Heterogeneity in Electricity Distribution Networks

Testing for the presence of latent classes

Date: 12 February 2015

Version: 1.00

Prepared for Ergon Energy

www.huegin.com.au

Level 10, 2 Elizabeth Plaza, North Sydney, NSW 2060

02 9409 2400

Introduction

Benchmarking in the presence of heterogeneity

Under the regulatory framework applied by the Australian Energy Regulator (AER) when evaluating electricity network expenditure forecasts, the AER must consider cost efficiency through benchmarking. The small sample size in Australia and diverse range of operating conditions.

In its most recent decision - the draft decision for NSW and ACT distribution - the AER introduced data from Ontario, Canada and New Zealand to facilitate the utilisation of a Stochastic Frontier Analysis (SFA) econometric model of efficiency (SFA requires more data than is available in Australia along). The SFA econometric model assumes a common opex cost function for the 68 networks in the sample population. Whilst some adjustments can be made to account for environmental factors that differ across the networks, unobserved heterogeneity remains perhaps the greatest challenge to the model efficacy.

Recent research in electricity network benchmarking suggests that in a population of electricity networks latent classes (groupings of networks with similar attributes within the group but differences to other groups) exist. The presence of these latent classes, if not recognised, leads to comparison of networks to an efficiency frontier that is not appropriate for its individual circumstances. This in turn leads to exaggeration of inefficiency.

In Huegin's opinion, latent classes exist in the data relied upon by the AER in the NSW and ACT draft decision. Our view is that the presence of these classes has led the AER to overestimate the efficiency gap of several Australian networks. This briefing note provides a summary of the issue.

How is Ergon Energy's opex predicted under the AER approach?

In the recent NSW and ACT draft decision for electricity distribution, the Australian Energy Regulator (AER) adjusted the base year expenditure of the NSW and ACT businesses based on the results of econometric modelling conducted by their consultants, Economic Insights. The AER utilised a Stochastic Frontier Analysis (SFA) model with a Cobb-Douglas functional form. SFA modelling requires many more data observations than that available in the Australian context, so Economic Insights introduced data from Ontario, Canada and New Zealand distribution networks. In all, data from 68 networks was used to estimate the coefficients of the SFA opex function - only 13 of which are Australian. Huegin contends that the likelihood of a single opex cost function across 68 networks in three different countries of significant environmental diversity is unlikely. The potential error in the model results from data incompatibility alone is significant, but of more concern is the irrelevance of the SFA model to Australian businesses (given that, by weight of numbers, the model form is influence by the many, very small Ontarian networks). If the AER continues to rely on the Economic Insights SFA model to predict efficient levels of base year expenditure, Ergon Energy faces the significant risk of having its base year opex set through a cost function that assumes:

1. All networks can be compared by the same, simple algebraic equation of cost; and

2. That the many, small yet relatively dense networks of Ontario are an appropriate comparison basis for Australian businesses.

This risk is particularly material for Ergon Energy - a network that is unique compared to most other Australian networks let alone Ontario and New Zealand.

Why it matters - unobserved heterogeneity

Electricity distribution networks differ through the legacy of their design, topographical and geographical attributes of their location and regulatory and legislative requirements associated with the local jurisdiction. It is therefore important to recognise the potential for unobserved heterogeneity when using economic benchmarking techniques to compare the expenditure, and efficiency, of electricity network businesses. Unobserved heterogeneity occurs when variables other than those included in the econometric model specification have influence on the dependent variable under study (in this case, opex). The absence of these variables in the model specification can lead to erroneous inferences about efficiency. That is, differences in the observed and predicted level of opex will include the influence of these material, but unaccounted for, variables. Because the variables are unobserved, the influence is often translated as inefficiency.

How it is mitigated - latent class modelling

There is a variant of econometric modelling known as latent class modelling, where algorithms classify groups or clusters of networks by categories defined by characteristic similarities within the group and differences between groups. Latent class modelling requires reasonably large datasets, and dividing networks in Australia into classes would tend to produce several groups of very small membership. However failure to recognise that there are legitimate differences between the networks renders the measurement of relative efficiency unreliable when efficiency scores are generated from a single model. Comparing Ergon Energy or Essential Energy to CitiPower, for example, on the basis of the raw results would not provide a meaningful assessment of relative productivity.

Huegin have in the past tested the sensitivity of individual DNSP scores to changes in the weightings on the variables in the AER models. That analysis demonstrated that not only are the results very sensitive to such changes, but it also showed that the magnitude and direction (i.e. negative or positive) of the change in the productivity score was similar for certain groups of businesses. That is, small groups of businesses exhibited similar patterns of change in the results with variations in the weightings, and these groups exhibited very different patterns from other groups. This is an indication of the existence of attribute based classes in the data. That is, the relationship between the combination of inputs and outputs is similar for certain businesses.

To test the existence of latent classes in the AER dataset, Huegin ran analysis using the AER's SFA model and the data for the 68 networks used by the AER as the input to its model. Latent class modelling requires an assumption of the number of classes before running the model. As such, Huegin ran analysis on the data set under the assumption of between one and five classes. The Akaike Information Criterion[1] was used to determine the optimal number of classes. We found the four-class assumption was strongest. These results suggest the presence of four distinct technological groups among the DNSPs in the dataset. The groupings of the networks (for the Australian businesses only) for each assumption is shown below.

Table 1: Latent class SFA model groupings - Australian networks only

DNSP	2 Classes estimated	3 Classes estimated	4 Classes estimated	5 Classes estimated
ActewAGL	1	1	1	1
Ausgrid	1	1	1	1
CitiPower	2	2	3	3
Endeavour Energy	1	1	4	1
Energex	1	1	4	1
Ergon Energy	1	1	1	1
Essential Energy	1	1	1	1
Jemena	2	3	4	4
Powercor	2	2	2	5
SA Power	2	3	2	4
AusNet Services	2	3	4	4
TasNetworks	2	3	4	4
United Energy	2	3	2	5

These groupings rely upon the actual variables chosen by the AER (customers, line length and ratcheted peak demand), but importantly they demonstrate that networks in Australia should not be considered in a single class. Further, at any assumption level between 2 and 5 classes, Ergon Energy does not sit in the same class as any of the frontier networks identified by the AER (CitiPower, United Energy, AusNet Services, SA Power Networks and Powercor).

Latent class modelling of the AER SFA model is still subject to the same issues of the underlying SFA model and the input data from the 68 networks. However the analysis conducted by Huegin shows that networks should not be compared as a single class or group, thereby casting doubt on the validity of measuring the efficient

[1] The Akaike Information Criterion is a measure of the statistical quality of a model relative to other models.

level of opex of businesses such as Ergon Energy against the frontier networks - all of which are in a different class cluster. To further test the existence of attributional classes, we used other segmentation methods on the data set. This analysis is discussed in the next section.

Confirming existence of classes - k-means clustering

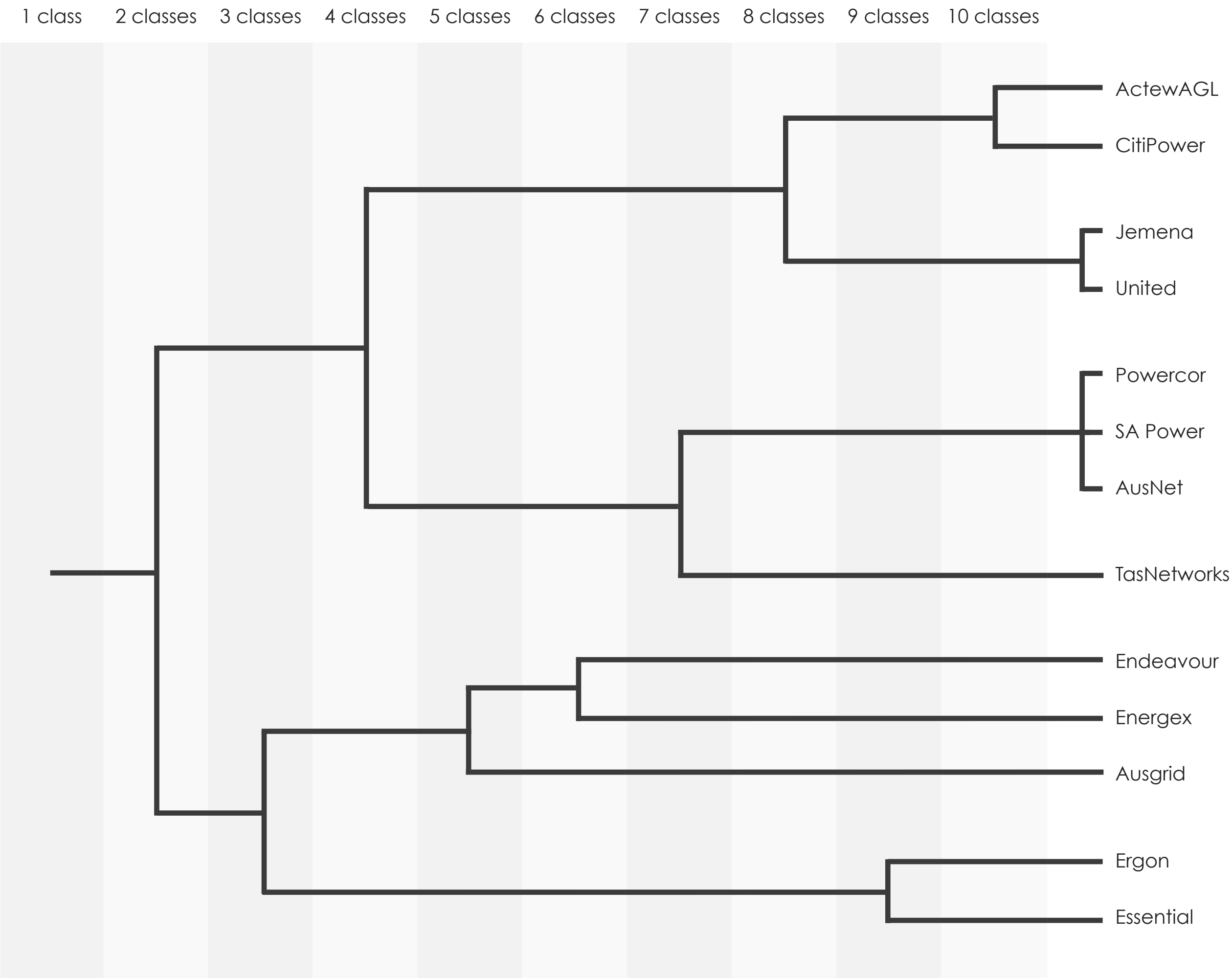
A more simple means of analysing the existence of classes in a data set is through clustering algorithms. These are computational methods capable of segmenting entities in a data set into classes of similar attributes. A common statistical clustering method is the k-means clustering technique which aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. We ran a k-means clustering algorithm over the 13 Australian networks using the variables outlined in the table below.

Table 2: Network variables used in cluster analysis

Electricity Supply	Physical Network Attributes	Performance	Density
Customer numbers	Overhead lines	Customer minutes off supply	Customer density
Maximum demand	Underground cables	SAIDI	Demand density
Energy distributed	Zone transformers	SAIFI	Energy density
	Distribution transformers		
	Circuit length		
	Share of single stage transformation		
	Transformers excluding first stage		
	System capacity		

Figure 1 on the following page illustrates the hierarchy of classes and membership of each group or cluster. Figure 1 has been constructed through the iterative application of a clustering algorithm. To create the hierarchy shown in figure 1, we ran analysis on the range of assumptions from two classes to ten classes. That is, we ran the clustering algorithm on the basis of an assumption that there are two classes of asset type amongst the 13 DNSPs, then an assumption of three classes and so on until the final run at the assumption of ten individual classes of DNSP. The result is the segmentation of businesses in clusters at each assumption, showing the nature and enduring strength of the similarities between networks based on the variables identified in the table above.

Figure 1: DNSP class hierarchy - k-means clustering



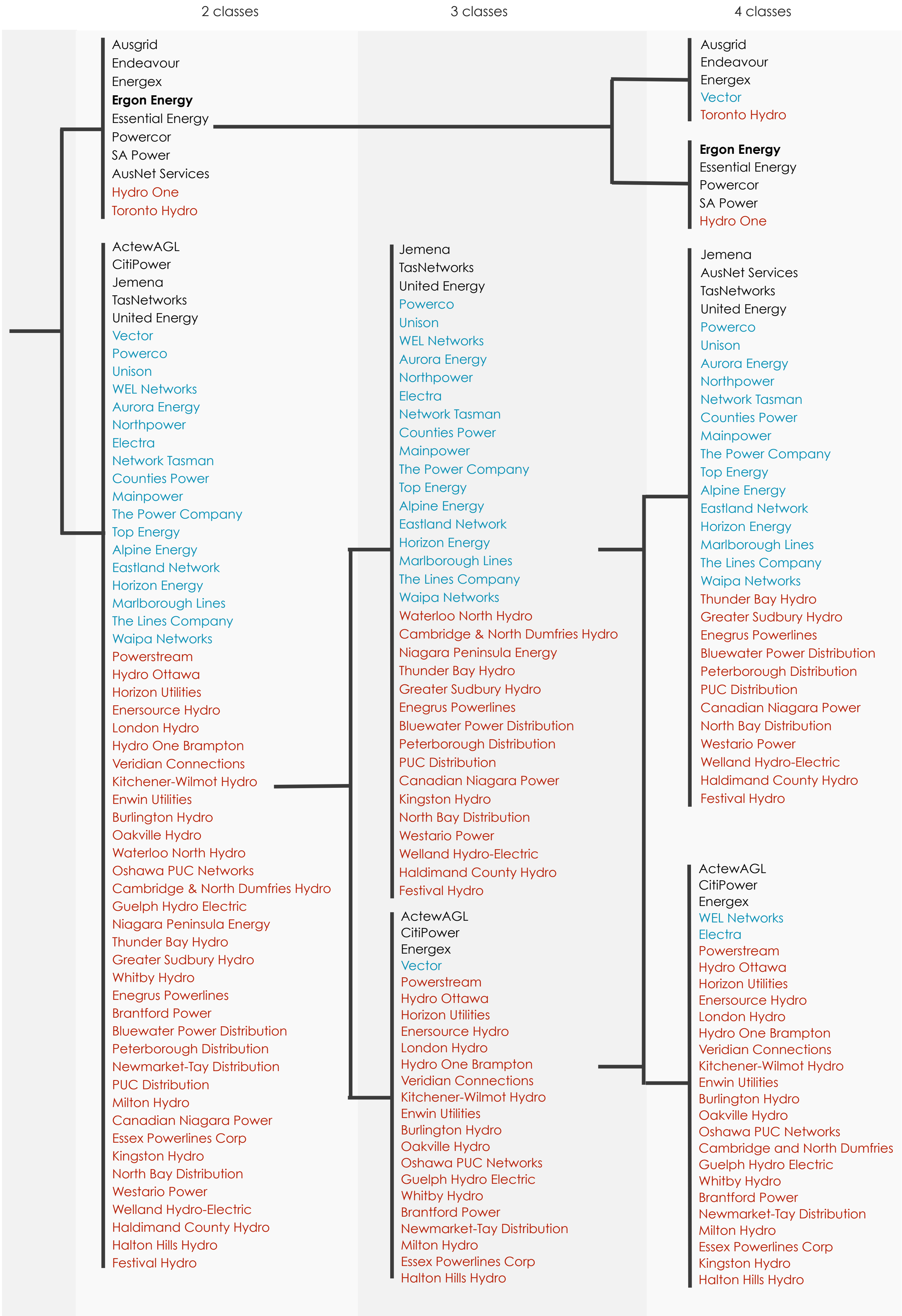
Once again, this analysis demonstrates that even at the two-class assumption, Ergon Energy should not be considered comparable to the frontier businesses in the AER analysis.

Figure 1 shows the clustering of networks using multiple variables and Australian data only. We conducted a final test of the presence of classes, which would influence the AER's efficiency results from the NSW and ACT draft decision, but has not been considered by the AER. We ran a k-means clustering analysis over the 68 networks in the AER data set and using the same variables as those included in the AER's SFA model. The results were quite revealing in terms of the validity of the SFA model to Ergon Energy's circumstances. At the two-class assumption level, the 68 businesses arrange themselves in a 58:10 split. That is, one of the two classes has only ten members, and:

1. Ergon Energy is in that class;
2. Only two international networks appear in that class - the two largest non-Australian networks, Hydro One and Toronto Hydro from Ontario; and
3. That group endures through the three class assumption also.

Figure 2 shows the arrangement of the networks at the two-class to four-class assumption levels. By the time four classes are assumed, Egon Energy has only four other networks in its group based on the AER's model and data. Whilst Powercor and SA Power Networks are in that group, the clustering only considers the variables in the AER SFA model. Addition or consideration of other variables not in the SFA model would produce different clustering results. The exercise of testing for clusters within the data demonstrates that the AER has erred in selecting a single cost function for all 68 businesses.

Figure 2: k-means clustering in the AER SFA data set - 2-class assumption



Contact Details

Jamie Blair

Level 10, 2 Elizabeth Plaza, North Sydney, NSW 2060

t: +612 9409 2400
e: info@huegin.com.au

huegin