SECTION III. CHOW AND WALD TESTS FOR THE STABILITY OF COEFFICIENT ESTIMATES

This section presents the results of Wald and Chow tests for the stability of the estimated coefficients of the Bretton Woods formula, using data from the Sixth to the Eleventh Reviews, including explanatory notes.

Part A presents the Chow test results on the stability of the coefficient estimates of the Bretton Woods formula, using data from the Sixth to the Eleventh Reviews, including the methodology used in performing the Chow tests.

The results of these statistical tests suggest instability in the coefficients of the Bretton Woods formula. The best result seems to come from the pair-wise test of the Seventh and Eighth Reviews, which could perhaps be attributed to two factors: (1) the short time period in between these reviews because the Eighth Review was accelerated; and (2) the relatively large selective element of the Eighth Review, which allows the underlying economic variables to have a somewhat greater influence on the outcome of the Eighth Review.

Part B presents the results of tests for the stability of the coefficient estimates of the Bretton Woods formula, using data from the Sixth to the Eleventh Reviews, while taking into account the systematic tendency of actual quotas to fall over time in relation to GDP or external trade.

The significance of the tendency of actual quotas to fall in relation to GDP or external trade over the Sixth to the Eleventh Reviews is examined by pair-wise testing (F-tests) of whether the variances of the error terms from the estimated Bretton Woods formula over these reviews are equal. If the error variance is the same over a pair of reviews, then the Chow test remains an appropriate test. If not, we need to use other statistics, like the Wald test.

The formal tests for the equality of variances between reviews show that the error variances for the rolling (or pair-wise comparisons of) Sixth and Seventh, Seventh and Eighth, Ninth and Tenth, and Tenth and Eleventh Reviews are statistically different. This implies that the Chow test for these pairs is inappropriate. Nonetheless, the alternative Wald test indicates that the coefficients generated by the Bretton Woods formula are not stable for these pair-wise comparisons of quota review periods. For the pair of the Eighth and Ninth Reviews, however, the results of the Chow test in Part A are valid.

In sum, regardless of whether we can use the Chow test or have to use the Wald test, the statistical tests suggest instability of the coefficient estimates of the Bretton Woods formula. A detailed list of the regressions performed is presented in Statistical Appendix, Part B, Section II.

A. Chow Tests

1. To test the stability of coefficients in the reduced Bretton Woods formula over the Sixth to Eleventh Reviews, we used the Chow test for (1) rolling consecutive pairs of reviews, starting with the Sixth Review (Table III.1.1), and (2) cumulative reviews, starting with the Sixth Review (Table III.1.2).

2. To perform the Chow tests, we used a constant-membership sample of 121 members participating in the Sixth Review plus China. To test, for example, the stability of the coefficient estimates under the Sixth and Seventh Reviews, we run the following regressions:

$$Q = (a^1 Y + b^1 R + c^1 P + d^1 VC) \times (1 + C/Y) \quad \text{for the Sixth Review}$$

$$Q = (a^2 Y + b^2 R + c^2 P + d^2 VC) \times (1 + C/Y) \quad \text{for the Seventh Review}$$

If $a^1 = a^2 \quad b^1 = b^2 \quad c^1 = c^2 \quad d^1 = d^2$, then we can estimate a common relationship for the entire (pooled) data, i.e.

$$Q = (aY + bR + cP + dVC) \times (1 + C/Y) \quad \text{for the Sixth and Seventh Reviews}$$

These four linear restrictions on a, b, c, and d can be tested using the F test. The F test is

$$F = \frac{(SSR_p - SSR_s)/(4+1)}{SSR_s/(122+122-2x4-2)}$$

where

SSR = Sum of squared residuals
$SSR_p = SSR$ of the pooled data $= SSR_{6thR\ and\ 7thR}$
$SSR_s = SSR_{6thR} + SSR_{7thR}$

which has an F distribution with degrees of freedom (4 + 1), (122 + 122 - 2 x 4 - 2).

3. If the F value is less than the critical value of 2.25 from the F tables at the 5 percent significance level, i.e., the calculated F value is not significant at the 5 percent level, we do not reject the null hypothesis that the relationship is stable (see Madala, G. S., Econometrics, 1977, McGrow-Hill, pp. 198–199).

4. The results shown in Tables III.1.1 and III.1.2 indicate, at the 5 percent significance level, rejection of the hypothesis that the coefficients of the Bretton Woods formula are stable over time.

Table III.1.1 : Chow Tests (Rolling)

	F-Statistic
Sixth and Seventh Reviews	14.22 *
Seventh and Eighth Reviews	3.67 *
Eighth and Ninth Reviews	20.65 *
Ninth and Tenth Reviews	40.27 *
Tenth and Eleventh Reviews	12.35 *

Note: The asterisk indicates significance of the 5% significance level.

Table III.1.2 : Chow Tests (Cumulative)

	F-Statistic
Sixth and Seventh Reviews	14.22 *
Seventh and Eighth Reviews	22.92 *
Eighth and Ninth Reviews	76.60 *
Ninth and Tenth Reviews	188.47 *
Tenth and Eleventh Reviews	127.58 *

Note: The asterisk indicates significance of the 5% significance level.

B. Wald Tests: Tests of Structural Change with Unequal Variances

In using the Chow test, an important assumption made is that the error variance is the same in all regressions. If this is not true, the error variance for one quota review period is σ_1^2, while that for the next review period is σ_2^2, and so on, in the restricted (two-reviews combined) model. The restricted model is, therefore, heteroscedastic, and the results from applying the non-linear Bretton Woods formula to such a model present problems of statistical inference.[1] In this case, it has been argued that it is likely that we overestimate the significance level of our test statistic.[2] In other words, the calculated F statistic is biased upward and indicates greater instability in the coefficient estimates than in fact exists.

To deal with this problem, we estimate all separate regressions and examine the estimates of the error variances. To test for significant differences, we use pairwise F-tests.[3] Without any significant difference, we proceed with Chow tests. If, however, there is evidence to suggest that the variances are actually different, we may explicitly estimate the model, accounting for the heteroscedasticity. However, if the sample is reasonably large, we may use the Wald test that is valid whether or not the error variances are the same. To set up this test, we suppose that $\hat{\theta}_1$ and $\hat{\theta}_2$ are two normally distributed estimators of a parameter based on independent samples,[4] with variance matrices V_1 and V_2. Then, under the null hypothesis that the two estimates have the same expected value, i.e., there is no structural change between the two quota reviews,

$$\hat{\theta}_1 - \hat{\theta}_2 \text{ has mean 0 and variance } V_1 + V_2$$

[1] In particular, heteroscedasticity exists whenever the variance of the error term changes across different segments of the population, which are determined by the different values of the explanatory variables (Wooldridge, Jeffrey, M., 2000, Introductory Econometrics: A Modern Approach, South-Western College Publishing, Thomson Learning, United States, p. 248). In the case of the estimated Bretton Woods formula, heteroscedasticity is present if the variance of the error term increases with the factors affecting actual quotas, i.e., GDP, trade, reserves, and variability.

[2] Toyoda, Toshihisa, 1974, "Use of the Chow Test Under Heteroscedasticity," Econometrica, Vol. 42, No. 3, May, pp. 601–8; and Schmidt, Peter and Robin Sickles, 1977, "Some Further Evidence on the Use of the Chow Test Under Heteroscedasticity," Econometrica, Vol. 45, No. 5, July, pp. 1293–98.

[3] The F-test is used for variance equality tests with two subgroups ($G = 2$). We compute the variance for each subgroup and denote the subgroup with the larger variance as L and the subgroup with the smaller variance as S. Then the F-statistic is given by $F = s_L^2 / s_S^2$ where s_g^2 is the variance in subgroup g=L, S. This F-statistic has an F-distribution with $n_L - 1$ numerator degrees of freedom and $n_S - 1$ denominator degrees of freedom under the null hypothesis of equal variance and independent normal samples.

[4] Without independence, this test fails.

and the Wald statistic,

$$W = (\hat{\theta}_1 - \hat{\theta}_2)' \ (V_1 + V_2)^{-1} (\hat{\theta}_1 - \hat{\theta}_2),$$

has a chi-squared distribution with K degrees of freedom. A test that the difference between the parameters is zero can be based on this statistic. It is straightforward to apply this to our test of common parameter vectors in our regressions. Large values of the statistic lead us to reject the hypothesis of no difference (or of stability in the coefficients). Note that we base such a test on estimates of V_1 and V_2. The test is valid in large samples, so we may use our least squares estimates of the two covariance matrices to compute W.[5]

As shown in the attached tables, the F-test results indicate that the null hypothesis (that the error variance from the estimated Bretton Woods formula is equal over rolling pairs of reviews, from the Sixth through the Eleventh Reviews) is rejected for the Sixth and Seventh, Seventh and Eighth, Ninth and Tenth, and Tenth and Eleventh Reviews (Table III.2.1). Therefore, Chow tests are not appropriate for testing the stability of coefficients of the Bretton Woods formula over these review periods. Nonetheless, application of the Wald test suggests that the coefficient estimates of the Bretton Woods formula are not stable over these reviews (Wald test values exceed the critical x^2 value (4 restrictions) of 9.49, at the 5 percent significance level, for all such pairs of reviews—Table III.2.2). However, the pair-wise comparison of the Eighth and Ninth Reviews suggests that the corresponding error variances are equal, and therefore the Chow test results are valid (Table III.2.3).

[5] See Greene, William, H., 1993, Econometric Analysis, 2d edition, Prentice Hall, Englewood Cliffs, N. J., pp. 215–6.

Table III.2.1 : Test for Equality of Variances of Error Terms (F-test)

	F-Statistic
Sixth and Seventh Reviews	3.85 *
Seventh and Eighth Reviews	1.58 *
Eighth and Ninth Reviews	1.16 *
Ninth and Tenth Reviews	1.59 *
Tenth and Eleventh Reviews	2.77 *

Note: An asterisk indicates significance at the 5% level. The critical values of the F-statistic (120.120) are 1.53, 1.35, and 1.26 at the 1% level, 5%, and 10% significance levels, respectively.

Table III.2.2 : Wald Test

	Chi-square Statistic
Sixth and Seventh Reviews	97.73 *
Seventh and Eighth Reviews	20.35 *
Ninth and Tenth Reviews	216.02 *
Tenth and Eleventh Reviews	59.26 *

Note: An asterisk indicates significance at the 5% level. The critical values of the Chi-square statistic for 4 restrictions are 13.28, 9.49, and 7.78 at the 1%, 5%, and 10% significance levels, respectively.

Table III.2.3 : Chow Tests

	F-Statistic
Eighth and Ninth Reviews	20.65 *

Note: An asterisk indicates significance at the 5% level. The critical values of the F-statistic (5.200) are 3.11, 2.66, and 1.88 for the 1%, 5%, and 10% significance levels, respectively.