nzier
authoritative analysis

Electricity Demand

Forecast model review

Report to the Electricity Commission

October 2004

Preface

NZIER is a specialist consulting firm that uses applied economic research and analysis to provide a wide range of strategic advice to clients in the public and private sectors, throughout New Zealand and Australia, and further afield.

NZIER is also known for its long-established *Quarterly Survey of Business Opinion* and *Quarterly Predictions*.

Our aim is to be the premier centre of applied economic research in New Zealand. We pride ourselves on our reputation for independence and delivering quality analysis in the right form, and at the right time, for our clients. We ensure quality through teamwork on individual projects, critical review at internal seminars, and by peer review at various stages through a project by a senior staff member otherwise not involved in the project.

NZIER was established in 1958.

Authorship

This report has been prepared at NZIER by Simon Hope and Preston Davies. The assistance of Doug Steel is gratefully acknowledged.

8 Halswell St, Thorndon

P O Box 3479, Wellington

Tel: +64 4 472 1880

Fax: +64 4 472 1211

econ@nzier.org.nz

www.nzier.org.nz

Executive Summary

Part F of the Electricity Governance Rules 2003 requires the Electricity Commission to publish demand forecasts as part of the grid planning assumptions - which underpin the Statement of Opportunities (SOO). The Electricity Commission has produced a paper outlining the process and methodology used to review and select the model to be used for the Commission's forecasts. NZIER was engaged to:

- critique the methodology used to evaluate and select alternative models;
- assess how reasonable the models are, and the validity of the inputs; and
- if required, suggest alternative models or approaches.

In reviewing the paper, we have kept in mind the purpose of the forecasts, and the audience to which they are aimed. This meant the report was prepared in cognisance of the trade-off between technical robustness and tractability/ease of use. As a result of this, the tone of the report is suggestive of the need for maintenance and continued improvement rather than substantive change.

In general, we did not detect substantial problems, omissions or errors which would require fundamental reconsideration of the methodologies used to evaluate and choose between alternative models. We do however, have some suggestions which could usefully augment the existing approach. These include additional specification tests such as testing for omitted and/or irrelevant variables, non-linearity's and overall model fit. We also highlight the need to be aware of the purpose for which the models are being built, and utilising only those evaluation criteria most suited to that purpose.

We reviewed three separate and distinct models covering residential, light industrial and commercial, and heavy industrial demand. While it is desirable to construct models that suit the conditions for each particular sector, this has resulted in three models with differing structures. We suggest that this might make interpretation and understanding more complex for the target audience, which may detract from the desire for simplicity and transparency. Our model specific suggestions may assist in increasing the level of commonality across the models, if the Commission agreed that simplicity and transparency may be compromised otherwise.

While all preferred models performed relatively well against the evaluation criteria suggested, some possible points for consideration include:

- Testing for stationarity/cointegration – to avoid the possibility of spurious results, we should ensure that we do not regress one non-stationary variable against another. Tests for non-stationarity and/or cointegration ensure we can make valid inferences about model coefficients. This is a consideration for the residential and commercial and light industrial models.

- Underlying parameter trends – we suggest some testing of the possibility of underlying trends in the relationship between the dependent and independent variables, which is assumed fixed at an average in the regression approach. This would aid in detecting the possibility of under or over-estimating future demand based on fixed point parameter estimates. Again, this is a consideration for the residential and commercial and light industrial models.
- Explanatory variable suitability – there may be additional value from substituting domestic consumption or household income for total GDP in the residential demand model. Similarly, industry GDP and some representation of price could help enhance the fit of the commercial and light industrial model. The heavy industry model may benefit from testing of trading partner GDP and real exchange rates in a regression approach.
- Smoothing demand via lagged dependent variables – it would be useful to reassess the use of a lagged dependent variable in the commercial and light industrial model. As noted in the Commission's report, it tends to dominate the regression and there are both intuitive and theoretical reasons to question its inclusion. Models with a lag included require specific diagnostic tests.
- The use of naïve forecasts across all models is a good test of how models perform against simple alternatives. Using them consistently would allow 'triangulation' between existing (Transpower) models, the preferred Commission models and the naïve forecast - to boost confidence in model choice.

Our report also considered the validity of the Commission's testing of model and forecast uncertainty. We are comfortable that the Commission used a well established and theoretically robust method, resulting in plausible bounds for selected inputs.

Finally we considered some long-term approaches to demand modelling, that could be applied when resources and commitments allow. These were exploratory and investigative rather than a signal of any structural weakness in the existing approaches. This reflects our view that the Commissions general approach is such that incremental improvement is favoured over substantial rebuilding.

Contents

1. Introduction

Part F of the Electricity Governance Rules 2003 requires the Electricity Commission ('the Commission') to publish demand forecasts as part of the grid planning assumptions - which underpin the Statement of Opportunities (SOO). The Electricity Commission has produced a paper[1] outlining the process and methodology used to review and select the model to be used for the Commission's forecasts.

The Commission has asked NZIER to:

- critique the methodology used to evaluate and select alternative models;
- assess how reasonable the models are, and the validity of the inputs; and
- if required, suggest alternative models or approaches.

The Commission has indicated that one of the major intended audiences for the forecasts is the general public. Having a readily available electricity demand forecast will enable interested parties to undertake submissions and prepare information using a common base. This will make the comparison/contrasting of submissions and other information more efficient and practical.

Thus, in terms of the trade-off between robustness/realism and tractability/transparency often encountered in modelling work, the Commission's preference is for the latter. We have prepared this review on this understanding.

In addition, attention is restricted to forecasting electricity demand only, rather than more general specifications involving combinations of energy types such as multi-industry or partial equilibrium models. These will be considered as part of a separate process to the compilation of these forecasts.

This review was prepared in the absence of any of the datasets or models themselves, and was conducted on the basis that possible policy changes that affect demand are outside the scope of consideration. They are considered to be inherent in the actual data, and are not allowed for in predictions of electricity demand.

[1] "Electricity Demand Forecast Model Review (Draft)"

2. Approach

This report commences with some overview comments about the general approach taken by the Commission and then proceeds to comment in more detail about the model selection criteria and the individual models themselves. Where there are obvious corrections or alternatives, these will be clearly spelt out, however, much of the commentary will be suggestive in nature. Accordingly, the report is written in a ‘conversational’ tone, rather than using formal notation and representations.

In accordance with the brief, our focus is on brevity. We have prepared the report almost on an “exceptions” basis, meaning that we comment mostly on areas where attention may be required, rather than all areas in the paper. This reflects our overall level of comfort with the general approach taken and the recursive nature of modelling. As a result of this approach, some of the discussion will seem negative. This is not our intention, as we have some empathy for the position of the Commission. Robust, defensible and universally useable forecasts are desired and required, yet they must be produced in as transparent and easily understood manner as possible. This is a difficult task. We have erred on the side of inclusion, and thus much of the material discussed may already be familiar to the Commission.

3. General comments

Model structure

The first general comment concerns the type of estimation method used. We see some merit in using a “regression-based” approach where independent variables thought to be drivers of demand for a dependent variable are used to explain the relationship between the variables. The obvious appeal of such a method is that it is transparent and grounded in both logic and theory. That is, it makes sense intuitively and there is some theoretical underpinning to the notion that demand for electricity will be influenced by variables that represent for example, the price, economic conditions, population and household size.

A possible issue is that there is little consistency across the respective preferred models in the paper. The residential model is a “pure” regression-based model; the commercial and light industrial model is something of a hybrid in the sense that it includes lagged values of the dependent variable, while the heavy industrial model is basically a trend model. If the Commission intend to present forecast of each sector using the separate models, it could be argued that this would entail more complexity in understanding than the use of a common model structure. While the paper acknowledges that no attempt was made to obtain a common structure, it is perhaps worth considering the costs and benefits of such an approach. We will return to this point in more detail below.

We note that, in terms of model selection, there is a tendency in the paper towards the models used by Transpower. While it is useful to have such benchmarks, we would suggest that some caution should be exercised in abandoning the search for useful models once replication of the Transpower results have been achieved. This is not a criticism of the Transpower models or their results, but rather a plea for some form of triangulation. We would suggest that the use of a naïve forecast would act as a relatively simple ‘reasonableness check’ for both the Transpower models and those used by the Commission.

We concur with the comments about more 'black box' type models such as neural networks, which are likely to be useful validation tools, but do not provide the level of transparency required under the Part F rules through the centralised data set. Utilisation of these models is not widespread or common enough to justify their use in presenting a base electricity demand forecast.

A final point concerns the utility of including the shortage variable in the models. While there are arguments for and against its use, we think more consideration should be given to its real value, particularly as the purpose of the modelling is forecasting, and by its very nature, it is difficult to forecast shortage years.

Residential modelling period

We are interested in the testing procedures used in deciding the modelling period chosen for the residential model. As is implicitly acknowledged in the paper, the start point of the model has a significant bearing on the estimated coefficients of the model (and ergo the forecasts), and it is desirable to include as much data as possible. Information that might be particularly useful in explaining long term relationships is lost by the choice of a later start period. The subsequent influence on demand of changes that occurred in a previous period (so called hysteresis effects) is also discounted by cutting off the data in such a way. There are relatively simple methods (such as using time dummy's, for example) available to adjust for what might be considered ‘structural breaks’ in the data and some testing of sensitivity of parameter estimates to different start points would be useful as a diagnostic tool. While sensitivity testing may have occurred and trade-offs made between the value of additional information and the ease of estimation, these were not reported in the paper.

We do not believe that truncating the data at the mid-seventies is necessarily problematic and concur with the general reasons proposed in the paper for doing so. However, in order to effectively ‘cover all bases’ it would be useful to estimate the models with the extended data to see what the effects are, or be a little more explicit about the ‘saturation’ effects and why the previous data is of less importance.

Use of Kalman Filter

While we agree with the use of filtering techniques to prepare the data for use in regressions, we think the paper would benefit from more discussion of this process.

4. Model selection and evaluation

We consider the evaluation and selection criteria used in the paper to be appropriate. Our understanding of the criteria and the measures used are as follows:

- Fit to historical data (R squared)
- Stability given uncertainty in inputs (Monte Carlo simulation)
- Performance of model when input data is truncated (hold out analysis)
- T values of individual coefficients

A general comment worth making is that the *purpose* of the model is important in evaluating the model chosen. A model designed for forecasting purposes should have as small a standard error of forecast as possible, while t statistics are more important in a model designed to test a specific hypothesis. Thus, caution should be exercised in necessarily ascribing model qualities to those where t statistics are significant.

One concern we have is the absence any discussion of other possible tests that may or may not have been used in model selection. Model building often involves trade-offs. In econometrics the most common trade-off is between bias and efficiency. In certain instances, the econometrician would prefer a slightly biased, but efficient estimator, to one that was unbiased but inefficient. Often this is related to the number of observations in the series (as bias is less of a concern with longer series'), and highlights again the potentially important issue of truncating the data series.

Below we make some suggestions on criteria and/or tests that encompass elements of the trade-off and could be used to augment those criteria/tests currently used. Some are more 'classical' in nature and apply more broadly to econometric model building while others refer more specifically to the situation at hand.

- **Use of evaluation measures** (and simulation measures). If forecast accuracy is the main objective, then an estimator that minimises *mean square error* (MSE) is potentially desirable. In terms of the efficiency-bias trade-off, MSE is particularly useful in that it encompasses both: $MSE = [Bias(\hat{\beta})]^2 + Var(\hat{\beta})$. Reporting of this, and other measures such as mean absolute errors or mean absolute percent errors would be useful.
- **F test**. It is often useful to perform statistical tests on numerous variables, as well as on the R^2 of the model itself.
- **Tests for omitted variable bias**. Omission of a relevant variable from a regression results in the estimates being both biased and inconsistent. This combination means that the bias will not disappear as the sample size grows larger. An iterative process is often useful, involving tests for irrelevant variables.

- **Tests for irrelevant variables**. The presence of irrelevant variables does not bias the parameter estimates but does affect the efficiency of the estimator, as the variance of the correctly included variable/s will be larger as a result of the irrelevant variable.[2] The loss of efficiency makes it more difficult to reject the null hypothesis of a zero parameter. Using F-tests on models thought to contain irrelevant variables allows one to test for omitted and/or irrelevant variables in an iterative sense.
- **Test for non-linearity**. Estimating a model that is linear in the explanatory variables when the true regression model is nonlinear is also a specification error. It is a special case of omitted variables and would be tested for in the same manner. A test for nonlinearity would involve first estimating a polynomial equation and test the significance of the individual variables.
- **Heteroscedasticity** (non constant error variances). While not usually a problem in studies involving time series, it is always good practice to test for its presence, as heteroscedasticity results in a loss of efficiency. This could lead to incorrect statistical inference (i.e. failing to reject the null hypothesis when it should be).
- **Residual plots**. The detection of data points (or indeed a variable) with an unusual influence is useful in terms of regression diagnostics. This is particularly relevant in the residential model where it is thought shortage years may exert downward bias on the forecasts of demand. Correcting for this possibility is a response to this influential data issue. In other cases, it may be the case that there has been an error in the coding or transcription of the data. Detection and evaluation of influential data points and influential variables is complex, but the use of DFBETAS and Studentised Residuals would assist (Pindyck and Rubinfeld, 1991).
- **Naïve forecast**. As mentioned in the general comments further above, the use of naïve forecasting techniques provide a test of reasonableness in most forecasting applications. It is relatively simple to project the series from the last data point, and then use this as a basis of comparison with more sophisticated techniques.

[2] While it is possible that a loss of efficiency won't occur, if the irrelevant and the correctly included variable/s are uncorrelated, this is very rare.

5. The models – model validity

This section discusses how reasonable the models chosen for use are and covers the validity of the inputs chosen. It is more specific in nature than the selection and evaluation section above.

5.1 Residential

5.1.1 The model and suggested investigation

In terms of the models reported in the paper, we believe the correct decision was made to use the Single Stage Log V2 model. We believe that the inclusion of the lag variable in the 2 Stage log V1 (Transpower) model is potentially problematic, and more importantly, adds an element of complexity to the interpretation of the equation. While the R^2 in the Transpower model is increased relative to the Single Stage Log V2 model, it comes at the cost of precision. The lack of significance (at conventional levels) of the other variables in the Transpower equation raises some questions which will be canvassed further as part of section 5.2.1 on the commercial and light industrial model.

While the significance of most of the variables in the Single Stage Log V2 model is encouraging, we would suggest that household consumption or household income would potentially be a useful inclusion. The GDP variable may contain too much non-residential noise to be as useful (notwithstanding it being highly significant).

We also suggest that some testing take place to assess the ‘fixed’ nature of the coefficient estimates. By virtue of using the regression approach we are assuming that the relationship between the dependent and independent variables stays fixed over time, but we may be missing trends in the coefficients themselves. For instance, the price elasticity of demand for electricity may be entirely different in 10 years time to what it is now, and may have been quite different 10 years ago as well.

Using a variation of a Chow test to estimate the coefficients in say, 10-year blocks and comparing these will give some idea as to how the coefficient is changing over time. If it is relatively stable/fixed, then there is no need for concern. However, if there are big discrepancies it may be useful to include an interactive term, such as GDP*time in the model. While this interactive term would be slightly more complex to interpret (i.e. the overall influence of, say, GDP would be the sum of the coefficient on the ‘ordinary’ GDP term, and the coefficient on the GDP* time term), it may capture more correctly any movements over time in the elasticities, and more accurately predict future demand based on the elasticity.

We suggest that the influence of the shortage variables should explicitly be tested for. The concern in the paper was that failure to account properly for shortage years would lead to downward bias in the forecasts. The question is whether it is sensible to exclude shortage years or not. Given that shortages are 1-in-60 year events, there is a possibility that for the forecast period we would not witness another shortage year. However, we have seen two in the last decade and so there may be some justification for their inclusion in the modelling process.

More importantly, it is not clear from the paper whether any tests were conducted for the possibility of non-stationarity in any of the variables in the models (i.e. that they follow random walks). Regressing one random walk against another can lead to spurious results, in that conventional significance tests will tend to indicate a relationship between the variables when in fact none exists. The usual approach to random walks is to difference a variable before using it in a regression (e.g. in the case of two variables integrated of order 1, use y_t-y_{t-1} rather than y_t as the dependent variable, or x_t-x_{t-1} as an independent variable).

However, this often results in a loss of information about the long run relationship between two variables (Pindyck and Rubinfeld, 1991). This differencing process (and consequent loss of information) need not take place if the variables in question are co-integrated. That is, a linear combination of the non-stationary variables is stationary (i.e. does not follow a random walk).

We would suggest that if tests for co-integration and/or random walks have been undertaken that the paper make this explicit. If no such tests have been undertaken, we would recommend that they were. While the variables used as drivers of demand for electricity have some theoretical basis for inclusion, and thus would tend to minimise the likelihood of spurious regression, it is perhaps more than merely an academic point to test for co-integration and/or random walks.

5.2 Light industrial and commercial

5.2.1 The model, and suggested investigation

Given the relative magnitude of this sector, careful consideration should be applied to the preferred methodology.

The paper notes the similarities between the explanatory variables used in this model and those used in the residential model i.e. GDP, price and a shortage flag. One area which raised uncertainty is the use of a per capita measure of Total Commercial and Light Industrial demand. The logic behind introducing a per capita measure is not made clear – and suggests a potential relationship between demand and population, in relation to a commercial demand variable. A per capita measure seems logical for the residential model but for this category demand is likely to reflect output and price rather than being driven by population.

Consideration should also be given to replacing total GDP as a commonly used explanatory variable for this group, with industry GDP. This will remove the demand relating to households and thus should increase the amount of the variation in the dependent variable explained by the GDP variable. The comment about using a per capita measure for the dependent variable also applies to the measure of GDP i.e. the logic behind such a ratio is not made clear, and is unlikely to be as useful as the direct level (or growth) in national industrial GDP.

The various models outlined via single and two stage, linear and log combinations are evaluated in terms of the model selection criteria outlined earlier, with four being selected for additional scrutiny. Particular focus is paid to t-statistics for individual coefficients, the fit of the model to historical data (presumably through examining the R^2 statistic) and the fit of data when using truncated out of sample testing (hold out analysis). Consideration could also be given to detailing any other model selection criteria considered, as well as those outlined as potential criteria earlier in section 4.

For the models where no lagged dependent variable is considered as explanatory, the relevant statistical tests applied should be accurate. The explanations of suitability here seem intuitive and reasonable.

Where we have some concern is in the validity of the same statistical tests in the presence of a lagged dependent variable, as considered in all the two stage linear and log models. In terms of the four models selected for further investigation, three contain lagged dependent variables. The earlier explanation contends that the first stage of the two stage process generates the lagged variable and 'smoothes' the demand so it can be used in the second stage of the regression. This is suggested as a solution for variable data which is primarily caused by factors which are impractical to model.

Once the three models containing lags are run the lagged variable is highly significant (in terms of its t statistic) with coefficients of close to, or over 0.9 in all cases and with low standard deviations. This is a point to note in terms of the implications on the explanatory power of each of the independent variables, and the link with the reasoning behind having a two stage process to generate the lagged dependent variables.

The first stage regresses electricity demand in the previous period on explanatory variables of time (Year), GDP and a shortage flag. The regression would generate an error term, which we would expect to be fairly sizeable given that the independent variables would not do a good job of explaining the fluctuations in demand – the causes of which were noted as being impractical to model. Once the resulting estimates were generated and substituted into the second stage equation, the variation and error would not be reduced any further. Basically, the error term in the current period regression would be unaffected by the smoothing process, and therefore raises questions around the utility of the lagged dependent variable.

In effect, the two stage process won't help to explain the variance that couldn't be modelled via a single equation using the same explanatory variables. We note however, that the variables in stage one differ slightly from those used in stage 2 – a time variable being present in the former. We are uncertain as to the reasoning behind the slightly different specification, and the potential implications (although they are likely to be small – it is likely to just make the demand series smoother).

Additionally, while using the estimated values (as opposed to the actual values) for the lagged demand variable would stop the coefficient estimates being biased, the coefficient on the lagged variable tends to be so significant that it dominates the influence of the other explanatory variables. As noted in the paper, this variable tends to dominate the other, more intuitive variables. For example, in the Transpower 2 stage linear V3 model and the 2 stage linear V1 model, GDP per capita and GDP (respectively) are shown to explain little that the lagged dependent variable couldn't explain, via low t statistics and small coefficients. The R^2 values are still high, as the combination of the lagged variable and the other explanatory variables are explaining much of the variation in demand, but the composition of explanatory power is swayed heavily toward the lagged variable.

The paper explains this effect as being caused by the relatively stable nature of commercial and industrial demand but it is not immediately clear why this period's demand should fundamentally be related to that of the previous period. In part, the effect of the lagged variable should be somewhat picked up in the GDP variable itself, as production can shift in an inter-temporal manner in response to shocks.

By having such a significant variable (with a coefficient close to 1), it should be carefully considered as to what this implies in terms of considering a more ‘pure’ time series model (such as a simple centred moving average) – where the relationships between demand in subsequent periods can be better modelled.

In contrast to this, consideration could be given to seeking better specification from a more simple regression equation with no lagged dependent variable, but GDP and an error which will represent the remaining volatile element to demand as mentioned in the report. This would allow the significance of the remaining explanatory variables to be better identified, with transparency and consistency about the inability to model the remaining variability.

By having essentially a combination of these two approaches the explanatory power of the other explanatory variables is somewhat reduced by the lagged variable – where the relationship between the adjacent demand periods is not necessarily firm. Having such a dominant lagged variable also introduces a considerable dynamic year-on-year element to the model, which may not be ideal in a long-term forecast model with a purpose such as prescribed here.

The 2 stage linear V5 model is slightly different in this regard, in that the GDP coefficient is relatively significant (different from zero). The lagged variable could be removed to assess the significance of the remaining variables without it, in a similar fashion to the single stage linear V1 model. It was noted that the GDP coefficient was significant but there was still an unexplained component which made the model appear less viable. We suggest that, as noted above, this unexplained component can be attributed to an error term and made transparent. Although it may reduce the R^2 value (not provided) the relationship may be more logical and intuitive with the presence of the variance which was impractical to model. We also suggest that the relatively low level of sensitivity to input variation may be caused by the use of smoothed demand. Even though the sensitivity of a model without the lagged variable may appear higher, it is likely to better reflect the known structural relationship between GDP and commercial and light industrial demand plus an error component.

These comments should not be seen as severe criticism of the current approach; they are more points to consider and to take into account when interpreting coefficients, forecasts and in ensuring that the need for transparency and clarity is measured against keeping it simple and robust.

Three of the four models shown are shown (via a Durbin Watson statistic) to exhibit autocorrelation in the error terms, including two of the two stage models. It is not known for these two stage models, whether the Durbin Watson statistic is a pure DW statistic or the Durbin *h* statistic. When one or more lagged dependent variables are present, the pure Durbin Watson statistic may be misleading and the Durbin *h* statistic should be used. The Durbin *h* statistic is approximately normally distributed with unit variance. The statistic can be defined as[3]

$$h = \hat{\rho}\sqrt{\frac{T}{1 - T[Var(\hat{\beta})]}}$$

Where

$\hat{\rho}$ = estimated first order serial correlation coefficient (can be estimated from the DW statistic)

$Var(\hat{\beta})$ = square of the standard error of the coefficient of the lagged dependent variable

T = number of observations

If the Durbin *h* statistic is not used, it should be calculated and the resulting statistics interpreted to see if the inferred implications differ from those stated in the report. The results are not expected to differ significantly, although the inconclusiveness of the null hypothesis of no serial correlation may need revising. This would have implications for the other criteria used to evaluate the 2 Stage linear V5 model.

The report notes that a number of the models assessed included a price variable, which turned out in most cases not to be significant, or in some cases it produced a non-intuitive sign. We concur with the reports note that a price variable is likely to be important in relation to demand in this group. Consideration could be given to including a relative price variable, taking into account changes and levels of electricity prices compared to other input prices.

Another possible source for further model refinement could come from looking into the changing income elasticity of demand over time for this category. The regression models as stated essentially use the average value of GDP in creating the estimates, whereas there may be a changing trend in income/GDP to electricity demand for this group over time.[4] A version of the Chow test is able to test for structural change that could be evident over the time period being considered. If a trend in the income elasticity of demand is identified, it may be possible to introduce a GDP/time combined variable into the equation to take account of the change in slope over time.

---

[3] Pindyck and Rubinfeld, 1991

[4] There could also be difference in the intercept values in different time periods as opposed to just different slopes

5.3 Heavy industrial

5.3.1 The model, and suggested investigation

The paper identifies problematic and sketchy relationships between heavy industrial electricity demand and GDP and population as explanatory variables (although no rejected models are presented). The report also notes that growth in the sector is likely to be driven by factor costs and by conditions in the markets which they operate in.

In forming the model recommended in the paper, the Electricity Commission has fitted a trend variable, along with a shortage flag variable. All coefficients are statistically significant. The Durbin Watson statistic indicates significant positive autocorrelation is not an issue. This is probably not an unreasonable method for estimation, given that demand in this category is fairly lumpy. However, we do feel there is some scope for further investigation and testing.

We concur with the statements about the difficult in identifying relationships for this group, but suggest the likelihood of relationships between heavy industrial demand and factors relating to the export market could be identified in a regression model. For example, the economic growth of the major trading partners for buyers of heavy industrial output from New Zealand is likely to be influential on how much electricity these users demand. As most of the producers in this category are likely to be exporters, the real exchange rate will be another factor that could potentially explain some of the variation in heavy industrial electricity demand. Data on these variables should be readily available.

If variables such as these were considered, there may be scope for including domestic GDP to help explain a small amount of the variation in demand, from the point of view of the Think Big projects etc. which were identified as reasons for shortening the time period considered. Increased production in such projects would feed through domestic GDP, and including this variable may help improve the fit of the model to the actual historical data. Industry GDP would be preferable to total domestic GDP, as it would help reflect the varying growth rates of different industries. This would be particularly useful if we expect significant changes in the composition of heavy industry electricity demand (for separate industries within the heavy industry grouping) in the future.

Intuitively, we would also expect that price would be a major factor for heavy industrial users. Indeed the report itself notes that "changes in price are more of an issue for those large industrials where electricity is a significant contributor to total production costs". Availability of data may be an issue in this case, but given that heavy industry are likely to be generally large users, they may be able to respond to significant price changes by closing plant/winding down production. This would only occur when prices were prohibitively large such as in a dry year.

6. Modelling uncertainty

The Commission has wisely looked to assess the level of modelling and forecast uncertainty associated with electricity demand. They note that the underlying model accuracy sought needs to be balanced with an ease of intuition and practicality.

In assessing the level of uncertainty in both the estimates of the historical relationships and of forecast estimates, the Commission has chosen to utilise Monte Carlo simulation. This process uses estimated distributions for the independent variables, and runs random samples of these input distributions in place of the actual input data in an iterative process. Running this process a number of times produces bounds of uncertainty based on the inherent uncertainty in the inputs, and thus the models themselves. Confidence levels can then be applied to these bounds. The Commission has included useful charts of the bounds of uncertainty resulting from Monte Carlo simulation for the historical estimates, and for the forecast estimates.

In terms of the overall approach, we agree with the use of Monte Carlo simulation to assess uncertainty, as it gives practical, interpretable results and means that individuals can determine their own views on how acceptable they feel the risk around a central model may be.

The Commission uses synthetic distributions for each of the input series in assessing model error. In doing so, they examine the variation of the input against a short term trend. In this case they use a 5 year moving average. This approach seems reasonable, and the only suggestion we make here is to vary this moving average and assess how this alters the level of variation from this trend i.e. will a 4 or 6 or 7 year moving average trend significantly alter the composition of the synthetic distribution?

In terms of forecast input uncertainty, the explanations for each of the various inputs seem in general quite sound. There will always be uncertainty about the distribution of the variation in the inputs, but the Commission seems to have made a concerted effort to allow for uncertainty and to make their allowances clear.

The separation of uncertainty in GDP is a useful idea, and having a component which incorporates uncertainty around the external environment is particularly important. The household uncertainty discussion is likewise quite sound, covering the two important factors of household size and population. It is agreed that the Statistics New Zealand scenarios would be effective at providing variation in the population input, and it is not thought that significant additional value would be gained by refining it to include birth rates etc. The report does not allow us to get a good insight into the actual distribution for the price input, but the effective price elasticities do not appear excessive or understated. It would be valuable to try and investigate the points where light industrial and commercial and major industrial users will exit the industry because of prohibitive price increases.

The point about the forecast data effectively incorporating an underlying rate of efficiency improvement is agreed, and for the purposes of these forecasts, the current approach seems reasonable (as long as it is clearly stated as is done in the report). We agree that only widely accepted future changes in these levels should be allowed for in forecasts.

The method behind the combination of the total forecast uncertainty is not elaborated in the report, however it is important that the implications of the aggregation are known – as taking the sum of the absolute values would provide a different consolidated level of total uncertainty (and may weight certain levels of uncertainty differently) than would other methods. This is not expected to make a considerable difference to overall uncertainty however.

One issue to consider in relation to this method for assessing uncertainty is the effect of the two stage process used in some models, on the Monte Carlo simulation bounds. If the lagged dependent variable is created to smooth the series, then it may create smaller bounds of uncertainty than would otherwise be the case (if there was no lagged dependent variable). It is not clear whether this may bias the bounds and make the level of uncertainty appear smaller, but it is an issue the Commission may wish to consider.

7. Longer term modelling possibilities

The discussion above looks at the models investigated by the Electricity Commission, and in doing so makes some incremental suggestions for further consideration. These suggestions should be interpreted as such, and not major criticisms of the underlying methodologies recommended. Indeed some of the suggestions are more to rule out alternatives rather than to discount the method preferred by the Commission.

In following on from these shorter term suggestions, there is some scope for recommending longer term considerations. Given additional time, modelling capacity and possibly additional information, the Commission may be able to further progress and refine their forecasts by:

- Looking into the possibility of using a systems dynamics type approach to modelling demand, whereby complex systems of dynamic interactions (such as those involved with electricity demand) can be modelled to incorporate feedback loops (particularly in terms of the impact of price).

They are also useful in that they incorporate stock and flow elements and are very practical and intuitive in nature.

- Considering partial or general equilibrium models. This would allow the Commission to model multiple interacting agents where individual behaviour is based on optimisation, and where interactions between agents are mediated by markets and prices. This would be a resource intensive option, but may provide considerable long-term value for planning purposes.
- Investigate further the possibility of using industry panel data to model electricity demand over time. It may be possible to break the light industrial and commercial, and heavy industrial sectors down by industry which may increase the ability to model industry specific expectations. There may be constraints in terms of data, but a lower level of aggregation could increase the flexibility of the model.