e)\\ &= \gamma'(n^{-1}Z_*'Z_*)\gamma/(n^{-1}e'e) + o_p(1)\end{aligned}$$

Since $n^{-1}(\beta + \alpha\gamma - \hat{\beta})'X'Z_*$ and $n^{-1}\epsilon'Z_*$ are $o_p(1)$. Further, as

$$\gamma'(n^{-1}Z_*'Z_*)\gamma/(n^{-1}e'e) \xrightarrow{p} \frac{\gamma'M_z\gamma}{\sigma^2+\gamma'M_z\gamma},$$

we have

$$R_B^2 \xrightarrow{p} \frac{\gamma'M_z\gamma}{\sigma^2+\gamma'M_z\gamma}. \tag{2.2}$$

Now compare R_A^2, equation (2.1), calculated excluding X in the auxiliary regression, to R_B^2, equation (2.2), calculated including X in the auxiliary regression, by contrasting $M_z(M_z+\alpha'M_x\alpha)^{-1}M_z$ and M_z. As $\alpha'M_x\alpha$ is positive definite, $M_z(M_z+\alpha'M_x\alpha)^{-1}M_z < M_z$ so that $R_B^2 - R_A^2 = O_p(1) > 0$. Hence the test statistic nR_A^2 must under-reject when the null is false, so that erroneously omitting X from the auxiliary regression lowers the power of the test. To yield the result with a strict inequality, rule out the degenerate cases $\alpha = 0$ and $Z_* = 0$.

2.3 THE TEST UNDER H_0: SIZE CONSIDERATIONS

Now $\gamma = 0$ so that $y = X\beta + \epsilon$. Define $e = y - X\hat{\beta} = \epsilon + X(\beta - \hat{\beta})$. Consider first the regression of e on Z alone. Call this regression (C). Define

$$\begin{aligned}
\hat{\gamma}_C &= (Z'Z)^{-1}Z'e \\
&= (Z'Z)^{-1}Z'\left(\epsilon + X(\beta-\hat{\beta})\right) \\
&= (Z'Z)^{-1}Z'\left(I - X(X'X)^{-1}X'\right)\epsilon \\
&= (Z'Z)^{-1}\left(Z' - Z'X(X'X)^{-1}X'\right)\epsilon \\
&= (Z'Z)^{-1}\left(Z_*' + \alpha'X' - (Z_* + X\alpha)'X(X'X)^{-1}X'\right)\epsilon \\
&= (Z'Z)^{-1}\left(Z_*' - Z_*'X(X'X)^{-1}X')\right)\epsilon \\
&= (Z'Z)^{-1}Z_*'\left(I - X(X'X)^{-1}X'\right)\epsilon \\
&= (Z'Z)^{-1}Z_*'\left(\epsilon + X(\beta-\hat{\beta})\right) \\
&= (Z'Z)^{-1}Z_*'e.
\end{aligned}$$

The R^2 for regression (C) is

$$\begin{aligned} R_C^2 &= \hat{\gamma}_C' Z' Z \hat{\gamma}_C / e'e \\ &= e' Z_* (Z'Z)^{-1} (Z'Z) (Z'Z)^{-1} Z_*' e / e'e \qquad (2.3) \\ &= e' Z_* (Z'Z)^{-1} Z_*' e / e'e. \end{aligned}$$

Now consider the regression of e on both Z and X. First residualize Z with respect to X, and define this residual to be $\hat{Z}_*$, as above. Recall that e is orthogonal to X by construction, so that e regressed on Z and X is equivalent to e regressed on $\hat{Z}_*$ alone. Call the regression of e on $\hat{Z}_*$ regression (D). Define $\hat{\gamma}_D = (\hat{Z}_*'\hat{Z}_*)^{-1}\hat{Z}_*'e$. The R^2 for regression (D) is

$$\begin{aligned} R_D^2 &= \hat{\gamma}_D' \hat{Z}_*' \hat{Z}_* \hat{\gamma}_D / e'e \\ &= e' \hat{Z}_* (\hat{Z}_*' \hat{Z}_*)^{-1} \hat{Z}_*' e / e'e. \end{aligned}$$

Note that $\hat{Z}_*'e = Z_*'(I - X(X'X)^{-1}X')e$ and $X(X'X)^{-1}X'e = 0$ by construction, so that $\hat{Z}_*'e = Z_*'e$. Hence

$$R_D^2 = e' Z_* (\hat{Z}_*' \hat{Z}_*)^{-1} Z_*' e / e'e. \qquad (2.4)$$

Now compare R_C^2, equation (2.3), calculated excluding X in the auxiliary regression, to R_D^2, equation (2.4), calculated including X in the auxiliary regression, by contrasting $(Z'Z)^{-1}$ and $(\hat{Z}_*'\hat{Z}_*)^{-1}$. Recall that $n^{-1}Z_*'Z_* \xrightarrow{p} M_Z$ *and* $\hat{Z}_* = Z_* + o_p(1)$ so that $n^{-1}\hat{Z}_*'\hat{Z}_* \xrightarrow{p} M_Z$. Also recall that $n^{-1}(Z'Z) \xrightarrow{p} M_Z + \alpha' M_X \alpha$ and M_X is $O_p(1)$. Hence $n^{-1}Z'Z - n^{-1}\hat{Z}_*'\hat{Z}_* \xrightarrow{p} O_p(1)$ and $nR_D^2 > nR_C^2$. As $nR_D^2 \sim \chi_q^2$, nR_C^2 must under-reject when the null is true, so that erroneously omitting X from the auxiliary regression biases the size of the test. The χ_q^2 distribution of R_D^2 follows under standard regularity conditions and can be derived in the manner of Koenker (1981). To yield the result with a strict inequality, rule out the degenerate cases $\alpha = 0$ and $Z_* = 0$.

The preceding calculations reveal that the exclusion of X from the set of regressors in the auxiliary regression will downward bias the R^2 of the procedure under both the null and alternative hypotheses if and only if $\alpha \neq 0$ and $Z_* \neq 0$, that is, if Z is not orthogonal to X and Z is also not a perfect linear function of X. The inclusion of X in the auxiliary regression is required to control for linear correlation between Z and X, and provides the correct size and power of the LM test. It is appropriate to include X in the regression because e *is*, and Z *may not be*, orthogonal to X. Omission of X in the auxiliary regression systematically biases the test statistic downwards under both the null and alternative.

3. APPLICATIONS

An interesting question is, when might we encounter such a problem? That is, when do we need to worry about the inclusion of auxiliary regressors in conducting regression-based specification tests? Working with economic data, in particular macro economic data, presents us with a great deal of instances where manipulated data sets are the only ones to work with. Price indices, labour force data series, growth rates of sectors and the entire economy itself, are all routinely deseasonalized, and pre-whitened. Many data series are cleaned by the removal of outliers, and data is detrended, and all these manipulations can inadvertently lead to exactly the sort of problem outlined here. This sort of data manipulation is not the exception in empirical work. In fact, in using macroeconomic data, such adjustments are the rule. See, for instance, Harvey (1997) and Nelson and Kang (1981), and the *Journal of Econometrics* Annals 1993 issue on seasonality and econometric models.

If all the data have been similarly orthogonalized, the problems outlined in this paper are not applicable. But if this is not the case - if we have a deseasonalized dependent variable and we are testing for inclusion of a raw explanatory variable series for instance - the concerns raised here are valid, and this seasonality must be controlled for in the auxiliary regression.

4. SUMMARY

The nR^2 form of the LM test is often suitable for standard hypotheses, as well as for hypotheses which are difficult to test with LR or Wald tests, such as tests of nonlinear restrictions with a linear null model. Together with the asymptotic equivalence of the LM, LR and Wald tests under fairly general conditions, this provides powerful motivation for the usefulness of the LM test. There are also a large collection of similar regression-based specification and diagnostic tests. These tests, as well as the LM test, often include regressors which are orthogonal by construction to the dependent variable of the regression test. A regression of the dependent variable on these regressors alone would have an R^2 of 0. The variables used in the auxiliary regression of these tests are not well motivated by a simple mechanical derivation of the test form. In this paper, the regressors which are orthogonal by construction to the dependent variable of the auxiliary regression are demonstrated to be necessary for both power *and* size considerations.

It must be emphasized that the sort of situation which leads to this problem is not the exception in applied work, but the rule. When one manipulates data prior to empirical analysis, one must be careful to conduct regression-based specification tests with proper attention to the inclusion of auxiliary regressors. Failure to do so has unambiguous impact on size and power.

ACKNOWLEDGMENTS

The author is grateful for the comments and suggestions of R. Glen Donaldson, Peter Kennedy, Lisa Kramer, and James MacKinnon, and the financial aid of the Social Sciences and Humanities Research Council of Canada. Any remaining errors are the author's responsibility.

BIBLIOGRAPHY

Breusch, T.S., and A.E. Pagan (1980). "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics." *Review of Economic Studies*, 97, 239-253.

Chow, G. (1983). *Econometrics*, New York: McGraw-Hill.

Davidson, R., and J.G. MacKinnon (1993). *Estimation and Inference in Econometrics*, New York: Oxford University Press.

Harvey, A. (1997). "Trends, Cycles and Autoregressions." *The Economic Journal*, 107, 192-201.

Judge, G., W.E. Griffiths, R.C. Hill H. Lutkepohl, and T. Lee (1985). *The Theory and Practice of Econometrics*, New York: J.Wiley.

Koenker, R. (1981). "A Note on Studentizing a Test for Heteroskedasticity." *Journal of Econometrics*, 17, 107-112.

Lehmann, E.L. (1986). *Testing Statistical Hypotheses*, New York: J. Wiley.

MacKinnon, J.G. (1992). "Model Specification Tests and Artificial Regressions." *Journal of Economic Literature*, 30, 102-146.

Nelson, C.R. and C.I. Plosser (1981). "Spurious Periodicity in Inappropriately detrended Time Series." *Econometrica*, 49, 741-751.

White, H. (1984). *Asymptotic Theory for Econometricians*, Florida: Academic Press.

Received May, 1998; Revised December, 1998.