Appendix N: Diebold Mariano tests

In this appendix, we report on DBP's analysis of predictability in respect of different methods of forming estimates of the appropriate risk-free rate during an access period. We make use of exactly the same dataset (data from Bloomberg on the 10-year CGS from July 1979 to February 2008) as the ERA uses in Appendix 5 of the Guidelines, and we use exactly the same form of the Dickie-Fuller tests in the same R-software package that the ERA used, as can be seen by the fact that we obtain the same critical values in the tables below. We have not, however, been able to replicate the ERA's results. We have had one meeting with the ERA to discuss its approach (May 1st) which clarified some matters, but our four follow-up emails from this meeting have yet to result in clarification of exactly what it is that we are doing differently to the ERA; although it is apparent that there is some difference in the way raw bond data are converted into averages before the error vectors are formed. It would be a relatively simple matter to compare spreadsheets, and to that end, in the interests of transparency, we sent the ERA all of our work on September 5th. We look forward to co-operating with the ERA to finalise this issue going forward.

The results of our analysis are presented in Tables 1 and 2 below. We do not present results for the Diebold Mariano test itself, but rather focus on the characteristics of the forecast errors and the DM test statistic. In particular, if the forecast errors are non-stationary, then the relevant model is essentially a poor model which will not be able to produce good forecasts because the forecasts will differ depending upon when in the time series they are made. There would appear to be little point in comparing two models that are poor to see which is worse than the other.

Perhaps more importantly, if what the ERA terms the "loss differential" (Guidelines Appendix 5 paragraph 24) is non-stationary, then the conditions implementing the Diebold-Mariano test are not meet, and any inferences made in respect of the test are invalid.

We examine two time periods; the full sample from July 1979 and a shortened sample from July 1984. The shorter sample is arguably more relevant because it begins just after the float of the Australian dollar and a change in the policy of the Federal Government whereby the price of its bonds began to be set in the marketplace. In practical terms, bond prices are much more stable before 1984 than after 1984. We form our averages by making use of the "EDATE" function in Excel, which allows us to ensure that, even if the numbers of trading days differ from year to year, the one year average from July 25th 1987 will go to July 24th 1988 and not some other date. We understand from discussions with the ERA that it simply added one year's worth of trading days to a given date to form averages, but experimentation with this approach shows only a small difference in results. Bloomberg do not have data for every trading day (we do not include weekends and public holidays), and where data are missing, we leave the relevant day blank. This gave us results closest to the ERA's answers.

We examine two different tests of stationarity; the Dickie-Fuller (1979) test employed by the ERA, and the Phillips-Peron (1988) test. Tests of stationarity are notoriously low in power, and it is common practice in econometrics to use more than one test (in our investigation of MRP, for example, we use three) to ascertain whether the test results agree with each other or not.

We examine the same two cases that the ERA investigates; its fixed model and its annual update model. We note that these are not directly comparable because the "actuals" against which predictions are compared are different. In the ERA's "fixed" model, an average formed over 20 days, 60 days, five years or ten years prior to the access period is compared with the average 10 year risk-free rate during the five years of the access period. In the ERA's "annual update" model, the relevant average before an access period is compared with an average of five similar updates during the access period. Thus, a 20-day average formed prior to the access period is compared with five annual updates of 20 day averages during the access period. As the ERA pointed out to us in discussions, the "fixed" case compares against actual market outcomes, whilst the "annual update" case compares forecasts made prior to the access period with regulatory outcomes during the access period.

The annual update case, as developed by the ERA is potentially useful as a test for regulated service providers of whether the expectation they have prior to an access period about the interest rates the will face during an access period which incorporates annual updates will be met or not. It is thus a test of the predictability of the regulatory regime from the perspective of the regulated firms themselves. However, it is not a test of whether the regulatory regime itself produces efficient forecasts; for that, one needs to test regulatory outcomes against actual outcomes, as the ERA does with its fixed model.

We examine a proper "apples with apples" comparison by comparing the 40-day forecast made prior to the access period (ERA in Tables 1 and 2 below) with the actual five year average of the 10-year risk-free rate during the access period, and comparing a the average of five tranches of debt, each of which comprises ten 40-day averages from ten previous years, calculated at the outset and during the access period with the average interest rate during the access period (AER in Tables 1 and 2 below). This reflects what the ERA and AER (respectively) propose to do in respect of the risk-free rate in forthcoming access arrangements (we ignore the fact that the ERA proposes to use the five-year rate, as it has not done so in its Diebold Mariano tests, and the two rates are highly correlated); the ERA proposes to fix the risk-free rate and the AER proposes to form an average of ten-years' worth of past debt (in ten tranches each of a length up to a year; we choose 40 days as this is what the ERA used in its ATCO Draft Decision). We note that our approach does not cover the debt-risk premium, for which data are not available as a long time series.

The results of our assessment are contained in Tables 1 and 2 below. The cells highlighted orange show the cases where the relevant variable is stationary at the five percent level.

Table 1: Phillip-Perron unit root test results

	test stat 1979	test stat 1984	1% crit val	5% crit val	10% crit val
20-day fixed	-2.90	-2.62	-3.43	-2.86	-2.57
20-day annual update	-3.05	-2.80	-3.43	-2.86	-2.57
60-day fixed	-2.92	-2.29	-3.43	-2.86	-2.57
60-day annual update	-3.23	-2.58	-3.43	-2.86	-2.57
5-year fixed	-1.45	-0.17	-3.43	-2.86	-2.57
5-year annual update	-2.74	0.45	-3.43	-2.86	-2.57
10-year fixed	3.01	2.03	-3.43	-2.86	-2.57
10-year annual update	1.24	6.01	-3.43	-2.86	-2.57
AER	-0.15	-0.46	-3.43	-2.86	-2.57
ERA	-2.08	-2.08	-3.43	-2.86	-2.57
20-day minus 60 day fixed	-7.06	-6.43	-3.43	-2.86	-2.57
20-day minus five-year fixed	-2.03	-1.83	-3.43	-2.86	-2.57
20-day minus ten-year fixed	-1.16	-1.84	-3.43	-2.86	-2.57
20-day minus 60 day annual update	-7.48	-6.70	-3.43	-2.86	-2.57
20-day minus five-year annual update	-2.02	-1.83	-3.43	-2.86	-2.57
20-day minus ten-year annual update	-2.49	-3.58	-3.43	-2.86	-2.57
AER minus ERA	-1.17	-1.38	-3.43	-2.86	-2.57

Source: DBP analysis

Table 2: Dickie-Fuller unit root test results

	test stat 1979	test stat 1984	1% crit val	5% crit val	10% crit val
20-day fixed	-2.83	-2.51	-2.58	-1.95	-1.62
20-day annual update	-2.48	-2.11	-2.58	-1.95	-1.62
60-day fixed	-2.60	-2.24	-2.58	-1.95	-1.62
60-day annual update	-2.51	-1.87	-2.58	-1.95	-1.62
5-year fixed	-0.15	-0.24	-2.58	-1.95	-1.62
5-year annual update	-0.54	-0.26	-2.58	-1.95	-1.62
10-year fixed	-1.74	-2.53	-2.58	-1.95	-1.62
10-year annual update	-0.30	-3.68	-2.58	-1.95	-1.62
AER	-4.13	-4.77	-2.58	-1.95	-1.62
ERA	-2.18	-3.35	-2.58	-1.95	-1.62
20-day minus 60 day fixed	-10.13	-9.15	-2.58	-1.95	-1.62
20-day minus five-year fixed	-2.09	-1.80	-2.58	-1.95	-1.62
20-day minus ten-year fixed	-0.84	-0.97	-2.58	-1.95	-1.62
20-day minus 60 day annual update	-7.61	-7.10	-2.58	-1.95	-1.62
20-day minus five-year annual update	-1.92	-1.74	-2.58	-1.95	-1.62
20-day minus ten-year annual update	-2.02	-1.87	-2.58	-1.95	-1.62
AER minus ERA	-1.14	-1.26	-2.58	-1.95	-1.62

Source: DBP analysis

For the Dickie Fuller test results for 1979, our results do not differ greatly, in terms of the conclusions drawn, compared with the ERA; the actual test statistics differ, but we find that the error vectors are stationary in most cases (the ERA finds they are stationary in all cases) and we find that the DM-test statistic is usually stationary (the ERA finds that it is always stationary). However, there is less support for both these findings when one examines the results of the Phillips-Perron tests, and less still when one examines data from 1984.

Note that we are not computing the DM-test statistics in Tables 1 and 2 above, but rather examining the conditions under which these could be computed and valid inferences drawn. At best, this would be in the 20 versus 60 day cases, with somewhat less support concerning the question of whether one could even test the 20-day average against multiple-year averages.

We note that DBP has not proposed to use long averages, and indeed, we believe that the debate has moved on from this question post the Western Power decision. The most important question, given the two models for debt actually being proposed by the ERA and the AER in their respective Guidelines, if one is concerned about efficiency and predictions matching subsequent actual outcomes as best they can (which the ERA professes to be in its Guidelines), then the relevant comparison is between the ERA and AER models. The Dickie-Fuller test results find that each error vector examined in isolation is stationary, but that the DM-test statistic (AER-ERA in Tables 1 and 2 above) is not. This means that the conditions of the DM test are not met in respect of comparing these two models. The Phillips-Perron test indicates that neither the test statistic nor the error vectors are stationary, which suggests that neither model is likely to predict well.

The practical upshot of our investigations is that one cannot draw conclusions about whether the ERA or the AER's models are likely to form better predictions of rates likely to prevail during the access period.