Comments on NZCC approach for forecasting opex

Jeff Borland
Department of Economics
University of Melbourne
June 26, 2014

Introduction

I have been provided with short summaries of results, data files, and program files relating to the estimation of parameters in models for opex being undertaken by the NZ Commerce Commission (NZCC).

I have also reviewed submissions relating to forecasting opex for electricity distribution in New Zealand:
• Frontier Economics (2014a), 'Output 1: Top-down approaches for forecasting EDB costs under a DPP framework', April, Report prepared for the Electricity Networks Association of NZ.
• Frontier Economics (2014b), 'Output 3: Development of approaches to forecast EDB costs under a DPP framework', April, Report prepared for the Electricity Networks Association of NZ.
• Network Strategies (2014), 'Final report for Vector: Forecasting key inputs to DPP reset decision for electricity distribution businesses', Report number 33021.

On the basis of this information, in my opinion the empirical approach being followed by the NZCC is appropriate and the analysis has been done in a way that reflects standard practice in undertaking econometric analysis.

Commentary on material in the submissions relating to forecasting opex for electricity distribution in New Zealand.

My commentary is organised according to what I interpret to be the main themes from the submissions:
• Appropriate specification of the econometric models for network and non-network opex;
• Choice of observations to include/exclude from use in estimation of the econometric models; and
• Other issues.

1.Appropriate specification of econometric models for network and non-network opex

The issues

Two major issues are raised with regard to model specification. First, it is suggested by Frontier Economics (for example, 2014b, p.20) that it would be worth investigating econometric models for network and non-network opex that include substantially expanded sets of explanatory variables. Second, both Frontier Economics (for example, 2014a, p.40) and Network Strategies (2014, pp.28-29) propose alternative econometric models for network and non-network opex that it is argued should be used in the current DPP exercise in preference to the models used by the NZCC in its DPP reset exercise for 2010-15.

General commentary

My opinion is that the preferred econometric model to be used in forecasting opex should reflect a benefit-cost calculation of the effect on the quality of forecasts of opex from adding extra explanatory variables to the econometric models.
The degree of accuracy of forecasts of opex will depend on:
(i)The degree of accuracy in the description of the relation between opex and its driver(s) from the econometric model; and
(ii) The degree of accuracy of forecasts of the drivers of opex.

Therefore, introducing an extra driver of opex involves a tradeoff. On the one hand, including the extra driver may improve the accuracy in the description of the drivers of opex. This will improve the accuracy of the forecast of opex. On the other hand, because it will be necessary to forecast the future value of the extra driver in order to forecast opex, this introduces an added potential source of inaccuracy in forecasting opex. Whether it is optimal to incorporate the extra driver of opex will depend on the relative size of the increase and decrease in accuracy in forecasting opex.

In making this argument I understand that I am disagreeing with the position put by Frontier Economics (2014a, pp.29-30):
'This suggests that the preferred model from a statistical point of view may not always be the best model for regulatory purposes. Statistical selection procedures tend to favour parsimony. Including as few variables in the model as necessary to get a good fit in the data. By contrast, a regulatory model should lean towards including more cost drivers if this can be justified statistically. Since all the coefficients in the model [preferred by Frontier] are statistically highly significant, [that] model fits that criterion.'
My view is that the best econometric model for regulation is the model that will lead to the most accurate forecast of opex. This means that it is not the best approach for regulation to include extra cost drivers whenever they are statistically significant. An extra cost driver may be statistically significant in the econometric model for opex, but may still worsen the accuracy of the forecast of opex. This happens where the need to include a forecast of that driver in order to forecast opex becomes an extra source of inaccuracy in the forecast of opex.

Specific responses to issues

It is evident from the 2010-15 DPP reset exercise that forecasts of the drivers used to forecast opex do exhibit some degree of error. Figures 1 and 2 show respectively the forecast errors for circuit length and number of icps, expressed as the percentage deviation between the forecast and actual outcome. For circuit length, most of the forecasts are within a 5 per cent band. However, there are also some more sizable errors. For the number of icps, all forecasts are within a 5 per cent band. Circuit length appears to have been systematically over-predicted, whereas predominantly the number of icps has been under-predicted.

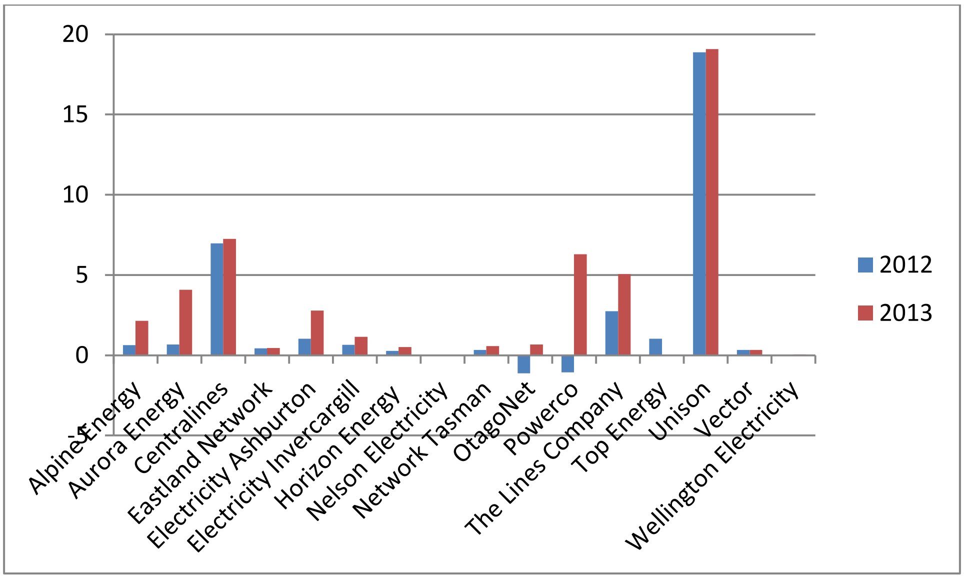

Figure 1: Forecast error of circuit length, Percent deviation of forecast from actual, 2012 and 2013

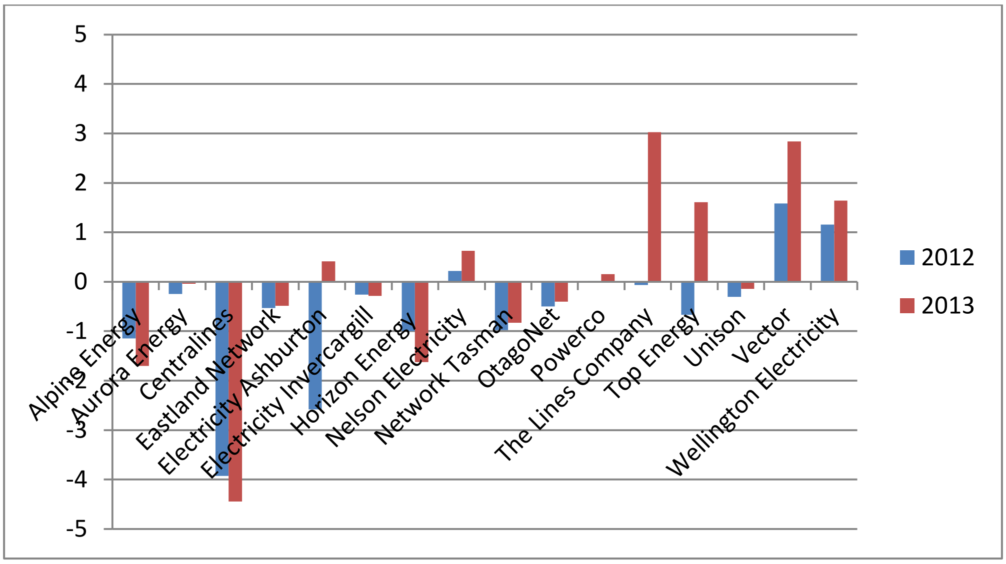

Figure 2: Forecast error in number of icps, Percent deviation of forecast from actual, 2012 and 2013

This error in forecasting the drivers of opex means that it cannot be assumed that the benefit of adding extra drivers to the econometric models for opex will outweigh the costs. This must be demonstrated.

<u>Issue 1: Adding a substantially expanded set of drivers for opex</u>

My opinion is that the benefits of adding a substantially expanded set of drivers to the econometric models of opex would be most likely to be outweighed by the costs. This is because there would only be (at best) a minor improvement on the accuracy of the description of the relation between opex and its driver(s), but the potential for significant extra inaccuracy in forecasts of the drivers of opex.

The econometric models for opex with the expanded set of drivers (Frontier Economics, 2014a, Table 21) do have superior explanatory power to the models used by the NZCC for the 2010-15 DPP reset exercise. However, the increase in explanatory power is relatively minor. For example, the adjusted R-squared for network opex is increased by 0.027 and for non-network opex is increased by 0.031. This must be compared with the costs of introducing extra errors to forecasts of opex that would come from having five extra drivers in both the models for network and non-network opex. This problem is made particularly acute by some of the new drivers depending on factors that it would be very difficult to forecast – such as the dependence of SAIDI on weather. A further difficulty with including the extra drivers is that, as recognised by Frontier Economics (2014a, pp.57-58), the theoretical basis for including some of the suggested drivers is problematic; for example, it is difficult to interpret the signs of effect of some of the drivers.

<u>Issue 2: Alternative specifications of econometric models for network and non-network opex</u>

Frontier Economics (2014a, Tables 5-7) and Network Strategies (2014, Exhibits 4.4, 4.5) make suggestions for econometric models for network and non-network opex that it is argued would be preferable to the models used by the NZCC in its DPP reset exercise for 2010-15. Here I address the suggestions that have been developed in the most detail – those made by Frontier Economics.

Frontier Economics suggest that both the models for network and non-network opex should include circuit length and icps/circuit length as explanatory variables. This is based on arguments that: (i) It is preferable for interpretation to distinguish drivers into size and density variables; and (ii) The suggested model has greater explanatory power for non-network opex than the NZCC model.

Network opex: As is recognised by Frontier Economics (2014a, p.28) the model they propose is statistically identical to the NZCC model for network opex. Hence the forecasts derived from both models are identical.

Non-network opex: For non-network opex the suggestion made by Frontier Economics consists of adding an extra explanatory variable, circuit length, compared to the NZCC model. To compare these models I have: (1) Estimated each model using data from 2010-2011; (2) Used each model to forecast non-network opex for each supplier for 2012 and 2013. The forecast is made using the parameters estimated from the econometric models and forecasts of circuit length and icps for 2012 and 2013 for each supplier made by the NZCC in the 2010-15 reset exercise ; and (3) Calculated summary measures of the size of forecast error for supplier for each model in 2012 and 2013. The forecast error is equal to the difference between a supplier's actual reported opex and the forecast of opex.

To estimate the NZCC model I exclude observations for Orion and Buller for 2011, and to estimate the Frontier model I also exclude the observation for TCL for 2011. (Estimates of the NZCC model are hardly changed if it is estimated also excluding the TCL observation for 2011.) Results from the models are reported in Table 1.

	NZCC model	**Frontier model**
Constant	0.056278 (0.369)	0.011008 (0.374)
Circuit length		0.831807 (0.040)
Icps/Circuit length		0.773071 (0.062)
Icps	0.814739 (0.035)	
Adjusted R-squared	0.906	0.906
NOBS	56	55

Table 1: Models for non-network opex, 2010-2011

Summary measures of forecast errors for the NZCC and Frontier models for non-network opex are presented in Table 2. It is evident that the models are very close in the accuracy of their forecasts of non-network opex in 2012 and 2013. For both years the Frontier model has slightly better forecast accuracy than the NZCC model using the criteria of minimum mean absolute deviation and mean squared error.

	2012		**2013**	
	NZCC	**Frontier**	**NZCC**	**Frontier**
Mean absolute deviation	0.2291	0.2241	0.2859	0.2733
Mean squared error	0.0834	0.0799	0.1256	0.1196
Correlation between forecast and actual opex	0.965	0.967	0.958	0.962

Table 2: Measures of forecast accuracy of NZCC and Frontier models for non-network opex

My recommendations

• Retain the same econometric model for network opex that was used in the 2010-15 DPP reset exercise.

• At present retain the same econometric model for non-network opex that was used in the 2010-15 DPP reset exercise. Review this decision when an extra year of data on opex (for 2014) becomes available later in the year.

In my opinion the argument in favour of adopting the Frontier model for non-network opex is that the analysis I have undertaken does indicate that adopting the Frontier model might provide a slight increase in forecast accuracy.

In my opinion the arguments against adopting the Frontier model for non-network opex are that: (a) Moving to the Frontier model does involve adding an extra driver of non-network opex which then must also be forecast, which can be a source of forecast error. For 2012 and 2013 it appears that this effect has been outweighed by the benefit of the model providing a better description of the relation between opex and its drivers. However I am concerned that, as the length of time since the reset exercise lengthens, there may be a larger forecast error in circuit length, which might imply that the overall effect of adding an extra driver could be to reduce forecast accuracy; and (b) Making a change to the econometric model for non-network opex introduces a change to the regulatory environment. Given the likely benefits that stability in the regulatory environment has for suppliers' decision-making on investment, it seems preferable to make changes to the econometric models for opex only when it is likely to result in a significant improvement in forecast accuracy.

In my opinion, at this stage there is not a sufficiently strong basis to conclude that it is worth changing the econometric model for non-network opex to the Frontier model. However, I do believe that this is an issue that should be kept under active consideration. It would be worth evaluating again once data on suppliers' opex for 2014 becomes available – That data will enable the forecast accuracy of the models for that extra year to be assessed.

2. Choice of observations to include/exclude from use in estimation of the econometric models

Issues

The approach adopted by the NZCC in the 2010-15 DPP exercise was to apply 4 tests to identify outlier observations and to exclude observations that were identified as outliers by at least 3 of the tests. Frontier Economics (2014a, pp.31-32) propose an alternative approach: (1) Excluding all observations for Buller and Nelson. This is done on the basis of an inspection of graphical presentations of data on the relation between log(opex) and log(circuit length) and log(icp) which it is argued suggests that these suppliers (both small suppliers) do not exhibit the same linear relationship between log(opex) and log(circuit length) and log(icp) that holds for other suppliers; and (2) Undertake further investigation to establish whether observations identified as outliers might in fact involve incorrect data entry by a supplier. Frontier Economics (2014a, p.30) adopt the same approach as the NZCC in deleting the 2011 observation for Orion due to the Canterbury earthquake. (Network Strategies, 2014, pp.29-30 indicate that with their econometric models there was not evidence of outliers.)

Specific responses to issues

Issue 1: Exclude all observations for Buller and Nelson

I disagree with this suggestion by Frontier Economics. Suppose that there is indeed a different relation between opex and its drivers for small suppliers compared to medium and large suppliers. Dropping the observations for the small suppliers would then mean that the econometric models for the relation between opex and its drivers is appropriate for forecasting opex for medium and large suppliers but not for small suppliers. This would be likely to lower overall forecast accuracy.

Tables 3 and 4 below show the forecast accuracy of the models for network and non-network opex including and excluding observations for Buller and Nelson (as well as observations for Orion and TLC in 2011).

	2012		**2013**	
	Including Buller and Nelson	**Excluding Buller and Nelson**	**Including Buller and Nelson**	**Excluding Buller and Nelson**
Mean absolute deviation	0.2179	0.2012	0.2186	0.2286
Mean squared error	0.0768	0.0684	0.0746	0.0758

Table 3: Measures of forecast accuracy of the NZCC preferred model for network opex, Including and excluding observations for Buller and Nelson

	NZCC model			
	2012		**2013**	
	Including Buller and Nelson	**Excluding Buller and Nelson**	**Including Buller and Nelson**	**Excluding Buller and Nelson**
Mean absolute deviation	0.2291	0.2294	0.2859	0.2888
Mean squared error	0.0834	0.0851	0.1256	0.1282
	Frontier model			
	2012		**2013**	
	Including Buller and Nelson	**Excluding Buller and Nelson**	**Including Buller and Nelson**	**Excluding Buller and Nelson**
Mean absolute deviation	0.2241	0.2251	0.2733	0.2785
Mean squared error	0.0799	0.0820	0.1196	0.1222

Table 4: Measures of forecast accuracy of the NZCC and Frontier models for non-network opex, Including and excluding observations for Buller and Nelson

For the model for network opex there are mixed findings regarding the effect of excluding observations for Buller and Nelson. Excluding those observations improves forecast accuracy slightly for 2012, but worsens forecast accuracy for 2013. For both the NZCC and Frontier models for non-network opex the effect of excluding observations for Buller and Nelson is to lower forecast accuracy in both years. Overall there does not appear to be a strong case that excluding observations for Buller and Nelson will improve forecast accuracy.

An alternative approach to dealing with potential differences between small and large suppliers is to attempt to incorporate the different relation between opex and its drivers for small suppliers into the econometric models for opex. This means that the econometric model can then be used as an appropriate basis for forecasting opex for small as well as medium and large suppliers. Hence I have sought to test whether there does appear to be a significantly different relation between opex and its drivers for small suppliers, and whether incorporating that into forecasts of opex causes an improvement in forecast accuracy.

To apply the alternative approach I begin by estimating econometric models for network and non-network opex that add to the NZCC preferred specifications a dummy variable for Buller and Nelson and the interaction of the dummy variable for Buller and Nelson with the explanatory variable(s). Table 5 reports the results from estimating these alternative regression models using data for 2010-2011.

	Network opex	**Non-network opex**
Constant	-0.7658 (0.491)	-0.0077 (0.399)
Circuit length	0.4763 (0.104)	
Icps	0.4991 (0.079)	0.8206 (0.037)
Circuit length*(Buller and Nelson)	-16.40 (11.7)	
Icps*(Buller and Nelson)	-22.10 (15.0)	-0.8846 (0.403)
Buller and Nelson	291.67 (202.0)	7.954 (3.54)
Adjusted R-squared	0.897	0.905
NOBS	56	57

Table 5: Models for network and non-network opex, 2010-2011

There is not evidence of a significantly different relation between small and medium/large suppliers in the drivers of network opex. The extra explanatory variables in the model are not individually significant, and an F-test finds that the variables are not jointly significant. There is evidence of a significantly different relation between small and medium/large suppliers in the drivers of non-network opex. The extra variables are individually and jointly significant.

Next I test the forecast accuracy of the alternative models for both network and non-network opex. Tables 6 and 7 present summary measures of the forecast errors for network and non-network opex respectively for the NZCC preferred model and the alternative model incorporating the drivers for Buller and Nelson.

	2012		**2013**	
	Preferred	**Alternative**	**Preferred**	**Alternative**
Mean absolute deviation	0.2179	0.2111	0.2186	0.2932
Mean squared error	0.0768	0.0727	0.0746	0.1498
Correlation between forecast and actual opex	0.976	0.973	0.967	0.953

Table 6: Measures of forecast accuracy of NZCC and alternative models for network opex

	2012		**2013**	
	Preferred	**Alternative**	**Preferred**	**Alternative**
Mean absolute deviation	0.2291	0.2218	0.2859	0.2865
Mean squared error	0.0834	0.0833	0.1256	0.1265
Correlation between forecast and actual opex	0.965	0.964	0.958	0.960

Table 7: Measures of forecast accuracy of NZCC and alternative models for non-network opex

For network opex it appears that the alternative model has slightly better forecast accuracy than the NZCC preferred model for 2012, but much worse forecast accuracy for 2013. For non-network opex it appears that the NZCC preferred model and the alternative model have similar forecast accuracy. On the basis of these results, in my opinion it is appropriate to retain the preferred NZCC econometric models rather than seeking to directly model differences in the drivers of opex for small suppliers. However this is an issue that should be revisited as extra years of data on opex become available for testing between the NZCC preferred econometric models and the alternative models that incorporate the separate drivers for Buller and Nelson.

My recommendations

• I do not believe it is appropriate to exclude observations for Buller and Nelson from estimating the econometric models for Opex.
• I do not believe it is appropriate to adopt an approach of incorporating differences in the relation between opex and its drivers for small and medium/large suppliers into the econometric models for opex. However, it would be important to investigate this issue further once extra years of data on opex for testing the forecasting performance of such an alternative model become available.
• NZCC should retain the approach of testing for and deleting outliers.

<u>Issue 2: Undertake further investigation to establish whether observations identified as outliers might in fact involve incorrect data entry by a supplier</u>

My recommendation

I agree with part (2) of the approach suggested by Frontier Economics. It seems that this point applies just to the observation for TLC in 2011. This would be worth checking with the supplier as to whether incorrect data entry might have occurred.

<u>**3. Other issues**</u>

<u>(a)Time effects</u>

Frontier Economics (2014a, p.38; 2014b, p.19) provide some evidence to suggest that there may be omitted time effects from the NZCC econometric models for opex. The main evidence for this at present is the finding of a significant year effect for 2013 in a model for non-network opex. As is acknowledged in the discussion in the Frontier Economics report, at this stage it would be difficult to incorporate time effects into the econometric models for opex in a way that could be applied in forecasting. Incorporating time effects would become important if it was found that, in addition to drivers for opex currently included in the econometric models, there was a systematic time effect on opex – for example, a time trend or a finding that there has been a fixed structural shift in opex in recent years. Therefore the appropriate response is to maintain a 'watching brief' on this issue, undertaking further investigation in the future when extra years of data will be available.

My recommendation

It is not necessary to seek to incorporate time effects in the current econometric modelling of opex. However, this is an issue that merits on-going consideration as extra years of data become available.

(b) Post-estimation adjustment

Frontier Economics (2014a, pp.74-76; 2014b, pp.21-22) suggests that it may be appropriate to use a forecast of opex that is partly based on the forecast from an econometric model and partly based on supplier forecasts:
Adjusted forecast = w*Model forecast + (1-w)*Supplier forecast.
The major problem with this proposed methodology is that, as is acknowledged in the discussion by Frontier Economics, it relies on suppliers having an incentive not to over-state their forecasts of costs. In the absence of a mechanism to guarantee that incentive, it does not seem appropriate to proceed with this proposal.

My recommendation

I would recommend against adopting the post-estimation adjustment approach.

(c) Is Vector different?

In the conclusion of its report Network Strategies (2014, p.37) suggests:
'In the course of our analysis, it became evident that there are a number of clear difference between Vector and the other EDBs. Vector is the largest of all the EDBs and its Auckland market...is unique within New Zealand... ...as each of the EDBs are given equal weighting within the analysis, the models may not deliver an outcome that is a best fit for Vector's unique situation.'

There does not seem to be evidence to support the conclusion that the NZCC preferred models for opex perform significantly differently in forecasting opex for Vector compared to other suppliers. First, the forecast errors for Vector do not seem significantly different to those for other suppliers. Figure 3 below shows the absolute forecast errors for network and non-network opex by supplier for 2012 and 2013 (where forecasts of opex are derived using the NZCC preferred models). It is apparent that the forecast errors for Vector are not notable; in fact, they are at the lower end of the distribution of errors by size. Second, econometric analysis shows that there is no significant difference in the relation between opex and its drivers for Vector compared to other suppliers. This is shown in Table 8 below using data for 2010 to 2013 (excluding observations for Orion and TLC in 2011).

My recommendation

It is not necessary to modify the econometric modelling to seek to account for differences between Vector and other suppliers.

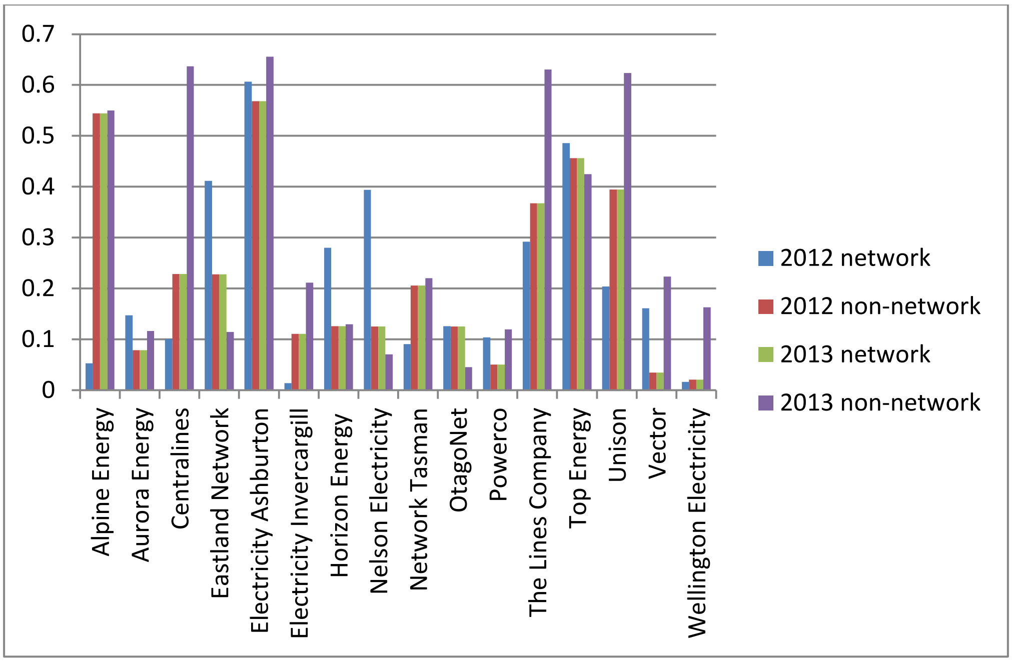

Figure 3: Absolute forecast error for network and non-network opex by supplier, 2012 and 2013

	Network opex	Non-network opex
Constant	-0.2936 (0.332)	0.5145 (0.309)
Circuit length	0.5196 (0.058)	0.7778 (0.029)
Icps	0.4188 (0.055)	
Circuit length*(Vector)	-8.063 (117.09)	0.0047 (0.013)
Icps*(Vector)	6.002 (86.88)	
Adjusted R-squared	0.891	0.886
NOBS	113	115

Table 8: Models for network and non-network opex, 2010-2011