On the formulation of Wald tests on long-run parameters

Boswijk, H.P.

Publication date
1993

Published in
Oxford Bulletin of Economics and Statistics

OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 55, 1 (1993)
0305-9049 $3.00

PRACTITIONERS CORNER

On the Formulation of Wald Tests on Long-Run Parameters

Peter Boswijk

I. INTRODUCTION*

The single-equation error correction model has become one of the most important tools in the econometric analysis of time series. It provides a clear distinction between short-run dynamic adjustments and long-run equilibrium relations. The strong connection between *cointegration* and error correction models that follows from the Granger representation theorem see Engle and Granger (1987) has only increased its use. Because economic theory usually concerns long-run properties, hypotheses of interest often lead to restrictions on the long-run parameters. In this note, we shall analyse two Wald test statistics of linear restrictions on these parameters. The first statistic utilizes an estimated covariance matrix of the long-run parameter estimators, obtained from a first-order Taylor series expansion. The second statistic uses a reformulation of the null hypothesis, leading to a linear restriction on a model that is linear in the parameters. It is argued that the latter statistic has to be favoured because of its invariance properties, and because it is not affected by the lack of moments of the long-run parameter estimators. The use of asymptotic standard errors for constructing confidence intervals will also be addressed. An application to the UK consumption function illustrates the issues.

II. THE MODEL AND THE TEST STATISTICS

Consider the single-equation error corection model:

$$\Delta y_t = \lambda(y_{t-1} - \theta' x_{t-1}) + \sum_{j=1}^{p-1} \gamma_j \Delta y_{t-j} + \sum_{j=0}^{p-1} \beta_j' \Delta x_{t-j} + \varepsilon_t, \qquad t=1,\ldots,T \quad (1)$$

*This paper was completed while the author was visiting Nuffield College, University of Oxford. Helpful comments and suggestions from Anindya Banerjee, David Hendry, Jan Kiviet and Michael Veall are gratefully acknowledged.

where y_t is the dependent variable, z_t is a k-vector of explanatory variables, and ε_t is a white noise disturbance with variance σ^2. The lag length p is chosen the same for all variables for notational ease, but none of the results in this paper will depend on this, nor on the absence of a constant term in (1). We shall be concerned with inference on θ, the $k\times 1$ vector of long-run parameters. If the explanatory variables in z_t are not Granger-caused by y_t, these are long-run multipliers. If the components of z_t are integrated of order 1 and the model is stable, then the vector $(1, -\theta')'$ is a *cointegrating vector*, see Engle and Granger (1987), because $(y_t-\theta' z_t)$ is stationary although y_t and z_t are not. The model is non-linear in the parameters, but an obvious linear reformulation is

$$\Delta y_t = \pi_1 y_{t-1} + \pi_2' z_{t-1} + \sum_{j=1}^{p-1} \gamma_j \Delta y_{t-j} + \sum_{j=0}^{p-1} \beta_j' \Delta z_{t-j} + \varepsilon_t \tag{2}$$

where $\pi_1 = \lambda$ and $\pi_2 = -\lambda\theta$ so that $\theta = -\pi_2/\pi_1$..

The hypothesis of interest states that the long-run parameters obey the linear restrictions

$$H_0 : R\theta = r$$

where R and r are of order $h\times k$ and $h\times 1$, respectively $(h\leq k)$. In the parametrization of (2), the null hypothesis $(r-R\theta)=0$ corresponds to

$$H_0' : r\pi_1 + R\pi_2 = Q\pi = 0,$$

where $Q=[r \vdots R]$ and $\pi=(\pi_1, \pi_2')'$.

Let $\hat{\pi}$ denote the ordinary least-squares (OLS) estimator of π in (2). The indirect least-squares (ILS) estimator of θ is defined by

$$\hat{\theta} = -\frac{1}{\hat{\pi}_1}\cdot\hat{\pi}_2. \tag{3}$$

If the disturbances are normally distributed, and z_t is weakly exogenous for θ, then $\hat{\theta}$ is also the maximum likelihood estimator (conditional upon the starting values). A consistent estimator of the covariance matrix[1] of $\hat{\theta}$ is (see also Bårdsen, 1989 and Banerjee *et al.*, 1990).

$$\hat{V}[\hat{\theta}] = \hat{J}\hat{V}[\hat{\pi}]\hat{J}', \tag{4}$$

where

$$J = \frac{\partial\theta}{\partial\pi'} = \left[\frac{1}{(\pi_1)^2}\pi_2 \; \vdots \; -\frac{1}{\pi_1} Ik\right] = -\frac{1}{\pi_1}[\theta \vdots I_k] \tag{5}$$

[1] Bewley (1979) shows that θ can be estimated directly by instrumental variables estimation in a reparametrization of (1), the so-called *pseudo-structural form*, and Wickens and Breusch (1988) have shown that the same holds for the covariance matrix estimator.

is the Jacobian matrix or derivative of θ with respect to π, and $\hat{J}$ is obtained from (5) with π_1 and π_2 (and hence θ) replaced by their OLS estimators, and where $\hat{V}[\hat{\pi}]$ is the OLS covariance matrix estimator of $\hat{\pi}$. The Wald test statistic for H_0 is given by

$$W_1 = (R\hat{\theta} - r)'[R\hat{J}\hat{V}[\hat{\pi}]\hat{J}'[^{-1}(R\hat{\theta} - r).. \tag{6}$$

Alternatively, the Wald statistic for H_0' in (2) is

$$W_2 = (Q\hat{\pi})'[Q\hat{V}[\hat{\pi}]Q']^{-1}(Q\hat{\pi}). \tag{7}$$

Note that W_2/h is equal to the F-statistic for H_0' in (2).

If y_t and z_i are stationary and the model (1) is stable, then both statistics are asymptotically equivalent and distributed as $\chi^2(h)$ under the null hypothesis. Moreover, the same asymptotic null distribution applies if y_t and z_t are cointegrated, provided that z_t is weakly exogenous for θ, see e.g. Johansen (1992). Below it is assuymed that either of these conditions is satisfied.

Although the test statistics are asymptotically equivalent, they may differ substantially in finite samples, so that the problem arises of choosing between the alternative formulations. This will be considered next.

III. THE CHOICE OF A TEST STATISTIC

In order to facilitate comparison of the two statistics, we derived some alternative expressions for W_1 and W_2. Observe that

$$R\hat{\theta} - r = \frac{1}{\hat{\pi}_1}(-R\hat{\pi}_2 - r\hat{\pi}_1) = -\frac{1}{\hat{\pi}_1}Q\hat{\pi}. \tag{8}$$

or $Q\hat{\pi} = -\hat{\pi}_1(R\hat{\theta} - r)$. Now W_2 may be expressed similarly to W_1 in (6) by dividing $Q\hat{\pi}$ and its covariance matrix by $-\hat{\pi}_1$ and $(\hat{\pi}_1)^2$, respectively. Define θ_0, to be any k-vector that satisfies $R\theta_0 = r$. Then the implicit covariance matrix of $(R\hat{\theta} - r)$ used for W_2 is

$$\begin{aligned}\frac{1}{(\hat{\pi}_1)^2}Q\hat{V}[\hat{\pi}]Q' &= \frac{-1}{\hat{\pi}_1}[R\theta_0 \vdots R]\hat{V}[\hat{\pi}][R\theta_0 \vdots R]'\frac{-1}{\hat{\pi}_1}\Big) \\ &= R\hat{J}_0\hat{V}[\hat{\pi}]\hat{J}_0'R',\end{aligned} \tag{9}$$

where the notation

$$J_0 = -\frac{1}{\hat{\pi}_1}[\theta_0 \vdots I_k] \tag{10}$$

is inspired by its similarity to (5). Thus we have

$$W_2 = (R\hat{\theta} - r)'[R\hat{J}_0\hat{V}[\hat{\pi}]\hat{J}_0'R']^{-1}(R\hat{\theta} - r). \tag{11}$$

Note that θ_0 and hence $\hat{J}_0$ are not unique, because unless $h = k$, there are infinitely many vectors θ_0 that satisfy $R\theta_0 = r$. However, by definition $R\theta_0$ and hence $R\hat{J}_0$ are uniquely defined, which is required in (9) and (11). The reason for introducing θ_0 is to show that the two statistics differ in the way the covariance matrix of θ is estimated. Where W_2 uses the null hypothesis (in the form of θ_0, W_1 does not utilize this information but replaces θ by θ. The same difference arises when W_1 is expressed similarly to W_2. Using again $Q\hat{\pi} = -\hat{\pi}(R\hat{\theta} - r)$, we have

$$W_1 = (Q\hat{\pi})'[\hat{Q}\hat{V}[\hat{\pi}]\hat{Q}']^{-1}(Q\hat{\pi}). \tag{12}$$

where we have defined

$$\hat{Q} = -\hat{\pi}_1 R\hat{J} = [R\hat{\theta} \vdots R]. \tag{13}$$

Because under the null hypothesis the available information is used more efficiently for the construction of W_2, we may expect this statistic to display better size properties than W_1.

A major disadvantage of the Wald testing principle is its lack of invariance with respect to reformulations of the null hypothesis. The difference between W_1 and W_2 is an illustration of this property, which is analysed by Gregory and Veall (1985, 1986). The likelihood ratio (LR) test does not suffer from this problem and is therefore to be preferred over the Wald test. However, it is well-known that in a linear regression model with normally distributed errors, the F- or Wald test statistic for a set of linear restrictions is (a monotonic transformation of) the LR test statistic. Therefore, under the assumption that $\{\varepsilon_t\} \sim \mathrm{IN}(0, \sigma^2)$ in (1), the statistic W_2 corresponds to the LR test of H_0. Note that W^1 does not lead to an LR test, because it is not a monotonic function of W_2 (or the LR statistic).

An alternative approach to the invariance question is provided by Critchley *et al.* (1989). In their analysis, the essential problem with a Wald statistic for a non-linear restriction $g(\theta) = 0$ (on a general parameter vector θ) is that the asymptotic covariance matrix of $g(\hat{\theta})$ varies with θ. The dependency of $V[\hat{\theta}]$ on.θ via the Jacobian matrix J in (4) and)5) is an example of this. However, in this case the restricting function $g(\hat{\theta}) = (R\hat{\theta} - r)$ can be reformulated into $Q\hat{\pi}$, the covariance matrix of which does not depend upon any unknown parameters (except for σ^2). In this so-called constant metric case, the Wald test coincides with the Geodesic test that Critchley *et al.* propose as a solution to the invariance problem. This gives a second argument for the use of W_2 statistic. Our recommendation corresponds to that of Gregory and Veall (1987, p. 66) in the context of the partial adjustment model.

A final point in favour of the W_2 statistic is the following. Because the ILS estimator $\hat{\theta}$ is defined as a ratio of OLS estimators, it can be shown to have no finite moments. The factor that $\hat{\pi}_1$ will have a positive density at the origin causes the occurrence of large outliers in the distribution of $\hat{\theta}$. As can be seen from (12), the statistic W_1 depends on the estimator $\hat{\theta}$ and thus its distribution will be affected by these outliers, whereas this is not the case for W_2.

To see whether these theoretical arguments are reflected in a better finite sample behaviour of the W^2 statistic, we perform a small Monte Carlo experiment. The generating mechanism is

$$\Delta y_t = \alpha + \beta_0 \Delta z_t + \lambda(y_{t-1} - \theta z_{t-1}) + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{IN}(0, \sigma^2), \qquad t = 1, \ldots, T \quad (14)$$

where $T = 50$, $\alpha = 0$, $\beta_0 = 0.5$, $\sigma = 0.5$, $\lambda \in \{-0.1, -0.2, -0.5\}$ and $\theta \in \{0.5, 0.8, 1, 1.2, 1.5\}$. The explanatory variable z_t is fixed at one realization of a Gaussian random walk with innovation variance 1. For each of the 10,000 replications, we estimate (14) by OLS and wse compute the Wald statistics W_1 and W_2 for the hypothesis $\theta = 1$.

In the first row of Table 1 the rejection frequencies under the null ($\theta = 1$) at the 5 percent critical value of the $F(1, 50/4)$ distribution are given. In all cases the actual size of the tests is larger than the nominal 5 percent level, but the difference becomes smaller as the error correction coefficient gets larger in absolute value. The anticipated better size performance of the W_2 statistic is apparent, although at $\lambda = -0.1$ both tests are quite unsatisfactory in this respect. In the bottom row of Table 1 the exact (i.e. simulated) 5 percent critical values are given. The rejection frequencies at these critical values provide estimates of the power of the tests, which is given in Table 1 for $\theta \in \{0.5, 0.8, 1.2, 1.5\}$. No test statistic uniformly dominates the other in power performance: for values of θ smaller than 1, W_1 has higher power, whereas the opposite holds for $\theta > 1$. In the latter case, the differences in power can be quite substantial. In almost all cases the power increases with $-\lambda$; a similar effect can be expected from increasing the sample size or decreasing the error variance. In summary, this experiment corroborates the theoretical arguments concerning the behaviour of the statistics under the

TABLE 1
Size and Power of the Wald Tests

	$\lambda = -0.1$		$\lambda = -0.2$		$\lambda = -0.5$	
	W_1	W_2	W_1	W_2	W_1	W_2
Size						
$\theta = 1$	0.259	0.124	0.150	0.087	0.071	0.060
Power						
$\theta = 1.2$	0.016	0.039	0.010	0.151	0.772	0.847
$\theta = 1.5$	0.001	0.127	0.191	0.770	1.000	1.000
$\theta = 0.8$	0.109	0.097	0.250	0.173	0.805	0.709
$\theta = 0.5$	0.237	0.230	0.723	0.40	1.000	0.995
5% critical value	23.119	6.520	10.390	5.337	5.077	4.384

Note: The size is estimated by the rejection frequencies at the 5% critical value of the $F(1,46)$ distribution (4.05). The power is estimated by the rejection frequencies at the exact 5% critical values.

null hypothesis, but provides no unamiguous conclusions about the relative power performance of the tests.

IV. THE USE OF STANDARD ERRORS

The covariance estimator given in (4) provides the asymptotic standard errors of the long-run parameters. The calculation of these standard errors is discussed by Bårdsen (1989) and Banerjee *et al.* (1990), and is implemented in Hendry's (1989) *PC-GIVE*. Following the arguments in the previous section, these standard errors should be used with caution. On the one hand, the lack of moments of the estimators suggests that they may provide a poor measure of dispersion. On the other hand, the fact that the second version of the Wald test (W_2) has to be preferred over W_1 indicates that their role in constructing t-ratios is also limited. An alternative would be to compute standard errors from the covariance matrix corresponding to W_2. However, as can be seen from (9) these depend upon the null hypothesis and therefore cannot provide a measure of dispersion in general.

A possible solution to this problem is to construct confidence intervals corresponding to W_2. Define $W_2(\theta_{0j})$ to be the W_2 statistic for the hypothesis that the jth component of θ has a particular value θ_{0j}. Then a $100(1-\alpha)\%$ confidence interval C for θ_j is the set of hypothesized values that are not rejected by W_2 at a given significance level α:

$$C=\{\theta_{0j}\in\mathbf{R}: W_2(\theta_{0j})<\chi^2_{\alpha}(1)\}, \tag{15}$$

where $\chi^2_{\alpha}(1)$ is the $100)1-\alpha)\%$ quantile of the $\chi^2(1)$ distribution. Given the dependence of W_2 on θ_{0j}, the calculation of the boundaries of these intervals will amount to solving non-linear equations, and hence will require some numerical routine.

V. AN APPLICATION[2]

In this section we shall apply the tests to a consumption function for the UK, using annual data (1948–83) on the log of consumption c_t. the log of income y_t and the inflation rate Δp_t. Following Hendry (1983), but allowing for a general long-run income elasticity, the following specification is obtained

$$\underset{(0.255)}{\Delta c_t=0.329}+\underset{(0.041)}{0.528\Delta y_t}-\underset{(0.027)}{0.132\mathrm{p}_t}-\underset{(0.101)}{0.205}(c_{t-1}-\underset{(0.045)}{0.854}y_{t-1}).$$

$$\hat{\sigma}=0.507\% \qquad \mathrm{AR2}-F(2,28)=0.65 \qquad \mathrm{ARCH1}-F(1,28)=1.22$$
$$\mathrm{Norm}-\chi^2(2)=0.87 \qquad \mathrm{Chow}-F(10,20)=0.70$$

Diagnostic testing against second-order serial correlation, first-order autoregressive conditional heteroskedasticity, non-normality, and predictive failure over the last 10 years, reveals no misspecification of (16). The long-run income elasticity of consumption is quite far from unity, and significantly so, judging from the t-statistic $t_1=(0.854-1)/0.045=-3.21$ (or $W_1=(3.21)^2$). However, an equivalent formulation of (16) is

$$\underset{(0.255)}{\Delta c_t=0.329}+\underset{(0.041)}{0.528\Delta y_t}-\underset{(0.027)}{0.132\Delta p_t}\underset{(0.101)}{c_{t-1}-y_{t-1})}-\underset{(0.023)}{0.030y_{t-1}}.\quad(17)$$

leading to a t-statistic of $t_2=-0.030/0.023=-1.28$, i.e. $W_2=(1.28)^2$. In this case we see that the two formulations of the null hypothesis lead to conflicting inferences. Imposing the restriction yields

$$\underset{(0.003)}{\Delta c_t=0.004}+\underset{(0.039)}{0.510\Delta y_t}-\underset{(0.026)}{0.141\Delta p_t}-\underset{(0.018)}{0.078(c_{t-1}-y_{t-1})}. \qquad (18)$$

$\hat{\sigma}=0.512\%$ $\quad$ AR2 $-F(2,29)=0.53$ $\quad$ ARCH1 $-F(1,29)=0.02$
norm $-\chi^2(2)=0.97$ $\quad$ chow $-F(10,21)=0.58$

A marked effect is the strong decrease in the error correction coefficient, accompanied by an even stronger reduction in its standard error. However, the t_2 test and diagnostics indicate that the restricted model parsimoniously encompasses the general model.

The difference between the two approaches becomes even more apparent if we consider 95% confidence intervals. The 'naive' confidence interval for θ, based on (16) is equal to (0.765, 0.943), obviously excluding the long-run unit elasticity hypothesis. On the other hand, the confidence interval bsed on W_2, as suggested in the previous section, equals (0.804, 3.166). The considerable asymmetry of the interval is another reflection of the dependence of the variance of $\hat{\theta}$ on θ; apparently this variance increases with θ, so that quite large values of θ are included in the interval.

VI. CONCLUSION

In this note we have shown that a linear restriction on the long-run parameters of an error correction model may be conveniently reformulated as a linear restriction on a linear regression model. It has been argued that the corresponding Wald or F-test has more favourable finite sample properties than a test based on the asymptotic covariance matrix of the long-run parameter estimators.

Institute of Actuarial Science and Econometrics,
University of Amsterdam

Date of Receipt of Final Manuscript: August 1992

[2] The computations in this section were performed using *PC-GIVE*, see Hendry (1989).

REFERENCES

Banerjee, A., Galbraith, J. W. and Dolado, J. J. (1990). 'Dynamic Specification and Linear Transformations of the Autoregressive-Distributed Lag Model', *BULLETIN*, Vol. 52, pp. 95–104.

Bårdsen, G. (1989). 'Estimation of Long-Run Coefficients in Error Correction Models', *BULLETIN*, Vol. 51, pp. 345–50.

Bewley, R. A. (1979). 'The Direct Estimation of the Equilibrium Response in a Linear Dynamic Model', *Economics Letters*, Vol. 3, pp. 357–61.

Critchley, F., Marriott, P. and Salmon, M. (1989). 'On the Differential Geometry of the Wald Test with Nonlinear Restrictions', EUI Working Paper No. 89/424, European University Institute, Florence.

Engle, R. F. and Granger, C. W. J. (1987). 'Cointegration and Error Correction: Representation, Estimation, and Testing', *Econometrica*, Vol. 55, pp. 251–76.

Gregory, A. W. and Veall, M. R. (1985). 'Formulating Wald Tests of Nonlinear Restrictions', *Econometrica*, Vol. 53, pp. 1465–68.

Gregory, A. W. and Veall, M. R. (1986). 'Wald Tests of Common Factor Restrictions', *Economics Letters*, Vol. 22, pp. 203–08.

Gregory, A. W. and Veall, M. R. (1987). 'Formulating Wald Tests of the Restrictions Implied by the Rational Expectations Hypothesis', *Journal of Applied Econometrics*, Vol. 2, pp. 61–8.

Hendry, D. F. (1983). 'Economic Modelling: The Consumption Function in Retrospect', *Scottish Journal of Political Economy*, Vol. 30, pp. 193–200.

Hendry, D. F. (1989). *PC-GIVE: An Interactive Econometric Modelling System, Version* 6.01, Institute of Economics and Statistics, Oxford.

Johansen, S. (1992). 'Cointegration in Partial Systems and the Efficiency of Single-Equation Analysis', *Journal of Econometrics*, Vol. 52, pp. 389–402.

Wickens, M. R. and Breusch, T. S. (1988). 'Dynamic Specification, The Long-Run and the Estimation of Transformed Regression Models', *The Economic Journal*, Vol. 98, pp. 189–205.