A review of the methodology and models used by the New Zealand Electricity Commission for national electricity demand forecasting

undertaken for the
New Zealand Electricity Commission
by

Peter Thomson
Statistics Research Associates Ltd
Wellington, New Zealand

8 February 2010
Revised 17 February 2010

Contents

Summary and recommendations

A review of the current methodology and underpinning models used by the New Zealand Electricity Commission to forecast national residential and commercial/industrial electricity demand has been undertaken. This includes an independent replication of the existing residential and commercial/industrial econometric regression models, and suggestions for improvements to the models. Other models and procedures have also been proposed and developed, particularly for growth rates of demand. A selective literature review is used to support the findings, and discussion given on the nature and objectives of a more general modelling framework suitable for long-term national electricity demand forecasting.

The review makes the following recommendations.

Recommendation 1 *With regard to the Commission's current long-term forecasting models and strategies, it is recommended that:*

(a) more comprehensive, systematic, technical documentation be developed that provides full details of the forecasting models adopted, their objectives and assumptions, the data used and any pre-processing undertaken, the validation procedures used, and full mathematical formulations of the models within an overall statistical framework;

(b) a unified modelling framework based on log transformed data be adopted for the current residential and commercial/industrial demand models, with a view to simplifying the former and improving the accuracy of the latter;

(c) more comprehensive validation procedures be developed to assess the accuracy and reliability of demand and driver forecasts, the quality of sample paths, sensitivity to driver perturbations, forecast error attribution, and the maintenance of any historical long-term cointegrating relationships in the forecasts;

(d) a more secure, systematic and consistent framework for generating synthetic realisations of future demand and driver time series be developed based on the Commission's existing general strategy.

Recommendation 2 *With regard to other long-term forecasting models and strategies, it is recommended that the Commission:*

(a) develop a suite of competing forecasting models based on growth rates as well as levels, and use the combination of these forecasts for long-term forecasting and risk assessment;

(b) develop and assess forecasting models which systematically model slowly evolving parameters and relationships over time, such as those given by equations (9) and (11) in the review;

(c) explore the development of vector forecasting models with common trends that take explicit account of cointegrating relationships between the various demand and driver series.

Further details are given in the report.

1 Scope of review and terms of reference

The New Zealand Electricity Commission (Commission) seeks an independent review of the current methodology and models it uses to forecast national residential and commercial/industrial electricity demand. This review of the underpinning methodology and models, rather than the forecasts themselves, was prompted by recent updates to New Zealand population projections used as drivers in the Commission's residential electricity demand model which have resulted in a significant jump in forecast residential demand.

Statistics Research Associates Ltd (SRA) has been commissioned to undertake this review which should include:

(a) independent replication of the existing residential and commercial/industrial econometric national demand models;

(b) development and assessment of alternative model formulations considered appropriate by SRA including, but not limited to, models incorporating time evolution;

(c) provision of advice and recommendations with respect to the continued use of the existing Commission models and possible alternatives.

Information provided for the review included Kirtlan (2008), NZIER (2009) and a number of earlier documents on demand forecasting available from the Commission's website (see *http://www.electricitycommission.govt.nz/opdev/modelling* in particular). Brian Kirtlan (Commission) also provided access to the most recent forecasts based on data to March 2009, and the extensive MATLAB code used for generating the Commission's demand forecasts and forecast distributions.

The review is documented in the following sections together with further comments and recommendations.

2 Background

The New Zealand Electricity Commission (Commission) is required to publish a Statement of Opportunities (SOO) which includes, among other things, national electricity demand forecasts. These and other requirements have led it to develop long-term econometric point forecasting models for mean national electricity demand (point forecasts), and also for the distribution of future demand about the mean forecasts (predictive distributions or density forecasts).

A primary purpose of these long-term electricity demand forecasts is to enable identification of potential opportunities for efficient management, including investment in upgrades and transmission alternatives, and long-range capacity planning for New Zealand's electric power system and transmission network. National electricity demand forecasts are key inputs to such long-term planning exercises. Since the latter will need to account for a variety of overarching views (scenarios) of New Zealand's future energy requirements, the electricity demand forecasting models used need to be amenable to the development of such scenario forecasts.

Long-term forecasting of this kind is based on the historical records of electricity demand, and any key macroeconomic or demographic explanatory variables that demand may depend on. The latter include New Zealand's Gross Domestic Product (GDP) and population among others. Such predictors or drivers can be used to provide conditional forecasts of electricity demand and inputs to scenario forecasts. The Commission's demand forecasts and models have been reviewed and refined several times over recent years, particularly in a sequence of reviews by the New Zealand Institute of Economic Research (NZIER) (see NZIER, 2009, for example). The appropriate choice of relevant predictor variables has been thoroughly considered in these and other reviews and is not explicitly considered in this review which is focussed on the methodology and models used by the Commission, and the general modelling framework adopted.

The electricity demand forecasting models also need to take proper account of other quantitative and judgemental information on the likely path of New Zealand's long-term future electricity demands and the economic and demographic drivers on which they are based. The views of an informed consensus of relevant experts (economists, energy planners etc.) will inevitably moderate and shape the long-term forecasts and the nature of the various scenarios considered. As a consequence, the forecasting models adopted need to be amenable to the incorporation of such judgemental information, and be sufficiently transparent that any forecast errors that result can be accurately attributed to the source of the error, either the formal model adopted or any judgemental or other input used as a driver.

Currently the Commission produces forecasts of annual electricity demand out to 2050 using econometric regression models fitted to around 30 years of annual historical data (35 years for residential demand and 23 years for commercial/industrial demand). The forecasting horizon of approximately 40 years is very long by comparison to the length of the estimation period (for models fitted to trending data of this length, horizons up to 5 years might more generally be regarded as short to medium-term, and 5–10 years as medium to long-term). However, although forecast horizons of 5 years and 10 years are important to the Commission, its national electricity demand forecasts are primarily used as key inputs to the Generation Expansion Model (GEM) used to schedule the commissioning and construction of new electricity generation. Currently GEM minimises discounted capital expenditure and ongoing operating and maintenance costs over a time horizon of 32 years. Such very long-term or far-term forecasting poses many challenges, not least validation.

Quantitative and qualitative validation of long-term forecasts and the models they are based on is a very important, but challenging, exercise. The longer the forecast horizon the more difficult this becomes due to many factors including, but not limited to, historical data limitations, the impact of technological innovation, evolving demographics, changing national and global economies incorporating evolution as well as unforeseen structural breaks, and a changing physical environment, including climate change. In practice long-term forecasting methods and models can usually be validated reasonably adequately for horizons up to 5 or even 10 years, but become increasingly more difficult, if not impossible, to validate over longer forecast horizons.

In summary, key principles governing the choice of long-term forecasting models and a suitable modelling framework include the following.

- **Well-defined forecast objectives** and forecast horizons.
- **Quality data** over a reasonable time period which should preferably exceed the forecast horizon.
- **Transparent forecasting models**, within an overall modelling framework, that are sufficiently simple and open that they are well-understood by all stakeholders, encompass sufficiently broad views of the future, and which capture the major long-term historical joint variation of all relevant variables over time.
- **Careful validation** of the forecasts produced at key horizons, both quantitatively and qualitatively.

In practice, forecast models based on such principles are more likely to lead to an informed consensus of long-term views on future electricity demand and, as a consequence, useful long-term forecasts and enlightened forward planning.

2.1 Selective literature review

In general, medium and long-term forecasting of energy demand has received considerably less attention in the peer-reviewed academic literature than short-term forecasting. What literature exists is sometimes of uneven quality from both methodology and validation points of view, with more of a focus on new models and methods rather than careful comparative evaluation. When it comes to very long-term horizons such as those considered by the Commission, the literature is even less informative. Nevertheless there are a number of useful reviews and papers that are relevant to the forecasting and modelling strategy adopted by the Commission. A selective review of these papers and recent literature follows.

The book Armstrong (1985) provides a general reference to long-term forecasting. In the context of long-term energy forecasting, a useful starting point is Craig et al. (2002) which provides an insightful overview of how very long-term energy forecasts (those covering two or more decades) are created and why they are useful. It focuses on energy demand in the US, but the general thrust is highly relevant to many other countries including New Zealand. The paper makes the following pertinent observations, among many others.

- Very long-term energy forecasts (two or more decades) are not able to be validated in any conventional quantitative sense, so that confidence in the resulting forecasts rests on the nature of the forecast construction, its fitness for purpose, and the underpinning models.
- Forecast models and assumptions need to be clearly and comprehensively documented so that the models are readily understood, evaluated and reproduced by others.

- Accurate data compilation and careful construction of relevant scenarios are more important to achieving forecasting objectives than complex models and methodology. The importance of model transparency cannot be overemphasised.
- Do not rely on one forecasting model. Combinations of forecasts or forecast averaging is a simple, pragmatic and instructive way of generating more robust and accurate long-term predictions than any single method. Granger (1989) provides arguments as to why this is generally the case, and Huss (1985) shows that, in the context of medium to long-term energy forecasting and a range of forecasting methods, forecast averaging generally outperformed all the other methods.

A primary finding of Craig et al. (2002) was that the forecasters in the period 1950–1980 underestimated the importance of unforeseen surprises such as the ability of the US economy to respond to the 1970s oil embargo by increasing efficiency. Not only were the forecasts of that period systematically high, but their uncertainty was also systematically underestimated.

Huss (1985) compares the results of medium to long-term electricity demand forecasting methods applied across a selection of forecasting models applied to data from 49 large US utilities for forecast horizons of 2, 4, 6 and 11 years. Although somewhat dated and the data available limited, the conclusions of this study are still relevant since the methods compared continue to be used in practice. These methods included four univariate time series methods (no drivers apart from time itself) and linear regression with much the same drivers as those considered by the Commission. The time series methods considered were a simple linear time trend estimated using ordinary least squares, exponential smoothing, adaptive trend estimation, and a forecast combination that averaged the previous three methods. For the regression method, the drivers were either forecast using simple exponential smoothing (providing ex ante demand forecasts), or the actual values of the driver were used (providing ex post demand forecasts with perfectly predicted drivers, a situation normally impossible to achieve in practice). Among other conclusions, it was found that the univariate time series forecasting models considered performed better than the econometric linear regression model, although the performance of the latter improved with longer forecast horizons, and forecast combination or forecast averaging outperformed the other time series methods. Of the methods used by the utilities considered, Huss (1985) also found that econometric techniques failed to outperform trend extrapolation/judgemental techniques.

Over the last two decades, other energy forecasting models and methodologies have been developed and considered, most being a variant of those considered by Huss (1985), but some are sufficiently different in concept and form to warrant further investigation. The latter category include vector time series models which attempt to simultaneously model both dynamic and long-run equilibrium relationships between trending variables. Among these are vector autoregressions (VAR models) and vector error correction (VECM) models both of which are commonly used for short to medium-term econometric forecasting. In particular, VECM models take account of cointegration where, for example, demand and its drivers are linearly related to a common trending variable (scalar or vector) that is typically unobserved and must be estimated from the data. Allen and Morzuch (2006) provide a general perspective of progress and problems with these econometric forecasting models since the 1980s. Dergiades and Tsoulfidis (2008) give an example of fitting such a model to US residential electricity demand to explore the interrelationships between demand and its drivers, although not necessarily in the context of long-term forecasting.

A recent paper by Hyndman and Fan (2009) consider long-term (up to 10 years ahead) point and density forecasting of South Australian peak electricity demand based on half-hourly measurements made over the summer months November to March. Semi-parametric additive models are used to fit relationships between the logarithms of demand (excluding major industrial demand) and drivers that include temperature, calendar, demographic and economic variables. Forecasts out to 10 years ahead were made using a combination of temperature simulation, economic scenarios, and residual bootstrapping. Their model decomposes half-hour variation in the logarithms of demand from annual variation with the annual base demand levels estimated using linear regression. For the latter, the best model included South Australian GDP, lagged average price, and cooling degree days. Model selection procedures and diagnostic plots were used to check the goodness of fit of the model in-sample, and the evaluation of 1 year ahead out-of-sample forecasts was also undertaken. Although the Hyndman and Fan (2009) model proposes a number of interesting innovations, especially for seasonal time scales and the construction of forecast densities, its base annual demand model is not dissimilar from that proposed by the Commission.

As noted in NZIER (2009), econometric linear regression models similar to those adopted by the Commission are widely used in practice and continue to be discussed in the academic literature. See, for example, Chui et al. (2009), who consider forecasting long-term electricity demand in Ontario, Canada, and Mohammed and Bodger (2005) who consider forecasting long-term electricity demand in New Zealand. In such studies, electricity demand and any economic drivers such as GDP, are typically transformed by applying the logarithm transform. The justification for this transformation comes mainly from the time series properties of electricity demand, but such a transformation is also consistent with considering a multiplicative Cobb-Douglas production function (see von Hirschhausen and Andres, 2000, and Dergiades and Tsoulfidis, 2008, for example) which can be fitted using linear regression methods after taking logarithms.

As has already been mentioned, validation of long-term energy forecasting models is a challenging, but very important, exercise. A case for thorough retrospective analysis of past performance of long-term energy forecasting procedures is given in Koomey et al. (2003). More recently, Lady (2009) considers the accuracy of long-term forecasts from a complex model, the US National Energy Modeling System (NEMS), using quantitative methods and an approximate model. The approach advocated is, in essence, a designed statistical experiment which can elucidate a number of useful properties of the NEMS model, including the nature of its response (sensitivity) to changes in driver forecasts. Lady (2008) also considers the important issue of forecast error attribution. What part of the forecast error in electricity demand is due to driver forecast errors, and what part is due to the model itself? A New Zealand study of this nature concerning tax revenue forecasting is given in Keene and Thomson (2007). Other important validation issues include the impact of constructing models and forecasts using recent data that is subsequently revised (for example, New Zealand's GDP and population size). Given that the most recent data is typically most heavily weighted in any forecasting exercise, data revisions have the potential to introduce significant distortions in long-term forecasts. In the context of short to medium-term forecasting, Faust and Wright (2009) consider the impact of such revisions on forecasting quarterly US macroeconomic data, and for a variety of forecasting methodologies.

3 Commission's national demand forecasting models

The long-term forecasting models used by the Commission are described in Kirtlan (2008) with supporting documentation available from the Commission's website. The MATLAB code used to generate the forecasts and the predictive distributions was also made available. Further written and verbal comments were provided by Brian Kirtlan in response to queries. This review is based on these information sources which are collectively referred to as the background documentation.

While the general objectives of the point and density long-term forecasts required by the Commission appear to be reasonably well covered in the background documentation (either explicitly or implicitly), they are not all in one document and there is a lack of sufficient detail to better understand the technical requirements of any forecasting model adopted. As noted in the previous section, clear, accurate and comprehensive documentation of the forecasting models and their objectives are key requirements of any forecasting exercise, particularly long-term forecasting.

Issues that deserve a fuller discussion and resolution include, but are not limited to, the following.

- **Objectives** The forecasts appear to be used for many purposes and a variety of horizons. These separate requirements need to be carefully elucidated from a forecasting perspective. Is point forecasting (forecasting mean demand) or the forecasting of high quantiles all that is required, or are synthetic realisations (sample paths or trajectories) of annual demand needed. If the latter is true, then the dynamics of annual demand may well need to be more accurately modelled; if not then simpler models can be used.

- **Forecast horizon** The current horizon of 2050 (42 years) exceeds the requirements for GEM (32 years) by a decade and is very long by comparison to the estimation periods used. Given that quantitative validation is normally only possible for horizons of 5–10 years, it would seem prudent to limit the effective horizon to 20 years (similar to NEMS described in Lady, 2009) with provision for longer horizons if necessary. The relative importance of intermediate lead times up to the forecast horizon also needs to be considered as this will have a bearing on forecast validation.

- **Econometric versus time series models** As noted in Section 2.1, the literature suggests that time series models perform at least as well, if not better, than econometric models which seek to determine the relationship between electricity demand and nationally important drivers such as GDP, population etc. The latter are typically useful for applications where economic, demographic and environmental scenarios need to be specified (high economic growth, high immigration, climate warming etc.) to better inform long-term planning. If these scenarios are not used or are less important than accuracy, then time series models may well be preferred.

Currently it would seem that the Commission documents each forecasting round (see Kirtlan, 2008, for example), but there is not a single, more technical, document that specifically provides details of the forecasting models adopted, their objectives and assumptions, why they are appropriate, full details of the data used and any pre-processing undertaken, details of the validation procedures, together with supporting arguments and appropriate (mathematical) model formulations. Such a document, essentially a purpose-built forecast manual, has many advantages. In particular, it helps ensure continuity of forecast quality, captures in-house technical knowledge so that it is no longer the preserve of any one individual, is essential for training new forecasting staff, and allows other interested parties (external or internal) the opportunity of providing informed comments as well as the opportunity to reproduce the forecasts or apply the procedures to other data. These comments are especially true here where it is clear that a significant amount of innovative work has been invested in a range of specialised forecasting procedures. These observations form part of the recommendations made in Section 3.4.

3.1 Model fitting and selection

Annual data spanning the period 1946 to 2008 were provided with national residential demand, commercial/industrial demand, residential electricity prices and domestic residences sourced from the New Zealand Ministry of Economic Development, and the national economic and demographic annual measurements (real GDP, population, CPI) sourced from Statistics New Zealand. For the most part, the annual data series were for March years (ending on 31 March of the given year) with the exception of population which was for June years and the CPI which was for calendar years. Although it is important to have all data sets on the same time scales, especially if dynamics are to be modelled, it is unlikely that small time shifts of the order of a quarter will make a major difference. Nevertheless, this is a potential source of error that can and should be eliminated.

However, to account for issues such as evolving relationships between demand and its drivers and any structural breaks in the data, the Commission has only used data from 1974 to 2008 (35 years) to fit the forecasting model for residential demand, and data from 1986 to 2008 (23 years) for fitting commercial/industrial demand. These intervals will be termed *estimation windows* in the discussion that follows. Accounting for evolution is necessary, but inevitably leads to the placing of more weight on the most recent observations and their interrelationships, effectively reducing the amount of data used for model fitting. The implication is that more recent relationships, rather than past relationships, are more likely to persist in the long-run, despite the fact that this may not have always been so in the past. This serves to throw yet more weight on the nature of the model and its relationships to any drivers.

For both residential demand and commercial/industrial demand, the Commission fits linear regression models of the form

$$Y(t) = \beta_0 + \sum_{j=1}^{p} \beta_j X_j(t) + \epsilon(t) \tag{1}$$

where t indexes years and the fits are obtained using ordinary least squares (OLS) over the respective estimation windows. For residential demand $p = 3$, $Y(t)$ is the logarithm of residential demand per capita, $X_1(t)$ is the logarithm of GDP per capita, $X_2(t)$ is the logarithm of domestic residences per capita, and $X_3(t)$ is the logarithm of residential electricity price in 2008 dollars (adjusted using the CPI). For commercial/industrial demand $p = 2$, $Y(t)$ is commercial/industrial demand, $X_1(t)$ is GDP, and $X_2(t)$ is a dummy variable indicating a year when an electricity shortage event occurred. The residual error $\epsilon(t)$ should represent non-systematic or random error with mean zero, constant variance and little, if any, serial correlation.

Note that the residential demand model uses the logarithm to transform demand and its drivers (called the log transform), whereas the commercial/industrial demand model works with the original (untransformed) variables. Given the range of the numbers involved, it will typically make little difference to the in-sample regression results whether the data are initially log-transformed or not. However, forecasts over the longer-term horizons will show a difference (typically the forecasts using the log transform will be lower, although any bias can be corrected for), and the model based on logarithms has useful economic interpretations that can lead to a better understanding of the model. In particular, the logarithm converts multiplicative and compound growth relationships that are prevalent in economics, finance and nature to additive relationships with annual differences that measure (continuously) compounding growth rates.

Using the log transform for residential demand, but not for commercial/industrial demand, seems an unnecessary dichotomy that would appear to lead to few, if any, forecasting gains. Although the graphical evidence is modest (see Figure 1, for example), plots of the data show that both demand series and their drivers have fluctuations about their long-term trends that would appear, for the most part, to be roughly proportional to the level of the series. This observation is an indication that the log transform would be a suitable transformation for both demand series and their drivers. In addition to aiding understanding, a common transformation also means that accounting for forecast errors in aggregate demand is a simpler and more straightforward exercise (see Keene and Thomson, 2007).

Figure 1 plots the original and log transformed demand series and their drivers where all series have been scaled by their respective values in 2008. All series show increasing trends from 1946 with the exception of commercial/industrial demand and the adjusted residential electricity price which have increasing trends from the mid 1970s. It would appear that there are significant linear relationships between these variables (original or transformed), particularly over the estimation periods, which implies that the demand

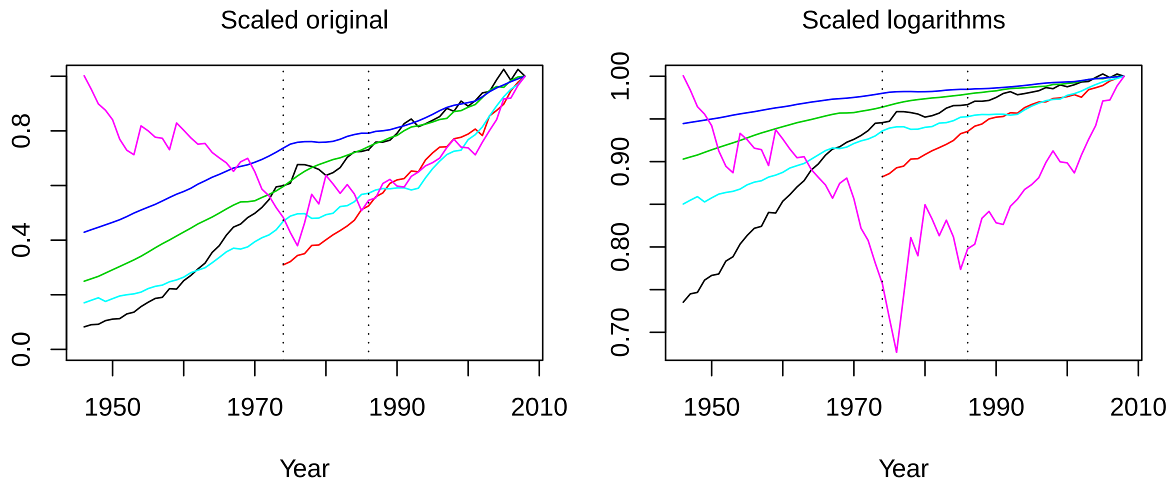

Figure 1: Plots of residential demand (black), commercial/industrial demand (red), domestic residences (green), population (blue), GDP (cyan) and residential electricity price in 2008 dollars (magenta). The left panel shows the original variables, the right panel shows the log transforms, and all series have been scaled by their respective values in 2008. The vertical dotted lines mark the starts of the residential (1974) and commercial/industrial (1986) estimation windows.

series and their drivers are likely to be linearly related to a common trend or trends. In particular, residential demand appears to be closely related to domestic residences since the early 1970s, and the logarithm of commercial/industrial demand appears to be closely related to the logarithm of GDP since 1990. Such considerations no doubt lay behind the choice of drivers for the demand forecasting models adopted by the Commission.

The results of fitting the 2009 residential and commercial/industrial demand forecasting models were replicated using R, a comprehensive statistical computing and graphical environment (see R Development Core Team, 2004), as were the 2006 and 2007 forecasting models described in Kirtlan (2008). In all cases and as expected, there was close agreement between the fitted models using R and those determined by the Commission who used MATLAB. The agreement was assessed by comparing the various numerical forecasts for 2009 provided by Brian Kirtlan and, for 2006 and 2007, by examining the coefficients listed in Tables 1 and 2 of Kirtlan (2008). The only exceptions were the values of the squared coefficient of multiple correlation (so-called R^2 coefficient) which measures the strength of the fitted linear relationship. The values determined by R were somewhat less than those quoted by the Commission due, perhaps, to a difference in definitions. Although not likely to be important, the cause of such differences needs to be identified and rectified if necessary.

The commercial/industrial demand model involves one driver (GDP), apart from the shortage dummy, whereas the residential demand forecasting model involves four drivers, although one (population) has been used to standardise demand, GDP and domestic residences into their per capita forms. The simplicity of the commercial/industrial model contrasts with the relative complexity of the residential demand model. As noted in NZIER (2009), the latter model is approaching the limits of what might be regarded as simple and transparent. For the 2009 residential demand forecasting model, the regression

Table 1: The Commission's 2009 demand forecasting models for the logarithms of residential demand and for commercial/industrial demand. OLS regression coefficients (coef.) and their standard errors (s.e.) are given.

Log residential demand			Commercial/industrial demand		
Variable	coef.	s.e.	Variable	coef.	s.e.
Constant	-3.531	0.408	Constant	1656	773
log(GDP per capita)	0.258	0.073	GDP	0.143	0.008
log(Residences per capita)	1.040	0.147	Shortage	-247	436
log(Price)	-0.133	0.052			

coefficients and their conventional OLS standard errors are given in Table 1. If the standard errors are any guide, then domestic electricity price does not play a significant role in the residential demand model (not surprising given the nature of its sample path in Figure 1), and the shortage years in the estimation window (1993, 2002 and 2004) have an insignificant impact on the commercial/industrial demand model. Omission of these variables would further simplify the respective models.

Using the basic properties of the logarithm, the Commission's residential demand model can be reformulated as

$$\log Y(t) = \beta_0 + \beta_1 \log G(t) + \beta_2 \log R(t) + \beta_3 \log P(t) + \beta_4 \log N(t) + \epsilon(t) \qquad (2)$$

where $Y(t)$ is residential demand, $G(t)$ is GDP, $R(t)$ is domestic residences, $P(t)$ is domestic electricity price, $N(t)$ is national population, $\epsilon(t)$ is non-systematic random error, and β_4 is constrained to satisfy $\beta_4 = 1-\beta_1-\beta_2$. Here the estimates of β_0, β_1, β_2 and β_3 are the same as those listed for the Commission's residential model given in Table 1. Evidently β_2 is very close to, and not significantly different from, one (a result also replicated for the 2006 and 2007 results). This implies that the Commission's residential model can be simplified to

$$\log(Y(t)/R(t)) = \beta_0 + \beta_1 \log(G(t)/N(t)) + \beta_3 \log P(t) + \epsilon(t) \qquad (3)$$

which involves fewer parameters and is just as easy, if not easier, to interpret. It is likely that this slightly simpler, more transparent, model will produce forecasts that are very similar to those of the Commission's residential demand model. Further support for this standardisation is provided by Figure 2 which plots the logarithms of residential demand per capita and residential demand per domestic residence over time. The latter is more stationary over the estimation period while the former shows greater evidence of trending.

For the Commission's residential demand model (2), the fitted values are given by

$$\log \hat{Y}(t) = \hat{\beta}_0 + \hat{\beta}_1 \log G(t) + \hat{\beta}_2 \log R(t) + \hat{\beta}_3 \log P(t) + (1 - \hat{\beta}_1 - \hat{\beta}_2) \log N(t)$$

where the estimated regression coefficients $\hat{\beta}_j$ $(j = 0, 1, 2, 3)$ are given in Table 1. Estimates of the errors $\epsilon(t)$ are given by $\log Y(t) - \log \hat{Y}(t)$ where these can be approximated as

$$\log Y(t) - \log \hat{Y}(t) \approx \frac{Y(t) - \hat{Y}(t)}{Y(t)}$$

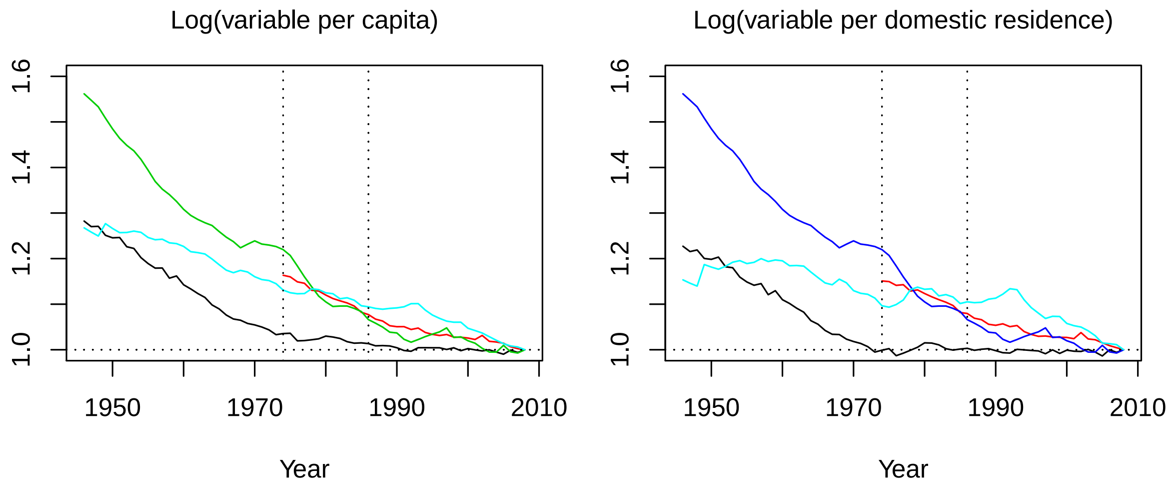

Figure 2: Plots of the logarithms of residential demand (black), commercial/industrial demand (red), domestic residences (green), population (blue), and GDP (cyan), per capita (left panel) and per domestic residence (right panel). All series have been scaled by their respective values in 2008 and the vertical dotted lines mark the starts of the residential (1974) and commercial/industrial (1986) estimation windows.

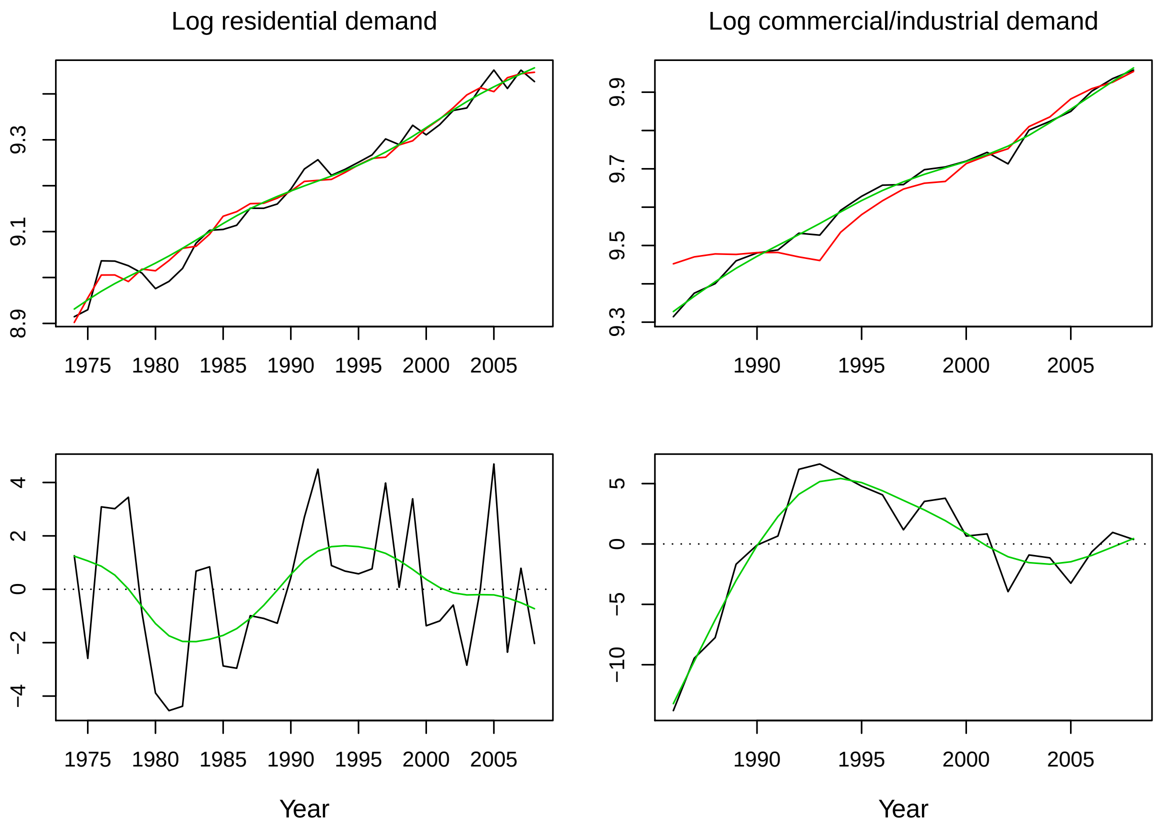

Figure 3: Plots of the logarithms of residential demand (black, top left) and commercial/industrial demand (black, top right), their respective forecasts (red, upper panels) and percentage forecast errors $100(\log Y(t) - \log \hat{Y}(t))$ (black, lower panels). Smoothed log demand series (green, upper panels) and smoothed percentage forecast errors (green, lower panels) have been superimposed for reference.

provided the right hand side is small. This follows from the fact that

$$\log(1+x) \approx x \tag{4}$$

for x small. In practice this approximation works sufficiently well that the $\log Y(t) - \log \hat{Y}(t)$ can be safely interpreted as proportionate differences (percentage differences if multiplied by 100) regardless of whether the forecasts of $Y(t)$ were obtained using log transformed data or not. Adopting this interpretation, the percentage forecast errors $100(\log Y(t) - \log \hat{Y}(t))$ for the Commission's residential and commercial/industrial demand models are plotted in Figure 3 over their respective estimation periods and with smoothed trends superimposed. The latter can be estimated using a variety of methods, but are estimated here using the Hodrick-Prescott filter (Hodrick and Prescott, 1997) which is widely used in economic applications, particular for the identification of business cycles, but also more generally. The main requirement of such a filter is that it should reliably produce nonparametric trends or local levels that run through the middle of the data.

From Figure 3 it would seem that the Commission's residential demand model produces fitted values (in-sample forecasts) for the log transformed series that are to a large extent unbiased and whose errors appear stationary, although possibly serially correlated. If serial correlation is present then this will impact on the accuracy of short to medium-term forecasts, but have little effect on the accuracy of medium to long-term forecasts. The percentage errors for the residential demand in-sample forecasts vary between -4.5% and 4.7% with a median absolute percentage error of 1.4%. The Commission's commercial/industrial demand model is not as good with evidence of forecast bias, particularly in the first half of the estimation period where the fit is poor, and non-stationary forecast errors. The percentage errors for the commercial/industrial demand in-sample forecasts vary between -13.8% and 6.6% with a median absolute percentage error of 3.3%. These in-sample results typically represent the best outcome that could be achieved when forecasting out-of-sample using the actual values of the drivers. Model error and the need to forecast the drivers will inevitably increase these values quite considerably, particularly for medium to long-term forecast horizons.

Many variants of these models can, and no doubt have, been tried. In each case, the in-sample regression diagnostics and forecast errors should be assessed in the usual way in addition to the checks on out-of-sample forecast performance discussed in Section 3.2. Information criteria such as Akaike's Information Criterion (AIC) are useful for selecting suitable candidate forecasting models that are more parsimonious, involving fewer parameters and associated drivers. The standard AIC criterion trades model fit against model complexity by selecting the model order p that minimises

$$AIC = -2 \text{ log likelihood} + p$$

where, in the case of standard linear regression models, the likelihood is often based on the assumption that the $\epsilon(t)$ are stationary, Gaussian, zero-mean time series. Hyndman and Fan (2010) use a modification of the AIC criterion (see Harrell, 2001) to discriminate between their long-term demand forecasting models.

3.2 Forecast validation

The quality and accuracy of long-term point forecasts are difficult to assess in any conventional quantitative sense (root mean squared forecast errors for example), although such exercises can and should be undertaken for shorter horizons up to 5 years or 10 years if possible. By contrast to short-term forecasting, this means that more emphasis must be placed on the nature of the forecasting model adopted, its fitness for purpose, and its ability to address the substantive long-term issues in view. The model should be as open and transparent as possible, to encourage input from informed stakeholders, and be consistent with relevant economic or physical considerations. The choice of drivers should encompass the main variables of interest (economy, weather, demographics etc.), but their number should be kept to a minimum to enhance interpretability and understanding. The way the model responds to changes in the drivers (high/low GDP or population trajectories for example) should suitably reflect current thinking as well as patterns in the historic data record. A good example is given in Lady (2009).

In the case of density forecasting, the validation task is more difficult. For short to medium-term horizons and a given econometric model, a moving fixed-length estimation window can be applied successively over the years within a suitable subset of the span of the historic data. Re-estimating the regression coefficients and using the forecast drivers available at the time (ex ante forecasts) generates successive out-of-sample demand forecasts that can then be assessed for accuracy using measures such as root mean squared error (RMSE), mean absolute error (MAE), root mean squared proportionate error (RMSPE), mean absolute proportionate error (MAPE) and more robust versions involving the median rather than the mean, for example. Given sufficient historical data of reasonable quality, predictive densities for each forecast horizon can also be built up in this way to assess the shape and key quantiles of the predictive distributions concerned.

If realistic synthetic realisations of annual demand need to be randomly generated then the validation task is even more difficult since the model now needs to account for the nature and properties of the sample paths of the series concerned. In such cases it may be important for regression models such as (1) to incorporate dynamic structure so that the $\epsilon(t)$ are serial correlated and yield simulated realisations that are realistically smooth. The same considerations also apply to the generation of synthetic realisations of drivers such as GDP. Then key properties of the historical data (static and dynamic) can be used to benchmark the same properties in the synthetic or simulated demand time series. Alternatively, bootstrap models can be used where the input data (demand and its drivers) have first been transformed to approximate stationarity, perhaps by taking time differences of the logarithms to give annual growth rates. However care must be taken with bootstrap methods to preserve the key features of the historical realisations.

In any validation exercise involving drivers, it is important to attribute the source of forecast errors as accurately as possible. Knowledge of where the primary sources of demand forecast error reside is the key to model and forecast improvement. For example, if the forecasts for a particular driver produce the greatest component of the overall demand forecast error, then better forecasts for that driver will need to be developed, if possible. On the other hand, if the primary source of forecast error resides in the model

Table 2: Percentage forecast errors for the 2007 and 2008 values of GDP, population and domestic residences based on forecasts made in 2007 using data up to and including 2006.

	Year	
Variable	2007	2008
GDP	2.51	4.22
Population	1.40	1.58
Domestic residences	0.35	-0.39

itself, then alternative models may well need to be developed. An example of such an analysis is given in Keene and Thomson (2007) who develop an appropriate forecast error decomposition in the context of New Zealand tax revenue forecasting.

The impact of regular data revisions (particularly for GDP and population) also needs to be quantitatively assessed within the context of a suitable forecast error decomposition. The recent more substantive revisions to New Zealand population were, in part, a reason for this review. The fact that these revisions were important is borne out by Table 2 which shows the percentage forecast errors between forecasts for the years 2007 and 2008 (one year and two years ahead respectively) made in 2007 and their outturns. For those two years, GDP had an average percentage forecast error of 3.4%, over twice that for the population (1.5%), and the average domestic residences percentage forecast error is approximately zero. From (2) or (3), increases in population $N(t)$ reduce forecast log residential demand ($\beta_4 < 0$), whereas increases in GDP $G(t)$ increase both log residential demand ($\beta_1 > 0$) and commercial/industrial demand. It would appear that the greater contribution to the overall percentage forecast error for total demand in 2006 and 2007 is coming from the GDP forecasts, rather than the population projections.

The Commission's residential and commercial/industrial demand forecasting models produce forecasts that are conditional on the associated future values of the drivers (GDP, domestic residences, population and residential electricity price). To make such conditional forecasts operational, forecasts of the drivers are needed. Rather than developing their own models for the drivers, the Commission use long-term forecasts provided by outside agencies such as Statistics New Zealand (population) and NZIER (GDP). It is not clear what checks the Commission applies to these externally provided forecasts, or how closely it monitors the models used. In the case of two or more drivers, any long-term forecasts will need to exhibit much the same long-run interrelationships (essentially cointegrating relationships) as those observed in the historical data. However, unless specifically built into the forecasts for the drivers (usually through a joint forecast model of the drivers concerned), there is no guarantee that any long-term historical cointegrating relationship will necessarily hold over medium to long-term forecast horizons.

Figure 4 plots the logarithms of the historical demand and driver data augmented by the Commission's forecasts. Each series has been scaled by its 2008 value to aid comparison. The forward evolution of log population, log domestic residences and log GDP from 2008 is not too different from the backward evolution from 2008 seen in the historic data, although there seems to be slightly less spread in the forecasts than the historic data over similar time spans. The evolution of log residential electricity price is anomalous

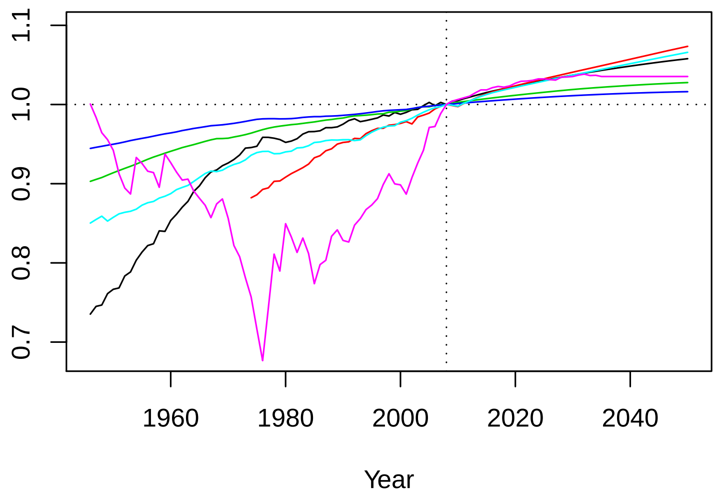

Figure 4: Plots of the logarithms of residential demand (black), commercial/industrial demand (red), domestic residences (green), population (blue), GDP (cyan) and residential electricity price in 2008 dollars (magenta) augmented by the Commission's forecasts. All series have been scaled by their respective values in 2008 and the vertical dotted line marks the end of the historical data (2008).

leading to further doubts as to its reliability both as a predictor or as a variable within any cointegrating relationship. Although Figure 4 is suggestive rather definitive, it does nevertheless throw some light on the nature of the interrelationships between the variables and, in particular, on how well these have been maintained in the long-term forecasts.

Plots of the percentage growth rates of the demand series and their drivers provide further information on the nature and quality of the forecasts. Using the approximation (4), growth rates for any time series $X(t)$ are conveniently given by

$$\log X(t) - \log X(t-1) \approx \frac{X(t) - X(t-1)}{X(t-1)}$$

provided the right hand side of the above (a simple growth rate) is small (note that $\log X(t) - \log X(t-1)$ is a continuously compounding growth rate). Figure 5 plots the percentage growth rates of the two demand series and their drivers with the historical series augmented by the Commission's forecasts. Trends (smoothed percentage growth rates) of the historical series are also plotted where these are calculated, as before, using the Hodrick-Prescott filter. All growth rates show evidence of slowly evolving nonstationary levels that are, for the most part, reasonably consistent with the forecast growth rates. The residential price growth rates are a notable exception since their forecast growth rates appear inconsistent with the trend of their recent historical growth rates. Most of the forecast growth rates are essentially simple linear or constant time series over medium to long-term horizons, as might be expected. This motivates the need to consider alternative models based on growth rates (see Section 4) alongside the current models based on levels. Note, however, that the GDP forecast growth rates over the short to medium-term (approximately 10 years) show relatively high variability which is, to a lesser extent, reflected

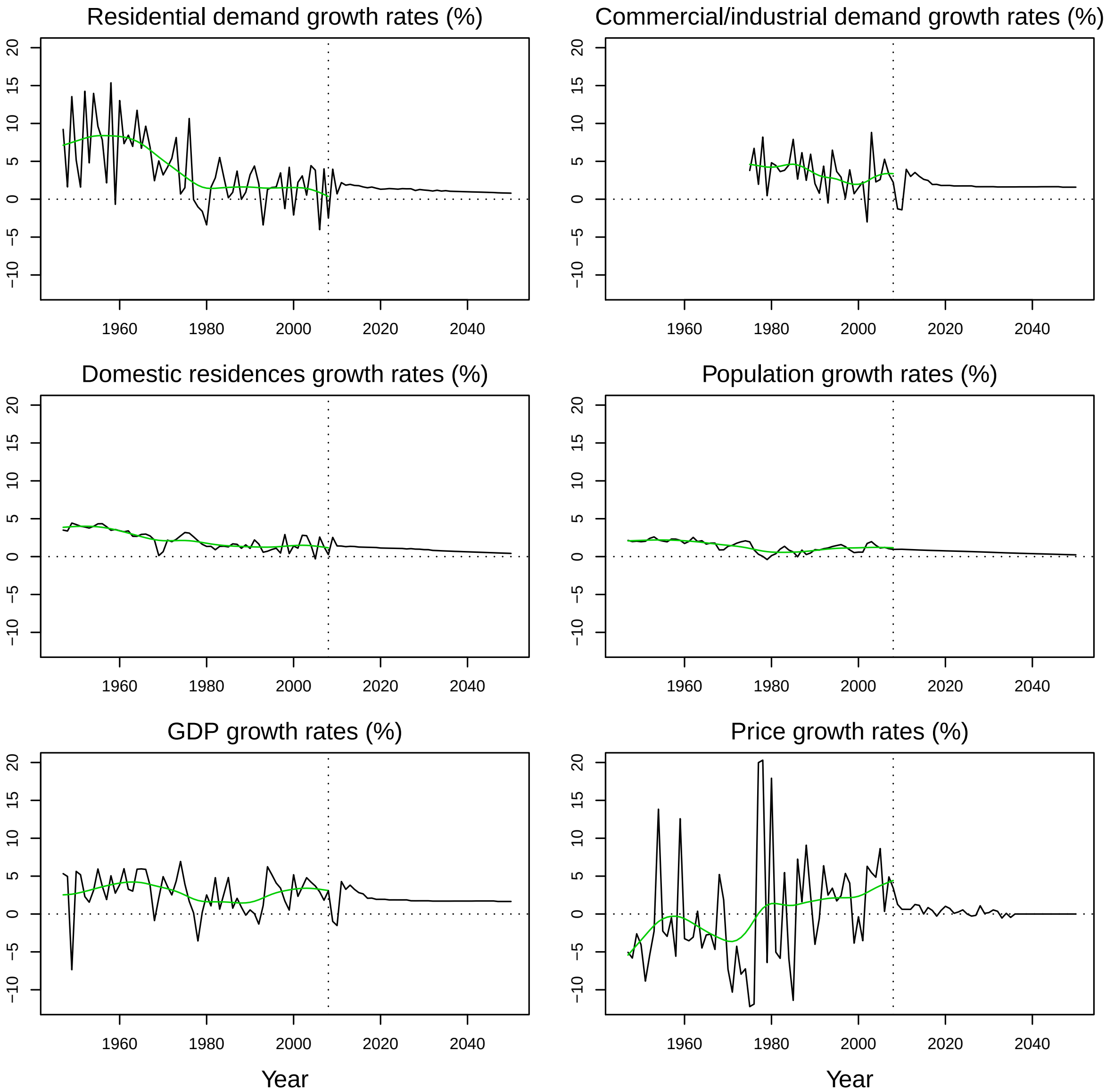

Figure 5: Plots of the percentage growth rates of residential demand, commercial/industrial demand, domestic residences, population, GDP and residential electricity price in 2008 dollars where the historical series have been augmented by the Commission's forecasts. Smoothed percentage growth rates of the historical series (green) have been superimposed for reference and the vertical dotted line marks the end of the historical data (2008).

in both demand forecast growth rates. The variability of the historical growth rates about their trends is least for the demographic variables (domestic residences and population), and greatest for price. The growth rate variability of GDP and the two demand series are comparable, particularly since the 1980s.

These and other validation exercises need to be undertaken to provide confidence in, and quantitative assessments of, long-term demand forecast accuracy, reliability, and fitness for purpose. Continuous model improvement is an important goal. Although the Commission undertakes a number of validation checks, it would appear that more can, and arguably should, be done. More systematic validation, even at the shorter horizons, builds confidence in the long-term performance of the forecasting models adopted, and lays a secure foundation for model development. In this regard, the use of forecast drivers from other agencies, prepared in isolation of the Commission's forecasting models and context, is seen as a potential weakness unless supported by suitable quality checks and validation studies.

3.3 Predictive distributions

The Commission currently uses a Monte Carlo procedure to generate independent realisations of future electricity demand based on their current residential and commercial/industrial forecasting models. These realisations are used to generate suitable quantiles, among other distributional measures, of the predictive distribution or forecast density at any given horizon out to 2050. An explanation of the procedure used is given in Section 3.7 of Kirtlan (2008). However this explanation proved insufficient to adequately grasp what is being done in any detail and so the MATLAB code that generated these synthetic realisations had to be carefully examined.

As noted earlier, density forecasting is a substantially more difficult exercise than point forecasting, and the generation of realistic sample paths for future demand is even more difficult, especially if the nature of the sample path dynamics is important. In this sense the Commission's approach is commendable, ambitious and, in many respects, innovative. However, although the Commission's Monte Carlo approach can no doubt be put on a more secure and systematic footing, the current lack of adequate documentation, and reliance on ad-hoc assumptions and procedures are major weaknesses which potentially undermine confidence in the uncertainty measures that result.

Furthermore, the variation explored is all within the framework of the Commission's existing models for residential and commercial/industrial demand (model specific forecast error or uncertainty), and fails to incorporate any error associated with the choice of models used (model error). The latter can be assessed by considering a range of different models and combining their forward realisations in one overall ensemble to produce a greater range and distribution of forecast error than can be achieved with a single model.

The MATLAB routines that underpin the generation of future demand and driver sample paths are based on multiplicative recursive models that are not always consistently applied or efficiently coded. As might be expected in code of this complexity and length, one can find many deficiencies, most of which are small and relatively unimportant. Some of the more important issues are discussed in the following subsections.

3.3.1 In-sample variation

To determine variation in the fitted regression coefficients of the two demand models due to the use of stochastic (rather than deterministic) drivers, each driver was, in effect, modelled as

$$D(t) = T_D(t)(1 + e_D(t)) \tag{5}$$

where $D(t)$ denotes the driver, $T_D(t)$ denotes a smooth trend, and the multiplicative error $e_D(t) = (D(t) - T_D(t))/T_D(t)$ measures the trend deviation as a proportion of the trend. The Commission use a simple 5 year moving average to estimate $T_D(t)$. This moving average is not as smooth and loses information at the ends of series by comparison to the Hodrick-Prescott filter used in Figures 3 and 5 for example. Alternative realisations of the various drivers are obtained by retaining the estimates of $T_D(t)$, and randomly generating independent values of $e_D(t)$ from their empirical distributions. Based on these realisations, regression estimates are then generated and used in the respective demand forecasts.

However, any correlation between the $e_D(t)$ across different drivers (say population and GDP) is not accounted for. More importantly, this procedure introduces errors in the independent variables which will typically result in biased regression coefficients (see Davies and Hutton, 1975, for example). Nevertheless, even if properly accounted for, this source of variation will be small by comparison to that of the forecasts themselves.

3.3.2 Sample path generation

To generate a future realisation of population, a population size for 2050 is generated from a suitable lognormal distribution calibrated against various Statistics New Zealand projections for 2050. Then a realisation of future population values over the intervening years is generated by adding a linear term to the original point forecasts. The linear term, in essence, interpolates the difference between the two 2050 population values and zero difference at the origin of the forecast interval. This is a linear model in forecast population levels with only one source of variation. Although simple, it doesn't reflect the cohort models used by Statistics New Zealand which are largely multiplicative in character.

To generate future realisations of GDP per capita, a multiplicative model is used of the form

$$\frac{D(t)}{D(t-1)} = \frac{\hat{D}(t)}{\hat{D}(t-1)}(1 + \eta)(1 + e_D(t))$$

where $D(t)$ denotes the future values of GDP per capita, $\hat{D}(t)$ is the original forecast of GDP per capita, η is a single zero-mean, Gaussian variable with known standard deviation (0.002) representing a proportionate productivity change, and the $e_D(t)$ are the GDP trend deviations determined from the in-sample variation described in Section 3.3.1. Using logarithms and the approximation (4) yields the practically equivalent growth rate model

$$\log D(t) - \log D(t-1) = \log \hat{D}(t) - \log \hat{D}(t-1) + e(t) \tag{6}$$

where $e(t) = \eta + e_D(t)$). This will give valid realisations of future values of GDP per capita if the underpinning model for $D(t)$ follows the simple growth rate model

$$\log D(t) - \log D(t-1) = \mu(t) + \epsilon(t) \tag{7}$$

where $\mu(t)$ is a slowly evolving trend, the stochastic properties of the error term $\epsilon(t)$ are determined from the historical growth rate trend deviations, and forecasts of $\mu(t)$ are given by $\log \hat{D}(t) - \log \hat{D}(t-1)$. This is a simple, transparent, and reasonable model for GDP per capita judging from the plots in Figure 5. However, the Commission's model uses in-sample trend deviations from the multiplicative levels model for GDP given by (5) and a productivity component, neither of which are related to the errors $\epsilon(t)$ in the GDP per capita growth rate model (7).

Future realisations of domestic residences per capita are generated in much the same way as GDP per capita, but with $e(t)$ in (6) equal to a zero-mean, Gaussian variable with known standard deviation (0.005) multiplied by a linear factor that varies from 0, at the beginning of the forecast interval, to 1 at the end. There is only one source of variation and, as before, the properties of $e(t)$ are not based on the historical growth rates of domestic residences per capita.

For residential electricity prices, future realisations are generated using a multiplicative model of the form (5). After taking logarithms, realisations are determined from

$$\log D(t) = \log \hat{D}(t) + e(t)$$

where now the $D(t)$ denote future residential electricity prices, the $\hat{D}(t)$ are the original forecasts, and $e(t)$ has the same structure as the $e(t)$ for domestic residences per capita. This time the single source of variation is a zero-mean, Gaussian variable with standard deviation 0.1.

The Commission's general strategy for generating synthetic forward realisations of the various drivers seems appropriate, and the procedures devised are innovative, if not always consistent within or across drivers. The mismatch of models, ad-hoc assumptions and procedures can, and should, be rectified and replaced by a more secure and systematic treatment that is likely to result in simpler computations. If simple models such as (7) are utilised, then it is also likely that the forecast densities can be determined analytically without recourse to simulation.

3.4 Recommendation

The analysis and discussion in the previous sections lead to the following recommendation.

Recommendation 1 *With regard to the Commission's current long-term forecasting models and strategies, it is recommended that:*

(a) more comprehensive, systematic, technical documentation be developed that provides full details of the forecasting models adopted, their objectives and assumptions, the data used and any pre-processing undertaken, the validation procedures used, and full mathematical formulations of the models within an overall statistical framework;

(b) a unified modelling framework based on log transformed data be adopted for the current residential and commercial/industrial demand models, with a view to simplifying the former and improving the accuracy of the latter;

(c) *more comprehensive validation procedures be developed to assess the accuracy and reliability of demand and driver forecasts, the quality of sample paths, sensitivity to driver perturbations, forecast error attribution, and the maintenance of any historical long-term cointegrating relationships in the forecasts;*

(d) *a more secure, systematic and consistent framework for generating synthetic realisations of future demand and driver time series be developed based on the Commission's existing general strategy.*

4 Other demand forecasting models

The Commission's current econometric regression models have the virtue that they involve economic and demographic drivers that can be used to set scenarios and moderate long-term forecasts of demand. Such models are more difficult to understand, less transparent, and therefore less useful, when they involve many drivers. The Commission's residential demand model involves three regressors (although this can be reduced to two if model (3) is adopted), whereas its commercial/industrial demand model involves one regressor (ignoring the shortage dummy which is, in essence, a data adjustment). In the case of the latter, the plots in Figure 3 suggest that the relationship between commercial/industrial demand and GDP is evolving over time and a shorter estimation window (perhaps from the early 1990s) would yield a better fit and more stationary residuals. This suggests that there may be merit in considering even simpler models of residential demand involving only one driver, or even no drivers (time series models), and more adaptive models for commercial/industrial demand that can better cope with the evolving relationship between commercial/industrial demand and GDP, especially over its most recent history. Given the nature of the trends evident in the log demand levels (Figure 3) and the demand and driver growth rates (Figure 5), evolution clearly plays a more general role in the demand and driver series which must be accommodated, either in the form of carefully selected estimation windows, or otherwise.

The suggestions made above will, if taken up, lead to a number of competing models, some based on levels and others on growth rates, with some involving drivers and others not. This will lead to a variety of point forecasts and the need for forecast averaging. As noted in Section 2.1, this is a strength, rather than a weakness, since it addresses model error, and is in accord with both the literature and best practice. Furthermore, generating sample paths of future electricity demand from each model and combining these within an overall ensemble is likely to yield a more accurate, reliable view of future demand than any single model and, as a result, more robust estimates of forward risk. These considerations underpin the development that follows.

Many of the models used implicitly or explicitly by the Commission can be put in a common framework of the form

$$Y(t) = \kappa(t)X_1(t)^{\beta_1(t)}X_2(t)^{\beta_2(t)}\dots X_p(t)^{\beta_p(t)}e(t)$$

where $Y(t)$ denotes either residential or commercial/industrial electricity demand, the $X_j(t)$ $(j = 1, \dots, p)$ denote suitable drivers, and $e(t)$ denotes multiplicative random error that varies about a mean of unity. Here the coefficients $\kappa(t)$ and $\beta_j(t)$ ($j = 1, \ldots, p$) are assumed to be evolving slowly and smoothly over time. Taking logarithms of this multiplicative model gives the additive log levels model

$$\log Y(t) = \beta_0(t) + \beta_1(t) \log X_1(t) + \ldots + \beta_p(t) \log X_p(t) + \epsilon(t) \tag{8}$$

where $\beta_0(t) = \log \kappa_0(t)$ and the $\epsilon(t)$ are now random errors with mean zero. The evolution in the parameters $\beta_j(t)$ can be handled through a judicious choice of estimation windows (the Commission's current approach), or modelled more directly. Here an approach to the latter is considered which, although not entirely devoid of problems, has a more systematic and automatic way of accounting for evolutionary parameter variation.

Consider the simple case of (8) where $p = 1$, $X_1(t) = X(t)$, $\beta_0(t) = \alpha(t)$ and $\beta_1(t) = \beta$. Then the log levels model (8) becomes

$$\log Y(t) = \alpha(t) + \beta \log X(t) + \epsilon(t) \tag{9}$$

where the evolutionary component $\alpha(t)$ is assumed to be a slowly varying trend, the constant β describes the long-run linear relationship with $\log X(t)$, and any deviations from that long-run relationship are absorbed into the random error $\epsilon(t)$. An important special case of (9) is where $\beta = 1$ so that

$$\log(Y(t)/X(t)) = \alpha(t) + \epsilon(t) \tag{10}$$

which is an even simpler evolutionary model for the logarithms of the ratios $Y(t)/X(t)$. Taking time differences of (9) yields the growth rate model

$$\log Y(t) - \log Y(t-1) = \mu(t) + \beta(\log X(t) - \log X(t-1)) + \eta(t) \tag{11}$$

where $\mu(t) = \alpha(t) - \alpha(t-1)$, and the $\eta(t)$ are random errors with zero mean. This is essentially a regression model of demand growth rates against driver growth rates with a slowly evolving intercept $\mu(t)$ whose nature will depend on that of $\alpha(t)$. If the latter is a locally linear trend over time, then $\mu(t)$ should be a trend that is locally constant over time. Note that the coefficient β can be interpreted as an elasticity. When $\beta = 0$ the model becomes a simple time series trend plus error model, and when $\beta = 1$

$$\log(Y(t)/X(t)) - \log(Y(t-1)/X(t-1)) = \mu(t) + \eta(t) \tag{12}$$

gives a simple growth rate model for the ratios $Y(t)/X(t)$. These simple evolutionary models are readily interpreted and relatively easy to apply in practice.

There are many ways to model the trends $\alpha(t)$ and $\mu(t)$, but here the focus is on stochastic models suitable for evolutionary trends and for forecasting. Consider a conventional trend plus error model for time series observations $O(t)$ of the form

$$O(t) = T(t) + e(t) \tag{13}$$

where $T(t)$ is an unobserved stochastic trend and the $e(t)$ are additive random errors. Following Harvey (1989) or Durbin and Koopman (2001), the stochastic trend $T(t)$ follows a *local linear trend model* if

$$T(t) = T(t-1) + \delta(t-1) + \nu(t), \qquad \delta(t) = \delta(t-1) + \xi(t) \tag{14}$$

where $\nu(t)$, $\xi(t)$ are mutually independent Gaussian white noise processes each with zero means and standard deviations σ_ν, σ_ξ respectively. To ensure that the trend $T(t)$ evolves smoothly over time, both σ_ν and σ_ξ will need to be small by comparison to σ_e, the standard deviation of the error $e(t)$. Trend forecasts based on the observations $O(t)$ and the local linear trend model (14) are given by the Kalman filter, and are linear functions of time unless the $\delta(t)$ are zero for all t, in which case the forecast is a constant value. More generally, the model specified by (13) and (14) is a special case of a *structural time series model* which decomposes an observed time series into a sum of unobserved stochastic components such as a trend, seasonal, business cycle and residual error. See Durbin and Koopman (2001) for further details.

Despite being governed by only two parameters σ_ν and σ_ξ, the trend model (14) provides a relatively flexible range of options. For example, setting $\sigma_\nu = \sigma_\xi = 0$ (equivalently $\nu(t) = \xi(t) = 0$) in (14) yields a simple regression model with a non-evolutionary, straight-line, time trend, whereas setting $\sigma_\nu = 0$ (equivalently $\nu(t) = 0$) yields the so-called *smooth trend model* which tends to give smoother, more flexible, trends than other variants of (14). In particular, the smooth trend model with $\sigma_e/\sigma_\xi = 40$ gives in-sample estimates of $T(t)$ that are the same as those given by the Hodrick-Prescott filter. The *local level trend model* is a special case of (14) with $\delta(t) = 0$ for all t. Its forecasts are a constant value that is, in essence, estimated as an exponentially weighted mean of the observations $O(t)$ with greatest weight placed on the most recent observations. As a consequence it provides a sensible estimate (an exponentially smoothed value) of the last local level of the time series $O(t)$. Weights can be selected using optimal statistical methods such as maximum likelihood, or pre-specified. Given that any adaptive linear time trend may well prove to be too unstable in practice for long-term forecasting, it is the local level trend model that is likely to prove more useful, particularly for data (possibly transformed or scaled) that appear to follow a locally evolving level.

While the log levels model (10) could be used to model residential demand per domestic residence since the early 1970s and, less convincingly, commercial/industrial demand as a ratio of GDP from the early 1990s, it is the growth rates of demand and driver series where these models are likely to apply to greater effect. From Figure 5 it would appear that the growth rates generally vary about a smooth local level, particularly from the mid 1970s, and the trend deviations appear reasonably homogeneous in terms of their statistical properties. This leads to consideration of growth rate models such as (11) and (12) with $\mu(t)$ modelled as a local level trend. Figure 6 shows the forecasts obtained by fitting the time series model (11) with $\beta = 0$ to the growth rates of residential and commercial/industrial demand, and fitting the model (12) to the growth rates of residential demand per domestic residence and commercial/industrial demand as a ratio of GDP. For residential demand, both growth rate models produce forecasts that are less than those from the Commission's econometric regression model in log levels. For commercial/industrial demand the converse is true, with both growth rate models producing forecasts that are greater than those from the Commission's model, especially the time series model based only on the growth rates of demand. Although the latter forecast is considerably higher than the others (a difference accentuated by the 42 year forecast horizon), it is nevertheless based on a reasonable assessment of the most recent local level of the historical growth rates that is only 1.5% higher than the long-run rate given by the

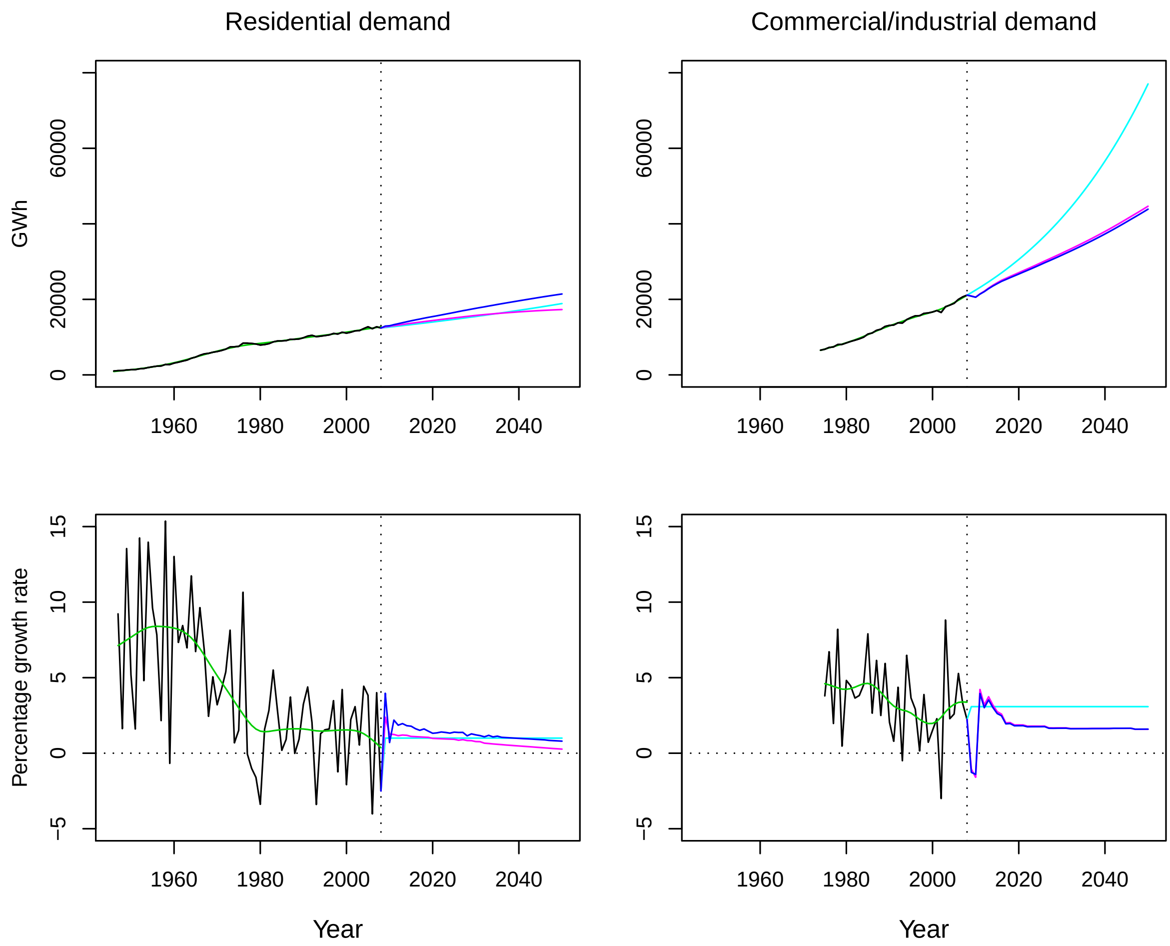

Figure 6: Plots of levels (black, upper panels) and percentage growth rates (black, lower panels) of residential and commercial/industrial demand with smoothed series (green) superimposed. Forecasts based on a local level trend model fitted to the growth rates of demand (cyan) and demand ratios (magenta) are plotted together with the Commission's forecasts (blue). The demand ratios are residential demand per domestic residence and commercial/industrial demand as a ratio of GDP. The vertical dotted line marks the end of the historical data (2008).

Commission's model. If growth in commercial/industrial electricity demand were maintained at current levels then this forecast represents what is likely to occur. Although these forecasts are illustrative only and have not been validated in any way, they do indicate the forecast variation that may need to be accounted for, but which is absent from the Commission's current forecasting procedures.

Models (9) and (11) with β estimated from the data can also be considered, as can the more general variants of the local linear trend model (14). Suitable computational procedures for fitting and forecasting such models are now readily available in a number of statistical computing systems including R (R Development Core Team, 2004).

So far discussion has focussed primarily on forecasting univariate time series, typically conditional on suitable drivers that also need to be forecast. Currently the Commission obtains driver forecasts from external providers whose forecast objectives may not necessarily align with those of the Commission, and whose methods may also be univariate in character. However the demand and driver time series are highly interrelated showing cointegration relationships that need to be maintained over the long-term. Even if driver series are not involved, it is likely that forecasting models that capture any interrelationship between residential demand and commercial/industrial demand would improve the forecast accuracy of both, particularly the latter. If such covariation issues are important then suitable vector models would need to be considered including, but not limited to, vector forms of the structural time series models (9) and (11) with common trends, and the VECM models mentioned in Section 2.1 applied to growth rates. The development of suitable vector models for forecasting New Zealand electricity demand should lead to forecasting gains and a better, more robust, understanding of any long-term cointegrating relationships. As part of such an exercise, the Commission would need to develop its own forecasting models of the associated drivers.

4.1 Recommendation

The analysis and discussion in the previous section lead to the following recommendation.

Recommendation 2 *With regard to other long-term forecasting models and strategies, it is recommended that the Commission:*

(a) develop a suite of competing forecasting models based on growth rates as well as levels, and use the combination of these forecasts for long-term forecasting and risk assessment;

(b) develop and assess forecasting models which systematically model slowly evolving parameters and relationships over time, such as those given by equations (9) and (11) in the review;

(c) explore the development of vector forecasting models with common trends that take explicit account of cointegrating relationships between the various demand and driver series.

Acknowledgements

The reviewer is grateful to Brian Kirtlan (Electricity Commission) for advice, guidance and the provision of background information provided over the course of this review.

References

Allen, P.G. and Morzuch, B.J. (2006) Twenty-five years of progress, problems, and conflicting evidence in econometric forecasting. What about the next 25 years? *International Journal of Forecasting* **22**, 475–492.

Armstrong, J.S. (1985) *Long-range Forecasting: from crystal ball to computer (2nd ed).* Wiley, New York.

Chui, F., Elkamel, A., Surit, R., Croiset, E. and Douglas, P.L. (2009) Long-term electricity demand forecasting for power system planning using economic, demographic and climate variables. *European Journal of Industrial Engineering* **3**, 277–304.

Craig, P.P., Gadgil, A. and Koomey, J.G. (2002) What can history teach us? A retrospective examination of long-term energy forecasts for the United States. *Annual Review of Energy and the Environment* **27**, 83–118.

Davies, R.B. and Hutton, B. (1975) The effects of errors in the independent variables in linear regression. *Biometrika* **62**, 383–391.

Durbin, J. and Koopman, S.J. (2001) *Time Series Analysis by State Space Methods.* Oxford University Press.

Dergiades, T. and Tsoulfidis, L. (2008) Estimating residential demand for electricity in the United States, 1965–2006. *Energy Economics* **30**, 2722–2730.

Faust, J. and Wright, J.H. (2009) Comparing Greenbook and reduced-form forecasts using a large realtime dataset. *Journal of Business and Economic Statistics* **27**, 468–479.

Granger, C.W.J. (1989) Combining forecasts - twenty years later. *Journal of Forecasting* **8**,167–173.

Harrell, Jr., F.E. (2001) *Regression Modeling Strategies: with applications to linear models, logistic regression, and survival analysis.* Springer, New York.

Harvey, A.C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press.

Hodrick, R.J. and Prescott, E.C. (1997) Post-war US business cycles: an empirical investigation. *Journal of Money, Credit and Banking* **29**, 1–16.

Huss, W.R. (1985) Comparative analysis of company forecasts and advanced time series techniques using annual electric utility energy sales data. *International Journal of Forecasting* **1**, 217–239.

Hyndman, R.J. and Fan, S. (2009) Density forecasting for long-term peak electricity demand. *IEEE Transactions on Power Systems*, doi:10.1109/TPWRS.2009.2036017

Keene, M. and Thomson, P.J. (2007) An analysis of tax revenue forecast errors. Working Paper 07/02, New Zealand Treasury, Wellington, New Zealand.

Kirtlan, B. (2008) Electricity demand forecast review February 2008. New Zealand Electricity Commission report. *http://www.electricitycommission.govt.nz/consultation/GPA*

Koomey, J., Craig, P., Gadgil, A. and Lorenzetti, D. (2003) Improving long-range energy modelling: a plea for historical retrospectives. *The Energy Journal* **24**, 75–92.

Lady, G.M. (2009) Evaluating long term forecasts. *Energy Economics*, doi:10.1016/j.eneco.2009.10.006

Mohamed, Z. and Bodger, P. (2003) Forecasting electricity consumption in New Zealand using economic and demographic variables. *Energy* **30**, 1833–1843.

NZIER (2009) Review of electricity demand forecast model. Report commissioned by the New Zealand Electricity Commission. *http://www.electricitycommission.govt.nz/opdev/modelling/gpas/2005SOO/Demand/Review*

R Development Core Team (2004) *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna. *http://www.R-project.org*

von Hirschhausen, C. and Andres, M. (2000) Long-term electricity demand in China – from quantitative to qualitative growth. *Energy Policy* **28**, 231–241.