Review of Statistical Analysis of Starting Price Adjustments

Report prepared for
Commerce Commission

5 April 2011

Tim Coelli and Denis Lawrence

Economic Insights Pty Ltd
6 Kurundi Place, Hawker, ACT 2614, AUSTRALIA
Ph +61 2 6278 3628 Fax +61 2 6278 5358
WEB www.economicinsights.com.au
ABN 52 060 723 631

CONTENTS

EXECUTIVE SUMMARY

The Commerce Commission has received submissions from interested parties on its Discussion Paper on default price–quality path (DPP) starting price adjustments, including a supplementary submission package received from the Electricity Networks Association (ENA). ENA's supplementary submission included a report by Thomson and van Zijl (2010) (hereafter 'TvZ') which proposed an approach for forecasting a supplier's ROI and determining a point estimate of ROI for use as part of starting price adjustments.

The Commission has engaged Economic Insights to prepare this report assessing the appropriateness of the statistical analysis and proposals for starting price adjustments submitted by the ENA. Our primary findings are as follows:

- Most of the normalisation adjustments applied to Scenario 2 and Scenario 5 data in PwC (2010) appear to be reasonable for the purpose at hand. However, adjusting for Price Path Breach/Headroom in the Scenario 5 data seems unusual given that data that did not make this adjustment are likely to better reflect the ROIs that can be achieved from the associated price levels.
- The extent of data normalisations made in the early years highlight the problematic nature of using currently available Information Disclosure Data for this type of analysis.
- The degree of complexity of the models chosen by TvZ is generally appropriate for the limited amount of data that is available at this point in time. However, we believe that some alternative (simple) models with stationary means and short run dynamics modelled using autoregressive (AR) processes should also be investigated.
- The usefulness of any tests for stationarity is very much constrained by the small sample size in this study.
- TvZ make use of a homogeneity of variance test that relies upon the existence of independent samples of data of reasonable length (neither of which exist in these data) and hence the parameter homogeneity (stability) question has not been resolved.
- TvZ examine the out–of–sample properties of three models – a constant level (CL) model, a random walk (RW) model, and a local level (LL) model – and find the LL model performs marginally better than the other two. As noted above, we consider that simple AR models also warrant investigation as they might perform better than all three of the models TvZ examine.
- TvZ do not do formally estimate their LL model but instead provide a rough calibration of the model. It would be premature for the Commission to consider adopting the proposed LL model until the model has been formally estimated and tested.
- In our view TvZ's proposed LL model (in its current form) is not fit for purpose due to the quality and extent of available data, a number of concerns with the estimation and testing procedures, and adverse incentive properties. With regard to the latter, the model forecast places 73 per cent of its weight on the most recent ROI observation. This provides substantial incentives for gaming on the part of EDBs.
- TvZ's Scenario 5 data shows many EDBs with a ROI consistently above the target ROI

over the past 6 years, along with decreases in ROI for many EDBs in the most recent years. Thus, those models with prediction equations that allocate more weight to the most recent observations (LL and RW) would tend to have fewer EDBs subject to starting price adjustments (in this particular reset).

- For this price reset, if the Commission were to adopt such a statistical approach, one way of overcoming the estimation, inconsistency and incentive problems noted above would be to use the prediction and CI derived from the CL model (with T=4) and then reconsider the best model prior to the next DPP price reset decision in 2015. This would be an appropriately conservative approach because:
    - it reduces the risk that an EDB will be the subject of a price reset because they have an 'unusual' year in year 4,
    - it reduces incentives for gaming, and
    - it allows a reconsideration once a longer period of more consistent data become available.
- The choice of the CL model is also supported by the conceptual discussion of the ROI data generating process (DGP) in appendix A, where we note that:
    - there is little reason to expect a long term trend (either stochastic or deterministic) in ROI data
    - there are likely to be various random shocks affecting the ROI series, but the effects of these shocks will be transitory, with mean reversion (at varying rates) expected due to the effects of the regulator and market forces, and
    - two types of semi–regular cycles are likely to be present in the series: one due to macroeconomic effects, that will be essentially common across the EDBs, and one due to asset age effects, that could be expected to differ across EDBs.
- ROI forecasts and forecast confidence intervals (CIs) should ideally be derived from the one model if one wishes to accurately assess the probability of incorrectly rejecting the null hypothesis that the ROI of the *i*–th EDB equals the target ROI. We agree with TvZ that the method proposed in the Commission's Discussion Paper is not consistent with this ideal (and does not explicitly claim to be).
- The TvZ recommendation that the z_α used in setting the control limits should be at least 2 is said to be based on the statistical process control (SPC) literature. It is not clear why SPC should have relevance to price regulation of EDBs.
- The ENA (2010, p.3) have also made an additional suggestion that any EDB with a ROI below the *upper* control limit should be allowed to raise its starting prices so that its ROI reaches this *upper* control limit. However, the target ROI is already set at the 75th percentile of the WACC range. If we assume that the WACC and the ROI distributions are normally distributed and have the same standard deviations, then this would be equivalent to suggesting that the target ROI should actually be 2.67 standard deviations above the mean WACC, which is the 99.62 percentile of this distribution. We note that while the Commission has indicated it is interested in promoting dynamic efficiency, the degree of generosity implicit in this ENA suggestion would be very difficult to justify.

- The TvZ report is a useful piece of statistical analysis that provides a first step in identifying forecasting models that could have the potential to be of use in the future.
- In determining what the most appropriate methodology for determining starting price adjustments is, the Commission will have to take the broader regulatory context into account which involves recognising that EDBs have recourse to a building blocks–based customised price path (CPP) option as well as the productivity–based DPP. That is, the DPP, the CPP and associated Input Methodologies will likely need to be considered as a package in determining appropriate starting price adjustments.
- Neither the TvZ report nor the alternative CL method suggested above address this broader package of regulatory options and issues.
- Statistical analysis is unlikely to provide the primary method of determining appropriate starting price adjustments in this broader regulatory context, but it may have a role to play in determining future scenarios and forecasts of key output and input variables used.

1 INTRODUCTION

Under Part 4 of the Commerce Act 1986 the Commerce Commission ('the Commission') is required to set default price-quality paths (DPPs) for suppliers subject to default/customised price–quality regulation. These suppliers include non–exempt electricity distribution businesses (EDBs), gas distribution businesses (GDBs) and gas transmission businesses (GTBs).

As part of setting a DPP the Commission must specify the starting prices applying to each supplier, which must be either:

a) the prices that applied at the end of the preceding regulatory period; or

b) prices, determined by the Commission, that are based on the current and projected profitability of each supplier.

The Act does not expressly state a particular type of process that the Commission must employ to make starting price adjustments.

A DPP has been set for EDBs for the period 2010–2015. However, the Act provides that the Commission may reset this DPP if an input methodology (determined on 23 December 2010) had applied at the time the DPP were reset (ie 30 November 2009) and that input methodology would have resulted in a materially different path being set. The Commission has until October 2011 to undertake the reset.

The Commission (2010) consulted on its views and a proposed approach for making starting price adjustments in its starting price adjustments for DPPs Discussion Paper. The paper set out a proposed framework for making starting price adjustments and discussed the components of that framework. Specifically, under the framework the Commission would calculate a return on investment (ROI) value for each supplier and compare this value with an ROI band centred on the calculated industry weighted average cost of capital (WACC) estimate.

The Commission received submissions from interested parties on the paper, including a supplementary submission package received from the Electricity Networks Association (ENA). ENA's supplementary submission included a report by Thomson and van Zijl (2010) (hereafter 'TvZ') which proposed an approach for forecasting a supplier's ROI and determining a point estimate of ROI for use as part of starting price adjustments.

TvZ (2010, p.1) recommended that, given a suitable forecasting model fitted to ROI time series over all but the last year of a DPP:

a) optimal forecasts of the underlying ROI for the last year of the DPP be determined using this model;

b) the decision that the underlying ROI for an EDB exceeds the target ROI is made when its forecast exceeds the control limit given by:

$$\text{target ROI} + z_\alpha \text{ standard deviation of the forecast error}$$

where z_α controls the false positive rate α and z_α should be at least 2.

TvZ argued that the non–stationary local level model identified during their exploratory analysis was a suitable forecasting model for ROI panel time series that was 'fit for purpose'.

The Commission has engaged Economic Insights Pty Ltd ('Economic Insights') to prepare a report that assesses the appropriateness of the statistical analysis and proposals for starting price adjustments submitted by ENA. The Commission requested that the report include assessments of:

- the validity of the proposed normalisation adjustments, including the data normalisation techniques (in particular the appropriateness of adjusting for Price Path Breach / Headroom);
- the choice of models;
- the statistical testing of model validity;
- the conclusions and robustness of the conclusions drawn from the preferred model in light of any assumptions made;
- whether the model is fit for purpose in the context of setting starting prices (ie the applicability of the proposed approach for setting starting prices in a default/customised price–quality path regulatory setting that is incentive–based); and
- the conclusions drawn in the ENA's supplementary submission package and whether alternative interpretations might also be consistent with the evidence presented.

The Commission also requested a brief commentary on the use of models in similar contexts (ie in assessing and forecasting financial performance) and the extent to which the models proposed in TvZ are consistent with those models.

In the following section of this report we assess each of these aspects of the TvZ analysis.

About the authors of this review

Tim Coelli is a PhD trained econometrician who has lectured in econometrics (theory and application) for 19 years in universities. He was a Professor of Economics (teaching econometrics) at the University of Queensland from 2002 to 2009, and is currently an Adjunct Professor there and an Associate of Economic Insights. Between 1994 and 2002 he was Associate Professor of Econometrics and Senior Lecturer in Econometrics at the University of New England. Tim has authored five books and has over 40 articles in leading academic journals. He is ranked among the top 3 per cent of academic economists in Repec (http://ideas.repec.org/cgi-bin/rank.cgi?pco188&hEyL).

Tim has taught graduate classes in time series econometrics and in panel data econometrics. He has also presented many training courses on productivity and efficiency analysis and has refereed articles in over 70 leading academic journals. He has held visiting positions at universities in Belgium, Spain and the UK.

Denis Lawrence is also a PhD trained econometrician and has undertaken many applied regulatory econometric analyses. He is a Director of Economic Insights and has extensive experience in the application of productivity–based regulation in New Zealand, Australia and North America.

2 REVIEW OF THE TVZ STATISTICAL ANALYSIS AND FRAMEWORK

2.1 Normalisation adjustments including data normalisation

TvZ use various EDB data sets constructed by PwC (2010). The data sets include ROI normalisations corresponding to six scenarios, three of which cover the 12 years from 1999 to 2010, and three of which cover the 6 years from 2005 to 2010. TvZ (2010, pp.4–5) provide a brief description of the six scenarios as follows:

- Scenario 0 comprises the disclosed ROIs for the years ending 31 March 1999 to 31 March 2010. Apart from naming conventions, these data sets have not been adjusted (normalised) and reflect the relevant regulatory Information Disclosure Requirements (IDRs) for each year concerned.
- Scenario 1 is an adjustment of Scenario 0 that smooths the impact of the irregular revaluations that took place over 1999–2004, and uses CPI–based revaluations over 2005–2007 in line with those for 2008–2009.
- Scenario 2 adjusts Scenario 0 by replacing all the disclosed revaluations over 1999–2007 by CPI–based revaluations. All valuations over 1999–2010 are now made on the same basis.
- Scenario 3 is an adjustment of Scenario 0, but only over the years ending 31 March 2005 to 31 March 2010. Significant changes to the construction of ROI data were introduced in 2008 with the publication of the 2008 IDRs. Scenario 3 ROI restate the regulatory asset base of each EDB over 2005–2007 in a consistent manner to that used over 2008–2010.
- Scenario 4 is a modification to Scenario 3 with revenue adjusted for any price path breaches and headroom over 2005–2009.
- Scenario 5 is a modification of Scenario 4 that excludes capital contributions from regulated revenue and deducts them from the regulatory asset base. This is in accord with the draft input methodologies issued by the Commission on 22 October 2010.

Most of the TvZ analyses use the Scenario 2 and Scenario 5 normalised data sets.

Most of the normalisation adjustments applied to Scenario 2 and Scenario 5 data in PwC (2010) appear to be reasonable for the purpose at hand. However, the nature and extent of normalisations listed above serve to highlight the relatively poor quality and consistency of currently available data, particularly in the early years of the series. This particularly applies to asset valuation series which are critical to analyses of profitability. Combined with the relatively limited number of observations available, this will make the application of sophisticated statistical methods somewhat problematic at this point in time.

There have also been extensive EDB amalgamations and restructurings since 1999 and some major one–off events, such as the aftermath of the Auckland CBD outage. Amalgamations have caused reporting issues due to a flaw in the IDRs which can lead to some of the data for the absorbed entity not being reported in the year of amalgamation (see Lawrence 2003). It is not clear whether the data used by TvZ includes adjustment for these reporting anomalies.

Data quality and consistency issues aside, adjusting for Price Path Breach/Headroom in the Scenario 5 data seems unusual given that data that did not make this adjustment are likely to better reflect the ROIs that can be achieved from the associated price levels. This adjustment is likely to have two effects on the empirical exercise. First, it is likely to cause EDBs with large revenue adjustments to "come back to the pack" and hence have a lower probability of exceeding the control limits. However, the estimate of the variance of the ROI series is also likely to decrease and hence reduce the widths of the confidence intervals (ie control limits). The net effect is likely to be no change in the probability that an arbitrarily selected EDB will exceed the control limit. However, we may find that those EDBs with small adjustments to revenues could now have a higher probability of exceeding the (new) bounds. This will balance up the fact that those EDBs with large adjustments to revenues now have a lower probability of exceeding the (new) bounds.

We recommend that the empirical process be repeated without the *Price Path Breach/Headroom* adjustments so that the impact of this particular data normalisation can be assessed.

2.2 Choice of model(s)

In terms of the choice of model(s) to investigate, we believe that the degree of complexity of the models chosen by TvZ is generally appropriate for the amount of data that is available at this point in time (T=4, N=16).

TvZ (2010) consider three models:

- a constant level (CL) model (no trend)
- a random walk (RW) model (stochastic trend), and
- a local level (LL) model.

However, we believe that some alternative (simple) models with stationary means and short run dynamics modelled using autoregressive (AR) processes should also be investigated. We provide arguments for the use of these models in the sections below. Note that these types of models have been widely applied in panel data settings in economics and business in recent years. For example, see the seminal paper of Arellano and Bond (1991), a recent survey in Binder et al (2005) and an example of a simple AR(1) model in Bond (2002).

2.3 Statistical testing

The statistical testing of model validity in TvZ (2010) involved three main activities:

- testing for stationarity in the ROI data;
- testing for parameter stability across EDBs; and
- evaluation of out–of–sample forecasting performance.

We shall now deal with each of these in turn.

Stationarity tests

The usefulness of any tests for stationarity is very much constrained by the small sample size in this study. TvZ (2010) note that having only 12 or 6 time series observations means that these tests are 'challenging' and 'problematic', respectively (TvZ p8). We agree with this.

Traditional unit root tests, such as the Augmented Dickey–Fuller (ADF), are known to have low power, especially in small samples (Kennedy 1998, p.285). Samples of size 6 and 12 are *very* small samples. We have been unable to find any references to papers that apply these tests to such short time series.

We believe that the ADF tests that were conducted are particularly unreliable because of the small sample size and also for various additional reasons, which we now discuss.

First, the box–plots in Figure 1 in TvZ (2010) show evidence of non–normal distributions in many cases. Tests of normality should be conducted and appropriate data transformations applied if required. Franses and McAleer (1998) show that the conclusions of ADF tests can alter depending on the data transformations that are applied. Logarithmic transformations are often applied to financial data (eg Tsay 2010, p.83). We note that the application of log transformations to Scenario 2 ROI data may have been problematic because of the existence of a small number of negative values.

Second, TvZ (2010, p.9) describe the ADF test applied to model (1) as using 'the simplest of these procedures'. By this we assume that this means that one lag of the differenced dependent variable is used in the testing procedure (see Enders 2005, p.190). If this is correct, it is not clear whether an appropriate lag–length testing procedure (eg Enders 2005, p.193) was used to select this lag of one. Enders (2005, p.191) notes that if too few lags are used the key parameters (and their standard errors) 'will not be well estimated'. Harris (1992) recommends a lag length of $12(T/100)^{0.25}$, which implies a lag length of 7 (not 1) when T=12.

An additional complication is that TvZ (2010) use Scenario 2 data (involving 12 observations) to test the stochastic properties of the data and validate their models (using out of sample predictions) and then use Scenario 5 data (involving 6 observations) for estimation and prediction. This is done because Scenario 5 data are judged to be the best quality data set among the six data sets considered, while Scenario 2 data are regarded as being the best quality data among the three data sets that have 12 observations. Furthermore, due to unusual changes in means and variances, the Scenario 2 data are also 'adjusted' by subtracting annual means and scaling the first 9 observations by (an arbitrary) 50 per cent. This does not leave one with substantial faith in the Scenario 2 data.

TvZ (2010, p.8) supplement their ADF tests with a visual inspection of plots of the ROI data and conclude that 'a simple trend plus error model may well be appropriate'. We would argue that one could also look at the plots and conclude that a mean reversion model may well be appropriate. Given that typical business cycles and investment cycles could easily be 10 years or longer in duration, an inspection of plots of 6 or 12 time series observations could provide evidence of a wide variety of possible models.

TvZ (2010, p.10) also conduct NH tests for stationarity. This test differs from the ADF test in a number of ways, but most importantly the NH test has stationarity as the null hypothesis while the ADF has non-stationarity as the null. Given that traditional hypothesis tests generally favour the null, in the sense that we need to get extreme information (eg in the 5 per cent tails of the distribution under the null) to reject the null, it is not a surprise to observe that the ADF results indicate non–stationarity in all 16 cases while the NH results find *stationarity* in approximately 50 per cent of cases. Thus, we have limited confidence in the NH and ADF results.

TvZ (2010, p.10) also make use of Z tests that are based on aggregations of the individual p–values to jointly test the null hypotheses across the 16 EDBs. The results from these two tests do provide support for the case of non–stationarity, however it is not clear why these particular panel data stationarity tests were chosen in preference to the many other options available (eg see Binder et al 2005 for a general discussion and Harris and Tzavalis 1999 for a panel unit root test that is developed for small T cases).

Furthermore, TvZ (2010, p.11) state that 'perhaps the greatest weakness of the Z tests is the assumption that the components of each Z test are independent'. The importance of this point is illustrated in Peel et al (2004) who apply a panel unit root test that can account for cross sectional dependencies (Chang 2002) to data on financial ratios and find that it rejects the null of a joint unit root (stationarity) even when all of the individual ADF tests do not reject stationarity.

Overall, our view is that statistical tests and visual inspect of plots is insufficient for one to argue confidently for non–stationarity trends or otherwise.

Parameter stability tests

As is noted by TvZ (2010), tests for parameter stability should ideally be based upon a likelihood ratio (LR) test or similar, based upon the maximum likelihood (ML) estimation of the LL model. However, TvZ (2010) have not yet formally estimated the preferred LL model.

TvZ (2010, p.14) instead make use of a homogeneity of variance test due to Fligner and Killeen. These tests rely upon the existence of independent samples of data of reasonable length (neither of which exist in these data) and hence TvZ (2010, p.14) note that 'the results are more indicative than definitive'.

Given these issues, we would argue that the parameter homogeneity (stability) question has not been resolved.

However, in terms of the practical implementation of the proposed LL model in a regulatory setting, the existence of heterogeneous parameters could cause some challenges. This is because if each EDB has different parameters then each EDB will have a ROI forecasting equation where differing weighting schemes apply (ie some predictions may put a lot of weight on recent observations while others may not). Furthermore, the forecast variances and hence control limits would differ for each EDB. This could be problematic from a regulatory perspective.

Out–of–sample predictions

The use of out–of–sample predictions to evaluate the quality of alternative forecasting models is a useful exercise that is commonly applied in the time series forecasting literature

(Kennedy 1998, p.295).

TvZ (2010, p.19) conduct an analysis of three models:

- a constant level (CL) model (no trend)
- a random walk (RW) model (stochastic trend), and
- a local level (LL) model.

They find that all three models are similar in terms of forecast bias (close to zero), but that the root mean square error (RMSE) of the LL model is marginally lower than the RW RMSE which is marginally lower than the CL RMSE (see their Figure 8).

These results indicate that all three models perform in a similar manner, and that in the absence of any other relevant factors (see discussion below of regulatory issues) one could argue that the LL model could be *marginally* preferred (among these three models). As noted above, we consider that simple AR models also warrant investigation as they might perform better than all three of the models TvZ examine.

2.4 Robustness of the conclusions

Next we discuss robustness of the conclusions drawn from the preferred model in light of any assumptions made.

TvZ (2010, p.21) state that the LL model has not actually been formally estimated, even though it would be a ‘relatively straightforward exercise’. The authors explain that maximum likelihood (ML) techniques could be used and the initial trend values could be ‘estimated by conventional smoothing techniques’. However, they do not do this and instead provide a rough calibration of the model, as one would normally use when constructing a simple exponentially weighted moving average model (see Levine et al, 2002, p.662).

It would be desirable for the model TvZ propose to be estimated. Statistical software packages such as Stata 11 (http://www.stata.com/stata11/) have routines which can estimate a range of State Space models using MLE, including the simple LL model.

We suspect that the estimation of an LL model when T=4 could actually be a very challenging exercise, especially because the properties of estimators can depend to a large degree upon the assumed initial starting values for the trends, and that the importance of this increases as the time dimension of the model reduces. We have been unable to find any examples in the literature where non–stationary panel data models (of any type) have been estimated using data where there are only four time periods available.

In our view, it would be premature for the Commission to consider adopting the proposed LL model until the model has been formally estimated and tested. As part of this process, we recommend the Commission ask to see the detail of the ML estimation method, in particular detail regarding what is meant by ‘conventional smoothing techniques’ and how these methods can be applied in a way that is transparent to all stakeholders and provide ML estimators that have well defined properties.

The ML estimation of the LL model will also allow one to test a number of the key assumptions underlying the model using likelihood ratio (LR) tests. This could include testing for: the existence of stochastic trends; the existence of a common stochastic trend; and, the degree to which the parameters are common across EDBs.

2.5 Fitness for purpose

Next we discuss whether the model is fit for purpose in the context of setting starting prices (ie the applicability of the proposed approach for setting starting prices in a default/ customised price–quality path regulatory setting that is incentive–based).

This discussion is to a large degree independent of the above statistical issues.

The statistical models are designed to be used for two purposes:

1) predicting the ROI for the i-th firm in year T using ROI data from previous years, where year T is the year immediately preceding the new five–year DPP period, and
2) obtaining an estimate of the standard deviation of the ROI series for use in constructing a 'band' around the target ROI (the 75th percentile WACC measure), which is used to assess if the predicted ROI of the i-th firm differs enough from the target ROI to warrant a starting price adjustment.

In our view the proposed LL model (in its current form) is not fit for purpose.

The model forecast places 73 per cent of its weight on the most recent ROI observation. This provides substantial incentives for gaming on the part of EDBs.

Having said this, the Commission (2010) suggestion of putting a 100 per cent weight on the most recent ROI observation may provide even greater incentives for gaming.

Discussion of the Commission (2010) proposed method

We would argue that forecasts and forecast confidence intervals (CIs) should ideally be derived from the one model if one wishes to accurately assess the probability of incorrectly rejecting the null hypothesis that the ROI of the *i*-th EDB equals the target ROI. The method proposed by the Commission (2010) is not consistent with this ideal (and does not explicitly claim to be).

TvZ (2010, p.21) are correct in the comments that they make in this regard. They argue that the standard deviations that are used to construct the control limits are implicitly derived from a CL model (where T=4) while the prediction used implicitly comes from the RW model.[1]

One implication of this is that it then becomes difficult to determine (for the proposed method) the probability of incorrectly rejecting the null hypothesis that the ROI of the *i*-th EDB equals the target ROI.

If the RW model is the 'true' model then the use of a 95 per cent CI from the CL model will produce control limits that are generally too narrow and hence cause one to incorrectly reject the null more than 5 per cent of times. This will be problematic if the 5 per cent level of

[1] They also correctly point out that the estimator of the pooled standard deviation should be the square root of the mean of the variances and not the mean of the standard deviations. The proposed Commission (2010) estimator is non-standard in the context of the CL model.

significance was desired.

An alternative approach

The Commerce Act requires starting prices to be based on 'the current and projected profitability of each supplier'. An argument implicit in TvZ's preference for the LL model is that more recent data is more relevant and more likely to predict future ROIs than older data. However, apart from this method having adverse incentive effects, it is necessary to carefully consider the data generating process (trend plus cycles plus noise). If the objective is to forecast the underlying profitability trend, then putting a lot of weight on one observation likely means running the risk of random noise or cyclical effects having a heavy influence on the forecast. To forecast the underlying profitability *trend*, the implied optimal forecast is likely to give substantial weight to a number of observations. This also has attractive incentive properties by making it harder for businesses to game the outcome.

For this price reset, if the Commission were to adopt such a statistical approach, one way of overcoming the inconsistency and incentive problems noted above would be to use the prediction and CI derived from the CL model (with T=4) and then reconsider the best model prior to the next DPP price reset decision in 2015.[2]

This would be an appropriately conservative approach because:

- it reduces the risk that an EDB will be the subject of a price reset because they have an 'unusual' year in year 4, and
- it reduces incentives for gaming.

The LL (and RW) models arguably provide more opportunity for gaming because their prediction equations allocate most (or all) weight to the most recent observation. Some EDBs could be motivated to organise their activities so that they incur extra expenses in year 4 and hence have a low ROI in that year. The CL model would reduce these incentives substantially.

For the current price reset, the three different models have different implications regarding the number, distribution and sizes of price resets applied to the 16 EDBs. The examples provided in TvZ (2010, pp.19–22) produce 10, 4 and 5 resets for models CL, RW and LL, respectively.

The Scenario 5 data (see Figure 3 in TvZ 2010) shows many EDBs with a ROI consistently above the target ROI over the past 6 years, along with decreases in ROI for many EDBs in the most recent years. Thus, those models with prediction equations that allocate more weight to the most recent observations (LL and RW) would tend to have fewer EDBs subject to starting price adjustments (in this particular reset).

However, if the majority of ROI data series had instead shown *increases* in ROI in recent years, the opposite would be the case with the CL model tending to have fewer EDBs subject to starting price adjustments than the LL model.

If it could be guaranteed that in the medium to long term the same model would be applied to

[2] This suggestion is complicated by the fact that, for the initial reset, the actual ROI values are available for year 5.

price resets in every 5 year period, one could then argue that the effects of these favourable and unfavourable ROI movements would eventually average out.

The choice of the CL model is also supported by a conceptual discussion of the ROI data generating process (DGP) in appendix A, where we note that:

- there is little reason to expect a long term trend (either stochastic or deterministic) in ROI data
- there are likely to be various random shocks affecting the ROI series, but the effects of these shocks will be transitory, with mean reversion (at varying rates) expected due to the effects of the regulator and market forces, and
- two types of semi–regular cycles are likely to be present in the series: one due to macroeconomic effects, that will be essentially common across the EDBs, and one due to asset age effects, that could be expected to differ across EDBs.

In addition we observe that:

- some simple stationary autoregressive process could be considered in future models to attempt to capture the mean reversion and cyclical patterns;
- the target ROI is a WACC measure, which will vary to some degree with the macroeconomic cycle. This will tend to dampen the degree to which an EDB is likely to exceed the control limits because of macroeconomic factors;
- regulatory processes generally aim to mimic a competitive situation. Industries with long–lived assets (eg motor vehicle production, iron and steel production, etc.) would expect their ROI to rise and fall with their individual investment patterns. Hence, there is no reason why this should not also occur with regulated firms (to some extent); and,
- all three models will tend to trigger price resets in some EDBs as their ROI values are influenced by changes in their asset age profiles over time. This is probably unavoidable unless a weighting scheme that uses data over 10 years or more is used.

Setting the probability of a Type I Error

The TvZ (2010, p.19) recommendation that the z_α used in setting the control limits should be at least 2 is said to be based on the statistical process control (SPC) literature. It is not made clear why SPC should have relevance to price regulation of EDBs.

The value of 2 approximates the value of 1.96, which is the z–value corresponding to the 2.5 per cent upper tail of the standard normal distribution.

Thus, using a z–value of 2 corresponds to the situation where the probability of a Type I Error (incorrectly rejecting the null when it is true) is approximately 5 per cent (assuming normality).

Note that the null hypothesis is that the ROI of the *i*-th EDB equals the target ROI.

The Commission may choose a 5 per cent Type I Error or some other percentage that it judges will correctly reflect the trade–off between the interests of EDBs and consumers, in terms of static and dynamic efficiency.

This could involve an assessment of the corresponding probabilities of a Type II Error (incorrectly accepting the null when it is false). This value will vary according to the assumed 'true value' of the ROI of the *i*-th EDB. For further information see Larsen (1982, p.414).

Industry structure issues

TvZ's statistical analysis uses data for 16 of the 17 EDBs subject to default price path regulation (Wellington Electricity Lines Limited is excluded). Structural change has been an ongoing feature of the New Zealand electricity and gas distribution industries with substantial amalgamations progressively reducing the number of distribution businesses. With the number of EDBs, in particular, still being relatively high in relation to New Zealand's population and area, it is likely more amalgamations will occur. It is unclear how the proposed statistical method would deal with changes in industry structure (mergers, divestments, etc) over time.

2.6 ENA Conclusions and alternative interpretations

Here we discuss the conclusions drawn in ENA's supplementary submission package and whether alternative interpretations might also be consistent with the evidence presented.

The conclusions in ENA (2010) essentially coincide with those presented in TvZ (2010). In the discussion above we have pointed out a number of instances where we think that alternative conclusions could be drawn.

ENA (2010, p.3) have also made an additional suggestion that any EDB with a ROI below the *upper* control limit should be allowed to raise its starting prices so that its ROI reaches this *upper* control limit. In our assessment, this would not be wise. The target ROI is already set at the 75th percentile of the WACC range. If we assume that the WACC and the ROI distributions are normally distributed and have the same standard deviations, then this would be equivalent to suggesting that the target ROI should actually be 2.67 standard deviations above the mean WACC, which is the 99.62 percentile of this distribution. We note that the Commission has indicated it is interested in promoting dynamic efficiency, but the degree of generosity implicit in this ENA suggestion would be very difficult to justify.

2.7 Use of models in similar contexts

Now we discuss the use of models in similar contexts (ie in assessing and forecasting financial performance) and the extent to which the models proposed in TvZ are consistent with those models.

In our assessment the proposed models are reasonably consistent with some of the statistical models that are currently used by private sector financial analysts. It is difficult to provide empirical evidence to support this without actually conducting a survey of forecasting teams in financial institutions.

We have conducted a search of the academic literature but the degree to which the academic literature and 'real-world practice' coincide can be small, especially in the short run. This is because academics are motivated to produce newer and more complicated models so that they can satisfy journal editors, while financial forecasters are motivated to use reliable established models that do the job with minimum fuss.

The literature on the statistical analysis of financial time series data is growing at a fast rate. New textbooks are being published every year with some recent publications being Mills and Markellos (2008), Taylor (2008) and Tsay (2010). In these books various statistical methods are applied to financial time series data, including ARIMA models, State Space models (of which the LL model is one simple form), GARCH models (which allow for time varying volatility), exponential smoothing models, vector autoregression (VAR) models, Error Correction models, etc.

A search of recent journal articles produces an extensive array of new dynamic panel data models that are generally very complex. For example, refer to De Silva et al (2010) for some state space models; Binder et al (2005) for some non-stationary VAR models and Baltagi (2008) for a comparison of the forecast performance of a number of stationary dynamic panel data models. The latter paper comes to the interesting conclusion that:

> 'homogeneous panel data estimators perform well in forecast performance mostly due to their simplicity, their parsimonious representation, and the stability of the parameter estimates. Average heterogeneous estimators perform badly due to parameter estimate instability caused by the estimation of several parameters with short time series.'

This lends some support to the decision by TvZ (2010) to focus on parsimonious model formulations that assume common parameters across EDBs.

3 CONCLUSIONS

The TvZ report is a useful piece of statistical analysis that provides a first step in identifying forecasting models that could have the potential to be of use in the future. However, the nature and extent of normalisations that had to be made to the data used serve to highlight the relatively poor quality and consistency of currently available data, particularly in the early years of the series. This particularly applies to asset valuation series which are critical to analyses of profitability. Combined with the relatively limited number of observations available, this will make the application of sophisticated statistical methods somewhat problematic at this point in time.

A specific data issue relates to adjusting for Price Path Breach/Headroom in the Scenario 5 data, which seems unusual given that data that did not make this adjustment are likely to better reflect the ROIs that can be achieved from the associated price levels.

TvZ do not formally estimate their preferred LL model but instead provide a rough calibration of the model. It would be premature for the Commission to consider adopting the proposed LL model until the model has been formally estimated and tested.

In our view TvZ's proposed LL model (in its current form) is not fit for purpose due to the quality and extent of available data, a number of concerns with the estimation and testing procedures, and adverse incentive properties. With regard to the latter, the model forecast places 73 per cent of its weight on the most recent ROI observation. This provides substantial incentives for gaming on the part of EDBs.

The TvZ LL model and our proposed alternative CL model both look at the issue of starting prices adjustments from a productivity–based regulation perspective. That is, the question in the context of the DPP is what is a reasonable approach to determining appropriate starting price adjustments to apply at a point in time that reflect a reasonable trade–off between the interests of EDBs and consumers and which do not involve excessive Type I or Type II errors? However, in determining what the most appropriate methodology for determining starting price adjustments is, the Commission will have to take the broader regulatory context into account. This involves recognising that EDBs have recourse to a building blocks–based customised price path (CPP) option as well as the productivity–based DPP. That is, the DPP, the CPP and associated Input Methodologies will likely need to be considered as a package in determining appropriate starting price adjustments.

The TvZ report does not address this broader package of regulatory options and issues that will need to be considered in developing an appropriate starting price methodology. Statistical analysis is unlikely to provide the primary method of determining appropriate starting price adjustments in this broader regulatory context but it may have a role to play in determining future scenarios and forecasts of key output and input variables used.

APPENDIX A: WHAT IS THE DATA GENERATING PROCESS?

What does it mean when we say that we wish to predict the 'underlying ROI' (TvZ 2010, p.26)? The statement appears to imply that we wish to remove the effects of some 'random' factors (eg unusual climatic events, strikes, etc) so as to get at the true value that we would expect to exist in a 'normal" year.

It is therefore useful to consider carefully what factors we believe to be generating the data that we observe, and then assess the degree to which the forecasting model we use is likely to be able to disentangle these effects.

One way to conceptualise a data generating process (DGP) is to first consider the conditions under which all data observations on the ROI (in all firms and years) will be identical. We then add in the factors that may cause them to differ (one by one) and discuss what effects we expect each of them to have.

So let us assume that we have a world where there are a number of firms operating an EDB and that all firms are the same size and face identical environments (geography, population density, climate, regulations, ownership structures, etc.); identical input and output markets (and hence prices); have identical asset mixes and ages; and that every year is identical to the previous one (ie no business cycle, climatic variation, earthquakes, etc).

Furthermore, we assume we have all the conditions required for perfect competition, such as costless entry and exit, perfect information, identical management skills, and so forth.

In this situation we may expect that all firms in all years achieve identical ROIs, and let us say that this ROI is 8 per cent (for example).

Now let us discuss the effects of relaxing some of these assumptions.

- First, we can introduce a macro cycle (of roughly 10 years' duration, for example). This may mean that output demand and (to some extent) input supply conditions may change over this cycle, with the possibility that ROI will be higher at the peak of the cycle due to greater demand for electricity from industrial customers.
- Additionally, if we allow some firms to have a higher percentage of industrial customers (relative to other firms) then the effect of the business cycle on ROI is also likely to be greater for those firms with more industrial customers.
- Next we can allow climate to vary over time. Thus, in cold years we could expect increased demand for electricity for heating and hence higher revenues leading to higher ROIs. Or in years when there are big storms, maintenance costs could rise and hence ROIs fall.
- We could also allow climatic conditions to vary across geographical regions as well. Some regions could have colder winters and hence have higher demand per customer (which could increase ROI) or this could increase the "peakiness" of demand which could decrease ROI (because of the higher capital requirement per unit of electricity delivered).

Relaxing all the other assumptions, we could have various reasons why a particular firm in a particular year could now have a higher or lower ROI.

- A higher population density could reduce capital requirements per customer and hence increase ROI.
- A more challenging terrain could increase capital construction and maintenance costs and hence reduce ROI.
- A period of light–handed regulation could increase ROIs.
- Different ownership structures (eg private firm versus cooperative) could affect the incentives to set tariffs so as to recover full long run costs and hence could influence ROIs.
- Some firms could have older assets and hence have a lower regulatory asset base and hence have a higher ROI.
- Also, over the longer term, these assets would need to be replaced/refurbished, leading to firm–specific cycles in investment and hence ROI.
- Some firms may be larger and hence be able to reap scale economies and thus have higher ROIs.
- Some firms may have lower quality management and hence have inefficiency that leads to lower ROIs.
- Other random factors may affect things – such as measurement errors, earthquakes and conflicts.

This is not an exhaustive list, but we have built up a picture of the many factors which can influence the ROI of a particular firm in a particular year.

Components of a time series

It is useful to now look at what is normally written about forecasting time series in a first year business statistics texts, such as Levine et al (2002), where movements in a time series are broken up into four main components;

Time series = f(trend, cycle, seasonal, irregular)

Given that we are dealing with annual data in this case, we can omit the seasonal component to obtain:

Time series = f(trend, cycle, irregular)

Various functional forms are possible, such as multiplicative or additive. Assuming the latter we can write:

$$Y_t = TR_t + C_t + I_t \qquad ,t = 1,2,...,T \,. \tag{1}$$

Trends (TR_t) can take various forms, such as deterministic (eg linear, quadratic, exponential) or stochastic (eg a random walk).

Cycles (C_t) can be modelled in various ways (eg using autoregressive functions or trigonometric functions).

Irregular (I_t) components are the residual factors that are often assumed to be generated from some type of independent and identically distributed (iid) random variable when a formal statistical model is specified.

When one has panel data one can rewrite (1) as:

$$Y_{it} = TR_{it} + C_{it} + I_{it} \qquad , i = 1, 2, \ldots, N, \;\; t = 1, 2, \ldots, T \,, \tag{2}$$

for the *i*–th firm in the *t*–th year, where the degree to which the components are common/correlated across firms is also of interest.

From our discussion of the DGP above, one could argue that there is little justification for a long term trend in ROI data on EDBs. One could reasonably expect that the market and/or the regulator would act so as to prevent any long run trends in ROI. A random shock (eg an earthquake or a large storm) could cause ROI to deviate from its long run mean for a period of time, but one would expect mean–reversion to then occur over subsequent periods.

Our DGP discussion suggests some possible semi–regular cycles, such as a macro cycle, which would be arguably common to all EDBs and an 'asset age' cycle, which would be EDB specific because different EDBs would have built assets at different points in time and also face different rates of population growth which would also affect the pattern of variations in (average) asset ages for a particular EDB over time.

Our DGP discussion also suggests a number of possible irregular factors. We can distinguish between those that we could expect to be reasonably fixed over time for a particular EDB (eg topography, population density, firm size, spatial climatic differences, etc) and those which vary across time and firms (eg temporal climatic differences, earthquakes, strikes, etc).

We would expect that EDB–specific time–invariant fixed factors (eg topography, population density) are likely to have a persistent effect on things such as costs per unit output, but are unlikely to have a persistent effect on ROI because regulators and/or market forces will eventually ensure that the ROI approaches a reasonable level.

Given the above discussion, we would expect that a time series model with a constant mean and short run cyclical dynamics (eg modelled using AR or ARMA processes) is likely to provide a good approximation to ROI time series data.

APPENDIX B: ADDITIONAL MINOR COMMENTS ON TVZ (2010)

Pages 3 to 4: A brief discussion is provided of four past studies that have looked at the stochastic properties of financial data in samples of companies. Three of the studies do find evidence of stochastic trends. However, it should be noted that none of these studies explicitly look at ROI measures. They instead look at corporate earnings or something similar. Also, the data sets used are generally rather dated (some pre–1980) and mostly involve listed companies and not samples of regulated utilities.

Page 6: Scales on plots should be identical to aid visual comparisons.

Page 7: Again compare scales on Fig 2 and Fig 3.

Page 13 para 2: It is stated that 'model (2) … is commonly used for short term prediction'. Support needs to be provided for this statement, both in general and in particular for a time series of length 4 and/or for ROI prediction.

Page 13 para 2: It needs to be made clear whether these simulations rely on the assumption of Gaussian distributions?

Page 17 para 3: References need to be provided to support this 'widely used in practice' assertion, with particular reference to time series of length 4.

Page 19: It needs to be explained in what sense these forecasts are 'optimal', whether this is a BLUP argument and whether it is this based on finite sample theory or asymptotic theory. References need to be provided.

Page 20 para 1: We assume that TvZ are making the point that the estimator of the pooled standard deviation should be the square root of the mean variance and not the mean of the standard deviations? A reference needs to be provided to support the assertion that the Commission's estimator is less efficient.

Page 20: The discussion refers to a 'control limit' which is the *upper* limit of the bounds. Why is no mention made of the *lower* limit?

Page 24 para 2: ML will be 'optimal' (ie achieve consistency and asymptotic efficiency) if the model assumptions are correct (under certain regularity conditions) – see Greene (2003, p.473).

Page 25 para 2: Parameter homogeneity across EDBs or subsets of EDBs can be tested with a LR test.

Page 26: 'Optimal' needs to be defined and support provided for the claim that these methods are 'widely used in practice'.

Page 26: It is not clear what 'both recursions are suitably initialised' means and what the properties of these estimators in very small samples are.

REFERENCES

Arellano, M., and S.R. Bond (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *Review of Economic Studies*, 58, 277–297.

Baltagi, B. (2008), "Forecasting with Panel Data", *Journal of Forecasting*, 27, 153–173.

Binder, M., Hsiao, C., Pesaran, M.H. (2005), "Estimation and inference in short panel vector autoregressions with unit roots and cointegration", *Econometric Theory*, 21, 795–837.

Bond, S.R. (2002), "Dynamic panel data models: a guide to micro data methods and practice", *Portuguese Economics Journal*, 1, 141–162.

ComCom (2010), *Starting Price Adjustments for Default Price-Quality Paths*, Discussion Paper, Commerce Commission, August 2010

Chang, Y. (2002), "Nonlinear IV unit root tests in panels with cross-sectional dependency", *Journal of Econometrics*, 110, 261–92.

ENA (2010), *Further Submission on the Method to Adjust DPP Starting Prices*, Electricity Networks Association, 10 December 2010.

Enders, W. (2005), *Applied Econometric Time Series*, 2nd Ed., Wiley, New Jersey.

Franses, P.H. and M. McAleer (1998), "Testing for Unit Roots and Non-linear Transformations", *Journal of Time Series Analysis*, 19, 147–164.

Greene, W.H. (2003), *Econometric Analysis*, 5th edition, Prentice Hall, New Jersey.

Harris, R.D.F. and E. Tzavalis (1999), "Inference for unit roots in dynamic panels where the time dimension is fixed", *Journal of Econometrics* 91: 201–226

Kennedy, P. (1998), *A Guide to Econometrics*, 4th Ed., Blackwell, Oxford.

Larson, H.J. (1982), *Introduction to Probability Theory and Statistical Inference*, 3rd Ed., Wiley, New York.

Lawrence, Denis (2003), *Regulation of Electricity Lines Businesses, Analysis of Lines Business Performance – 1996–2003*, Report by Meyrick and Associates to the Commerce Commission, Canberra, 19 December.

Levine, D.M., D. Stephan, T.C. Krehbiel and M.L. Berenson (2002), *Statistics for Managers using Microsoft Excel*, Prentice Hall, New Jersey.

Mills, T.C. and R.N. Markellos (2008), *The Econometric Modelling of Financial Time Series*, Cambridge University Press, Cambridge.

Peel, D.A, M.J. Peel and I.A. Venetis (2004), "Further empirical analysis of the time series properties of financial ratios based on a panel data approach", *Applied Financial Economics*, 14, 155-163.

PWC (2010), *Electricity Networks Association Return on Investment Data*, Price-Waterhouse-Cooper, October 2010.

de Silva, A., R. Hyndman and R. Snyder, (2010) "The vector innovation structural times series framework: a simple approach to multivariate forecasting", *Statistical Modelling* 10 (4), 353-374

Taylor, S.J. (2008), *Modelling Financial Time Series*, World Scientific, Singapore.

Tsay, R.S. (2010), *Analysis of Financial Time Series*, Wiley, New York.

Thomson, P. and T. van Zijl (2010), *A statistical forecasting framework and models for the determination of starting price adjustments for default price-quality paths*, Report prepared for the Electricity Networks Association, 8 December 2010;