Evaluating Forecasts: Response to the ERA's Draft Decision on Proposed Revisions to the Access Arrangement for the Dampier to Bunbury Natural Gas Pipeline 2016 - 2020

A Report for DBP

February 2016

Report Author

Simon Wheatley

Contact Us

Level 40, 161 Castlereagh Street
Sydney NSW 2000

Phone: +61 2 8880 4800

Disclaimer

This report is for the exclusive use of the HoustonKemp client named herein. There are no third party beneficiaries with respect to this report, and HoustonKemp does not accept any liability to any third party. Information furnished by others, upon which all or portions of this report are based, is believed to be reliable but has not been independently verified, unless otherwise expressly indicated. Public information and industry and statistical data are from sources we deem to be reliable; however, we make no representation as to the accuracy or completeness of such information. The opinions expressed in this report are valid only for the purpose stated herein and as of the date of this report. No obligations is assumed to revise this report to reflect changes, events or conditions, which occur subsequent to the date hereof. All decisions in connection with the implementation or use of advice or recommendations contained in this report are the sole responsibility of the client.

Contents

Executive Summary v

1. Introduction 1

2. *t* Tests and Wald Tests 3

2.1 *t* Test 4

2.2 Wald Test 5

2.3 Multiple Comparisons 6

2.4 Paired *t* Tests 6

2.5 Power 7

3. Cross-Validation 10

3.1 Analysis 10

3.2 Literature Search 13

A1. Terms of Reference 16

A2. Federal Court Guidelines 18

A3. Curriculum Vitae 20

Figures

Figure 1: Power of Tests that the Mean of a Series of Return Forecasts is Zero that Use Daily and Monthly Data, Computed Analytically 8

Tables

Table 1: Power of Tests that the Mean of a Series of Return Forecasts is Zero that Use Daily and Monthly Data, Computed by Simulation 9

Table 2: Articles in which the Phrase 'Cross-Validation' Appears for a Cross-Section of Journals...... 14

Table 3: Articles in which the Phrase 'K-Fold Cross-Validation' Appears for a Cross-Section of Journals........ 15

Table 4: Articles in which the Phrase 'Wald Test' Appears for a Cross-Section of Journals 15

Executive Summary

DBP has asked HoustonKemp to critically review the analysis of the Economic Regulation Authority (ERA) in its December 2015 draft decision on the Dampier to Bunbury Natural Gas Pipeline.[1] In particular, DBP has asked that HoustonKemp critically review the analysis of the ERA of DBP's tests for unbiasedness.

Regulators of firms must ensure that holders of shares in the firms expect to earn at least the returns that the market requires on other assets of similar risk if the firms are to attract capital. The returns that investors expect to earn on shares, however, are not something that one can easily observe. So, in practice, estimating the returns requires that one use either a capital asset pricing model or a method for extracting estimates of returns from market prices. One way of assessing a capital asset pricing model or a method for extracting estimates of the returns that investors expect to earn from market prices is to test whether the model or method can deliver forecasts of returns that are unbiased.

DBP in its December 2014 submission to the ERA uses *t* tests and Wald tests to assess whether out-of-sample forecasts generated by a number of capital asset pricing models are unbiased and the ERA, in its draft decision, makes a number of criticisms of DBP's use of these tests. DBP has asked HoustonKemp to assess the arguments that the ERA makes.

t Tests and Wald Tests

The ERA, in its December 2015 draft decision, criticises DBP's use of *t* tests and Wald tests. In particular, the ERA states that: [2]

> 'A proposed model adequacy test based on a Wald test or t-test is not explicitly referenced in the statistical literature. Typically t-tests and Wald tests, within the context of model fitting, are used to identify whether a hypothesized parameter value falls within the sampling distribution of a sample estimate of that parameter, thereby forming a goodness-of-fit test. Wald tests and t-tests are generally not applied in the statistical literature to a comparison of predictions and their equivalent out-of-sample observations.'

The ERA also states that: [3]

- 'the … test … is not de rigueur in the statistical literature for assessing model performance;'
- 'the (tests do) not evaluate prediction bias as claimed by DBP;'
- 'the testing of each portfolio through the use of a t-test will suffer from the multiple comparison problem;'
- 'a paired t-test is a uniformly more powerful test.'

Finally, the ERA states that: [4]

> 'Pseudo-replication arises when measures are taken at a finer temporal scale, but the change in support is not explicitly reflected in the model. By applying predictions at a finer time scale then a larger sample can be generated for the DBP t-test, thereby artificially increasing the power of the test with the increased sample size.'

[1] ERA, *Draft Decision on Proposed Revisions to the Access Arrangement for the Dampier to Bunbury Natural Gas Pipeline 2016 – 2020 Appendix 4 Rate of return*, December 2015.

[2] ERA, *Draft Decision on Proposed Revisions to the Access Arrangement for the Dampier to Bunbury Natural Gas Pipeline 2016 – 2020 Appendix 4 Rate of return*, December 2015, page 230.

[3] ERA, *Draft Decision on Proposed Revisions to the Access Arrangement for the Dampier to Bunbury Natural Gas Pipeline 2016 – 2020 Appendix 4 Rate of return*, December 2015, page 46.

[4] ERA, *Draft Decision on Proposed Revisions to the Access Arrangement for the Dampier to Bunbury Natural Gas Pipeline 2016 – 2020 Appendix 4 Rate of return*, December 2015, page 230.

We show that:

- *t*-tests can be used to test for predictive bias;
- *t*-tests have been used in the literature to test for predictive bias;
- the Wald test can be used to test whether a set of forecasts are simultaneously unbiased and so does not suffer from the multiple comparison problem;
- paired *t*-tests that pair forecast errors are of little use in evaluating whether a number of forecasts are unbiased; and
- the use of weekly returns rather than monthly returns or daily returns rather than weekly returns in an attempt to raise the number of observations that one employs will have virtually no impact on power.

We also note that:

- the square of a *t* statistic is really a Wald statistic and so *t*-tests and Wald tests are closely related.

Cross-Validation

The ERA argues in its December 2015 DBP draft decision that there are better tests than *t* tests and Wald tests that one can use in comparing series of forecasts. In particular, the ERA states that: [5]

> 'The Authority is of the view that a more appropriate framework for assessing prediction accuracy, and hence model adequacy, is to utilise the cross-validation measure of prediction error.'
>
> 'Cross-validation estimates expected prediction error. There are different cross-validation schemes which aim to estimate the out-of-sample prediction error purely from within-sample data. Other schemes for estimating out-of-sample prediction error include splitting the sample into discrete training and test sets (i.e., a holdout scheme).'
>
> 'A month-ahead, moving window forecast is a form of cross-validation, as applied in the DBP model adequacy test. Such a cross-validation scheme is not considered as efficient as K-fold schemes, in terms of the number of predictions they generate for the same sample of data. Step-ahead forecasting is designed to reduce the impact of non-stationary effects on estimates of the out-of-sample prediction error. However, there is little evidence (from Bergmeir and Benitez (2012)) that K-fold schemes perform less well for non-stationary time series than step-ahead forecasting. Importantly, deploying K-fold schemes largely voids the concerns expressed (sic) DBP, where a month-ahead scheme is recommended so as to generate sufficient out-of-sample data for the model adequacy test.'

Hyndman, the current editor of the International Journal of Forecasting, states on his web site that: [6]

> 'When the data are not independent cross-validation becomes more difficult as leaving out an observation does not remove all the associated information due to the correlations with other observations.'

In work that the ERA cites in the passage that we provide above, Bergmeir and Benitez (2012) examine whether the theoretical problems to which Hyndman alludes are sufficient, as a practical matter, to invalidate

[5] ERA, *Draft Decision on Proposed Revisions to the Access Arrangement for the Dampier to Bunbury Natural Gas Pipeline 2016 – 2020 Appendix 4 Rate of return*, December 2015, pages 46 and 232-233.

[6] http://robjhyndman.com/hyndsight/crossvalidation/

the use of K-fold and similar cross-validation procedures to evaluate forecasts of a single series and they conclude that: [7]

> 'Using standard 5-fold cross-validation, no practical effect of the dependencies within the data could be found.'
>
> 'Regarding time-evolving effects, no differences could be found … This is not surprising, as we limited the study to stationary time series.'

Thus, while it is clear that, contrary to the claim made by the ERA, Bergmeir and Benitez do not examine the impact of non-stationarity on cross-validation tests, Bergmeir and Benitez find that the theoretical problems to which Hyndman refers are not sufficient, in the series that they examine, to invalidate the use of K-fold and similar cross-validation procedures.

In more recent work, Bergmeir, Hyndman and Koo (2015) examine analytically whether K-fold cross-validation is valid for assessing forecasts of a single series when the data to be forecast follow an autoregressive model. [8] They conclude that: [9]

> 'In this work we have investigated the use of cross-validation procedures for time series prediction evaluation when purely autoregressive models are used, which is a very common use-case when using Machine Learning procedures for time series forecasting. In a theoretical proof, we showed that a normal K-fold cross-validation procedure can be used if the lag structure of the models is adequately specified. In the experiments, we showed empirically that even if the lag structure is not correct, as long as the data are fitted well by the model, cross-validation without any modification is a better choice than OOS evaluation. Only if the models are heavily misspecified, are the cross-validation procedures to be avoided as in such a case they may yield a systematic underestimation of the error.'

Like Bergmeir and Benitez (2012), Bergmeir, Hyndman and Koo limit their attention to stationary series and find that the problems to which Hyndman refers are not sufficient to invalidate the use of K-fold cross-validation – at least when the single series to be forecast follows an autoregressive process. [10]

While the results of Bergmeir and Benitez (2012) and Bergmeir, Hyndman and Koo (2015) are interesting, it is not clear that DBP can simply rely on the results to implement K-fold cross-validation tests of the pricing models that it considers. [11] The focus of DBP's work is on comparing the forecasts that a number of pricing models generate. We conclude that to show that K-fold cross-validation is an appropriate tool with which to compare the forecasts would require one address issues that Bergmeir and Benitez and Bergmeir, Hyndman and Koo do not consider.

DBP also requested that we conduct a literature search to gauge the extent to which cross-validation is used by finance academics. If cross-validation procedures offer important advantages over alternative methods of evaluating forecasts, there are few disadvantages to using the procedures and the market for academic research is efficient, then one should expect to find evidence of the frequent use of the procedures in published work.

---

[7] Bergmeir, C. and J. Benitez, On the use of cross-validation for time series predictor evaluation, Information Sciences, 2012, pages 192-213.

[8] Bergmeir, C., R. Hyndman and B. Koo, *A note on the validity of cross-validation for evaluating time series prediction*, Monash University, 2015.

[9] Bergmeir, C., R. Hyndman and B. Koo, *A note on the validity of cross-validation for evaluating time series prediction*, Monash University, 2015, page 15.

[10] Bergmeir, C. and J. Benitez, On the use of cross-validation for time series predictor evaluation, Information Sciences, 2012, pages 192-213.

[11] Bergmeir, C. and J. Benitez, On the use of cross-validation for time series predictor evaluation, Information Sciences, 2012, pages 192-213.

Bergmeir, C., R. Hyndman and B. Koo, *A note on the validity of cross-validation for evaluating time series prediction*, Monash University, 2015.

The evidence that we provide indicates that K-fold cross-validation is used infrequently in high quality finance research. We find only one publication in a top-four finance journal that uses K-fold cross-validation.

While cross-validation is used and Wald tests are used frequently by finance academics, we do not know for what purposes they are used. Nevertheless, the frequent use of Wald tests suggests that a Wald test is a natural tool for determining whether the evidence indicates that an array of forecasts are simultaneously unbiased.

1. Introduction

DBP has asked HoustonKemp to critically review the analysis of the Economic Regulation Authority (ERA) in its December 2015 draft decision on the Dampier to Bunbury Natural Gas Pipeline.[12] In particular, DBP has asked that HoustonKemp critically review the analysis of the ERA of DBP's tests for unbiasedness.

DBP in its December 2014 submission to the ERA uses *t* tests and Wald tests to assess whether out-of-sample forecasts generated by a number of capital asset pricing models are unbiased. The ERA, in its December 2015 draft decision, makes a number of criticisms of DBP's use of these tests. In particular, the ERA argues that: [13]

(a) *t* tests and Wald tests are generally not applied to assess whether forecasts are unbiased;

(b) the use of *t* tests will give rise to a multiple comparison problem;

(c) paired *t* tests are uniformly more powerful;

(d) the power of *t* tests can be increased arbitrarily by using more frequently sampled data; and

(e) K-fold cross-validation is a more appropriate method for assessing whether forecasts are unbiased.

DBP has asked HoustonKemp to assess these arguments. The rest of the report is organised as follows:

- section 2 examines the use of *t* statistics and Wald statistics to test whether forecasts are unbiased; and
- section 3 examines the use cross-validation to assess whether forecasts are unbiased.

In addition:

- Appendix A1 provides the terms of reference for this report;
- Appendix A2 provides a copy of the Federal Court of Australia's *Guidelines for Expert Witnesses in Proceeding in the Federal Court of Australia*; and
- Appendix A3 provides the curriculum vitae of the author of the report.

Statement of Credentials

This report has been prepared by **Simon Wheatley**.

Simon Wheatley is a Special Adviser to HoustonKemp and was until 2008 a Professor of Finance at the University of Melbourne. Since 2008, Simon has applied his finance expertise in investment management and consulting outside the university sector. Simon's interests and expertise are in individual portfolio choice theory, testing asset-pricing models and determining the extent to which returns are predictable. Prior to joining the University of Melbourne, Simon taught finance at the Universities of British Columbia, Chicago, New South Wales, Rochester and Washington.

In preparing this report, the author (herein after referred to as 'I' or 'my' or 'me') confirms that I have made all the inquiries that I believe are desirable and appropriate and that no matters of significance that I regard as relevant have, to my knowledge, been withheld from this report. I acknowledge that I have read, understood

[12] ERA, *Draft Decision on Proposed Revisions to the Access Arrangement for the Dampier to Bunbury Natural Gas Pipeline 2016 – 2020 Appendix 4 Rate of return*, December 2015.

[13] ERA, *Draft Decision on Proposed Revisions to the Access Arrangement for the Dampier to Bunbury Natural Gas Pipeline 2016 – 2020 Appendix 4 Rate of return*, December 2015, pages 46 and 230.

and complied with the Federal Court of Australia's *Practice Note CM 7, Expert Witnesses in Proceedings in the Federal Court of Australia*. I have been provided with a copy of the Federal Court of Australia's *Practice Note CM 7, Expert Witnesses in Proceedings in the Federal Court of Australia*, dated 4 June 2013, and my report has been prepared in accordance with those guidelines.

I have undertaken consultancy assignments for DBP in the past. However, I remain at arm's length, and as an independent consultant. Within the last three years, besides undertaking consultancy assignments for DBP, I have undertaken assignments for ActewAGL Distribution, Ausgrid, AusNet Services, Australian Gas Networks, APA, CitiPower, Endeavour Energy, the Energy Networks Association, Energex, Ergon Energy, Essential Energy, Jemena Electricity Networks, Jemena Gas Networks, Powercor, SA Power Networks, Sydney Water and United Energy. Assignments for the Energy Networks Association and these other companies have provided 85 per cent of my gross income.

2. *t* Tests and Wald Tests

Regulators of firms must ensure that holders of shares in the firms expect to earn at least the returns that the market requires on other assets of similar risk if the firms are to attract capital. The returns that investors expect to earn on shares, however, are not something that one can easily observe. So, in practice, estimating the returns requires one use either a capital asset pricing model or a method for extracting estimates of returns from market prices. Clearly, it is important that any model or method be subject to an empirical assessment. As Nobel Prize-winner Granger and his co-author Newbold state:[14]

> 'It is important that economic forecasts be critically evaluated. An evaluation exercise, as well as providing information about the relative worth of a set of forecasts, may well suggest directions in which the forecast-generating mechanism can be improved.'

One way of assessing a capital asset pricing model or a method for extracting estimates of the returns that investors expect to earn from market prices is to test whether the model or method can deliver forecasts of returns that are unbiased. If a regulator restricts the holders of shares in the firms that it regulates to earn a return that is on average less than the return that investors expect to earn on other assets of similar risk, then the firms will be unable to attract capital. Thus if a regulator uses a model or method to estimate the return that investors expect to earn on shares that is downwardly biased, then the firms that it regulates will be unable to attract capital.

DBP in its December 2014 submission to the ERA uses *t* tests and Wald tests to assess whether out-of-sample forecasts generated by a number of capital asset pricing models are unbiased. The ERA, in its December 2015 draft decision, makes a number of criticisms of DBP's use of these tests. In particular, the ERA states that: [15]

> 'A proposed model adequacy test based on a Wald test or t-test is not explicitly referenced in the statistical literature. Typically t-tests and Wald tests, within the context of model fitting, are used to identify whether a hypothesized parameter value falls within the sampling distribution of a sample estimate of that parameter, thereby forming a goodness-of-fit test. Wald tests and t-tests are generally not applied in the statistical literature to a comparison of predictions and their equivalent out-of-sample observations.'

The ERA also states that: [16]

- 'the … test … is not de rigueur in the statistical literature for assessing model performance;'
- 'the (tests do) not evaluate prediction bias as claimed by DBP;'
- 'the testing of each portfolio through the use of a t-test will suffer from the multiple comparison problem;'
- 'a paired t-test is a uniformly more powerful test.'

Finally, the ERA states that: [17]

> 'Pseudo-replication arises when measures are taken at a finer temporal scale, but the change in support is not explicitly reflected in the model. By applying predictions at a finer time scale then a larger sample can be generated for the DBP t-test, thereby artificially increasing the power of the test with the increased sample size.'

---

[14] Granger, C. and P. Newbold, Forecasting economic time series, Academic Press, 1977, page 269.

[15] ERA, *Draft Decision on Proposed Revisions to the Access Arrangement for the Dampier to Bunbury Natural Gas Pipeline 2016 – 2020 Appendix 4 Rate of return*, December 2015, page 230.

[16] ERA, *Draft Decision on Proposed Revisions to the Access Arrangement for the Dampier to Bunbury Natural Gas Pipeline 2016 – 2020 Appendix 4 Rate of return*, December 2015, page 46.

[17] ERA, *Draft Decision on Proposed Revisions to the Access Arrangement for the Dampier to Bunbury Natural Gas Pipeline 2016 – 2020 Appendix 4 Rate of return*, December 2015, page 230.

We show that:

- *t*-tests can be used to test for predictive bias;
- *t*-tests have be used in the literature to test for predictive bias;
- the Wald test can be used to test whether a set of forecasts are simultaneously unbiased and so does not suffer from the multiple comparison problem;
- paired *t*-tests that pair forecast errors are of little use in evaluating whether a number of forecasts are unbiased; and
- the use of weekly returns rather than monthly returns or daily returns rather than weekly returns in an attempt to raise the number of observations that one employs will have virtually no impact on power.

We also note that:

- the square of a *t* statistic is really a Wald statistic and so *t*-tests and Wald tests are closely related.

2.1 *t* Test

Consider a time series of forecast errors $x_1, x_2, ..., x_T$ and let the sample mean of the series be $\bar{x}$ and the sample standard deviation:

$$s = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(x_t - \bar{x})^2} \tag{1}$$

If the forecast errors are normally and independently distributed with mean μ and standard deviation σ, then the *t* statistic:[18]

$$\frac{\bar{x}}{s/\sqrt{T}} \tag{2}$$

will be noncentral *t* distributed with $T-1$ degrees of freedom and noncentrality parameter $\mu/\left(\sigma/\sqrt{T}\right)$. So if the forecast errors are normally and independently distributed, the statistic can be used to test the hypothesis that the forecasts are unbiased, that is, that $\mu = 0$, by comparing the statistic to tabulated values of the central *t* distribution. This test is called a *t* test.

Under the null hypothesis that $\mu = 0$, the *t* statistic will, in large samples, be distributed, approximately, as a standard normal. In fact, under the null hypothesis that the mean forecast error is zero, the *t* statistic, in large samples, will be distributed as a standard normal even when the forecast errors are not normally distributed so long as they are independently and identically distributed through time.[19] So, in large samples, one can use a *t* statistic to test whether the mean of a series of forecast errors is zero even when the errors are not normally distributed. A test of the hypothesis that the mean forecast error is zero that compares the *t* statistic to tabulated values of the standard normal distribution is called an asymptotic *t* test.[20]

[18] Davidson and McKinnon, *Estimation and inference in econometrics*, Oxford University Press, 1993, page 810.
Freund, J., *Mathematical statistics*, Prentice-Hall, 1972, pages 220-223.

[19] White, H., *Asymptotic theory for econometricians*, 1984, Academic Press, San Diego, CA, pages 107-131.

[20] Davidson and McKinnon, *Econometric theory and methods*, Oxford University Press, 1999, page 147.

Mincer and Zarnowitz (1969) use a *t* test to assess whether four series of forecasts are unbiased.[21] Mincer and Zarnowitz were two influential Polish-American economists. Mincer was a professor at Columbia University while Zarnowitz was a professor at the University of Chicago. The 1969 paper of Mincer and Zarnowitz is one of the most widely cited and influential papers on forecasting and DBP indicate that the ERA suggested it use one of the tests that Mincer and Zarnowitz propose.[22, 23] Besides Mincer and Zarnowitz, Holden and Peel (1989), Gavin and Mandal (2003), Mankiw, Reis and Wolfers (2003) and Weber (2010), among others, use *t* tests to judge whether series of forecasts are unbiased.[24] In using *t* tests, Mankiw, Reis and Wolfers employ standard errors that are heteroscedasticity and autocorrelation consistent.

2.2 Wald Test

If the forecast errors $x_1, x_2, ..., x_T$ are normally and independently distributed and $\mu = 0$, then the square of the *t* statistic:[25]

$$\left(\frac{s^2}{T}\right)^{-1} \bar{x}^2 \tag{3}$$

will be *F* distributed with one and $T-1$ degrees of freedom. If $\mu = 0$, then in large samples, the square of the *t* statistic will be chi-square distributed with one degree of freedom. In fact, under the null hypothesis that $\mu = 0$, the square of the *t* statistic, in large samples, will be chi-square distributed even when the forecast errors are not normally distributed so long as they are independently and identically distributed through time.[26] A test of the hypothesis that the mean forecast error is zero that compares the square of the *t* statistic to tabulated values of a chi-square with one degree of freedom is called a Wald test.[27] It follows that, in large samples, a two-tailed *t* test and a Wald test will reach approximately the same conclusion.

Mincer and Zarnowitz (1969) do not use a Wald test to assess whether the forecasts that they examine are unbiased because the sample size that they employ is small.[28] Were their sample size to have been larger, however, they could have used a Wald test and in doing so would have drawn the same inference from their sample as the inference that they would have drawn in using a two-sided *t* test.

[21] Mincer, J. and V. Zarnowitz, *The evaluation of economic forecasts*, 1969, in Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance, NBER, page 13.

[22] Google scholar indicates that the paper has been cited 790 times.

[23] DBP, *Proposed Revisions DBNGP Access Arrangement 2016-2020 Regulatory Period Rate of Return Supporting Submission: 12*, December 2014, page 58.

[24] Gavin, W. and R. Mandal, *Evaluating FOMC forecasts*, International Journal of Forecasting, 2003, pages 655-667.

Holden, K. and D. Peel, *Unbiasedness, efficiency and the combination of economic forecasts*, Journal of Forecasting, 1989, pages 175-188.

Mankiw, N., R. Reis and J. Wolfers, Disagreement about inflation expectations, NBER Macroeconomics Annual 2003, pages 209-248.

Weber, A., *Heterogeneous expectations, learning and European inflation dynamics*, in Twenty Years of Inflation Targeting, Cambridge University Press, 2010, chapter 12.

[25] Davidson and McKinnon, *Estimation and inference in econometrics*, Oxford University Press, 1993, page 810.

[26] White, H., *Asymptotic theory for econometricians*, 1984, Academic Press, San Diego, CA, pages 107-131.

[27] Davidson and McKinnon, *Econometric theory and methods*, Oxford University Press, 1999, page 416.

[28] Mincer, J. and V. Zarnowitz, *The evaluation of economic forecasts*, 1969, in Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance, NBER, page 13.

Wald tests can also be used to test whether more than one set of forecasts is unbiased. Consider a time series of forecast errors $X_1, X_2, ..., X_T$, where X_t is an $N \times 1$ vector of forecast errors, and let the sample mean of the series be $\bar{X}$ and the sample covariance matrix:

$$S = \frac{1}{T-1} \sum_{t=1}^{T} \left(X_t - \bar{X} \right) \left(X_t - \bar{X} \right)' \tag{4}$$

In large samples and so long as the forecast errors are independently and identically distributed through time, the Wald statistic: [29]

$$\bar{X}' \left(\frac{1}{T} S \right)^{-1} \bar{X} \tag{5}$$

will be chi-square distributed with N degrees of freedom under the null hypothesis that the set of forecasts is unbiased. This will also be true if S is replaced in (5) by:

$$\hat{\Omega} = \frac{1}{T} \sum_{t=1}^{T} \left(X_t - \bar{X} \right) \left(X_t - \bar{X} \right)' \tag{6}$$

because S and $\hat{\Omega}$ will differ little in large samples

2.3 Multiple Comparisons

When one tests a number of different hypotheses the probability that one will reject one or more of the hypotheses will rise as the number of hypotheses rises. For example, the probability that in 10 independent tests one will reject the null hypothesis at least once when the null hypothesis is true and the size of each test is 0.05 will be:

$$1 - (1 - 0.05)^{10} = 0.40 > 0.05 \tag{7}$$

This problem is sometimes referred to as the problem of making multiple comparisons. When the 10 tests are not independent of one another, the probability that one will reject the null hypothesis at least once can be even higher.

A Wald test of the joint hypothesis that each of 10 null hypotheses is simultaneously true, on the other hand, will not suffer from this problem. This is because the test examines simultaneously whether each of the 10 hypotheses are true rather than considering the hypotheses one at a time. In addition, the test takes into account that the 10 tests, were they to be conducted separately, might not be independent of one another.

2.4 Paired *t* Tests

Diebold and Mariano (1995) suggest that one use an asymptotic paired *t* test to assess the relative performance of two alternative sets of forecasts. [30] In their framework, performance is assessed using a loss function whose form is left unstated. DBP's focus is simply on whether a forecast can be shown to be unbiased. Again, if a regulator uses a model or method to estimate the return that investors expect to earn on shares that is downwardly biased, then the firms that it regulates will be unable to attract capital. It is not

[29] Davidson and McKinnon, *Estimation and inference in econometrics*, Oxford University Press, 1993, page 810.

[30] Diebold, F. and R. Mariano, *Comparative predictive accuracy*, Journal of Business and Economic Statistics, 1995, pages 253-263.

clear that the framework of Diebold and Mariano and a paired *t* test are well suited to the problem of determining which of two sets of forecasts is closer to being unbiased.

Note that the difference between the forecast errors associated with two forecasts, f_t and g_t, of a return, r_t, will be given by:

$$(r_t - f_t) - (r_t - g_t) = g_t - f_t \tag{8}$$

Thus a paired *t* test that uses the difference between the forecast errors associated with two forecasts will not be able to judge which of the two forecasts is closer to being unbiased. The test will only be able to judge whether the two forecasts differ on average through time. If the two forecasts were to differ on average through time, then the most that one would be able to say is that both forecasts could not be unbiased. Either one or both forecasts would have to be biased. If the two forecasts were not to differ on average through time, then the most that one would be able to say is that one forecast could not be biased while the other was unbiased. Either both forecasts would have to be biased or both forecasts would have to be unbiased. It is difficult to see that conclusions like these would significantly advance our understanding of how a regulator should go about estimating the return on equity and so we do not recommend that a paired *t* test be employed.

2.5 Power

The ERA argues that the use of weekly returns rather than monthly returns or daily returns rather than weekly returns will raise the number of observations that one employs and so raise the power of tests for unbiasedness. While the use of weekly returns rather than monthly returns or daily returns rather than weekly returns will raise the number of observations that one employs, there will be virtually no impact on the power of tests for unbiasedness.

Suppose that we are interested in predicting the continuously compounded return to a portfolio and suppose that the annual continuously compounded return to the portfolio is normally and independently distributed through time with mean 10 per cent and standard deviation 20 per cent. Then the monthly return will be distributed with mean 10/12 per cent and standard deviation $20/\sqrt{12}$ per cent. The daily return will be distributed with mean $10/(21 \times 12)$ per cent and standard deviation $20/\sqrt{21 \times 12}$ per cent, where we assume that there are 21 trading days in each month. Suppose that a forecast of the annual return is five per cent, of the monthly return 5/12 per cent and of the daily return $5/(21 \times 12)$ per cent.

The *t* statistic for a test of the unbiasedness hypothesis that uses T months of monthly data will be noncentral *t* distributed with $T-1$ degrees of freedom and noncentrality parameter:

$$\frac{(10-5)/12}{\left(20/\sqrt{12}\right)/\sqrt{T}} = 0.0722 \times \sqrt{T} \tag{9}$$

The *t* statistic for a test of the unbiasedness hypothesis that uses T months of daily data will be noncentral *t* distributed with $21T-1$ degrees of freedom and a noncentrality parameter that is also:

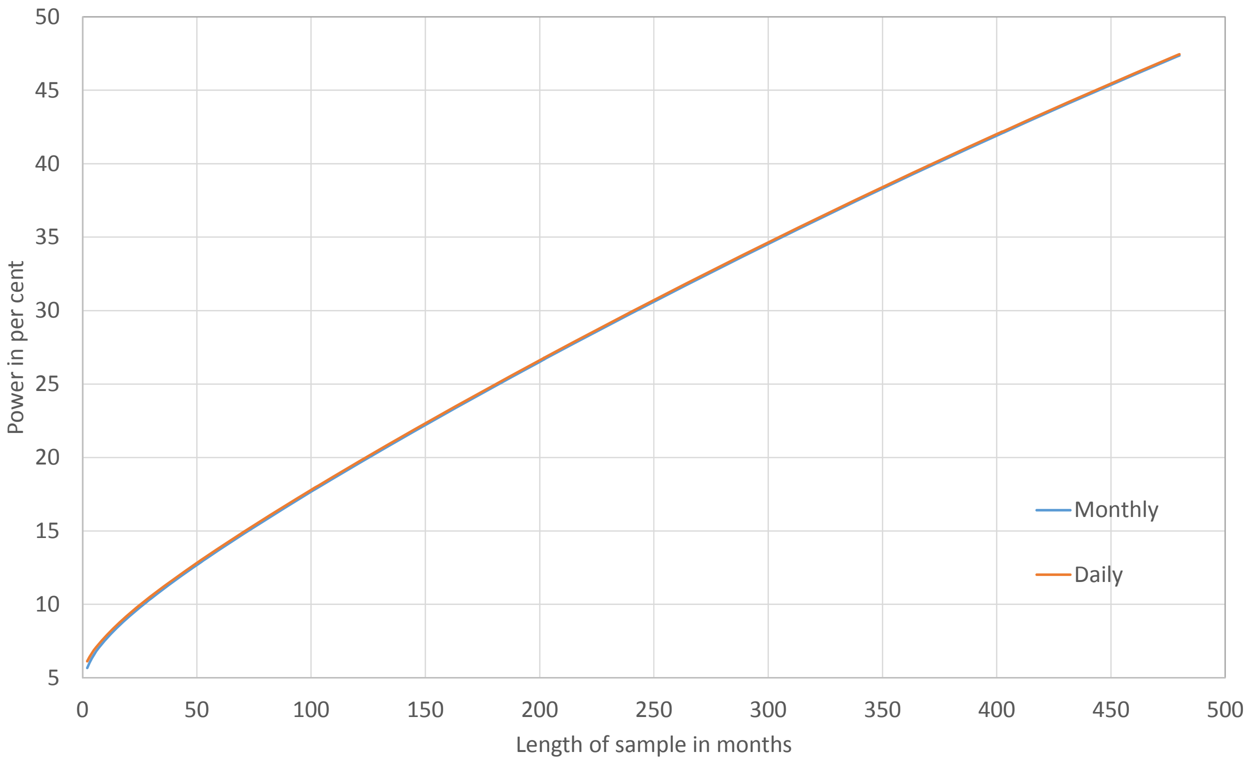
$$\frac{(10-5)/(21 \times 12)}{\left(20/\sqrt{21 \times 12}\right)/\sqrt{21 \times T}} = 0.0722 \times \sqrt{T} \tag{10}$$

Thus any difference in power will be solely due a difference between the larger number of degrees of freedom attached to the test that uses daily data. The impact of this larger number of degrees of freedom will be very small and will decline as the sample size rises.

Figure 1 plots the power of the tests that use monthly and daily data for a variety of sample lengths. The figure shows that any differences between the powers of the two tests are vanishingly small. Figure 1 presumes, for analytical tractability, that the returns being forecast are continuously compounded. We also conduct simulations to show that there are vanishingly small gains to be had from using daily not continuously compounded returns rather than monthly not continuously compounded returns.

In these simulations, we assume that the daily not continuously compounded return to a portfolio is normally and independently distributed through time with mean $10/(21\times 12)$ per cent and standard deviation $20/\sqrt{21\times 12}$ per cent, we generate series of daily not continuously compounded returns with these characteristics and from these series we produce series of monthly not continuously compounded returns. We presume that a forecast of the daily return is $5/(21\times 12)$ per cent and of the monthly return is $100\times((1+0.05/(21\times 12))^{21}-1)$ per cent.

Figure 1: Power of Tests that the Mean of a Series of Return Forecasts is Zero that Use Daily and Monthly Data, Computed Analytically



Notes: The figure assumes that the daily continuously compounded returns to a portfolio are normally distributed with mean $10/(21\times 12)$ per cent and standard deviation $20/\sqrt{21\times 12}$ per cent and are independently distributed through time, that a forecast of the daily return is $5/(21\times 12)$ per cent, a forecast of the monthly return is $5/12$ per cent and that t tests are used to judge whether the forecasts are unbiased.

Table 1 provides the results of the simulations and shows, like Figure 1, that any gains from using daily rather than monthly returns are minute. The intuition that the ERA has missed is as follows. While in moving from monthly to daily returns more data are produced and the volatility of each forecast error will decline, the mean of each forecast error will also decline. The impact of these facts – under a reasonable set of assumptions – is that there will be little gain to be had from using daily rather than monthly data.

Table 1: Power of Tests that the Mean of a Series of Return Forecasts is Zero that Use Daily and Monthly Data, Computed by Simulation

	Power in per cent	
Length of sample in months	Daily data	Monthly data
60	13.75	13.35
120	20.07	19.75
180	24.91	24.71
240	29.02	28.84
300	35.00	34.82
360	39.01	38.88
420	43.25	43.14
480	47.63	47.50

Notes: The figure assumes that the daily not continuously compounded returns to a portfolio are normally distributed with mean $10/(21\times 12)$ *per cent and standard deviation* $20/\sqrt{21\times 12}$ *per cent and are independently distributed through time, that a forecast of the daily return is* $5/(21\times 12)$ *per cent, that a forecast of the monthly return is* $100\times((1+0.05/(21\times 12))^{21}-1)$ *per cent and that t tests are used to judge whether the forecasts are unbiased.*

3. Cross-Validation

The ERA argues in its December 2015 DBP draft decision that there are better tests than *t* tests and Wald tests that one can use in comparing series of forecasts. In particular, the ERA states that: [31]

> 'The Authority is of the view that a more appropriate framework for assessing prediction accuracy, and hence model adequacy, is to utilise the cross-validation measure of prediction error.'
>
> 'Cross-validation estimates expected prediction error. There are different cross-validation schemes which aim to estimate the out-of-sample prediction error purely from within-sample data. Other schemes for estimating out-of-sample prediction error include splitting the sample into discrete training and test sets (i.e., a holdout scheme).'
>
> 'A month-ahead, moving window forecast is a form of cross-validation, as applied in the DBP model adequacy test. Such a cross-validation scheme is not considered as efficient as K-fold schemes, in terms of the number of predictions they generate for the same sample of data. Step-ahead forecasting is designed to reduce the impact of non-stationary effects on estimates of the out-of-sample prediction error. However, there is little evidence (from Bergmeir and Benitez (2012)) that K-fold schemes perform less well for non-stationary time series than step-ahead forecasting. Importantly, deploying K-fold schemes largely voids the concerns expressed (sic) DBP, where a month-ahead scheme is recommended so as to generate sufficient out-of-sample data for the model adequacy test.'

Cross-validation is a model evaluation method that is designed to judge how well a model will predict in data that the model has not already seen. [32] Typically, the data are split into a training set and a testing set. The parameters of the model are estimated using the training set and the performance of the model using these estimates is evaluated in the testing set.

The holdout method is perhaps the simplest kind of cross-validation. The data are split into a training set and a testing set and forecasts are only ever generated by the training set, which remains fixed.

The time series method allows the training set to expand as each successive forecast is made. Thus DBP use time series cross-validation.

K-fold cross-validation is designed to improve upon the holdout method. The data are split into K subsets and the holdout method is repeated K times. Each time, K-1 of the subsets are used as a training set and the remaining set is used as a testing set. When K is set equal to the number of observations, K-fold cross-validation is labelled leave-one-out cross-validation.

3.1 Analysis

Hyndman, the current editor of the International Journal of Forecasting, states on his web site that: [33]

> 'When the data are not independent cross-validation becomes more difficult as leaving out an observation does not remove all the associated information due to the correlations with other observations. For time series forecasting, a cross-validation statistic is obtained as follows

[31] ERA, *Draft Decision on Proposed Revisions to the Access Arrangement for the Dampier to Bunbury Natural Gas Pipeline 2016 – 2020 Appendix 4 Rate of return*, December 2015, pages 46 and 232-233.

[32] The discussion here uses arguments made at:

https://www.cs.cmu.edu/~schneide/tut5/node42.html

http://robjhyndman.com/hyndsight/tscvexample/

[33] http://robjhyndman.com/hyndsight/crossvalidation/

1. Fit the model to the data $y_1,...,y_t$ and let $\hat{y}_{t+1}$ denote the forecast of the next observation. Then compute the error $(e^*_{t+1} = y_{t+1} - \hat{y}_{t+1})$ for the forecast observation.
2. Repeat step 1 for $t = m,...,n-1$ where m is the minimum number of observations needed for fitting the model.
3. Compute the MSE from $e^*_{m+1},...,e^*_n$.'

In work that the ERA cites in the passage that we provide from its draft decision, Bergmeir and Benitez (2012) examine whether the theoretical problems to which Hyndman alludes are sufficient, as a practical matter, to invalidate the use of K-fold and similar cross-validation procedures to evaluate forecasts of a single series and they conclude that: [34]

> 'Using standard 5-fold cross-validation, no practical effect of the dependencies within the data could be found.'
>
> 'Regarding time-evolving effects, no differences could be found … This is not surprising, as we limited the study to stationary time series.'

Thus, while it is clear that, contrary to the claim made by the ERA, Bergmeir and Benitez do not examine the impact of non-stationarity on cross-validation tests, Bergmeir and Benitez find that the theoretical problems to which Hyndman refers are not sufficient, in the series that they examine, to invalidate the use of K-fold and similar cross-validation procedures.

In more recent work, Bergmeir, Hyndman and Koo (2015) examine analytically whether K-fold cross-validation is valid for assessing forecasts of a single series when the data to be forecast follow an autoregressive model. [35] They conclude that: [36]

> 'In this work we have investigated the use of cross-validation procedures for time series prediction evaluation when purely autoregressive models are used, which is a very common use-case when using Machine Learning procedures for time series forecasting. In a theoretical proof, we showed that a normal K-fold cross-validation procedure can be used if the lag structure of the models is adequately specified. In the experiments, we showed empirically that even if the lag structure is not correct, as long as the data are fitted well by the model, cross-validation without any modification is a better choice than OOS evaluation. Only if the models are heavily misspecified, are the cross-validation procedures to be avoided as in such a case they may yield a systematic underestimation of the error.'

Like Bergmeir and Benitez (2012), Bergmeir, Hyndman and Koo limit their attention to stationary series and find that the problems to which Hyndman refers are not sufficient to invalidate the use of K-fold cross-validation – at least when the single series to be forecast follows an autoregressive process. [37]

While the results of Bergmeir and Benitez (2012) and Bergmeir, Hyndman and Koo (2015) are interesting, it is not clear that DBP can simply rely on the results to implement K-fold cross-

[34] Bergmeir, C. and J. Benitez, On the use of cross-validation for time series predictor evaluation, Information Sciences, 2012, pages 192-213.

[35] Bergmeir, C., R. Hyndman and B. Koo, *A note on the validity of cross-validation for evaluating time series prediction*, Monash University, 2015.

[36] Bergmeir, C., R. Hyndman and B. Koo, *A note on the validity of cross-validation for evaluating time series prediction*, Monash University, 2015, page 15.

[37] Bergmeir, C. and J. Benitez, On the use of cross-validation for time series predictor evaluation, Information Sciences, 2012, pages 192-213.

validation tests of the pricing models that it considers.[38] It will be useful to consider a simple example.

The Sharpe-Lintner (SL) Capital Asset Pricing Model (CAPM), the model that the ERA employs to estimate the return on equity, imposes a zero intercept restriction on a regression of the return to a portfolio in excess of the risk-free rate on the excess return to the market portfolio. That is, the SL CAPM states that a regression of the excess return to a portfolio on the excess return to the market portfolio can be written:

$$z_{jt} = \beta_j z_{mt} + \varsigma_{jt} \tag{11}$$

where:

z_{jt}	=	the return on portfolio j in excess of the risk-free rate;
β_j	=	the beta of portfolio j;
z_{mt}	=	the return to the market portfolio of risky assets in excess of the risk-free rate; and
ς_{jt}	=	a regression disturbance.

The model DBP uses for estimating the mean of z_{mt} is:

$$z_{mt} = \mu_m + \eta_{mt} \tag{12}$$

where:

μ_m	=	the market risk premium; and
η_{mt}	=	a disturbance that is uncorrelated with ς_{jt}.

Combining (11) and (12) yields:

$$\begin{pmatrix} z_{mt} \\ z_{jt} \end{pmatrix} = \begin{pmatrix} 1 \\ \beta \end{pmatrix} \mu_m + \begin{pmatrix} \eta_{mt} \\ \beta_j \eta_{mt} + \varsigma_{jt} \end{pmatrix} \tag{13}$$

or using the notation of Bergmeir, Hyndman and Koo (2015):[39]

$$y_t = \theta + \varepsilon_t \tag{14}$$

where:

$$y_t = \begin{pmatrix} z_{mt} \\ z_{jt} \end{pmatrix}, \qquad \theta = \begin{pmatrix} 1 \\ \beta \end{pmatrix} \mu_m \qquad \text{and} \qquad \varepsilon_t = \begin{pmatrix} \eta_{mt} \\ \beta_j \eta_{mt} + \varsigma_{jt} \end{pmatrix}$$

Let the covariance matrix of ε_t be given by Ω. In (14), y_t is a vector rather than, as in Bergmeir and Benitez (2012) and Bergmeir, Hyndman and Koo, a scalar, and θ depends on Ω.[40] Bergmeir and Benitez and

[38] Bergmeir, C. and J. Benitez, On the use of cross-validation for time series predictor evaluation, Information Sciences, 2012, pages 192-213.

Bergmeir, C., R. Hyndman and B. Koo, *A note on the validity of cross-validation for evaluating time series prediction*, Monash University, 2015.

[39] Bergmeir, C., R. Hyndman and B. Koo, *A note on the validity of cross-validation for evaluating time series prediction*, Monash University, 2015.

[40] Bergmeir, C. and J. Benitez, On the use of cross-validation for time series predictor evaluation, Information Sciences, 2012, pages 192-213.

Bergmeir, Hyndman and Koo do not consider restrictions placed that link the elements of θ to the elements of Ω. DBP use time series of different lengths to estimate β and μ_m and Bergmeir and Benitez and Bergmeir, Hyndman and Koo do not consider forecasting policies that use series of different lengths – in part because they do not consider forecasts of more than one variable at a time. Finally, Bergmeir and Benitez and Bergmeir, Hyndman and Koo do not consider cases where a realisation of one dependent variable is used to forecast the contemporaneous value of another dependent variable – again, in part, because they do not consider forecasts of more than one variable at a time.

It should be clear from this discussion that some additional analysis would be required to determine whether the results of Bergmeir and Benitez (2012) and Bergmeir, Hyndman and Koo (2015) can be applied to the problem of comparing the forecasts that a number of pricing models generate that is the focus of DBP's work. [41]

3.2 Literature Search

DBP has also asked us to conduct a literature search so as to gauge the extent to which cross-validation is used by finance academics.

If cross-validation procedures offer important advantages over alternative methods of evaluating forecasts, there are few disadvantages to using the procedures and the market for academic research is efficient, then one should expect to find evidence of the frequent use of the procedures in published work.

Here, we conduct keyword searches of the four major finance journals as a way of discovering how frequently cross-validation procedures are used in high quality research in finance.

The four journals that we select are the Journal of Finance, the Journal of Financial Economics, the Journal of Financial and Quantitative Analysis and the Review of Financial Studies. These are the four finance journals included in the list of 45 journals used by the Financial Times in compiling its business school research rankings. [42] They are also the four journals which a recent study of finance journal rankings that Currie and Pandher (2010) conduct rate most highly in terms of their quality. [43]

We also conduct keyword searches of the two major forecasting journals, the International Journal of Forecasting and the Journal of Forecasting.

We search for references to the phrases 'cross-validation' and 'K-fold cross-validation' and, for the sake of comparison, for references to the phrase 'Wald test'. The results appear in Table 2, Table 3 and Table 4 below. Across the Journal of Finance, the Journal of Financial Economics and the Review of Financial Studies – the journals that Currie and Pandher rate most highly in terms of quality – we find 56 references to the phrase 'cross-validation' and one reference to the phrase 'K-fold cross-validation' indicating that this form of cross-validation is rarely used by academics in finance. In the Journal of Financial and Quantitative Analysis we find no references to the two phrases.

We find, not surprisingly, that in the two forecasting journals there are more references to the use of cross-validation. Across the two journals we find 168 references to the phrase 'cross-validation' and 13 references to the phrase 'K-fold cross-validation' indicating that this form of cross-validation is used occasionally by academics whose interest is in forecasting.

By way of a comparison, we find that across the four finance journals there are 341 references to the phrase 'Wald test' indicating that the test is used frequently by academics in finance. Across the two forecasting

[41] Bergmeir, C. and J. Benitez, On the use of cross-validation for time series predictor evaluation, Information Sciences, 2012, pages 192-213.

Bergmeir, C., R. Hyndman and B. Koo, *A note on the validity of cross-validation for evaluating time series prediction*, Monash University, 2015.

[42] http://www.ft.com/intl/cms/s/2/3405a512-5cbb-11e1-8f1f-00144feabdc0.html#axzz2DLrFMmod

[43] Currie, R.R. and G. S. Pandher, *Finance journal rankings and tiers: An active scholar assessment methodology*, Journal of Banking and Finance, 2011, pages 7-20.

journals we find 70 references to the phrase 'Wald test' indicating that this test is also used by academics whose interest is in forecasting.

To summarise, the evidence that we provide indicates that K-fold cross-validation is used infrequently in high quality finance research. While cross-validation is used and Wald tests are used frequently by finance academics, we do not know for what purposes they are used. Nevertheless, the frequent use of Wald tests suggests that a Wald test is a natural tool for determining whether the evidence indicates that an array of forecasts are simultaneously unbiased.

Table 2: Articles in which the Phrase 'Cross-Validation' Appears for a Cross-Section of Journals

Journal	Number of cites	First	Last
Panel A: Finance journals			
Journal of Finance	31	1972	2015
Journal of Financial Economics	12	1988	2015
Journal of Financial and Quantitative Analysis	0		
Review of Financial Studies	13	1991	2015
Panel B: Forecasting journals			
International Journal of Forecasting	92	1987	2016
Journal of Forecasting	76	1984	2016

Sources: http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1540-6261, http://www.sciencedirect.com/science/journal/0304405X, http://journals.cambridge.org/action/displayJournal?jid=JFQ, http://services.oxfordjournals.org/search.dtl, http://www.sciencedirect.com/science/journal/01692070, http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1099-131X

Note: Multiple references to the same phrase are treated as a single reference.

Table 3: Articles in which the Phrase 'K-Fold Cross-Validation' Appears for a Cross-Section of Journals

Journal	Number of cites	First	Last
Panel A: Finance journals			
Journal of Finance	0		
Journal of Financial Economics	1	2012	2012
Journal of Financial and Quantitative Analysis	0		
Review of Financial Studies	0		
Panel B: Forecasting journals			
International Journal of Forecasting	8	2011	2016
Journal of Forecasting	5	2006	2015

Sources: http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1540-6261, http://www.sciencedirect.com/science/journal/0304405X, http://journals.cambridge.org/action/displayJournal?jid=JFQ, http://services.oxfordjournals.org/search.dtl, http://www.sciencedirect.com/science/journal/01692070, http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1099-131X

Note: Multiple references to the same phrase are treated as a single reference.

Table 4: Articles in which the Phrase 'Wald Test' Appears for a Cross-Section of Journals

Journal	Number of cites	First	Last
Panel A: Finance journals			
Journal of Finance	112	1972	2016
Journal of Financial Economics	117	1982	2016
Journal of Financial and Quantitative Analysis	12	1990	2013
Review of Financial Studies	100	1988	2016
Panel B: Forecasting journals			
International Journal of Forecasting	36	1995	2016
Journal of Forecasting	34	1984	2015

Sources: http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1540-6261, http://www.sciencedirect.com/science/journal/0304405X, http://journals.cambridge.org/action/displayJournal?jid=JFQ, http://services.oxfordjournals.org/search.dtl, http://www.sciencedirect.com/science/journal/01692070, http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1099-131X

Note: Multiple references to the same phrase are treated as a single reference.

A1. Terms of Reference

Expert Terms of Reference

DBP Support for *t* and Wald Tests Consultant Brief

2016-20 DBNGP Access Arrangement 31 December 2015

In its *Draft Decision on Proposed Revisions to the Access Arrangement for the Dampier to Bunbury Natural Gas Pipeline 2016 -2020*, dated 22 December 2015 (**Draft Decision**),the ERA notes (p 231 to Appendix 4 to the Draft Decision) that DBP's approach to the "model adequacy test" for estimating the return on equity in its Access Arrangement Proposal dated 31 December 2014 (including DBP's Submission 12 concerning Rate of Return) is not tested and appears unsupported by any source, as well as not following standard financial/economical/statistical theory. The core of the Model Adequacy Test is the t-test (and Wald test). We require an expert to provide an opinion, with a particular focus on the finance literature, as to the following:

1. Whether there is a literature in finance/economics statistics looking at out of sample prediction bias using the t or Wald statistic in the same manner as has been advanced by DBP.
2. Whether there is a literature at a more general level using the t-test (and Wald test) to consider the difference between two means, and thus to consider some aspect of models not necessarily associated with forecast bias.

Further, the ERA also considers several other tests on p234 of Appendix 4 to the Draft Decision. DBP requires an expert to comment on:

3. the degree to which these tests are fit for the purpose of testing the bias associated with forecasts as DBP seeks to do with its model adequacy test; and
4. whether these tests are widespread in the finance field.

Finally DBP seeks an opinion on the use of the cross-validation method proposed by the ERA (see Appendix 4B (i) commencing at page 232). In particular, DBP requires an expert opinion as to whether the cross validation method is used within the field of finance and, if so, how widespread its use may be.

Please provide a short written, fixed fee, quotation responding to the points above by the 8th of January. Given the tight timeframes required for a response to the regulator, it is anticipated that all work will be completed to a Draft Report stage by February 5th, with comments back from DBP by February 12th and a Final Report by February 19th. Also having regard to those timeframes, this request for quotation, including the particular questions raised above, is provided to you as a draft in the first instance. We anticipate working closely with consultants during the project to address any additional issues as they arise. If, having regard to additional issues that arise, it becomes necessary to seek your opinion on additional or different matters, we will seek to agree that additional or revised scope with you. Accordingly, please also provide an hourly rate for relevant consultants to allow for an expansion of scope where this becomes necessary.

Since it is possible that your expert report may be relied on in future proceedings before the Australian Competition Tribunal, we require that the work be undertaken in accordance with the Federal Court Guidelines for Expert Witnesses (attached). Further, your report should contain a declaration that you have been given and have read, understood and complied with Practice Note CM7 issued by the Federal Court of Australia concerning guidelines for expert witnesses. It should also contain a declaration that you have made all the inquiries that you believe are desirable and appropriate and that no matters of significance that you regard as relevant have, to your knowledge, been withheld.

A2. Federal Court Guidelines

FEDERAL COURT OF AUSTRALIA
Practice Note CM 7

EXPERT WITNESSES IN PROCEEDINGS IN THE FEDERAL COURT OF AUSTRALIA

Practice Note CM 7 issued on 1 August 2011 is revoked with effect from midnight on 3 June 2013 and the following Practice Note is substituted.

Commencement

1. This Practice Note commences on 4 June 2013.

Introduction

2. Rule 23.12 of the Federal Court Rules 2011 requires a party to give a copy of the following guidelines to any witness they propose to retain for the purpose of preparing a report or giving evidence in a proceeding as to an opinion held by the witness that is wholly or substantially based on the specialised knowledge of the witness (see **Part 3.3 - Opinion** of the *Evidence Act 1995* (Cth)).

3. The guidelines are not intended to address all aspects of an expert witness's duties, but are intended to facilitate the admission of opinion evidence[44], and to assist experts to understand in general terms what the Court expects of them. Additionally, it is hoped that the guidelines will assist individual expert witnesses to avoid the criticism that is sometimes made (whether rightly or wrongly) that expert witnesses lack objectivity, or have coloured their evidence in favour of the party calling them.

Guidelines

1. General Duty to the Court[45]

1.1 An expert witness has an overriding duty to assist the Court on matters relevant to the expert's area of expertise.

1.2 An expert witness is not an advocate for a party even when giving testimony that is necessarily evaluative rather than inferential.

1.3 An expert witness's paramount duty is to the Court and not to the person retaining the expert.

2. The Form of the Expert's Report[46]

2.1 An expert's written report must comply with Rule 23.13 and therefore must

(a) be signed by the expert who prepared the report; and

[44] As to the distinction between expert opinion evidence and expert assistance see *Evans Deakin Pty Ltd v Sebel Furniture Ltd* [2003] FCA 171 per Allsop J at [676].

[45] The "*Ikarian Reefer*" (1993) 20 FSR 563 at 565-566.

[46] Rule 23.13.

(b) contain an acknowledgement at the beginning of the report that the expert has read, understood and complied with the Practice Note; and

(c) contain particulars of the training, study or experience by which the expert has acquired specialised knowledge; and

(d) identify the questions that the expert was asked to address; and

(e) set out separately each of the factual findings or assumptions on which the expert's opinion is based; and

(f) set out separately from the factual findings or assumptions each of the expert's opinions; and

(g) set out the reasons for each of the expert's opinions; and

(ga) contain an acknowledgment that the expert's opinions are based wholly or substantially on the specialised knowledge mentioned in paragraph (c) above[47]; and

(h) comply with the Practice Note.

2.2 At the end of the report the expert should declare that "[the expert] has *made all the inquiries that* [the expert] *believes are desirable and appropriate and that no matters of significance that* [the expert] *regards as relevant have, to* [the expert's] *knowledge, been withheld from the Court*."

2.3 There should be included in or attached to the report the documents and other materials that the expert has been instructed to consider.

2.4 If, after exchange of reports or at any other stage, an expert witness changes the expert's opinion, having read another expert's report or for any other reason, the change should be communicated as soon as practicable (through the party's lawyers) to each party to whom the expert witness's report has been provided and, when appropriate, to the Court[48].

2.5 If an expert's opinion is not fully researched because the expert considers that insufficient data are available, or for any other reason, this must be stated with an indication that the opinion is no more than a provisional one. Where an expert witness who has prepared a report believes that it may be incomplete or inaccurate without some qualification, that qualification must be stated in the report.

2.6 The expert should make it clear if a particular question or issue falls outside the relevant field of expertise.

2.7 Where an expert's report refers to photographs, plans, calculations, analyses, measurements, survey reports or other extrinsic matter, these must be provided to the opposite party at the same time as the exchange of reports[49].

3. Experts' Conference

3.1 If experts retained by the parties meet at the direction of the Court, it would be improper for an expert to be given, or to accept, instructions not to reach agreement. If, at a meeting directed by the Court, the experts cannot reach agreement about matters of expert opinion, they should specify their reasons for being unable to do so.

J L B ALLSOP
Chief Justice
4 June 2013

[47] See also *Dasreef Pty Limited v Nawaf Hawchar* [2011] HCA 21.

[48] The *"Ikarian Reefer"* [1993] 20 FSR 563 at 565

[49] The *"Ikarian Reefer"* [1993] 20 FSR 563 at 565-566. See also Ormrod *"Scientific Evidence in Court"* [1968] Crim LR 240

A3. Curriculum Vitae

Simon M. Wheatley

Tel: +61 3 9878 7985
E-mail: swhe4155@bigpond.net.au

Overview

Simon is a special adviser to HoustonKemp and was until 2008 a Professor of Finance at the University of Melbourne. Since 2008, Simon has applied his expertise outside the university sector to solving problems in consulting and in fund management. Prior to joining the University of Melbourne, Simon taught finance at the Universities of British Columbia, Chicago, New South Wales, Rochester and Washington. Simon's interests and expertise are in the theory of portfolio choice, testing asset-pricing models and determining the extent to which returns are predictable.

Employment

- Special Adviser, HoustonKemp, 2015-
- Affiliated Industry Expert, NERA Economic Consulting, 2014-2015
- Special Consultant, NERA Economic Consulting, 2009-2014
- External Consultant, NERA Economic Consulting, 2008-2009
- Quantitative Analyst, Victorian Funds Management Corporation, 2008-2009
- Adjunct, Melbourne Business School, 2008
- Professor, Department of Finance, University of Melbourne, 2001-2008
- Associate Professor, Department of Finance, University of Melbourne, 1999-2001
- Associate Professor, Australian Graduate School of Management, 1994-1999
- Visiting Assistant Professor, Graduate School of Business, University of Chicago, 1993-1994
- Visiting Assistant Professor, Faculty of Commerce, University of British Columbia, 1986
- Assistant Professor, Graduate School of Business, University of Washington, 1984-1993

Education

- Ph.D., University of Rochester, USA, 1986; Major area: Finance; Minor area: Applied statistics; Thesis topic: Some tests of international equity market integration; Dissertation committee: Charles I. Plosser (chairman), Peter Garber, Clifford W. Smith, Rene M. Stulz
- M.A., Economics, Simon Fraser University, Canada, 1979

- M.A., Economics, Aberdeen University, Scotland, 1977

Publicly Available Reports

HoustonKemp

- The Cost of Equity: Response to the AER's Draft Decisions for the Victorian Electricity Distributors, ActewAGL Distribution and Australian Gas Networks: A Report for ActewAGL Distribution, AusNet Services, Australian Gas Networks, CitiPower, Jemena Electricity Networks, Powercor and United Energy, January 2016
- Equity Beta for a Benchmark Australian Water Network Service Provider: A report for Sydney Water, June 2015 (with Greg Houston, Brendan Quach and Dale Yeats)

NERA

- Estimating Distribution and Redemption Rates: Response to the AER's Final Decisions for the NSW and ACT Electricity Distributors, and for Jemena Gas Networks: A report for ActewAGL Distribution, AGN, APA, AusNet Services, CitiPower, Ergon Energy, Jemena Electricity Networks, Powercor, SA Power Networks and United Energy, June 2015
- Further Assessment of the Historical MRP: Response to the AER's Final Decisions for the NSW and ACT Electricity Distributors: A report for ActewAGL Distribution, AGN, APA, AusNet Services, CitiPower, Energex, Ergon Energy, Jemena Electricity Networks, Powercor, SA Power Networks and United Energy, June 2015
- The Cost of Equity: Response to the AER's Final Decisions for the NSW and ACT Electricity Distributors, and for Jemena Gas Networks: A report for ActewAGL Distribution, AGN, APA, AusNet Services, CitiPower, Ergon Energy, Jemena Electricity Networks, Powercor, SA Power Networks and United Energy, June 2015
- The Cost of Equity: A Critical Review of the Analysis of the AER and its Advisors: A report for DBP, June 2015
- Do Imputation Credits Lower the Cost of Equity? Cross-Sectional Tests: A report for United Energy, April 2015
- The Relation Between the Market Risk Premium and Risk-Free Rate: Evidence from Independent Expert Reports: A report for United Energy, April 2015
- Review of the Literature in Support of the Sharpe-Lintner CAPM, the Black CAPM and the Fama-French Three-Factor Model A report for Jemena Gas Networks, Jemena Electricity Networks, AusNet Services, Australian Gas Networks, CitiPower, Ergon Energy, Powercor, SA PowerNetworks, and United Energy, March 2015
- Estimating Distribution and Redemption Rates from Taxation Statistics A report for Jemena Gas Networks, Jemena Electricity Networks, AusNet Services, Australian Gas Networks, CitiPower, Ergon Energy, Powercor, SA PowerNetworks and United Energy, March 2015
- Empirical performance of Sharpe-Lintner and Black CAPMs: A report for Jemena Gas Networks, Jemena Electricity Networks, ActewAGL, AusNet Services, CitiPower, Energex, Ergon Energy, Powercor, SA Power Networks, and United Energy, February 2015
- Historical estimates of the market risk premium: A report for Jemena Gas Networks, Jemena Electricity Networks, ActewAGL, Ausgrid, AusNet Services, Australian Gas Networks, CitiPower, Endeavour Energy, Energex, Ergon, Essential Energy, Powercor, SA Power Networks and United Energy, February 2015
- Robust regression techniques: A report for DBP, December 2014

- Imputation Credits and Equity Returns: A report for the Energy Networks Association, October 2013 (with Brendan Quach)
- The Fama-French Three-Factor Model: A report for the Energy Networks Association, October 2013 (with Brendan Quach)
- The Market Risk Premium: Analysis in Response to the AER's Draft Rate of Return Guidelines: A report for the Energy Networks Association, October 2013 (with Brendan Quach)
- The Market, Size and Value Premiums: A report for the Energy Networks Association, June 2013 (with Brendan Quach)
- Estimates of the Zero-Beta Premium: A report for the Energy Networks Association, June 2013 (with Brendan Quach)
- The Payout Ratio: A report for the Energy Networks Association, June 2013 (with Brendan Quach)
- Review of Cost of Equity Models: A report for the Energy Networks Association, June 2013 (with Brendan Quach)
- The Cost of Equity for a Regulated Energy Utility: A Response to the QCA Discussion Paper on the Risk-Free Rate and the MRP: A report for United Energy and Multinet Gas, March 2013 (with Brendan Quach)
- The Cost of Equity for a Regulated Energy Utility: A report for Multinet, February 2013 (with Brendan Quach)
- The Black CAPM: A report for APA Group, Envestra, Multinet & SP AusNet, March 2012 (with Brendan Quach)
- Prevailing Conditions and the Market Risk Premium: A report for APA Group, Envestra, Multinet & SP AusNet, March 2012 (with Brendan Quach)
- The Market Risk Premium: A report for CitiPower, Jemena, Powercor, SP AusNet and United Energy, 20 February 2012 (with Brendan Quach)
- Cost of Equity in the ERA DBNGP Draft Decision: A report for DBNGP, 17 May 2011 (with Brendan Quach)
- The Market Risk Premium: A report for Multinet Gas and SP AusNet, 29 April 2011 (with Brendan Quach)
- Cost of Capital for Water Infrastructure Company Report for the Queensland Competition Authority, 28 March 2011 (with Brendan Quach)
- The Cost of Equity: A report for Orion, 2 September 2010 (with Greg Houston and Brendan Quach)
- New Gamma Issues Raised by AER Expert Consultants: A report for JGN, 17 May 2010 (with Brendan Quach)
- The Required Rate of Return on Equity for a Gas Transmission Pipeline: A Report for DBP, 31 March 2010 (with Brendan Quach)
- Jemena Access Arrangement Proposal for the NSW Gas Networks: AER Draft Decision: A report for Jemena, 19 March 2010 (with Greg Houston and Brendan Quach)
- Payout Ratio of Regulated Firms: A report for Gilbert + Tobin, 5 January 2010 (with Brendan Quach)

- Review of Da, Guo and Jagannathan Empirical Evidence on the CAPM: A report for Jemena Gas Networks, 21 December 2009 (with Greg Houston and Brendan Quach)
- The Value of Imputation Credits for a Regulated Gas Distribution Business: A report for WA Gas Networks, 18 August 2009 (with Greg Houston, Brendan Quach and Tara D'Souza)
- Cost of Equity - Fama-French Three-Factor Model Jemena Gas Networks (NSW), 12 August 2009 (with Jeff Balchin, Greg Houston and Brendan Quach)
- Estimates of the Cost of Equity: A report for WAGN, 22 April 2009 (with Brendan Quach)
- AER's Proposed WACC Statement – Gamma: A report for the Joint Industry Associations, 30 January 2009 (with Greg Houston and Brendan Quach)
- The Value of Imputation Credits: A report for the ENA, Grid Australia and APIA, 11 September 2008 (with Greg Houston and Brendan Quach)

Consulting Experience

- HoustonKemp, 2015 -
- NERA, 2008 - 2015
- Lumina Foundation, Indianapolis, 2009
- Industry Funds Management, 2010

Academic Publications

- Imputation credits and equity returns, (with Paul Lajbcygier), 2012, *Economic Record* 88, 476-494.
- Do measures of investor sentiment predict returns? (with Robert Neal), 1998, *Journal of Financial and Quantitative Analysis* 33, 523-547.
- Adverse selection and bid-ask spreads: Evidence from closed-end funds (with Robert Neal), 1998, *Journal of Financial Markets* 1, 121-149.
- Shifts in the interest-rate response to money announcements: What can we say about when they occur? (with V. Vance Roley), 1996, *Journal of Business and Economic Statistics* 14, 135-138.
- International investment restrictions and closed-end country fund prices, (with Catherine Bonser-Neal, Greggory Brauer, and Robert Neal), 1990, *Journal of Finance* 45, 523-547 (reprinted in International Capital Markets Volume III, 2003, G. Andrew Karolyi and Rene M. Stulz, editors, Edward Elgar Publishing, Cheltenham, Glos).
- A critique of latent variable tests of asset pricing models, 1989, *Journal of Financial Economics* 21, 177-212.
- Some tests of international equity market integration, 1988, *Journal of Financial Economics* 21, 177-212 (reprinted in International Capital Markets Volume I, 2003, G. Andrew Karolyi and Rene M. Stulz, editors, Edward Elgar Publishing, Cheltenham, Glos).
- Some tests of the consumption-based asset pricing model, 1988, *Journal of Monetary Economics* 22, 193-215.

Working Papers

- An evaluation of some alternative models for pricing Australian stocks (with Paul Lajbcygier), 2009.
- Intertemporal substitution, small-sample bias, and the behaviour of U.S. household consumption (with Kogulakrishnan Maheswaran and Robert Porter), 2007.
- Keeping up with the Joneses, human capital, and the home-equity bias (with En Te Chen), 2003.
- Evaluating asset pricing models, 1998.
- Time-non-separable preferences or artifact of temporal aggregation? (with Robert Porter), 2002.
- Testing asset pricing models with infrequently measured factors, 1989.

Refereeing Experience

- Referee for Accounting and Finance, the Australian Journal of Management, Economic Letters, Financial Analysts Journal, Financial Management, Journal of Accounting and Economics, Journal of Business, Journal of Empirical Finance, Journal of Finance, Journal of Financial and Quantitative Analysis, Journal of Financial Economics, Journal of Futures Markets, Journal of International Economics, Journal of International Money and Finance, Journal of Money, Credit, and Banking, Journal of Monetary Economics, Management Science, National Science Foundation, Pacific-Basin Finance Journal, and the Review of Financial Studies.
- Program Committee for the Western Finance Association in 1989 and 2000.

Teaching Experience

- International Finance, Melbourne Business School, 2008
- Corporate Finance, International Finance, Investments, University of Melbourne, 1999-2008
- Corporate Finance, International Finance, Investments, Australian Graduate School of Management, 1994-1999
- Investments, University of Chicago, 1993-1994
- Investments, University of British Columbia, 1986
- International Finance, Investments, University of Washington, 1984-1993
- Investments, Macroeconomics, Statistics, University of Rochester, 1982
- Accounting, 1981, Australian Graduate School of Management, 1981

Teaching Awards

- MBA Professor of the Quarter, Summer 1991, University of Washington

Computing Skills

- User of SAS since 1980. EViews, Excel, LaTex, Matlab, R, Visual Basic. Familiar with the SIRCA SPPR, Compustat and CRSP databases. Some familiarity with Bloomberg, FactSet and IRESS.

Board Membership

- Anglican Funds Committee, Melbourne, 2008-2011

Honours

- Elected a member of Beta Gamma Sigma, June 1986.

Fellowships

- Earhart Foundation Award, 1982-1983
- University of Rochester Fellowship, 1979-1984
- Simon Fraser University Fellowship, 1979
- Inner London Education Authority Award, 1973-1977

HOUSTONKEMP
Economists

Level 40, 161 Castlereagh Street
Sydney NSW 2000 Phone: +61 2 8880 4800

HoustonKemp.com