
Some Further Evidence on the Use of the Chow Test under Heteroskedasticity

Author(s): Peter Schmidt and Robin Sickles

Source: *Econometrica*, Vol. 45, No. 5 (Jul., 1977), pp. 1293-1298

Published by: Econometric Society

Stable URL: <http://www.jstor.org/stable/1914076>

Accessed: 09-11-2015 16:43 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Wiley and Econometric Society are collaborating with JSTOR to digitize, preserve and extend access to *Econometrica*.

<http://www.jstor.org>

SOME FURTHER EVIDENCE ON THE USE OF THE CHOW TEST UNDER HETEROSKEDASTICITY

BY PETER SCHMIDT AND ROBIN SICKLES

1. INTRODUCTION

THE CHOW TEST IS A TEST of equality of sets of coefficients in two regressions. Part of the maintained hypothesis of the test is that the error variances be the same for the two regressions. If this is not the case, then the Chow test may be inaccurate, in the sense that the true size of the test (under the null hypothesis) may not equal the prescribed alpha-level.

Toyoda [3] has investigated the accuracy of the Chow test under conditions of heteroskedasticity, using an approximation to the distribution of the test statistic for the Chow test. However, the exact distribution, while cumbersome, can be calculated by the method of Imhof [2]. In this paper we redo some of Toyoda's calculations using the exact distribution, to test the adequacy of his approximation. We find that his approximation is sometimes rather inaccurate, especially when the two sample sizes and the two variances are very different. Some of Toyoda's conclusions are found not to be supported by our new evidence.

2. THE DISTRIBUTION OF THE TEST STATISTIC

Following the notation of Toyoda, we consider the model

$$(1) \quad Y_1 = X_1\beta_1 + u_1 = X_1\hat{\beta}_1 + e_1$$

and

$$(2) \quad Y_2 = X_2\beta_2 + u_2 = X_2\hat{\beta}_2 + e_2,$$

where Y_i and X_i are $T_i \times 1$ and $T_i \times k$ observation matrices, β_i and $\hat{\beta}_i$ are $k \times 1$ coefficient and least squares estimator vectors, and u_i and e_i are $T_i \times 1$ disturbance and residual vectors. Under the null hypothesis the model can be written as

$$(3) \quad Y = X\beta + u = X\hat{\beta} + e,$$

where

$$(4) \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix},$$

and where $\hat{\beta}$ and e are the least squares estimator and residual vector associated with (3).

The test statistic for the Chow test is

$$(5) \quad F = \frac{[e'e - e'_1e_1 - e'_2e_2]/k}{[e'_1e_1 + e'_2e_2]/(T_1 + T_2 - 2k)}.$$

If both (1) and (2) separately satisfy the usual ideal conditions, and if in addition $\sigma_1^2 = \sigma_2^2$, then the statistic in (5) has an F distribution with k and $T_1 + T_2 - 2k$ degrees of freedom under the null hypothesis that $\beta_1 = \beta_2$.

If $\sigma_1^2 \neq \sigma_2^2$, things are more complicated. Of course we still can write

$$(6) \quad F = C \frac{u'Mu - u'_1M_1u_1 - u'_2M_2u_2}{u'_1M_1u_1 + u'_2M_2u_2}$$

where $C = (T_1 + T_2 - 2k)/k$, and where

$$(7) \quad M = I - X(X'X)^{-1}X', \quad M_1 = I - X_1(X'_1X_1)^{-1}X'_1, \quad M_2 = I - X_2(X'_2X_2)^{-1}X'_2.$$

This in turn can be written as

$$(8) \quad F = C \frac{u' Au}{u' Bu}$$

where

$$(9) \quad B = \begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix}, \quad A = M - B.$$

Therefore, for a given critical point f , we have

$$(10) \quad P(F \geq f) = P(u' Hu \geq 0)$$

where

$$(11) \quad H = CA - fB.$$

Finally, we note that $u' Hu = u^{*'} H^* u^*$, where

$$(12) \quad u^* = \begin{bmatrix} \frac{1}{\sigma_1} u_1 \\ \frac{1}{\sigma_2} u_2 \end{bmatrix}, \quad H^* = \begin{bmatrix} \sigma_1^2 H_{11} & \sigma_1 \sigma_2 H_{12} \\ \sigma_1 \sigma_2 H_{21} & \sigma_2^2 H_{22} \end{bmatrix},$$

the H_{ij} being the appropriate submatrices of H . (H_{ij} is of dimension $T_i \times T_j$.) We therefore get

$$(13) \quad P(F \geq f) = P(u^{*'} H^* u^* \geq 0).$$

Since the elements of u^* are iid $N(0, 1)$, this probability can be calculated by the method given in Imhof [2]. Specifically,

$$(14) \quad P(F \geq f) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{\sin \left[\frac{1}{2} \sum_{t=1}^T \arctan(\eta_t r) \right]}{r \prod_{t=1}^T (1 + \eta_t^2 r^2)^{\frac{1}{2}}} dr$$

where $\eta_t, t = 1, \dots, T$, are the $T = T_1 + T_2$ eigenvalues of H^* . This can be evaluated numerically.

3. NUMERICAL RESULTS

In this section some values of (14) have been calculated to show the result of unequal variances on the true level of significance of the Chow test. (The nominal level of significance is set at 0.05.) These figures can then be compared with those given by Toyoda to check on his approximation.

These comparisons are complicated somewhat by the fact that Toyoda's approximate distribution depends only on T_1, T_2, k , and $\theta = \sigma_2^2/\sigma_1^2$. The exact distribution, on the other hand, depends on the form of X_1 and X_2 as well as T_1, T_2, k , and $\theta = \sigma_2^2/\sigma_1^2$. The exact distribution's dependence on θ instead of σ_2^2 and σ_1^2 can be seen most easily by noting that equation (13) is not affected by scalar multiplication. Certainly the particular X_1 and X_2 that one chooses may affect the quality of Toyoda's approximation.

We have therefore concentrated on the case $k = 2$ and tried four different types of X_1 and X_2 matrices. In each case the first column of X_1 and X_2 was a constant term. The four cases differed in taking the second column of X_1 and X_2 as: (i) a linear trend; (ii) iid $N(0, 1)$ deviates; (iii) $N(0, 1)$ variables with first-order autocorrelation coefficient $\rho = .5$; (iv) same as (iii), except $\rho = -.5$. The results are given in Table I, corresponding to the cases:

TABLE I
TRUE SIGNIFICANCE LEVEL OF CHOW TEST

T_1	T_2	k	θ	Linear Trend	iid $N(0, 1)$	$\rho = 0.5$	$\rho = -0.5$
10	10	2	.01	.087	.074	.074	.074
			.10	.076	.066	.066	.066
			1	.050	.050	.050	.050
			10	.076	.066	.066	.066
			100	.087	.074	.074	.074
25	25	2	.01	.084	.059	.059	.059
			.10	.063	.056	.056	.056
			1	.050	.050	.050	.050
			10	.063	.056	.056	.056
			100	.084	.059	.059	.059
20	30	2	.01	.130	.143	.142	.138
			.10	.111	.121	.119	.116
			1	.050	.050	.050	.050
			10	.050	.021	.022	.022
			100	.054	.018	.018	.020
40	10	2	.01	.010	.0001	.0002	.0001
			.10	.011	.0001	.0001	.0001
			1	.050	.050	.050	.050
			10	.306	.381	.375	.382
			100	.411	.502	.497	.504

$T_1 = T_2 = 10$; $T_1 = T_2 = 25$; $T_1 = 20, T_2 = 30$; $T_1 = 40, T_2 = 10$. Three of these correspond to cases considered by Toyoda. Toyoda considered the case $T_1 = T_2 = 50$; however, the cost of extracting eigenvalues of a 100×100 matrix was excessive, so we used $T_1 = T_2 = 25$. Also, Toyoda considered the case $T_1 = 30, T_2 = 20$, but these results are essentially the same as for $T_1 = 20, T_2 = 30$ —e.g., $T_1 = 20, T_2 = 30, \theta = 10$ corresponds to $T_1 = 30, T_2 = 20, \theta = .1$.

Looking at Table I, we see that the form of the X matrix sometimes makes a noticeable difference. More significantly, comparing our results with the results corresponding to $k = 2$ in Toyoda's Table I shows that his approximation is not always very accurate. For example, in the case $T_1 = T_2 = 10, k = 2, \theta = .1$, Toyoda gives a significance level of .168, while for our four X matrices we get .076, .066, .066, and .066. In the same case, but with $\theta = .01$, Toyoda gives .218, while we get .087, .074, .074, and .074.

In terms of the differences between our results and Toyoda's, four points stand out:

(i) We typically find the true level of significance to be less than that given by the approximation. As a result, we find Toyoda's results to be generally too pessimistic about the effect of heteroskedasticity on the Chow test.

(ii) We sometimes find a true level of significance less than the nominal level of .05, which Toyoda did not. This happens only when the two sample sizes are markedly different ($T_1 = 40, T_2 = 10$) and when the variance is larger in the larger-sized sample.

(iii) The approximation becomes less accurate as the two variances become more different.

(iv) We do not find that increasing one of the sample sizes (e.g., in going from $T_1 = 10, T_2 = 10$ to $T_1 = 40, T_2 = 10$) improves the reliability of the test—indeed, the opposite is true. This is in agreement with Toyoda's numerical results, but is not in agreement with his conclusion (based on analytic examination of the approximating distribution) that the test is accurate as long as at least one sample size is large.

TABLE II
TRUE SIGNIFICANCE LEVEL OF CHOW TEST
 $T_1 = T_2 = 10; k$ Variables

θ	$k = 3$	$k = 4$	$k = 5$
.01	.084	.093	.103
.10	.071	.078	.083
1	.050	.050	.050
10	.071	.078	.083
100	.084	.093	.103

Finally, to make sure that the case $k = 2$ is not radically different from the cases $k = 3, 4, 5$ (which Toyoda also considered), we have redone our calculations with $k = 3, 4, 5$. This was done only for the case $T_1 = T_2 = 10$, with all columns of X except the constant being made up of iid $N(0, 1)$ variables. The results are given in Table II. They are not very different from the corresponding results of Table I.

4. AN EXAMINATION OF TOYODA'S APPROXIMATION

Given that our exact results for some particular cases differ substantially from Toyoda's approximation, it is worth inquiring as to the reason for these differences. This section will attempt to provide at least a partial answer.

The test statistic for the Chow test is, as given in equation (5),

$$(15) \quad F = \frac{(e'e - e'_1e_1 - e'_2e_2)/k}{(e'_1e_1 + e'_2e_2)/(T_1 + T_2 - 2k)}$$

In his Lemma 1 [3, p. 602], Toyoda approximates the distribution of $(e'_1e_1 + e'_2e_2)$ as a scalar multiple of a chi-squared distribution. The scalar multiple and the number of degrees of freedom are chosen so as to make the first two moments of the exact distribution and the approximate distribution equal. We feel that this is a reasonable approximation, and is not apt to be the major source of inaccuracy.

This leaves the task of arriving at a reasonable approximation for the numerator of F , $(e'e - e'_1e_1 - e'_2e_2)$. The same method might suggest itself. However, we find

$$(16) \quad E(e'e - e'_1e_1 - e'_2e_2) = k(\sigma_1^2 + \sigma_2^2) - \text{trace} [(X'_1X_1 + X'_2X_2)^{-1} \cdot (\sigma_1^2X'_1X_1 + \sigma_2^2X'_2X_2)],$$

which depends on X_1 and X_2 ; any such approximation would be complicated by its dependence on X_1 and X_2 .

As an alternative, Toyoda notes that

$$(17) \quad e'e - e'_1e_1 - e'_2e_2 = u'Au$$

where A is as defined in our equations (8) and (9). The matrix A is idempotent of rank k . If $\text{cov}(u) = \sigma^2I$, then $u'Au$ would be distributed as $\sigma^2\chi^2_k$. Since, in fact,

$$(18) \quad \text{cov}(u) = \begin{bmatrix} \sigma_1^2I_{T_1} & 0 \\ 0 & \sigma_2^2I_{T_2} \end{bmatrix},$$

Toyoda concludes that the distribution of $e'e - e'_1e_1 - e'_2e_2$ is approximately $\sigma^2\chi^2_k$, where " σ^2 is any well-chosen weighted average of σ_1^2 and σ_2^2 " [3, p. 602].

Two things about this are clear. First, the approximation is exact when $\sigma_1^2 = \sigma_2^2$. Secondly, the approximation will be better when σ_1^2 and σ_2^2 are approximately equal; it may not be very good when σ_1^2 and σ_2^2 are far from equal. We feel that this is the probable reason for one of our major findings of the previous section—namely, that the accuracy of the approximation decreases as the variances become more different.

We next proceed to note that the actual weighted average of σ_1^2 and σ_2^2 that Toyoda used is

$$(19) \quad \sigma^2 \equiv \frac{(T_1 - k)\sigma_1^4 + (T_2 - k)\sigma_2^4}{(T_1 - k)\sigma_1^2 + (T_2 - k)\sigma_2^2}$$

(This is a convenient choice since it equals the scale factor used in the approximation of $e_1'e_1 + e_2'e_2$.)

This weighted average has the characteristic that, for fixed T_2 , $\sigma^2 \rightarrow \sigma_1^2$ as $T_1 \rightarrow \infty$. This may seem reasonable at first thought since, as $T_1 \rightarrow \infty$, more and more observations actually have a variance of σ_1^2 . However, in equation (16) note that as $T_1 \rightarrow \infty$,

$$(20) \quad \text{trace} [(X_1'X_1 + X_2'X_2)^{-1}(\sigma_1^2 X_1'X_1 + \sigma_2^2 X_2'X_2)] \rightarrow k\sigma_1^2$$

and, therefore,

$$(21) \quad E(e'e - e_1'e_1 - e_2'e_2) \rightarrow k\sigma_2^2.$$

Therefore, as $T_1 \rightarrow \infty$ (with T_2 fixed) we find that the expected value of the approximate distribution approaches $\sigma_1^2 k$, while the expected value of the exact distribution approaches $\sigma_2^2 k$.

In fact, it is not hard to show that as $T_1 \rightarrow \infty$, $e'e - e_1'e_1 - e_2'e_2$ converges in distribution to $\sigma_2^2 \chi_k^2$. To see this, note that

$$\begin{aligned} e'e - e_1'e_1 - e_2'e_2 &= u'Mu - u_1'M_1u_1 - u_2'M_2u_2 \\ &= u_1'u_1 - u_1'X_1(X_1'X_1 + X_2'X_2)^{-1}X_1'u_1 \\ &\quad + u_2'u_2 - u_2'X_2(X_1'X_1 + X_2'X_2)^{-1}X_2'u_2 - 2u_1'X_1(X_1'X_1 + X_2'X_2)^{-1}X_2'u_2 \\ &\quad - u_1'u_1 + u_1'X_1(X_1'X_1)^{-1}X_1'u_1 - u_2'u_2 + u_2'X_2(X_2'X_2)^{-1}X_2'u_2. \end{aligned}$$

Now, as $T_1 \rightarrow \infty$, we have

$$\begin{aligned} u_2'X_2(X_1'X_1 + X_2'X_2)^{-1}X_2'u_2 &\rightarrow 0, \\ u_1'X_1(X_1'X_1 + X_2'X_2)^{-1}X_2'u_2 &\rightarrow 0, \quad \text{and} \\ u_1'X_1(X_1'X_1)^{-1}X_1'u_1 - u_1'X_1(X_1'X_1 + X_2'X_2)^{-1}X_1'u_1 &\rightarrow 0, \end{aligned}$$

so that asymptotically we have $e'e - e_1'e_1 - e_2'e_2$ converging in probability to $u_2'X_2(X_2'X_2)^{-1}X_2'u_2$, whose distribution is $\sigma_2^2 \chi_k^2$.

What this means is that as $T_1 \rightarrow \infty$ (with T_2 fixed), the approximating distribution tends to $\sigma_1^2 \chi_k^2$, while the actual distribution tends to $\sigma_2^2 \chi_k^2$. Clearly, this raises the suspicion that the approximation should *not* behave well as one sample size becomes large relative to the other (as long as $\sigma_1^2 \neq \sigma_2^2$).

As a result, Toyoda's assertion [3, p. 605] that "if at least one of the two samples is of large size, the Chow test is robust for any finite variations of variances" is questionable, since this assertion is based on the approximation. It is, therefore, not surprising that our calculations show the Chow test to be quite nonrobust in the cases corresponding to $T_1 = 40, T_2 = 10$.

It is also not surprising that in the cases corresponding to $T_1 = 40, T_2 = 10$, we sometimes find a true significance level less than the nominal level of .05, whereas according to the approximation this should not be the case. Indeed, it is instructive that this happens (in our calculations) only in this case, when the two sample sizes are markedly different.

5. CONCLUSIONS

In this paper we have calculated the exact significance level of the Chow test under heteroskedasticity. The results obtained are specific to the cases considered, but they do serve to check the approximation suggested by Toyoda. Generally speaking, they do not show his approximation to be as accurate as might be hoped. In particular, our exact results suggest four main conclusions:

- (i) The level of significance is typically less than that given by the approximation.
- (ii) We sometimes find a true level of significance less than the nominal level of significance.
- (iii) The approximation becomes less accurate as the variances become more different.
- (iv) We do not find that increasing one sample size (with the other sample size fixed) increases the reliability of the test.

Finally, we examined Toyoda's approximation to see if we could account for these results. We did find reasons (in terms of the way that the approximation was constructed) for conclusions (ii), (iii), and (iv) above, though not for conclusion (i).

Based on our numerical results and on our examination of Toyoda's approximation procedure, we would have to conclude the approximation is of reasonable accuracy only when the two variances are of the same order of magnitude, and when the two sample sizes are also of the same order of magnitude.

*University of North Carolina
and
George Washington University*

Manuscript received February, 1975; final revision received August, 1976.

REFERENCES

- [1] CHOW, G. C.: "Tests of Equality between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28 (1960), 591-605.
- [2] IMHOF, J. P.: "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48 (1961), 419-426.
- [3] TOYODA, T.: "Use of the Chow Test under Heteroskedasticity," *Econometrica*, 42 (1974), 601-608.