A RESIDUALS-BASED WALD TEST FOR THE LINEAR SIMULTANEOUS EQUATION

By

Andrew W. Lo*
and
Whitney K. Newey**

Working Paper No. 1-85

RODNEY L. WHITE CENTER

FOR FINANCIAL RESEARCH

University of Pennsylvania
The Wharton School
Philadelphia, PA 19104

* The Wharton School, University of Pennsylvania
** Princeton University

A RESIDUALS-BASED WALD TEST FOR THE LINEAR SIMULTANEOUS EQUATION

Andrew W. Lo

University of Pennsylvania

and

Whitney K. Newey

Princeton University

September 1984

We are grateful to B. D. Berheim, J. Hausman, and R. W. Turner for helpful discussions.

1. Introduction

One of the most frequently used tests of linear restrictions on the coefficients of the standard linear regression model is the Wald (1943) test. In its original form, the Wald or W statistic is a quadratic form involving the estimated information matrix as well as the linear restrictions. However, it is well-known that the statistic may also be computed by using the sums of squared residuals from regressions with and without the restrictions imposed. It is the purpose of this note to present a similar residuals-based Wald statistic for testing linear restrictions on a single linear simultaneous equation estimated by two-stage least squares (2SLS). This test statistic is based on restricted and unrestricted sums of squared residuals but differs from the W-statistic for the standard linear model through the use of residuals transformed by the projection matrix of instrumental variables. The proposed statistic is equal to the "likelihood ratio" statistic suggested by Gallant and Jorgenson (1979) for the nonlinear 2SLS estimator, so that this note is partly expository in nature. We do, however, demonstrate the numerical equivalence of this statistic and the original Wald statistic for the linear 2SLS estimator. We also derive a Chow (1960) test for structural-parameter stability using the residuals-based W-statistic in the context of single-equation estimation by 2SLS.

2. The Wald Statistic

Consider the linear simultaneous equation with a priori excusion restrictions and normalization imposed:

$$y = X\delta_0 + \varepsilon \qquad (1)$$

where y is a $T\times 1$ vector of observations of the left-hand side variable, X is a Txq matrix of observations of right-hand side exogenous or predetermined and jointly endogenous variables, δ_0 is a $q\times 1$ vector of coefficients, and ε is a $T\times 1$ vector of disturbances. Let Z be a TxK matrix of observations of a set of instrumental variables. We make the identification assumptions that:

$$\operatorname*{plim}_{T\to\infty} \frac{Z'Z}{T} = Q \ , \qquad Q \text{ nonsingular}$$
$$\text{rank}\Big(\operatorname*{plim}_{T\to\infty} \frac{Z'X}{T}\Big) = q \ . \tag{2}$$

We also make the following assumption on the relation between the instruments Z and the disturbance ε:

$$\frac{Z'\varepsilon}{\sqrt{T}} \overset{A}{\sim} N(0, \ \sigma^2 Q^{-1}) \ . \tag{3}$$

We are interested in tests of the linear hypothesis:

$$H_0: \quad R\delta_0 = r \tag{4}$$

where R is a $p\times q$ matrix, rank (R) = $p < q$, and r is a $p\times 1$ vector.

The most commonly used statistic for testing (4) is the Wald statistic based on the 2SLS estimator of δ_0. Denote by P the projection matrix $Z(Z'Z)^{-1}Z'$ and let $\hat{y}$ and $\hat{X}$ be the predicted values from regressions of y on Z and X on Z respectively, i.e., $\hat{y} = Py$, $\hat{X} = PX$. The 2SLS estimator of δ_0 is given by:

$$\hat{\delta} = (\hat{X}'\hat{X})^{-1}\hat{X}'y = (\hat{X}'\hat{X})^{-1}\hat{X}'\hat{y} \ . \tag{5}$$

For any consistent estimator $\hat{\sigma}^2$ of σ^2, such as $\hat{\sigma}^2 = (y - X\hat{\delta})(y - X\hat{\delta})/(T - q)$, the Wald statistic is then:

$$W = \frac{(R\hat{\delta} - r)'\left[R(\hat{X}'\hat{X})^{-1}R'\right]^{-1}(R\hat{\delta} - r)}{\hat{\sigma}^2} \tag{6}$$

which is asymptotically chi-squared with p degrees of freedom under the null hypothesis $R\delta_0 = r$.

The form of the Wald statistic and the 2SLS estimator, and the equality of the Wald and sums-of-squared-residuals statistics in the standard linear model indicate how a residuals-based test statistic may be obtained. Consider a restricted 2SLS estimator $\tilde{\delta}$ which is obtained by solving either of the following constrained minimization problems:

$$\underset{\delta}{\text{Min}}(y - \hat{X}\delta)'(y - \hat{X}\delta) \quad \text{subject to } R\delta = r \tag{7a}$$

$$\underset{\delta}{\text{Min}}(\hat{y} - \hat{X}\delta)'(\hat{y} - \hat{X}\delta) = \underset{\delta}{\text{Min}}(y - X\delta)'P(y - X\delta) \tag{7b}$$

$$\text{subject to } R\delta = r \ .$$

That the solutions of (7a) and (7b) are identical follows from the symmetric idempotence of P. Using equations (5), (6), and the well-known constrained least squares relation[1] $\tilde{\delta} = \hat{\delta} - (\hat{X}'\hat{X})^{-1}R'[R(\hat{X}'\hat{X})^{-1}R']^{-1}(R\hat{\delta} - r)$ (which is functionally independent of y or $\hat{y}$), a straightforward calculation yields:

$$W = \frac{(y - \hat{X}\tilde{\delta})'(y - \hat{X}\tilde{\delta}) - (y - \hat{X}\hat{\delta})'(y - \hat{X}\hat{\delta})}{\hat{\sigma}^2} \tag{8a}$$

$$= \frac{(\hat{y} - \hat{X}\tilde{\delta})'(\hat{y} - \hat{X}\tilde{\delta}) - (\hat{y} - \hat{X}\hat{\delta})'(\hat{y} - \hat{X}\hat{\delta})}{\hat{\sigma}^2}$$

$$\therefore W = \frac{(y - X\tilde{\delta})'P(y - X\tilde{\delta}) - (y - X\hat{\delta})'P(y - X\hat{\delta})}{\hat{\sigma}^2} \tag{8b}$$

The Wald statistic is then the difference of the sum of squared residuals of the restricted and unrestricted regressions of y (or $\hat{y}$) on $\hat{X}$ divided by $\hat{\sigma}^2$. Note that in the case of 2SLS estimation, the numerator of the W-statistic uses the residuals from the "second-stage" regression while the denominator uses the structural residuals $\hat{\varepsilon}$ to form $\hat{\sigma}^2$, where $\hat{\varepsilon} = y - X\hat{\delta}$. Alternatively the Wald statistic may be obtained by forming the restricted structural residuals $\tilde{\varepsilon} = y - X\tilde{\delta}$, projecting both $\hat{\varepsilon}$ and $\tilde{\varepsilon}$ on the instruments Z, and forming the numerator of W as the difference of the restricted and unrestricted projected residuals. This is the form of the Wald statistic given in equation (8b).

Note that since $\tilde{\delta}$ is obtained from a restricted least-squares computation the same methods used in the standard linear regression model may be used here to obtain the restricted estimator. For example, consider the following model and linear hypothesis:

$$y = \delta_1 X_1 + \delta_2 X_2 + \varepsilon$$

$$H_0: \ \delta_1 + \delta_2 = 1 \ .$$

Then by adding and subtracting $X_1(\delta_1 + \delta_2 - 1)$ to both sides of the second-stage regression, we see that the restricted sum of squared residuals $(y - \hat{X}\tilde{\delta})'(y - \hat{X}\tilde{\delta})$ can be obtained from performing 2SLS on $y = X_1 = (X_2 - X_1)\tilde{\delta}_1 + \tilde{\varepsilon}$. One word of caution is in order. In calculating W, the set of instruments used in the restricted 2SLS estimation must be the <u>same</u> as the set of instruments used in the unrestricted 2SLS estimation. Our discussion of stability tests in the next section will illustrate this point.

3. The Chow Test for a Single Simultaneous Equation

A useful model validity check for a single simultaneous equation may be obtained by extending Chow's (1960) coefficient stability test to the simultaneous equations case. A "Wald" version of such a stabiilty test may be formulated as well.

Consider again the single simultaneous equation of Section 2. The Chow test is constructed in the following manner. Divide the sample of T observations into two subsamples, call them 1 and 2, each with T_1 and T_2 observations respectively where $T = T_1 + T_2$. Let y_i be a $T_i \times 1$ vector of observations of the left-hand side variable, X_i a $T_i \times q$ matrix of observations of right-hand side variables, and Z_i a $T_i \times K$ matrix of observations of instrumental variables, where rank $(Z_i) = K$ and $i = 1, 2$. Let the unrestricted version of the two subsample equations be given by the two equations:

$$y_i = X_i \delta_i + \varepsilon_i \,. \qquad (i = 1,\ 2) \qquad (9)$$

We make the following assumptions:

$$\operatorname*{plim}_{T_i \to \infty} \frac{Z_i' Z_i}{T} = Q_i, \qquad Q_i \text{ nonsingular} \qquad (i = 1,\ 2)$$

$$\operatorname{rank}\left(\operatorname{plim} \frac{Z_i' X_i}{T}\right) = q \qquad (i = 1,\ 2) \qquad (10)$$

$$\frac{Z_i' \varepsilon_i}{\sqrt{T_i}} \overset{A}{\sim} N(0,\ \sigma_i^2 Q_i^{-1}) \,. \qquad (i = 1,\ 2)$$

For $i = 1, 2$ define the projection matrices $P_i = Z_i(Z_i'Z_i)^{-1}Z_i'$ and the predicted values $\hat{X}_i = P_i X_i$, $\hat{y}_i = P_i y_i$. The 2SLS estimator of δ_i for each subsample i is $\hat{\delta}_i = (\hat{X}_i'\hat{X}_i)^{-1}\hat{X}_i' y_i = (\hat{X}_i'\hat{X}_i)^{-1}\hat{X}_i'\hat{y}_i$. Let $\hat{\sigma}_i^2 = (y_i - X_i\hat{\delta}_i)'(y_i - X_i\hat{\delta}_i)/(T_i - q)$ be estimators of the σ_i^2's. Then a Wald test-statistic for structural stability is given by:

$$W_c = (\hat{\delta}_1 - \hat{\delta}_2)'\left[\hat{\sigma}_1^2(\hat{X}_1'\hat{X}_1)^{-1} + \hat{\sigma}_2^2(\hat{X}_2'\hat{X}_2)^{-1}\right]^{-1}(\hat{\delta}_1 - \hat{\delta}_2) \qquad (11)$$

which is asymptotically distributed as chi-square with q degrees of freedom as <u>both</u> T_1 and T_2 increase.

We may use the results discussed in Section 2 to obtain a sums-of-squared-residuals version of W_c. Normalization by the estimated standard deviation and stacking yield:

$$y^* = \begin{vmatrix} y_1/\hat{\sigma}_1 \\ y_2/\hat{\sigma}_2 \end{vmatrix} = \begin{vmatrix} X_1/\hat{\sigma}_1 & 0 \\ 0 & X_2/\hat{\sigma}_2 \end{vmatrix} \begin{vmatrix} \delta_1 \\ \delta_2 \end{vmatrix} + \begin{vmatrix} \varepsilon_1/\hat{\sigma}_1 \\ \varepsilon_2/\hat{\sigma}_2 \end{vmatrix} \tag{12}$$

$$= X^*\delta^* + \varepsilon^* \ .$$

Define $\hat{\delta}^*$ and Z^* as:

$$\hat{\delta}^* = \begin{vmatrix} \hat{\delta}_1 \\ \hat{\delta}_2 \end{vmatrix} , \qquad Z^* = \begin{vmatrix} Z_1/\hat{\sigma}_1 & 0 \\ 0 & Z_2/\hat{\sigma}_2 \end{vmatrix} . \tag{13}$$

Let P^* denote the projection matrix $P^* = Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime} = \text{diag}(P_1, P_2)$. Note that W_c may then be interpreted as a Wald statistic of the form given in equation (6) with $\hat{\sigma}^2 = 1$, $\hat{\delta} = \hat{\delta}^*$, $(\hat{X}'\hat{X})^{-1} = \text{diag}\left(\hat{\sigma}_1^2(\hat{X}_1'\hat{X}_1)^{-1}, \hat{\sigma}_2^2(\hat{X}_2'\hat{X}_2)^{-1}\right.$, $R = [I_q \ - I_q]$, and $r = 0$. Let

$$X_0 = \begin{vmatrix} X_1/\hat{\sigma}_1 \\ X_2/\hat{\sigma}_2 \end{vmatrix} , \qquad \hat{X}_0 = P^*X_0 = \begin{vmatrix} \hat{X}_1/\hat{\sigma}_1 \\ \hat{X}_2/\hat{\sigma}_2 \end{vmatrix} . \tag{14}$$

Then the restricted 2SLS estimator $\tilde{\delta}^*$ of δ^* is $[\tilde{\delta}_1' \tilde{\delta}_2']'$ where $\tilde{\delta}_1 = \tilde{\delta}_2 = (\hat{X}_0'\hat{X}_0)^{-1}\hat{X}_0'\hat{y}^*$. Equation (8) then implies that a form of W_c based on sums of squared residuals is given by:

$$W_c = (y^* - X^*\tilde{\delta}^*)'P^*(y^* - X^*\tilde{\delta}^*) - (y^* - X^*\hat{\delta}^*)'P^*(y^* - X^*\hat{\delta}^*) \quad (15a)$$

$$= \frac{(y_1 - X_1\tilde{\delta}_1)'P_1(y_1 - X_1\tilde{\delta}_1) - (y_1 - X_1\hat{\delta}_1)'P_1(y_1 - X_1\hat{\delta}_1)}{\hat{\sigma}_1^2} + \quad (15b)$$

$$\frac{(y_2 - X_2\tilde{\delta}_2)'P_2(y_2 - X_2\tilde{\delta}_2) - (y_2 - X_2\hat{\delta}_2)'P_2(y_2 - X_2\hat{\delta}_2)}{\hat{\sigma}_2^2} .$$

The appropriate test statistic has the same form as the standard linear regression case except, as before, the residuals here are first projected on to the instruments.

Note that the restricted estimator is obtained by using 2K instruments (the columns of Z^*) as instrumental variables for the q right-hand side variables X. That is, the restricted estimators $\tilde{\delta}_i$ uses the same instrumental variables as the unrestricted estimators $\hat{\delta}_i$. Also, for simplicity we have treated only the case where the variance of the disturbance is allowed to differ across subsamples. A Chow test which imposes $\sigma_1^2 = \sigma_2^2$ would be equal to a Wald statistic which used the same estimator of σ^2 in forming the estimates of each of the asymptotic covariance matrices of $\hat{\delta}_1$ and $\hat{\delta}_2$. Finally, we have not discussed the deficient observations case where $T_i < K$ or $T_i < q$ for some i. The asymptotic theory which we have appealed to above would most likely provide a poor approximation in this case.

4. Conclusion

To summarize, the suggested procedure for computing the Wald statistic is the following:

(i) Form the restricted and unrestricted residuals $\hat{\varepsilon} = y - X\hat{\delta}$, $\tilde{\varepsilon} = y - X\tilde{\delta}$.

(ii) Transform the residuals, using the projection matrix P, to $\hat{u} = P\hat{\varepsilon}$, $\tilde{u} = P\tilde{\varepsilon}$.

(iii) Calculate $W = (T - q)(\tilde{u}'\tilde{u} - \hat{u}'\hat{u})/\hat{e}'\hat{e} \overset{A}{\sim} \chi^2_{\mathrm{rank}(R)}$.

Provided that there are sufficient degrees of freedom, the procedure for the Chow test is to perfom steps (i) to (iii) for each subsample and then compute $W_c = W_1 + W_2 \overset{A}{\sim} \chi^2_q$, where W_i is the statistic obtained from step (iii) for subsample i.

FOOTNOTES

1. See, for example, Theil (1971), p. 44, equation (8.9).

REFERENCES

Chow, Gregory C., "Tests of Equality Between Subsets of Coefficients in Two Linear Regressions," Econometrica, 28 (1960), pp. 591-605.

Gallant, A. Ronald and Dale W. Jorgenson, "Statistical Inference for a System of Simultaneous, Non-linear, Implicit Equations in the Context of Instrumental Variables Estimation," Journal of Econometrics, 11 (1979), pp. 275-302.

Theil, Henri, Principles of Econometrics, New York: John Wiley and Sons, 1971.

Wald, Abraham, "Tests of Statistical Hypotheses Concerning Several Parameters when the Number of Observations is Large," Transactions of the American Mathematical Society, 54 (1943), pp. 426-482.