NZIER
NEW ZEALAND INSTITUTE OF ECONOMIC RESEARCH

Review of Transpower's peak demand forecast model

High-level review of fitness-for-purpose of forecasts to be used in MBIE Electricity Demand and Generation Scenarios

NZIER report to MBIE

March 2016

About NZIER

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.

Each year NZIER devotes resources to undertake and make freely available economic research and thinking aimed at promoting a better understanding of New Zealand's important economic challenges.

NZIER was established in 1958.

Authorship

This report was prepared at NZIER by John Stephenson.

L13 Grant Thornton House, 215 Lambton Quay | PO Box 3479, Wellington 6140
Tel +64 4 472 1880 | econ@nzier.org.nz

1. Objective and scope

This report presents the results of a review of Transpower's peak demand forecasting model.

The main conclusion of the report is that **we are satisfied that Transpower's forecast modelling framework is fit-for-purpose** in terms of using the model for the peak-demand forecasts for the EDGS.

MBIE sought this review in response to a request from stakeholders to verify the Transpower forecasts used in MBIE's 2015 draft Electricity and Demand Generation Scenarios (EDGS).

Scope

This review focusses on regional, island and national winter peak demand forecasts. The Transpower model forecasts peak demand for winter, summer and shoulder (mid-season) peaks the forecasts used in the EDGS are annual peak demands.[1]

In terms of the model and its methods, the scope of the review included:

- forecast methods and theoretical soundness
- appropriateness of assumptions including any model interventions
- methods for determining ranges of possible future peak demands:
  - expected peak demand
  - so-called prudent peak forecasts (the 90th percentile of possible peak demand, used for capacity planning purposes)
- a check of the software (MATLAB) scripts used to produce the forecasts
- replicating Transpower's forecasts.

MBIE also sought our advice on the consistency of Transpower's forecasts with MBIE's other EDGS forecasts and the feasibility of reproducing Transpower's forecasts if input data was publicly available.

Review stages

The review took place in two stages. In the first stage we conducted a review of each of the above issues and made suggestions and recommendations for model improvements.

Transpower then considered our suggestions and recommendations and responded to them, including making revisions to their model as appropriate.

In stage two, we reviewed Transpower's responses and revisions.[2]

[1] Transpower also produces forecasts for Grid Exit Points (GXPs) but these were out of scope for this review.

[2] Stage two did not include a full review of modelling scripts or a replication of Transpower's forecasts.

2. Initial findings

2.1. Simplicity a real strength

Our stage 1 review commended the transparency of the overall method pursued by Transpower.

The purpose of longer term forecasting and scenario modelling puts a premium on the transparency, simplicity and explicability of the forecast model. This is because long-term forecasts (decades hence) are not able to be validated in any conventional quantitative sense. There simply isn't enough history in the data and too much can change in future. The usefulness of the model for decision making then rests on the nature of the forecast construction, how easy it is to use and to understand.

Transpower's method for forecasting peaks is based on 4 simple models of peak demand trends. The 4 models are:

- long-term time trends
- short-term time trends
- long-term trends accounting for changes in economic activity and population growth
- long-term trends accounting for changes in energy demand.

The detail of the modelling process is somewhat more complicated than this.[3] But the general approach is basically this simple.

The different trend models capture different relationships. For example, the short term model focussing on recent flattening in demand which can be hard to explain otherwise. The long term model that accounts economic activity allows forecasts to include the effects of anticipated changes in the size of the population or of the economy; as well as accounting for the extent to which GDP has helped to predict trends in peak demand in the past.

The overall simplicity of the Transpower forecast model was a real strength. This simplicity enabled the model to be implemented in a simple spreadsheet, for example, and this was useful for sharing the model with a wider audience.

2.2. The ensemble is complicated but is useful for explaining uncertainty

The final forecast, which combines the four models and which Transpower refers to as 'the ensemble forecast', are the most complicated part of the forecast. But this is a useful addition and better than simple model averaging.[4]

[3] For example, the long term time trend model includes a temperature value and peak demand that is modelled in the other equations is peak demand adjusted for typical changes to temperatures.

[4] The additional complication of the ensemble forecast is 'worth it' when there appear to be structural changes in the data and the different models produce widely different results. That said, if the ensemble forecast departed too much from a simple arithmetic average of the 4 models' forecasts then it would make sense to revisit the fit of each model.

Even if the exact mechanics are a bit tricky, the ensemble forecast can be thought of as weighted average of the trends picked up by each model, with weights reflecting errors in the fit each model.[5] This is useful information for a non-technical user. If 2 models have a high trend forecast and 2 have a low trend and the ensemble forecast is closer to the low forecast, this tells the user of the ensemble forecasts that the high rates are less certain – without having to check the detail of particular model errors.[6]

2.3. Trend models needed some attention

Our initial (stage 1) review raised questions about:

1. model estimation errors that persisted through time (auto-correlated errors)
   - 1.1 models with patterns in the errors can and should be improved upon
   - 1.2 the simplest approach to resolving this issue is to model patterns in model errors (using autoregressive terms)
2. whether there are model diagnostics that supported model choices, e.g.:
   - 2.1 we would have expected to see information on out-of-sample forecast performance[7]
   - 2.2 it was unclear why peak demand was modelled in levels and not, for example, growth rates which may have provided for better model fit
   - 2.3 it was unclear whether temperature adjustment of the data improved the modelling
   - 2.4 a large number of statistically insignificant coefficients were included in the models
3. reasons for data sample adjustments
   - 3.1 sample break points (2007, 2010) and
   - 3.2 well-developed justifications for excluding data for some years (2001 and 2003)
4. whether consideration had been given to using model forecast performance data as weights for the ensemble model.

2.4. Consistency issues for the MBIE EDGS

We considered whether there might be consistency issues when using the Transpower forecasts for the MBIE EDGS. We concluded that there is an issue in so far as Transpower includes industry demand growth assumptions in its peak demand forecasts. Alongside this, MBIE produces its own industry demand projections. These

[5] The central forecast from the 'ensemble' model is a number which is at the 50% percentile of combined cumulative probability densities for all 4 models. The ensemble model weights each model density equally (0.25 for each model) when combining the cumulative densities.

[6] There are weakness in the ensemble approach as implemented by Transpower. For example, the models themselves are weighted equally in the ensemble calculation and model errors are not weighted for, for example, forecast performance. However, these are not significant weaknesses and addressing them would not be costless and would further complicate the forecasting process.

[7] Where the model is fitted on a data sub-set and then a forecast is performed and evaluated using the actual data.

different views should align. This will not make a large difference to the peak demand forecasts – seeing as peaks tend to be driven by residential demand – but resolving this potential inconsistency should not be difficult.

3. Final findings and recommendations

3.1. Overall fit-for-purpose

We are satisfied that Transpower's forecast model is fit-for-purpose in terms of using the model for the EDGS. This is subject to non-technical communication and process issues that should be addressed as discussed below in the sub-section 3.4 on 'Outstanding issues that should be addressed'.

We caution that model performance should be monitored (sub-section 3.5) and we note that alternative forecast methods are worth investigating (sub-section 3.6).

3.2. Transpower responded positively to our initial review

Transpower has addressed a number of the findings from stage 1 of our review. Most importantly, Transpower addressed the questions we had which could mattered most for the quality of the trend models. In its revised model Transpower has:

1. resolved the problem of estimation errors that persist through time by modelling the errors explicitly with autoregressive terms
2. used model diagnostics, including model fit criterions (e.g. AIC) to check model fit and:
   - 2.1 conducted out-of-sample forecasts tests
   - 2.2 chosen to model changes in peak demands in most models where this improved fit
   - 2.3 chosen to no-longer include default temperature adjustment and include temperature as an explanatory factor only where it improves model fit (based on AIC)
3. the re-fitting of the model include all data, so that previous sample data adjustments were dropped
4. considered using model forecast performance data as weights for the ensemble model but decided that this is not necessary.

3.3. Where changes were not made the status quo has been justified

Many of the models still have insignificant coefficients and some of the regional models do not well at all (with negative R-squared values).[8]

Transpower has suggested that this is unavoidable if they are to maintain the simplicity of their modelling approach. This is because the poor fits are deemed to be related to using the same general forecast model structure for each area (region,

[8] Note that we are more concerned about the latter – poor model explanatory power – than the former because forecast performance is often enhanced by leaving factors in a model even if they are not statistically significant.

island and national) and keeping the same general model structure is part of the simplicity of their approach.

Transpower also note that the 'problematic' models are for smaller regions with little impact on island or national level peaks. Therefore the costs of resolving those problems are likely to outweigh the benefits in terms of forecast accuracy (at least for MBIE's purposes). This reasoning is sound.

Similarly, Transpower has said that they believe changing the weights in the ensemble forecasts would create complexity for not much if any gain in forecast accuracy. We understand that this might be the case and, on-balance, we think the ensemble weighting process is fit-for-purposes as it is.

3.4. Outstanding issues that should be addressed

3.4.1. Summary model or model results are needed

Additional summary information about the Transpower model is needed. This could take the form of a 'cut-down' model (possibly in a spreadsheet) or a summary report on model fit, the meaning of model coefficients (particularly trends) and standard model diagnostics. Ideally, it would include both.

The peak demand forecast model, as currently constructed, will not be easy for most people to interrogate. Even experienced modellers will find it difficult if they do not have experience using or access to MATLAB software.

The previous model included a 'cut-down' version that made most of the model's key ideas and workings accessible to an interested audience. However, even in the previous model the documentation had only limited information on model diagnostics or coefficient interpretation. This made it difficult for a lay audience or a technical audience to assess whether the modelled trends were reasonable.[9]

It is unclear who, precisely, should produce the summary information for the forecast model. We would encourage Transpower to publish more model diagnostics with its model documentation as a matter of course. However, Transpower's model produces a wide range of forecasts and summarising the models for a lay-audience could be very time-consuming. It may be that MBIE is best placed to summarise the meaning of the models that it uses and to provide that information to stakeholders.[10]

[9] A lay audience would struggle with the mathematical detail and lack of narrative while a technical audience would look straight for model fitting procedures and diagnostics for the constituent models and would not have found them.

[10] What we have in mind is information that can be interpreted and 'sense-checked' by informed participants in the EDGS process who may not have a modelling or statistics background. This might include listing the implied responsiveness of peak demand to GDP, for example, or providing a decomposition of forecasts into contributions from the different models.

3.4.2. MBIE modellers should 'run' the model themselves, if appropriate

MBIE modellers should be hands-on users of Transpower's peak demand forecast model; if it is appropriate from Transpower's perspective and assuming resources are available.

Furthermore, the forecasts that are used in the EDGS should, in our view, come from model output produced and 'owned' by MBIE modellers. This would ensure that MBIE is able to:

- interrogate and fully understand the model forecasts
- communicate the forecasts to its stakeholders
- understand any implications for their own forecasts such as energy demand
- readily update their view on peak demand when necessary
- cast an independent eye on the model
- ensure consistency of view in industry demand forecasts between the peak demand forecast and their own energy demand forecasts.

3.5. Issues to monitor

We recommend that Transpower and MBIE, if it is to use the peak demand forecasts, monitors the performance of the models carefully over time.

The poor fit of some of the models gives us pause for concern. Not because they are necessarily having a negative influence on forecasts at the present time but because they could in future. Model diagnostics should be checked each time a forecast is produced. A key thing to look for will be whether region-level model coefficients and model fits change significantly. If significant changes occur this could have material results on the forecast.

Transpower's out-of-sample forecasts also show a tendency to over-estimate peak demand in recent years. This is a concern and must be monitored. It is not enough of a concern to undermine our confidence in the forecasts[11] mainly because, as far as we are aware, no-one to date has fully explained the lull in demand growth we observed between, roughly, 2005 and 2015. In other words no one seems to have come up with a rigorous model that does not either over-estimate recent demand or underestimate demand growth in the early 2000s. Some models get close and there are lots of ex-post stories that seem to fit but are never tested rigorously. But, overall, empirical analysis has been unconvincing.

At the same time, Transpower's forecasts do include confidence intervals (including so-called 'prudent' or 90th percentile forecasts) and those intervals help to quantify uncertainty and to capture the extent to which our lack of knowledge of what caused past trends is translated into planning for the future. We are satisfied that the forecast intervals produced by Transpower are fit-for-purpose.

We also note that Transpower is well aware of 'out-of-model' sources of forecast error:

[11] Though it would be in other contexts.

> *We consider the forecasts in their present state to be fit-for-purpose assuming there is not a wholesale uptake of distributed generation and battery storage.*[12] [emphasis added]

We have been told that work is underway to address these questions of distributed generation and battery storage and how best to model their effects on peak demand growth. This will be a useful improvement to the overall forecasting process.

3.6. Future investment should consider alternative forecast methods

Transpower's forecast methods could potentially be improved upon by using quite different forecast methods. Any investigation of alternative methods should include evaluating the benefits of focussing on the predictability of peaks and the size of peaks using information from outside peak periods or, in the extreme, modelling the entire load duration curve.

Transpower's model currently focusses exclusively on peak demand periods. The models do not, therefore, account for information contained in demand from outside but near peak periods.

Information from outside peaks will be quite useful for predicting peaks and the magnitude of peaks. For example, if a year has many periods that are near peak but a peak is not particularly high in and of itself, the Transpower model will only measure a relatively low peak. Meanwhile, the presence of many high demand periods will increase the chances of a high single-period peak emerging. Conversely, many low demand periods plus a relatively isolated single-period peak would generally suggest a lower probability of increased peaks in future.

Another example of uses for information outside peaks is that if peaks are shifting through time – and there is some evidence that they may be – then the process of shifting will manifest itself in lower peaks or slowly growing peaks for a time. But, as the shift in peak settles-down, the probability of an increase in peak demand will get larger. A model based on trends in single-period peaks alone will not pick up on this sort of dynamic.

More sophisticated peak forecast models, ones which accounts for these sorts of dynamics and non-peak information, could be quite costly to implement for Transpower, particularly in light of the large number of forecasts that Transpower produces. Indeed one reason that we see the simplicity of Transpower's methods as a strength is because it means the methods are clear in spite of the large number of forecasts being produced. More data and more sophisticated models would seriously reduce the transparency of Transpower's peak demand forecasts

MBIE, however, only needs island and national peak demand forecasts for producing the EDGS. That being so, it would be less costly for MBIE to produce more sophisticated forecasts. With only 3 forecasts of interest, the forecasts could be produced in a way that is still reasonably transparent and may even provide for a richer description of forecasts.

[12] 'Transpower Peak Demand Forecast Updates', note received from Transpower c/- MBIE, 19 February 2016.

That said, a small scoping exercise would be advisable before any investment in MBIE's forecasting capability. This because the costs of a constructing a peak demand forecasting model would not be trivial. More work would need to be done to determine the size of potential benefits from a different forecast methodology to ensure that any investment has a positive expected net benefit.[13]

[13] Our discussion here is speculative in the sense that we have not been asked to analyse or advise on alternative forecasting methods. Our advice here is based on our own knowledge of peak demand forecasting and methods. We have not explicitly accounted for other important issues such as the scale of potential forecast performance gains or MBIE's internal forecasting resources and capability.