Attention: Mr. Warwick Anderson, General Manager, Network Investment and Pricing

An Independent Submission to the AER in response to ACTewAGL's Revised Regulatory Proposal, 2015-2019

Written by John Herbst

11 February, 2015

Feb 11, 2015

Dear Mr Anderson,

Thank you for the opportunity to make a submission regarding ACTewAGL's revised Regulatory Proposal 2015-2019. I have not studied the entire proposal, so I will comment only on the specific section I have studied thoroughly, Peak Demand forecasts. These forecasts appear to have been made with a primary goal of inflating predicted demand in order to justify expenses and increased revenue. Application of poor statistical methods and questionable judgment undermines the reliability of ACTewAGL's forecasts, as I will prove.

It is telling that all identified issues result in upward bias, never downward. This indicates that we are not looking at purely innocent mistakes. Having examined ACTewAGL's Peak Demand forecasts (Attachment E1 of the revised proposal) and taking into consideration the comments made by the firm Jacobs (Attachment E2), I am left entirely unconvinced by the methods and conclusions presented to the AER. Jacobs indicated some methodology problems that need to be 'revisited', and I believe this is a tremendous understatement. **The grave problems listed here render statistical inference (including forecasts and prediction intervals) biased and invalid. It is the AER's duty to reject the clearly biased forecasts resulting from this study.**

This submission necessarily has a 'statistical flavour' to it, thus it is important for the AER and others to weigh credentials when deciding who to trust. I hold a MS in Statistics, with a BA in Mathematics/Applied Statistics. I have lectured at the university-level in Mathematical Economics, Econometrics, Business Statistics and related courses for many years. I have taught thousands of Australian and international students about the practical and ethical issues surrounding statistical modelling, regression analysis, diagnostics, remedial measures and 'good statistical practices' in general. I have experience using all of the methods presented by ACTewAGL in its report, and I ask that the AER consider me an expert on the topics presented here.

The most obvious flaw is that **forecast intervals computed from Ordinary Least Squares (OLS) regression are invalid outside the range of the regressors used to estimate the regression coefficients.** As both Year and Population are expected to be larger in the future than any sampled values, the resulting forecasts for 2015-2020 suffer from bias. The extent of the bias is relative to the magnitude of violations of regression assumptions, which are unobservable. Forecasts made 'far from the mean of the regressors' are highly sensitive to deviations from regression assumptions, which is certainly the case here. **Forecasts which are based on absurd assumptions are absurd.**

It is unclear why ACTewAGL claims that widely accepted time-series techniques 'put too much weight on recent observations'. This is precisely what needs to be done to correct the forecasting problems inherent in OLS, and may solve other problems along the way. **OLS does not properly weigh observations for future forecasting.**

Results from as far back as 2004 are included in the model *with full weight*. Does Peak Demand in 2004 matter as much as Peak Demand in 2013 for the forecasts? Of course not! Therefore the observations should not have the same weight in the model. OLS does not handle different weights, thus another method must be employed when predicting future Demand.

It is safe to say that the full-weight inclusion of low Demand values for 2004 and 2005 has caused an inflated estimates of the coefficients on Year and/or Population. This has led to positive bias in Demand forecasts for future years. Again, a proper time-series approach to forecasting would serve to resolve this issue.

The use of 3 datapoints per year creates a serial correlation problem in the error terms. Serial correlation (also called autocorrelation) is a common problem with OLS when examining variables changing through time. A proper time-series approach would resolve this issue, as would simply choosing one datapoint per year rather than 3. Although it is possible that there could be serial correlation year over year, it would be far less pronounced that the intra-seasonal points that appear in the data. The QQ plot re-orders the data by y-value, thus does not show the correlation. The problems of uncorrected positive serial correlation include bias and incorrect calculation of error terms. **This miscalculation causes regression diagnostics to fail, causing models look like they fit better than they actually do.** It follows that forecast intervals will be biased and unreliable in the presence of uncontrolled serial correlation. **Model selection was based on an empirical analysis of incorrectly calculated error terms, thus claims that the chosen models are 'best' or even 'reasonable' are not supported by evidence.**

Forward Stepwise Model Selection ("SMS") is not appropriate for this study. SMS is an empirical, data-driven method for building a regression model. SMS, like all empirical techniques, is *methodology of last-resort.* It should be used only when choosing from a large pool of relatively homogeneous regressors, in the absence of theory. In this case, we have a small pool of different regressors, all with differing expected effects, and plenty of theory to support modelling decisions.

Forward SMS will not perform reliably when one variable in the pool is a reasonable proxy for a set of multiple other variables in the pool. It is no surprise that SMS selects Year (or Population) as the first regressor, because theory suggests these variables capture some of the effects of several of the economic variables in the pool. This is corrected by using a strong 'exclusion criterion', dropping Year or Population if a combination of other variables provides better information. All of this effort is unnecessary, as there is plenty of theory to support modelling decisions without empirical methods. Year is not a driver of change, thus the claimed interpretation of the estimated coefficient is invalid. **The true drivers of Demand should be the included, and should be the only variables included in the model.**

Stepwise model selection can also be used to find proper functional forms, cross-products and transformations, a step which is notably missing from this analysis. Without care, SMS will result in a model which is not compatible with theory, omitting important variables and substituting questionable proxies. **Even with care, one must verify that resulting models are not overly simplified, resulting in lower predictive power compared to an optimal model.**

Stepwise Model Selection contains a strong 'arbitrary' component which can be exploited by an unethical researcher. This means bias can be injected through SMS without visibility to the reader.

It appears that the purpose of SMS in this study is to justify the use of a poor set of explanatory variables which happens to provide the author's preferred results. It appears that the thresholds for inclusion were set very high, as the resulting models are very small. Similarly, it appears that the exclusion thresholds were set very high, an arbitrary and questionable choice in this case, as Year and Population will cause multicollinearity issues in models which include the other regressors.

An example of correct use of SMS: Years ago I advised a geologist attempting to model soil composition based on the colours of plants growing in it. Data was collected on thousands of different wavelengths of light, corresponding to thousands of slightly-different colours. As expected, similar colours provided mostly redundant data, so multicollinearity is a problem. Observe that the pool of potential wavelengths is large, homogeneous, and there is no theory to suggest one wavelength is a better choice than another. Where would one begin without SMS? R decided that the best regressor was a particular shade of Green, then continued to add a combination of the most useful and uncorrelated variables (a shade of red, purple, etc), stopping only when arbitrary criteria were met. The information contained in those thousands of variables was almost entirely captured by just 5 or 6 colours in the final model. However, one should never claim that an 'optimal model' has been created through this process. SMS guides the researcher, but is not to be used blindly.

If variables are known to be significant, they should be added to the model before running conditional SMS.

ACTewAGL correctly notes that inclusion of insignificant variables leads to increased variance in estimates, however it appears to have confused the problem of "inflated variance in estimated coefficients on regressors" with "inflated forecast error". Despite the inability to distinguish the relative contribution of each variable, **forecast error *improves* with the addition of theoretically significant variables to the model, even when the effect is not detectable through empirical analysis.** Population and Year are likely collinear with real drivers of peak demand like number of connections, business investment and growth, adoption of technology, etc. These specific economic variables should be in the model, rather than poor proxies. **Omitted Variable Bias is a serious problem in linear regression, and cannot be detected except through theory or outside confirmation.**

Bootstrapping is of limited use here, and is not considered a valid estimate due to the inability to estimate the higher moments of the sampling distribution. **This means bootstrapped estimates could have large or infinite variance and we would be none-the-wiser, and is the reason bootstrapping is not done in serious studies.** While it can help identify problems, it cannot be used to prove the absence of problems. We must rely on theory alone to claim absence of omitted variable bias. Likewise, the merits of back-forecasting are limited, and should never be used to override theory.

The assumption of linearity in the variables is not valid. This assumption is not necessary, but no efforts have been made to test transformations. The authors appear to have simply assumed that the fit is linear for each regressor. I would strongly recommend exploring non-linear functions of Year to improve fit and generate unbiased forecasts. **Theory suggests the relationship between Year and Demand should not be linear, and certainly not positively correlated in the future. This incorrect choice of functional form invalidates forecasts completely.** The relationship between Year and Peak Demand roughly resembles a downward parabola, so **a squared function of Year with**

a negative coefficient would be a good candidate for an additional variable. A model with indicator variables allowing for a change in trend could also help, for example using a positive slope for 2004-2012 and calculating a second (negative) slope for the future.

The dependent variable is not stochastic (not random), thus cannot be 'forecast' in the classical sense. Customers are increasingly gaining control over their Demand, and can now prevent high future Demand spikes. **This means all 10%PoE estimates are overstated (biased upward) under the classical interpretation.** Alternatively, we could re-interpret the forecasts as "the Peak Demand that would have occurred in the absence of technology, awareness and proper incentives to reduce it."

On page 24 of Attachment E1, ACTewAGL writes *"currently it has been established that PV has an inconsequential effect on peak demand."* **I challenge this claim and the validity of all forecasts based on it.** This is similar to the misleading claim made by another network that solar has 'shifted the peak' in rather than reducing it. The only way one 'shifts' a peak with solar is by lowering the original peak so far that it is no longer the peak. By definition there will always be a peak, and pv has served reduce the spikes in demand that were so feared in the last regulatory period. **The combination of load control + pv is certain to lower peak demand in the future, but this cross-product is not found in the model. Was it even considered?**

Perhaps we could find some use for these forecasts, though not for their original intention. The current forecasts assume that the PV effect is negligible. If this assumption is valid, then future forecasts will be unbiased (assuming ACTewAGL's forecasts were otherwise valid, as it claims). The PV Effect could then be estimated without bias using the formula:

Forecast Demand – Actual Demand = PV Effect

noting that Forecast Demand must be updated to use actual temperature data (and any other actual values of covariates) for the year, before applying the formula. In the absence of any other omitted variable bias, the entire difference in forecasts can be attributed to PV and relatively small amounts of sampling error, thus this estimator is the minimum variance unbiased estimator of the PV Effect, given that the regression model is valid and the error terms are normal, as ACTewAGL claims.

In small samples (and samples with low degrees of freedom), the assumption of Normal error terms is necessary to perform inference, including calculation of forecast intervals and coefficient estimates. **Since serial correlation has rendered the error terms invalid, we cannot rely on QQ-plots as evidence that the errors are normally distributed.** Therefore inference is questionable.

Arbitrary omission of outliers without unbiased *a priori* decision rules for exclusion will result in whatever bias the author wishes to inject into the study. ACTewAGL notes that it removed some low datapoints from 2012 results. **Removing low datapoints from recent periods results in unquestionably significant upward bias in future forecasts.**

The concept of 'fixing the model to produce better estimates for 2014' goes against good statistical practice. Models which focus too much on particular forecasts risk 'overfitting' or other functional form problems. In its revised proposal, ACTewAGL colour-coded results, then adjusted the model to fix the 'red ones'. This is just silly, and I am lost as to why Jacobs claims that the AER should be happy with the revised forecasts. It argued virtually the opposite point in its analysis of the original proposal (Attachment C2), noting that the discrepancy between 2013 forecasts and

results can happen in valid models and is not overly concerning when performing OLS. It chose to take no remedial action in that case, and I agree with that specific decision (ignoring the other problems). This new methodology is ethically questionable, akin to cherry-picking.

I hope the AER understands that ignoring any of the problems with OLS in a time-series regression results in unreliable and biased forecasts. The fact that all of these problems create upward bias is effectively proof that the resulting forecasts are also biased.

I would also ask that the AER examine the methodology of other networks to look for the issues I have pointed out here. In my opinion ACTewAGL's results should be read with extreme scepticism, especially given the recent discoveries that networks have been submitting misleading information to the Regulator for years.

Thank you again for this opportunity to respond to ACTewAGL's 2014-2019 Revised Regulatory Proposal.

Best regards,

[via email, 11/2/15]

John Herbst