Memorandum

From: Tim Coelli and Denis Lawrence **Date:** 28 February 2019

To: AER Opex Team

Subject: Appropriate specification for the inclusion of the share of underground cables variable in opex cost function models – technical issues

Before proceeding to technical modelling issues, we will first consider the reasons why underground cables are generally less costly to operate and maintain than overhead lines. The two primary areas of difference are vegetation management and general opex. Emergency management costs would also be impacted.

There are virtually zero vegetation management costs associated with underground assets. Given that vegetation management represents 30 per cent or more of opex for many DNSPs, this can represent a significant opex reduction.

Underground assets are very difficult to inspect and maintain. As such, the strategy for most distribution assets is to run to failure, or to rely on predicative failure analysis to drive replacement. On the other hand, overhead assets are inspected on a very regular basis. Most overhead lines are patrolled at least once a year (especially in bushfire prone areas) and most poles are inspected in 3–yearly intervals. Wooden pole inspection represents a material component of most opex budgets.

Emergency management costs are also reduced for underground assets as they are less exposed to contact with third parties including trees, animals, wind–borne debris and lightning. When a fault does occur on an underground asset, the repair costs can be significant but large repairs to underground assets are often treated as capex rather than opex.

The opex cost function economic benchmarking models used in Economic Insights (2014, 2018) include a *shareugc* variable defined as 'the share of underground cable length in total line and cable length' to account for the fact that underground lines are expected to be less costly to maintain relative to above ground lines for the reasons outlined above. This *shareugc* variable is included in log form in the opex cost function models and is expected to have an estimated coefficient with a negative sign.

NERA (2018, p.22) question the logic behind including this variable in log form and instead argue that it should be included in linear form in the model. Their main argument against the log form is summarised as follows:

> 'For example, consider two hypothetical DNSPs, each with 100km of total circuit length. DNSP A has 10km underground and 90km overhead, while DNSP B has 50km underground and 50km overhead.
>
> 'Both DNSPs propose to underground 1km of overhead line, increasing their underground share by 1 percentage point each. For both DNSPs, this would result

in opex savings associated with reduced maintenance and vegetation management on 1km of network, an equal savings for both networks.

‘However, according to the AER’s approach, DNSP A has increased its underground share by 10 per cent (1km undergrounded divided by existing 10km undergrounded), while DNSP B has only increased its underground share by 2 per cent (1 new km underground divided by existing 50km underground). The AER’s model thus assumes that DNSP A will be able to achieve a 1.6 per cent reduction in opex, while DNSP B will only be able to achieve a 0.32 per cent reduction in opex.’

This argument has some superficial attraction. However, the assertion that converting one kilometre of overhead line to underground line should result in equal savings for any particular kilometre of line can also be challenged. This assertion is unlikely to be correct as the proportion of underground lines changes for a particular DNSP. For example, in the absence of legislative restrictions it is likely that a DNSP will identify those parts of its overhead network which are most costly to maintain and make them a priority for conversion to underground network, as the opportunity arises (eg as overhead lines near the end of their lifespan). Hence, the first kilometre of line put underground is likely to produce larger *opex* savings than the next kilometre and so on. For example, a DNSP would most likely target those parts of its overhead network with large amounts of vegetation, higher probabilities of lightning strikes, older infrastructure or otherwise problematic outage histories and put them at the top of its priority list for new undergrounding. This is the classic ‘low hanging fruit’ argument, where there is diminishing marginal benefit (ie reduced marginal cost savings) from each additional km of line converted to underground.

On the other hand, it may be that much of the change in *shareugc* over the past decade in Australia has involved mostly greenfield rather than brownfield development. While each state has different regulations in terms of where an overhead line is acceptable and where not, in general all jurisdictions require new residential areas to be underground. These represent the greatest proportion of underground growth in the NEM. Public sentiment and requirements for public consultation may also represent ‘soft’ drivers for undergrounding. The time and cost associated with negotiating an overhead line in a built–up area may be so material as to provide an incentive for the DNSP to avoid the process altogether.

Thus, if the second of the above effects predominates, one could argue that an econometric model that provides estimates of cost savings per km of line undergrounded which do not vary substantially across sample observations may be preferable. At face value, the NERA argument may be seen to provide support for *shareugc* to be included in a linear form, however our analysis below suggests otherwise when all factors are taken into account.

In this memo we use a number of empirical methods to investigate the relative merits of including the *shareugc* variable in log form versus in linear form in the *opex* cost function model. Given that the dependant variable (*opex*) is in log form, we refer to these two options as ‘log–log’ and ‘log–lin’, respectively.

The two models may be defined as follows. First the log–log model is given by:

$$\ln(opex) = \beta_0 + \sum_{i=1}^{3} \beta_i \ln(y_i) + \beta_z \ln(shareugc) + \beta_t t$$

and then the log–lin model is given by:

$$\ln(opex) = \beta_0 + \sum_{i=1}^{3} \beta_i \ln(y_i) + \beta_z shareugc + \beta_t t$$

where the y_i denote the three output variables of customer numbers (*custnum*), circuit length (*circlen*) and ratcheted maximum demand (*rmdemand*), t is a time trend and the betas are unknown parameters to be estimated.

Note that *shareugc=ugc/circlen* where *ugc* is the underground circuit length.

In this memo we investigate the relative merits of these two alternative models by conducting the following two exercises:

1. We derive expressions for elasticities and marginal effects for the log–log and log–lin models and then investigate the degree to which these measures actually differ across the two models using sample data on the different DNSPs.
2. We use statistical criteria to attempt to distinguish between the two alternatives. This involves the use of non–nested hypothesis tests and also model selection criteria.

1. Elasticities and marginal effects

The log–log and log–lin models imply different elasticities and different marginal effects for both *shareugc* and for *ugc*. The fact that the models involve logs and ratios means that these measures are not always easy to identify at first glance.

Elasticities measure the percentage change in one variable when there is a one per cent change in another variable. In this case we are looking at the percentage change in costs resulting from a one per cent change in the share of underground itself (for the log–log case) or a one percentage point change in the value of this share (for the log–lin case).

Marginal effects (or marginal products), on the other hand, measure the change in the dollar value of costs when there is a one kilometre change in the length of underground cables.

We now derive these elasticities and marginal products. To simplify our algebra, we define

$$\alpha_0 = \beta_0 + \sum_{i=1}^{3} \beta_i \ln(y_i) + \beta_t t$$

so that we can then simplify the above two model expressions to obtain:

$$\ln(opex) = \alpha_0 + \beta_z \ln(shareugc)$$

for the log–log model and

$$\ln(opex) = \alpha_0 + \beta_z shareugc$$

for the log–lin model.

The marginal effect of a 1km change in underground circuit (*ugc*) on *opex* is of particular interest because it provides a tangible measure that can be easily interpreted. First, consider the log–log model:

$$\ln(opex) = \alpha_0 + \beta_z \ln(shareugc)$$

$$\ln(opex) = \alpha_0 + \beta_z \ln(ugc/circlen)$$

$$\ln(opex) = \alpha_0 + \beta_z[\ln(ugc) - \ln(circlen)]$$

$$\ln(opex) = \gamma_0 + \beta_z \ln(ugc)$$

where $\gamma_0 = \alpha_0 - \beta_z \ln(circlen)$.

The elasticity can be shown to be the partial derivatives in logs:

$$\varepsilon_{ugc}^{opex} = \frac{\partial opex}{\partial ugc} \times \frac{ugc}{opex} = \frac{\partial \ln(opex)}{\partial \ln(ugc)} = \beta_z$$

and hence the marginal effect is:

$$\frac{\partial opex}{\partial ugc} = \beta_z \frac{opex}{ugc}.$$

This process can be repeated for elasticities and marginal effects with respect to *shareugc*, where we obtain:

$$\varepsilon_{shareugc}^{opex} = \frac{\partial opex}{\partial ugc} \times \frac{shareugc}{opex} = \frac{\partial \ln(opex)}{\partial \ln(shareugc)} = \beta_z$$

$$\frac{\partial opex}{\partial shareugc} = \beta_z \frac{opex}{shareugc}$$

Furthermore, in the NERA example above they define a measure which is neither an elasticity nor a marginal effect. It is the percentage change in *opex* resulting from a one unit change in *shareugc*, which we will call '*pchange*'. This is a kind of hybrid mix of these two measures and is calculated as the elasticity divided by *shareugc*:

$$pchange_{shareugc}^{opex} = \frac{\varepsilon_{shareugc}^{opex}}{shareugc} = \frac{\beta_z}{shareugc}$$

Next, we repeat the above derivations for the case of log–lin model:

$$\ln(opex) = \alpha_0 + \beta_z shareugc$$

$$\ln(opex) = \alpha_0 + \beta_z\left(\frac{ugc}{circlen}\right)$$

$$opex = exp\left[\alpha_0 + \beta_z\left(\frac{ugc}{circlen}\right)\right]$$

$$\frac{\partial opex}{\partial ugc} = \frac{\beta_z}{circlen} \times exp\left[\alpha_0 + \beta_z\left(\frac{ugc}{circlen}\right)\right] = \beta_z \frac{opex}{circlen}$$

The elasticity is then defined as:

$$\varepsilon_{ugc}^{opex} = \frac{\partial opex}{\partial ugc} \times \frac{ugc}{opex} = \beta_z \frac{opex}{circlen} \times \frac{ugc}{opex} = \beta_z \frac{ugc}{circlen} = \beta_z shareugc$$

This process can be repeated for elasticities and marginal effects with respect to *shareugc*, where we obtain:

$$\frac{\partial opex}{\partial shareugc} = \beta_z opex$$

$$\varepsilon_{shareugc}^{opex} = \beta_z shareugc$$

Furthermore, noting that *pchange* is equivalent to the elasticity divided by *shareugc* we obtain:

$$pchange_{shareugc}^{opex} = \beta_z.$$

We summarise these various derived measures in the following table:

Table 1: **Summary of Derived Formulas**

Measure:	**log–log**	**log–lin**
ugc elasticity	β_z	$\beta_z shareugc$
ugc marginal effect	$\beta_z \frac{opex}{ugc}$	$\beta_z \frac{opex}{circlen}$
shareugc elasticity	β_z	$\beta_z shareugc$
shareugc marginal effect	$\beta_z \frac{opex}{shareugc}$	$\beta_z opex$
pchange	$\frac{\beta_z}{shareugc}$	β_z

The first thing we note is that the *pchange* formulae in Table 1 agree with the example given by NERA. That is, the log–lin model has a constant *pchange* value while the log–log model has a value that varies inversely with *shareugc*. In their example *shareugc* varies from 10% to 50% across the two example DNSPs and hence the value of *pchange* varies by a factor of 5 – from 0.32 to 1.60 (for the log–log model when assuming an estimated elasticity of –0.16).

This appears to be a large difference. <u>However</u>, one must keep in mind that *pchange* and the *ugc* marginal effect are not the same. Hence if we wish to know the actual *opex* savings (in dollars) resulting from converting one kilometre of overground circuit to underground circuit we need to calculate the *ugc* marginal effects.

The wording in the NERA example is not precise, but one could interpret it as implying that since *pchange* is constant across DNSPs A and B, then the marginal effects are also constant. Their example assumes the same *circlen* of 100kms across the two example DNSPs <u>but</u> it is incorrect to assume that they would also have the same *opex* values, because *opex* varies with *shareugc*, as is defined in the log–lin econometric model. For example, in our estimated SFA log–lin model reported below, the coefficient of *shareugc* is –0.35, which implies that *opex* will decrease as *shareugc* increases, all else held constant. Hence the *ugc* marginal effect will not be the same across the two example DNSPs, as one might incorrectly infer from the NERA example. This is because the NERA analysis fails to take account of the fact that two DNSPs that are otherwise identical but which have different shares of underground will also have different opex levels for the reasons outlined at the start of this memo. That is, the DNSP that has the higher share of underground will have a lower level of opex because it does not incur as much vegetation management and inspection costs as the other DNSP.

We now conduct an empirical exercise to investigate this issue.

In our analysis we estimate the Cobb–Douglas (CD) *opex* cost functions using Stochastic Frontier Analysis (SFA) and Least Squares Econometric (LSE) methods as described in Economic Insights (2014). We use these methods to estimate a model containing the *shareugc* variable in log form (log–log) and also another model with it in linear form (log–lin). The data set used is that described in Economic Insights (2018) which includes the most recent data from 2017 and involves a total of 804 observations.

The Stata computer output for four CD models are reproduced in Tables A1 to A4 in Appendix A.[1]

The four estimated CD models are:

1. SFA log–log
2. SFA log–lin
3. LSE log–log
4. LSE log–lin

From Tables A1 to A4 we see that the estimated coefficients of the ln(*shareugc*) or *shareugc* variables are:

1. SFA log–log $\beta_z = -0.15$
2. SFA log–lin $\beta_z = -0.35$
3. LSE log–log $\beta_z = -0.18$
4. LSE log–lin $\beta_z = -0.63$

The Australian DNSP data (within the total sample data set of 804 observations) involves 13 Australian DNSPs observed over 12 years, from 2006 to 2017. We evaluate the *ugc* elasticities, marginal effects and *pchange* for the 13 Australian DNSPs in the most recent year available (2017). These results are presented in Tables A9 and A10 in Appendix A.

The formulae used for the elasticities, marginal effects and *pchange* are outlined in Table 1 above. Note that in calculating the elasticities we define *opex* as predicted *opex* rather than observed *opex* (this is done so that the point of calculation lies on the fitted cost function).[2]

First, we discuss the SFA estimates in Table A9. The table includes values of the base data for each DNSP in 2017 (*opex*, *custnum*, etc.) and has been sorted by *custnum*/*circlen* because

[1] Stata output for the LSETLG log-log and log-lin regressions are presented in Tables A5 and A6, respectively. For convenience, subsequent analysis concentrates on CD results.

[2] Note that the LSE model was converted to a frontier model by using the intercept estimate from the most efficient DNSP (#9) as the global intercept. This provides opex predictions for each DNSP that reflect “efficient opex” instead of “inefficient opex”, which is more consistent with the SFA model results.

this density factor appears to be the main driver of differences in *ugc* marginal effects across the observations for the log–lin model (more on this later). We begin by observing that the log–log elasticity is constant across all observations, as expected. This estimate is –0.15, indicating that a 1 percent increase in *ugc* results in a 0.15 percent reduction in *opex*, all else held constant. The elasticity in the log–lin models varies across the different data points, from a minimum of –0.02 percent (for DNSP #7 when *shareugc* is 0.05) to a maximum of –0.20 percent (for DNSP #1 when *shareugc* is 0.56), with a median value of –0.09 percent (for DNSP #13 when *shareugc* is at its median of 0.25). Note that this median value of –0.09 is less than the estimated elasticity of –0.15 in the log–log model.

The *pchange* measures in the log–lin model are constant across observations as expected, with a value of –0.35. The *pchange* values vary for the log–log model, from a maximum of –3.31 percent (for DNSP #7 when *shareugc* is 0.05) to a minimum of –0.27 percent (for DNSP #1 when *shareugc* is 0.56), with a median value of –0.61 percent (for DNSP #13 when *shareugc* is 0.25).

These *pchange* differences appear to be very large. <u>However</u>, in terms of practical information, the estimated *ugc* marginal effects are much more useful measures. These measures are observed to vary across observations for <u>both</u> the log–log and log–lin models. The marginal effect estimates for the log–log models are mostly larger than those in the log–lin model. For example, the marginal effect estimates for DNSP #13 (which has the median *shareugc* value of 0.25 and hence is marked in yellow) are –4.31 for log–log versus –2.59 for the log–lin. That is, approximately $4,310 per kilometre versus $2,590 per kilometre, respectively.

A few important points need to be made regarding these marginal effects estimates. First, they are clearly <u>not</u> constant across DNSPs in the log–lin model, as may have been implied by the NERA example. Second, the marginal effects in the log–lin model vary from –0.40 to –3.37, or by more than 800%, while those in the log–log model vary from –1.28 to –4.31, or by less than 400%. Thus, the log–lin model produces <u>more</u> variation in estimated marginal effects rather than less. Third, the log–lin marginal effects appear to be increasing as density (the ratio of *custnum*/*circlen*) increases, which is not surprising given that *opex* per unit of *circlen* increases by over 800% as density increases across the sample data.

As noted earlier, the colour yellow is used to mark the DNSP with the median *shareugc* in Table A9. In addition to this we have marked DNSP #3 (*shareugc*=0.50) with green and DNSP #12 (shareugc=0.11) with blue, because these two DNSPs have the most similar *shareugc* ratios to those in the NERA example (0.50 and 0.10). The differences in log–log *pchange* measures are approximately as described in the NERA example, varying from –0.30 to –1.35, a ratio of roughly 5. However, what is of particular interest is that the *ugc* marginal effects for these two DNSPs are not that far apart, being –2.83 and –3.31, respectively. This shows that the log–log model is actually better behaved than the NERA example may superficially indicate.

The estimates of elasticities and marginal effects and *pchange* for the LSE models are reported in Table A10. The LSE estimates follow a similar pattern to the SFA estimates, except that they are generally larger and also the log–log and log–lin estimates are generally closer together. The elasticities at the median *shareugc* are –0.18 and –0.16 for log–log and log–lin, respectively, while the marginal effects at the median are –5.23 and –4.70 for log–log and log–lin, respectively. Again, we observe that these marginal effects are not that far apart.

Overall, we conclude that if the degree of stability of *ugc* marginal effects measures is the metric via which we are to select a model, the log–log model would be preferred on this basis. This is contrary to the conclusion made by NERA on the basis of their superficial example, which, by focusing on their *pchange* measure, has neglected to take into account the degree to which *opex* varies with *shareugc* in the econometric model, and the corresponding effect that this has on the *ugc* marginal effects measures.

2. Statistical criteria

In this exercise we carry out a number of statistical tests which examine which of the two alternative specifications (ie log–log versus log–lin) provides the better fit to the actual DNSP data. We use a 'non–nested' testing procedure to attempt to choose between the log–log and log–lin model options. We cannot use a standard 'nested' testing procedure because we cannot express one model as a restricted version of the other model. That is, we cannot 'nest' one model within the other.

Here we follow the non–nested 'F–test' procedure described in Kennedy (1998, p.89) and Maddala (1989, p.445). However, since we are only interested in one regressor variable, this test can be equivalently and more simply conducted with a t–test.

The procedure is as follows. We first specify the two competing models which have different sets of explanatory variables.

The log–log model can be approximately expressed as:

<u>Model 1:</u> $\ln(opex) = \beta_0 + \sum_{i=1}^{3} \beta_i \ln(y_i) + \beta_{z1} \ln(shareugc) + \beta_t t$

The log–lin model can be approximately expressed as:

<u>Model 2:</u> $\ln(opex) = \beta_0 + \sum_{i=1}^{3} \beta_i \ln(y_i) + \beta_{z2} shareugc + \beta_t t$

We then construct an artificial comprehensive model which has the two competing models embedded in it:

<u>Model C:</u> $\ln(opex) = \beta_0 + \sum_{i=1}^{3} \beta_i \ln(y_i) + \beta_{z1} \ln(shareugc) + \beta_{z2} shareugc + \beta_t t$

We test the null hypothesis of

H1: Model 1 (log–log) versus the alternate hypothesis of

HC: Model C

and also test the null hypothesis of

H2: Model 2 (log–lin) versus the alternate hypothesis of

HC: Model C

The decision process is then as follows:

If both null hypotheses are not rejected or both are rejected then neither model is preferred

If H1 is not rejected and H2 is rejected then Model 1 is preferred

If H1 is rejected and H2 is not rejected then Model 2 is preferred.

Model C has been estimated using both SFA and LSE methods and the results are reported in Tables A7 and A8 in Appendix A.

Given that SFA is estimated using Maximum Likelihood (MLE) methods, finite sample F–tests and t–tests are not applicable, but large–sample tests, such as Likelihood ratio tests (using the Chi–square distribution) and asymptotic t–tests (using the Standard Normal distribution) can be used instead.

For the case of the LSE models, we will also be using the Standard Normal distribution to obtain critical values for the t–tests because when sample size (and hence degrees of freedom) is very large (here the sample size is 804) the t–distribution approximates the Standard Normal distribution. This is why statistical tables rarely report t–distribution critical values for degrees of freedom larger than 100.

Critical values for a 2–tailed test using the Standard Normal distribution are 1.645, 1.960, and 2.326, for the 10 per cent, 5 per cent and 1 per cent significance levels, respectively.

Firstly, let us consider the SFA results from Table A7:[3]

The t–statistic for *ShareUGC* is 1.22 therefore we do not reject H1.

The t–statistic for ln(*ShareUGC)* is –3.89 therefore we do reject H2.

Hence, we conclude that Model 1 (log–log) is preferred using this test.

Sometimes the estimated standard errors and hence t–ratios can be poorly approximated due to the iterative nature of the MLE method used in SFA. Hence, as a check we can repeat this testing procedure using a likelihood ratio test which may be more reliable.

The likelihood ratio (LR) test statistic is calculated as the negative of twice the difference between the log likelihood function (LLF) values under the null and alternative hypotheses, and has a Chi–square distribution with degrees of freedom equal to the number of restrictions being tested (in our case just one restriction per test).

From the computer printouts in the tables in Appendix A we see that the LLF values for the three models are:

Model 1 (log–log): LLF1=554.68

Model 2 (log–lin): LLF2=547.89

Model C (comp): LLFC=555.42

Critical values for the Chi–square distribution with one degree of freedom are 2.71, 3.84 and 6.63 for the 10 per cent, 5 per cent and 1 per cent significance levels, respectively.

Calculating the LR test statistics we obtain:

LR1 = –2(LLF1–LLFC) = 1.45 therefore we do not reject H1.

LR2 = –2(LLF2–LLFC) = 13.58 therefore we do reject H2.

[3] Note that in the computer output, the variable *z* represents *ShareUGC* and *lz* represents ln(*ShareUGC*).

Hence, we conclude that Model 1 (log–log) is again preferred using this test.

We now repeat these non–nested tests for the case of the LSE models. These models do not produce LLF values and hence we will focus our attention on the t–tests. From the computer printout of Model C in Table A8 we obtain:

The t–statistic for *ShareUGC* is –0.92 therefore we do not reject H1.

The t–statistic for ln(*ShareUGC)* is –3.67 therefore we do reject H2

Hence, we again conclude that Model 1 (log–log) is preferred using this test.

In addition to conducting non–nested tests one can also use model selection criteria to attempt to discriminate between non–nested models. Two commonly used criteria are the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC) which are defined as:[4]

AIC = –2(LLF)+2k

and

BIC = –2(LLF)+ln(n)k,

where k is the number of parameters estimated and n is the sample size.

The AIC is an estimator of the relative quality of statistical models for a given set of data. It is founded on information theory. When a statistical model is used to represent the process that generated the data, the representation will almost never be exact – some information will be lost by using the model to represent the process. The AIC estimates the relative amount of information lost by a given model. The less information a model loses, the higher the quality of that model. The BIC is related to the AIC but includes a larger penalty for overfitting the model to the data.

Smaller values of the AIC and BIC are preferred. Since both Model 1 and Model 2 have the same numbers of parameters and use the same sample size we can simply compare the LLF values across the 2 models and prefer the one with the higher LLF value. As noted above, for the SFA models we obtained values of LLF1=554.68 and LLF2=547.89 for Models 1 and 2, respectively, and hence once again Model 1 (log–log) is preferred on this basis.[5]

Conclusion

In summary, the log–log model currently used in the AER’s economic benchmarking work provides a better statistical fit to the data and so is preferred on the basis of a range of model selection tests reported in section 2 above. Furthermore, from the analysis provided in section 1 of this memo, the log–log model also appears to produce more stable measures of *ugc* marginal effects across the sample data. Hence, we conclude that on the basis of our analysis, the log–log model is preferred and should be retained.

[4] Refer to Statacorp (2013) manual.

[5] Note that the LSE models are estimated using the xtpcse command in Stata which does not produce a LLF value and hence the AIC and BIC cannot be calculated for the LSE models.

3. Alternative ways of including undergrounding in the rate of change

Endeavour Energy (2018, p.22) argues that overhead lines and underground cables should be included as separate outputs in the rate of change rather than the AER (2018) proposal to include circuit length as an output and the share of undergrounding as part of the productivity factor. In this section we look at alternative ways in which undergrounding can be included in the rate of change.

The argument advanced by Endeavour Energy (2018, p.22) is as follows:

> 'Currently, circuit line length is one of the factors used to calculate the real output growth trend in the AER's base–step–trend opex model. Arguably, this output growth factor may overstate the impact of circuit line length growth on opex if it is calculated based on a historical proportion of undergrounding that is not maintained in the future. …
>
> 'If there is any deficiency in this measure it should be addressed directly rather than through a productivity factor. The most logical approach would be to split the circuit line length factor between overhead and underground with different elasticities.'

To assess the argument advanced by Endeavour Energy, it is necessary to examine the log–log opex cost function formulas outlined above. We start by rearranging the opex cost function formula as follows:

$$\ln(opex) = \beta_1 + \beta_{TL}\ln(TL) + \beta_Z \ln(shareugc)$$

where $\beta_1 = \beta_0 + \sum_{i=1}^{2} \beta_i \ln(y_i) + \beta_t t$, the y_i are the outputs other than circuit length and TL is total circuit length.

The opex cost function above can be rewritten as:

$$\ln(opex) = \beta_1 + \beta_{TL}\ln(TL) + \beta_Z \ln(ugc/TL)$$

where *ugc* is the circuit length of underground cables. This can be further expressed as follows:

$$\ln(opex) = \beta_1 + \beta_{TL}\ln(TL) + \beta_Z (\ln(ugc) - \ln(TL))$$

since the logarithm of a ratio is equivalent to the difference between the logarithms of the numerator and denominator. Rearranging terms we obtain:

$$\ln(opex) = \beta_1 + (\beta_{TL} - \beta_Z)\ln(TL) + \beta_Z \ln(ugc)$$

This shows that our current opex cost function specification can be rearranged to have 2 lines outputs (total length and underground length) and no OEF variable and it will produce the same result as our current specification with only one line output and the share of underground as an OEF variable. Similarly, the extent of undergrounding can be allowed for in either the output component of the rate of change or in the productivity component and it will produce the same result.

The estimational equivalence of these approaches is demonstrated in tables 2 and 3.

Table 2: **SFACD estimates, one line output and one OEF variable, 2012–2017**

Variable	*Coefficient*	*Standard error*	*t–ratio*
ln(Custnum)	0.708	0.096	7.350
ln(CircLen)	0.168	0.052	3.210
ln(RMDemand)	0.124	0.086	1.440
ln(ShareUGC)	–0.113	0.049	–2.290
Year	0.015	0.003	5.690
Country dummy variables:			
New Zealand	0.068	0.103	0.660
Ontario	0.316	0.094	3.370
Constant	–21.176	5.396	–3.920
Variance parameters:			
Mu	0.447	0.143	3.130
SigmaU squared	0.031	0.007	4.558
SigmaV squared	0.008	0.001	12.926
LLF			298.337

Table 3: **SFACD estimates, two line outputs and no OEF variable, 2012–2017**

Variable	*Coefficient*	*Standard error*	*t–ratio*
ln(Custnum)	0.708	0.096	7.350
ln(CircLen)	0.281	0.044	6.340
ln(UGC)	–0.113	0.049	–2.290
ln(RMDemand)	0.124	0.086	1.440
Year	0.015	0.003	5.690
Country dummy variables:			
New Zealand	0.068	0.103	0.660
Ontario	0.316	0.094	3.370
Constant	–21.027	5.382	–3.910
Variance parameters:			
Mu	0.447	0.143	3.130
SigmaU squared	0.031	0.007	4.558
SigmaV squared	0.008	0.001	12.926
LLF			298.337

We use the SFACD model to illustrate. Table 2 presents the results from Economic Insights (2018) with the specification including three outputs (of which total circuit length is one) and the share of underground as an OEF variable. Table 3 presents the results using four outputs (including total circuit length and underground circuit length) and no OEF variable. It can be seen that the coefficient on total length in table 3 is the coefficient on total length in table 2 minus the coefficient on the share of underground in table 2 while the coefficient on underground length in table 3 is the same as the coefficient on the share of underground in table 2.

Table 4: **DNSP line lengths and underground share growth rates, 2006–2017**

DNSP	*Total Length 2017 kms*	*U/G Length 2017 kms*	*Share UGC 2017 %*	*Share Growth 2006-12 %pa*	*Share Growth 2012-17 %pa*
ACT	5,333	2,972	55.7%	1.67%	1.03%
AGD	41,642	15,752	37.8%	1.55%	1.11%
CIT	4,550	2,274	50.0%	1.98%	0.74%
END	36,993	13,766	37.2%	2.43%	2.86%
ENX	53,757	18,638	34.7%	3.28%	1.70%
ERG	152,491	9,315	6.1%	10.55%	3.91%
ESS	192,103	8,717	4.5%	6.66%	3.24%
JEN	6,345	1,892	29.8%	2.68%	2.20%
PCR	75,121	6,326	8.4%	5.79%	4.99%
SAP	88,971	17,754	20.0%	2.48%	1.17%
AND	44,907	6,575	14.6%	4.47%	3.68%
TND	22,725	2,524	11.1%	2.67%	1.08%
UED	13,342	3,297	24.7%	1.85%	3.70%

In table 4 we present 2017 data on total circuit length, underground circuit length, the share of underground and the average annual growth rates in the share of underground for 2006 to 2012 and for 2012 to 2017 for each DNSP. The shares of underground ranged from a low of 4.5 per cent for Essential Energy to a high of 55.7 per cent for Evoenergy (ACT). For all but two DNSPs (END and UED), the average annual growth rate of the share of underground reduced noticeably for the 2012–2017 period compared to the period 2006–2012. For the period 2012–2017 average annual growth rates of the share of underground ranged from a low of 0.7 per cent for CIT to a high of 5.0 per cent for PCR. And, there is a broadly inverse relationship between the size of the share of underground and its average annual growth rate. That is, the highest growth rates in the share are generally observed for the DNSPs with the lowest shares of underground (ie remote and rural DNSPs), reflecting higher growth rates from smaller initial bases. The range in the size of the shares and the broadly inverse relationship between the size of the share and its annual growth rate point to the likely desirability of including more tailored growth rates of underground shares in the rate of change rather than an industry average growth rate.

In the accompanying spreadsheet ‘Economic Insights AER DNSP Output-Prod RoC Options 7Feb2019.xls’ we present five examples of how the AER could include the share(s) of undergrounding in the ‘rate of change’ component of its assessments of DNSPs’ proposed opex. To recap, the rate of change is the growth rate in real prices (assumed to be zero here for convenience) plus the growth rate of output minus the productivity growth rate. We use the results for 2012–2017 from Economic Insights (2018).

The five cases examined are as follows:

- Case 1: the status quo where the AER takes a weighted average of output component growth rates and assumes a zero productivity growth rate
- Case 2: the AER (2018) draft report preferred option of taking weighted average of output component growth rates and deducting a common one per cent productivity growth rate

(which accounts for the 2006–2016 industry average annual rate of undergrounding and technical change)

- Case 3: same as case 2 but we now allow for DNSP–specific undergrounding growth rates for the period 2012–2017 in the productivity component (ie the productivity component now varies by DNSP)
- Case 4: we account for DNSP–specific undergrounding for the period 2012–2017 in the output component by decomposing the share of undergrounding term – this gives the same answer as case 3 but would allow the AER to have all the DNSP–specific items in the output component while the productivity component is now limited to a common rate of technical change
- Case 5: similar to case 3 but instead of individual DNSP–specific undergrounding growth rates we divide the DNSPs into three groups (high undergrounding with over 40 per cent share, medium undergrounding with between 20 per cent and 40 per cent share and low undergrounding with less than 20 per cent share). We then take a weighted average undergrounding growth rate for the DNSPs in each of the three groups. The reason for considering this option is that it would reduce any potential incentive for DNSPs to game their undergrounding forecasts which could be present with cases 3 and 4. However, there may be other disadvantages with this option and it is presented simply for discussion purposes.

We further look at three different ways of implementing these options. These are labelled in the accompanying spreadsheet as:

- All: This uses the AER's current practice of taking an average across the results for the various opex cost function models (in this case the 4 opex cost function models reported in the 2018 ABR for 2012–2017) and corresponding opex PFP. This is straightforward for calculating the output growth rate. But an issue arises in this analysis in extending it to opex PFP as opex PFP does not separately allow for undergrounding. We have addressed this in the examples in the spreadsheet by assuming the undergrounding coefficient for opex PFP is zero and only applying the same 0.5 per cent rate of technical change to opex PFP. This is a relatively conservative treatment. An alternative approach would be to include a higher rate of technical change for opex PFP in recognition that it also includes the effect of increased undergrounding over time.
- CostFns: This only applies the method to the 4 opex cost function models and does not include opex MPFP.
- SFACD: This only includes the SFACD model in line with earlier practice.

The results using the 'All' method are presented in table 5, noting that the real price component is assumed to be zero for convenience. The results show the following:

- the status quo (Case 1) is relatively generous to the DNSPs (as expected as there is no allowance for positive productivity from any source)
- the AER (2018) draft report preferred option (Case 2) is relatively onerous but this in part reflects that the draft report used data starting in 2006 and going to 2016 and the growth

in undergrounding slowed noticeably for 2012–2017 compared to 2006–2012 for all but 2 of the 13 DNSPs

- it makes no difference whether DNSP–specific undergrounding is included via the productivity component (Case 3) or the output component (Case 4)
- in most cases using the DNSP–specific undergrounding shares for 2012 to 2017 (Cases 3 and 4) produce a less onerous result for DNSPs than the AER (2018) draft report result (noting the different time periods used plays an important role here), and
- as expected, whether the group average undergrounding share (Case 5) is better or worse than the DNSP–specific options for each DNSP depends on where the DNSP's own share lies relative to the group average.

Table 5: **Options for including undergrounding in the rate of change**

DNSP	*Case 1*	*Case 2*	*Case 3*	*Case 4*	*Case 5*
	Status quo	*AER (2018)*	*Prod=DNSP*	*Output incl UG,*	*Prod=Group*
	Prod = 0%	*Prod = 1%*	*SUG+0.5%*	*Prod=0.5%*	*SUG+0.5%*
ACT	1.40%	0.40%	0.79%	0.79%	0.81%
AGD	0.59%	–0.41%	–0.02%	–0.02%	–0.11%
CIT	1.00%	0.00%	0.43%	0.43%	0.41%
END	1.41%	0.41%	0.61%	0.61%	0.70%
ENX	1.02%	0.02%	0.34%	0.34%	0.31%
ERG	0.63%	–0.37%	–0.29%	–0.29%	–0.18%
ESS	1.27%	0.27%	0.42%	0.42%	0.47%
JEN	0.75%	–0.25%	0.02%	0.02%	0.05%
PCR	1.17%	0.17%	0.14%	0.14%	0.37%
SAP	0.45%	–0.55%	–0.17%	–0.17%	–0.35%
AND	1.16%	0.16%	0.27%	0.27%	0.36%
TND	0.44%	–0.56%	–0.17%	–0.17%	–0.36%
UED	0.65%	–0.35%	–0.24%	–0.24%	–0.06%

This analysis has shown that our current treatment of undergrounding already includes a different elasticity for the cost of adding additional underground and is equivalent to including an additional underground output variable and dropping the current OEF variable. There is therefore no need to include an additional underground cables output variable as suggested by Endeavour Energy (2018, p.22). Going forward, there may be some advantage from a practical standpoint in including growth in undergrounding in the output component of the rate of change (ie Case 4 above) so that all of the DNSP–specific components are together and the productivity component can then be the estimate of the rate of frontier shift which would be common across all DNSPs.

Appendix A: Econometric estimates and calculations

Table A1: **SFACD estimates of the log–log model**

```
. xtfrontier lvc ly2-ly4 lz1 yr cd2 cd3, ti cost

Iteration 0:   log likelihood =  501.44431
Iteration 1:   log likelihood =  530.09623
Iteration 2:   log likelihood =  551.31438
Iteration 3:   log likelihood =  552.89204
Iteration 4:   log likelihood =  554.42331
Iteration 5:   log likelihood =   554.6255
Iteration 6:   log likelihood =  554.67902
Iteration 7:   log likelihood =  554.68099
Iteration 8:   log likelihood =    554.681

Time-invariant inefficiency model               Number of obs      =       804
Group variable: DNSP                            Number of groups   =        67

                                                Obs per group: min =        12
                                                               avg =        12
                                                               max =        12

                                                Wald chi2(7)       =   3468.86
Log likelihood  =    554.681                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         lvc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ly2 |   .7089977   .0700974    10.11   0.000     .5716094     .846386
         ly3 |   .1261442   .0411287     3.07   0.002     .0455335     .206755
         ly4 |   .1642474   .0607989     2.70   0.007     .0450838     .283411
         lz1 |  -.1501092   .0330289    -4.54   0.000    -.2148447   -.0853738
          yr |   .0184356   .0012058    15.29   0.000     .0160722     .020799
         cd2 |   .0956584   .0970934     0.99   0.325    -.0946412     .285958
         cd3 |   .2894039   .0879543     3.29   0.001     .1170166    .4617912
       _cons |  -27.79952   2.460194   -11.30   0.000    -32.62141   -22.97763
-------------+----------------------------------------------------------------
         /mu |   .4548949   .1246796     3.65   0.000     .2105273    .6992624
   /lnsigma2 |  -3.193915   .1495179   -21.36   0.000    -3.486965   -2.900865
  /ilgtgamma |   1.000583   .2108838     4.74   0.000     .5872585    1.413908
-------------+----------------------------------------------------------------
      sigma2 |    .041011   .0061319                      .0305936    .0549756
       gamma |   .7311732   .0414511                      .6427359    .8043816
    sigma_u2 |   .0299861    .006111                      .0180088    .0419635
    sigma_v2 |   .0110249   .0005744                      .0098991    .0121506
------------------------------------------------------------------------------

. estat ic

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |    804           .     554.681     11    -1087.362   -1035.776
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

Table A2: **SFACD estimates of the log–lin model**

```
. xtfrontier lvc ly2-ly4 z1 yr cd2 cd3, ti cost

Iteration 0:   log likelihood =  510.88769
Iteration 1:   log likelihood =  515.61601  (backed up)
Iteration 2:   log likelihood =  531.95596
Iteration 3:   log likelihood =  544.77654
Iteration 4:   log likelihood =  547.05083
Iteration 5:   log likelihood =  547.62827
Iteration 6:   log likelihood =   547.8916
Iteration 7:   log likelihood =  547.89184
Iteration 8:   log likelihood =  547.89184

Time-invariant inefficiency model               Number of obs      =       804
Group variable: DNSP                            Number of groups   =        67

                                                Obs per group: min =        12
                                                               avg =        12
                                                               max =        12

                                                Wald chi2(7)       =   4059.60
Log likelihood  =  547.89184                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         lvc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ly2 |   .6952732   .0714445     9.73   0.000     .5552445     .835302
         ly3 |   .1853719    .038332     4.84   0.000     .1102426    .2605012
         ly4 |     .12055   .0608766     1.98   0.048     .0012339     .239866
          z1 |  -.3529766   .1306403    -2.70   0.007    -.6090268   -.0969265
          yr |   .0174565   .0011919    14.65   0.000     .0151204    .0197927
         cd2 |   .0437104    .092184     0.47   0.635    -.1369668    .2243877
         cd3 |   .3086493    .090206     3.42   0.001     .1318488    .4854498
       _cons |  -25.45947   2.388034   -10.66   0.000    -30.13993   -20.77901
-------------+----------------------------------------------------------------
         /mu |   .4002543   .0700395     5.71   0.000     .2629794    .5375292
   /lnsigma2 |  -3.121794   .1658693   -18.82   0.000    -3.446892   -2.796696
  /ilgtgamma |   1.077586   .2287645     4.71   0.000     .6292162    1.525956
-------------+----------------------------------------------------------------
      sigma2 |    .044078   .0073112                      .0318445    .0610113
       gamma |   .7460369    .043343                      .6523117    .8214139
    sigma_u2 |   .0328838   .0072973                      .0185814    .0471862
    sigma_v2 |   .0111942   .0005834                      .0100507    .0123377
------------------------------------------------------------------------------

. estat ic

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |    804           .    547.8918     11    -1073.784   -1022.198
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

Table A3: **LSECD estimates of the log–log model**

```
. xtpcse lvc ly2-ly4 lz1 yr cd2 cd3 d2-d13, c(a) het
(note: estimates of rho outside [-1,1] bounded to be in the range [-1,1])

Prais-Winsten regression, heteroskedastic panels corrected standard errors

Group variable:   DNSP                          Number of obs      =       804
Time variable:    year                          Number of groups   =        67
Panels:           heteroskedastic (unbalanced)  Obs per group: min =        12
Autocorrelation:  common AR(1)                                 avg =        12
                                                               max =        12
Estimated covariances      =        67          R-squared          =    0.9930
Estimated autocorrelations =         1          Wald chi2(19)      =  23841.59
Estimated coefficients     =        20          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |           Het-corrected
         lvc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ly2 |   .6840551   .0575285    11.89   0.000     .5713014    .7968088
         ly3 |    .107238    .026785     4.00   0.000     .0547404    .1597356
         ly4 |   .2068251   .0563772     3.67   0.000     .0963279    .3173223
         lz1 |  -.1818043   .0205419    -8.85   0.000    -.2220657   -.1415429
          yr |   .0193193    .002011     9.61   0.000     .0153779    .0232608
         cd2 |  -.2833338   .1318263    -2.15   0.032    -.5417086    -.024959
         cd3 |  -.1103041   .1310462    -0.84   0.400      -.36715    .1465418
          d2 |   .0300989   .1629404     0.18   0.853    -.2892583    .3494562
          d3 |  -.7211336   .1452657    -4.96   0.000    -1.005849   -.4364181
          d4 |  -.2356173   .1411181    -1.67   0.095    -.5122036     .040969
          d5 |  -.3177679   .1380494    -2.30   0.021    -.5883398    -.047196
          d6 |  -.1997316   .1533549    -1.30   0.193    -.5003016    .1008385
          d7 |  -.3667623   .1634424    -2.24   0.025    -.6871035   -.0464212
          d8 |  -.3749932    .141667    -2.65   0.008    -.6526555   -.0973309
          d9 |  -.8430181   .1510678    -5.58   0.000    -1.139106   -.5469307
         d10 |  -.5648979   .1457805    -3.87   0.000    -.8506223   -.2791734
         d11 |  -.5362143   .1450322    -3.70   0.000    -.8204723   -.2519563
         d12 |  -.5085928   .1526058    -3.33   0.001    -.8076947   -.2094909
         d13 |  -.5953842   .1424828    -4.18   0.000    -.8746453   -.3161231
       _cons |  -28.76936    4.04862    -7.11   0.000    -36.70451   -20.83421
-------------+----------------------------------------------------------------
         rho |    .641407
------------------------------------------------------------------------------
```

Table A4: **LSECD estimates of the log–lin model**

```
. xtpcse lvc ly2-ly4 z1 yr cd2 cd3 d2-d13, c(a) het
(note: estimates of rho outside [-1,1] bounded to be in the range [-1,1])

Prais-Winsten regression, heteroskedastic panels corrected standard errors

Group variable:   DNSP                           Number of obs      =       804
Time variable:    year                           Number of groups   =        67
Panels:           heteroskedastic (unbalanced)   Obs per group: min =        12
Autocorrelation:  common AR(1)                                  avg =        12
                                                                max =        12
Estimated covariances      =        67           R-squared          =    0.9930
Estimated autocorrelations =         1           Wald chi2(19)      =  22451.37
Estimated coefficients     =        20           Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |           Het-corrected
         lvc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ly2 |   .6381017    .058642    10.88   0.000     .5231655    .7530378
         ly3 |   .1614742   .0241645     6.68   0.000     .1141126    .2088358
         ly4 |   .2141661   .0567247     3.78   0.000     .1029877    .3253444
          z1 |  -.6320253   .0800284    -7.90   0.000     -.788878   -.4751725
          yr |   .0187294    .002035     9.20   0.000     .0147409    .0227178
         cd2 |  -.3291822   .1336984    -2.46   0.014    -.5912264   -.0671381
         cd3 |  -.0880971   .1325708    -0.66   0.506    -.3479312     .171737
          d2 |  -.0308929   .1667003    -0.19   0.853    -.3576195    .2958336
          d3 |  -.7021106   .1473433    -4.77   0.000    -.9908983    -.413323
          d4 |  -.3136042   .1443256    -2.17   0.030    -.5964772   -.0307313
          d5 |  -.4016087   .1412659    -2.84   0.004    -.6784847   -.1247327
          d6 |  -.1833789   .1586649    -1.16   0.248    -.4943563    .1275985
          d7 |  -.3215451    .165226    -1.95   0.052    -.6453822    .0022919
          d8 |  -.3965225   .1447853    -2.74   0.006    -.6802964   -.1127486
          d9 |  -.8364227   .1550759    -5.39   0.000    -1.140366   -.5324795
         d10 |  -.6807834   .1486623    -4.58   0.000    -.9721562   -.3894106
         d11 |  -.5815252   .1472787    -3.95   0.000    -.8701861   -.2928644
         d12 |  -.5364022   .1556005    -3.45   0.001    -.8413736   -.2314307
         d13 |  -.6220163   .1462362    -4.25   0.000     -.908634   -.3353987
       _cons |  -27.12844   4.091417    -6.63   0.000    -35.14748   -19.10941
-------------+----------------------------------------------------------------
         rho |   .6490583
------------------------------------------------------------------------------
```

Table A5: **LSETLG estimates of the log–log model**

```
. xtpcse lvc ly2-ly4 ly22 ly23 ly24 ly33 ly34 ly44 lz1 yr cd2 cd3 d2-d13, c(a) h
(note: estimates of rho outside [-1,1] bounded to be in the range [-1,1])

Prais-Winsten regression, heteroskedastic panels corrected standard errors

Group variable:   DNSP                          Number of obs      =       804
Time variable:    Year                          Number of groups   =        67
Panels:           heteroskedastic (unbalanced)  Obs per group: min =        12
Autocorrelation:  common AR(1)                                 avg =        12
                                                               max =        12
Estimated covariances      =        67          R-squared          =    0.9933
Estimated autocorrelations =         1          Wald chi2(25)      =  28779.95
Estimated coefficients     =        26          Prob > chi2        =    0.0000
```

lvc	Coef.	Het-corrected Std. Err.	z	P>|z|	[95% Conf.	Interval]
ly2	.557858	.0663948	8.40	0.000	.4277265	.6879895
ly3	.1102355	.0273148	4.04	0.000	.0566995	.1637716
ly4	.3051732	.0578296	5.28	0.000	.1918293	.4185172
ly22	-.3981699	.275118	-1.45	0.148	-.9373912	.1410514
ly23	.2060105	.0892356	2.31	0.021	.0311119	.380909
ly24	.1799136	.2093524	0.86	0.390	-.2304096	.5902367
ly33	-.011154	.036964	-0.30	0.763	-.0836021	.0612942
ly34	-.1831239	.0708317	-2.59	0.010	-.3219514	-.0442964
ly44	.0717242	.1686531	0.43	0.671	-.2588298	.4022782
lz1	-.1630637	.0247885	-6.58	0.000	-.2116483	-.1144791
yr	.0201084	.0019921	10.09	0.000	.0162039	.024013
cd2	-.3721261	.1291859	-2.88	0.004	-.6253257	-.1189264
cd3	-.2142207	.1280118	-1.67	0.094	-.4651193	.0366779
d2	-.1460403	.1680122	-0.87	0.385	-.4753383	.1832576
d3	-.7568452	.1415943	-5.35	0.000	-1.034365	-.4793254
d4	-.380846	.1407238	-2.71	0.007	-.6566596	-.1050325
d5	-.4689943	.1406457	-3.33	0.001	-.7446548	-.1933339
d6	-.2894695	.1658814	-1.75	0.081	-.6145911	.0356522
d7	-.5066862	.1763148	-2.87	0.004	-.8522569	-.1611156
d8	-.2960386	.1440227	-2.06	0.040	-.578318	-.0137593
d9	-.9271692	.1512062	-6.13	0.000	-1.223528	-.6308104
d10	-.6921147	.1472436	-4.70	0.000	-.9807069	-.4035225
d11	-.5724861	.1459178	-3.92	0.000	-.8584799	-.2864924
d12	-.5224307	.1492991	-3.50	0.000	-.8150515	-.2298099
d13	-.5563708	.1480471	-3.76	0.000	-.8465377	-.2662039
_cons	-30.32032	4.012043	-7.56	0.000	-38.18378	-22.45686
rho	.6292089					

Table A6: **LSETLG estimates of the log–lin model**

```
. xtpcse lvc ly2-ly4 ly22 ly23 ly24 ly33 ly34 ly44 z1 yr cd2 cd3 d2-d13, c(a) he
(note: estimates of rho outside [-1,1] bounded to be in the range [-1,1])

Prais-Winsten regression, heteroskedastic panels corrected standard errors

Group variable:   DNSP                          Number of obs      =       804
Time variable:    Year                          Number of groups   =        67
Panels:           heteroskedastic (unbalanced)  Obs per group: min =        12
Autocorrelation:  common AR(1)                                 avg =        12
                                                               max =        12
Estimated covariances      =        67          R-squared          =    0.9933
Estimated autocorrelations =         1          Wald chi2(25)      =  27463.75
Estimated coefficients     =        26          Prob > chi2        =    0.0000
```

lvc	Coef.	Het-corrected Std. Err.	z	P>|z|	[95% Conf.	Interval]
ly2	.5358283	.0672177	7.97	0.000	.404084	.6675727
ly3	.1508159	.0242387	6.22	0.000	.1033089	.1983229
ly4	.3014527	.0580567	5.19	0.000	.1876637	.4152416
ly22	-.3563588	.2852361	-1.25	0.212	-.9154112	.2026936
ly23	.1225232	.0889457	1.38	0.168	-.0518072	.2968536
ly24	.2106057	.2154967	0.98	0.328	-.21176	.6329714
ly33	.0531311	.0332523	1.60	0.110	-.0120422	.1183044
ly34	-.163892	.0719324	-2.28	0.023	-.3048768	-.0229071
ly44	.0346005	.1717232	0.20	0.840	-.3019708	.3711717
z1	-.5567483	.0879771	-6.33	0.000	-.7291804	-.3843163
yr	.0194091	.0019951	9.73	0.000	.0154989	.0233194
cd2	-.4281695	.1311643	-3.26	0.001	-.6852469	-.1710922
cd3	-.2001995	.1300115	-1.54	0.124	-.4550172	.0546183
d2	-.2112029	.1715949	-1.23	0.218	-.5475228	.1251169
d3	-.7798158	.1445561	-5.39	0.000	-1.06314	-.4964911
d4	-.4496818	.1434502	-3.13	0.002	-.730839	-.1685246
d5	-.5409187	.1432179	-3.78	0.000	-.8216207	-.2602167
d6	-.3523943	.170786	-2.06	0.039	-.6871288	-.0176599
d7	-.5296636	.1795892	-2.95	0.003	-.8816519	-.1776753
d8	-.3309359	.1481394	-2.23	0.025	-.6212838	-.0405879
d9	-.9253702	.1553314	-5.96	0.000	-1.229814	-.6209262
d10	-.8092906	.1488392	-5.44	0.000	-1.10101	-.5175712
d11	-.5976036	.1483443	-4.03	0.000	-.8883531	-.3068541
d12	-.5483721	.1527163	-3.59	0.000	-.8476906	-.2490536
d13	-.5935412	.1527958	-3.88	0.000	-.8930156	-.2940669
_cons	-28.51854	4.008501	-7.11	0.000	-36.37506	-20.66202
rho	.638118					

Table A7: **SFACD estimates of artificial Model C**

```
. xtfrontier lvc ly2-ly4 lz1 z1 yr cd2 cd3, ti cost

Iteration 0:   log likelihood =  525.71673
Iteration 1:   log likelihood =  526.23668
Iteration 2:   log likelihood =  546.21499
Iteration 3:   log likelihood =  548.82973
Iteration 4:   log likelihood =  554.82869
Iteration 5:   log likelihood =  555.38153
Iteration 6:   log likelihood =  555.41895
Iteration 7:   log likelihood =  555.42048
Iteration 8:   log likelihood =  555.42049

Time-invariant inefficiency model               Number of obs      =       804
Group variable: DNSP                            Number of groups   =        67

                                                Obs per group: min =        12
                                                               avg =        12
                                                               max =        12

                                                Wald chi2(8)       =   3451.49
Log likelihood  =  555.42049                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         lvc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ly2 |   .7026695    .070312     9.99   0.000     .5648605    .8404784
         ly3 |   .1212615   .0414161     2.93   0.003     .0400874    .2024356
         ly4 |   .1672816   .0608195     2.75   0.006     .0480775    .2864857
         lz1 |  -.1966823   .0505654    -3.89   0.000    -.2957886    -.097576
          z1 |   .2436451   .2001082     1.22   0.223    -.1485598    .6358499
          yr |   .0183115   .0012067    15.17   0.000     .0159464    .0206765
         cd2 |   .0861332   .0971751     0.89   0.375    -.1043266    .2765929
         cd3 |   .2596579   .0912898     2.84   0.004     .0807331    .4385827
       _cons |  -27.67065   2.453898   -11.28   0.000    -32.48021    -22.8611
-------------+----------------------------------------------------------------
         /mu |   .4536005   .1192861     3.80   0.000     .2198041    .6873968
   /lnsigma2 |  -3.185529   .1515031   -21.03   0.000     -3.48247   -2.888589
  /ilgtgamma |   1.015997   .2129462     4.77   0.000     .5986296    1.433363
-------------+----------------------------------------------------------------
      sigma2 |   .0413563   .0062656                      .0307314    .0556547
       gamma |    .734192   .0415573                      .6453427    .8074248
    sigma_u2 |   .0303635   .0062473                       .018119     .042608
    sigma_v2 |   .0109928   .0005728                      .0098701    .0121156
------------------------------------------------------------------------------
```

Table A8: **LSECD estimates of artificial Model C**

```
. xtpcse lvc ly2-ly4 lz1 z1 yr cd2 cd3 d2-d13, c(a) het
(note: estimates of rho outside [-1,1] bounded to be in the range [-1,1])

Prais-Winsten regression, heteroskedastic panels corrected standard errors

Group variable:   DNSP                          Number of obs      =       804
Time variable:    year                          Number of groups   =        67
Panels:           heteroskedastic (unbalanced)  Obs per group: min =        12
Autocorrelation:  common AR(1)                                 avg =        12
                                                               max =        12
Estimated covariances      =        67          R-squared          =    0.9930
Estimated autocorrelations =         1          Wald chi2(20)      =  23959.64
Estimated coefficients     =        21          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |           Het-corrected
         lvc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ly2 |   .6764598   .0579865    11.67   0.000     .5628083    .7901112
         ly3 |   .1129993   .0271616     4.16   0.000     .0597636     .166235
         ly4 |   .2130718   .0566405     3.76   0.000     .1020585    .3240851
         lz1 |   -.148492   .0404554    -3.67   0.000    -.2277831   -.0692009
          z1 |  -.1452175   .1574064    -0.92   0.356    -.4537283    .1632934
          yr |   .0192963   .0020071     9.61   0.000     .0153625    .0232301
         cd2 |  -.2928826   .1315632    -2.23   0.026    -.5507417   -.0350234
         cd3 |   -.111381     .130533    -0.85   0.394     -.367221    .1444589
          d2 |   .0095712   .1638997     0.06   0.953    -.3116663    .3308088
          d3 |  -.7241128   .1448141    -5.00   0.000    -1.007943   -.4402823
          d4 |  -.2581239     .142705    -1.81   0.070    -.5378206    .0215729
          d5 |  -.3414141   .1399028    -2.44   0.015    -.6156186   -.0672096
          d6 |  -.2055431   .1531595    -1.34   0.180    -.5057302    .0946441
          d7 |  -.3655651   .1625669    -2.25   0.025    -.6841904   -.0469398
          d8 |  -.3886171   .1420394    -2.74   0.006    -.6670093    -.110225
          d9 |  -.8518545   .1509496    -5.64   0.000     -1.14771   -.5559987
         d10 |  -.5932578   .1482425    -4.00   0.000    -.8838078   -.3027078
         d11 |  -.5535433   .1455237    -3.80   0.000    -.8387646   -.2683221
         d12 |  -.5233446   .1528493    -3.42   0.001    -.8229237   -.2237655
         d13 |  -.6125535   .1433618    -4.27   0.000    -.8935375   -.3315695
       _cons |  -28.62589   4.040933    -7.08   0.000    -36.54597   -20.70581
-------------+----------------------------------------------------------------
         rho |   .6401296
------------------------------------------------------------------------------
```

Table A9: SFACD model results

log-log

DNSP	opex	custnum	circlen	rmdemand	shareugc	ugc	cust/circ	opex per unit of: custnum	opex per unit of: circlen	opex per unit of: rmdem	elasticity	marg eff	pchange
7	217318	891935	192103	2968	0.05	8717	4.64	0.24	1.13	73.22	-0.15	-3.74	-3.31
6	180327	745501	152491	3238	0.06	9315	4.89	0.24	1.18	55.69	-0.15	-2.90	-2.46
10	158068	878300	88971	3193	0.20	17754	9.87	0.18	1.78	49.51	-0.15	-1.34	-0.75
9	162877	816349	75121	2571	0.08	6326	10.87	0.20	2.17	63.34	-0.15	-3.86	-1.78
12	55714	287652	22725	1154	0.11	2524	12.66	0.19	2.45	48.28	-0.15	-3.31	-1.35
11	124589	734644	44907	1952	0.15	6575	16.36	0.17	2.77	63.83	-0.15	-2.84	-1.02
4	146949	984230	36993	4344	0.37	13766	26.61	0.15	3.97	33.83	-0.15	-1.60	-0.40
5	211517	1448247	53757	5298	0.35	18638	26.94	0.15	3.93	39.93	-0.15	-1.70	-0.43
1	25284	191482	5333	724	0.56	2972	35.91	0.13	4.74	34.92	-0.15	-1.28	-0.27
2	235217	1706914	41642	6555	0.38	15752	40.99	0.14	5.65	35.88	-0.15	-2.24	-0.40
13	94684	676807	13342	2143	0.25	3297	50.73	0.14	7.10	44.19	-0.15	-4.31	-0.61
8	45045	334840	6345	1020	0.30	1892	52.77	0.13	7.10	44.18	-0.15	-3.57	-0.50
3	42892	339400	4550	1478	0.50	2274	74.60	0.13	9.43	29.02	-0.15	-2.83	-0.30

log-lin

DNSP	opex	custnum	circlen	rmdem	shareugc	ugc	cust/circ	opex per unit of: custnum	opex per unit of: circlen	opex per unit of: rmdem	elasticity	marg eff	pchange
7	217414	891935	192103	2968	0.05	8717	4.64	0.24	1.13	73.26	-0.02	-0.40	-0.35
6	184798	745501	152491	3238	0.06	9315	4.89	0.25	1.21	57.07	-0.02	-0.42	-0.35
10	178180	878300	88971	3193	0.20	17754	9.87	0.20	2.00	55.81	-0.07	-0.70	-0.35
9	167992	816349	75121	2571	0.08	6326	10.87	0.21	2.24	65.33	-0.03	-0.78	-0.35
12	58105	287652	22725	1154	0.11	2524	12.66	0.20	2.56	50.35	-0.04	-0.89	-0.35
11	134294	734644	44907	1952	0.15	6575	16.36	0.18	2.99	68.81	-0.05	-1.05	-0.35
4	160050	984230	36993	4344	0.37	13766	26.61	0.16	4.33	36.84	-0.13	-1.51	-0.35
5	231880	1448247	53757	5298	0.35	18638	26.94	0.16	4.31	43.77	-0.12	-1.51	-0.35
1	27029	191482	5333	724	0.56	2972	35.91	0.14	5.07	37.33	-0.20	-1.77	-0.35
2	251558	1706914	41642	6555	0.38	15752	40.99	0.15	6.04	38.37	-0.13	-2.11	-0.35
13	98865	676807	13342	2143	0.25	3297	50.73	0.15	7.41	46.14	-0.09	-2.59	-0.35
8	47424	334840	6345	1020	0.30	1892	52.77	0.14	7.47	46.51	-0.10	-2.62	-0.35
3	43834	339400	4550	1478	0.50	2274	74.60	0.13	9.63	29.66	-0.17	-3.37	-0.35

Table A10: LSECD model results

log-log

								opex per unit of					
DNSP	**opex**	**custnum**	**circlen**	**rmdemand**	**shareugc**	**ugc**	**cust/circ**	**custnum**	**circlen**	**rmdem**	**elasticity**	**marg eff**	**pchange**
7	222165	891935	192103	2968	0.05	8717	4.64	0.25	1.16	74.86	-0.18	-4.59	-3.97
6	184926	745501	152491	3238	0.06	9315	4.89	0.25	1.21	57.11	-0.18	-3.57	-2.95
10	156991	878300	88971	3193	0.20	17754	9.87	0.18	1.76	49.17	-0.18	-1.59	-0.90
9	165630	816349	75121	2571	0.08	6326	10.87	0.20	2.20	64.41	-0.18	-4.71	-2.14
12	56956	287652	22725	1154	0.11	2524	12.66	0.20	2.51	49.36	-0.18	-4.06	-1.62
11	124570	734644	44907	1952	0.15	6575	16.36	0.17	2.77	63.82	-0.18	-3.41	-1.23
4	146989	984230	36993	4344	0.37	13766	26.61	0.15	3.97	33.84	-0.18	-1.92	-0.48
5	210307	1448247	53757	5298	0.35	18638	26.94	0.15	3.91	39.70	-0.18	-2.03	-0.52
1	24998	191482	5333	724	0.56	2972	35.91	0.13	4.69	34.53	-0.18	-1.51	-0.32
2	235523	1706914	41642	6555	0.38	15752	40.99	0.14	5.66	35.93	-0.18	-2.69	-0.48
13	95849	676807	13342	2143	0.25	3297	50.73	0.14	7.18	44.73	-0.18	-5.23	-0.73
8	45327	334840	6345	1020	0.30	1892	52.77	0.14	7.14	44.45	-0.18	-4.31	-0.60
3	43394	339400	4550	1478	0.50	2274	74.60	0.13	9.54	29.36	-0.18	-3.43	-0.36

log-lin

								opex per unit of					
DNSP	**opex**	**custnum**	**circlen**	**rmdem**	**shareugc**	**ugc**	**cust/circ**	**custnum**	**circlen**	**rmdem**	**elasticity**	**marg eff**	**pchange**
7	220438	891935	192103	2968	0.05	8717	4.64	0.25	1.15	74.27	-0.03	-0.72	-0.63
6	191067	745501	152491	3238	0.06	9315	4.89	0.26	1.25	59.01	-0.04	-0.79	-0.63
10	177630	878300	88971	3193	0.20	17754	9.87	0.20	2.00	55.63	-0.13	-1.26	-0.63
9	170990	816349	75121	2571	0.08	6326	10.87	0.21	2.28	66.50	-0.05	-1.43	-0.63
12	59442	287652	22725	1154	0.11	2524	12.66	0.21	2.62	51.51	-0.07	-1.65	-0.63
11	133337	734644	44907	1952	0.15	6575	16.36	0.18	2.97	68.32	-0.09	-1.87	-0.63
4	158780	984230	36993	4344	0.37	13766	26.61	0.16	4.29	36.55	-0.23	-2.70	-0.63
5	228815	1448247	53757	5298	0.35	18638	26.94	0.16	4.26	43.19	-0.22	-2.68	-0.63
1	24764	191482	5333	724	0.56	2972	35.91	0.13	4.64	34.21	-0.35	-2.93	-0.63
2	250187	1706914	41642	6555	0.38	15752	40.99	0.15	6.01	38.17	-0.24	-3.79	-0.63
13	99577	676807	13342	2143	0.25	3297	50.73	0.15	7.46	46.47	-0.16	-4.70	-0.63
8	46550	334840	6345	1020	0.30	1892	52.77	0.14	7.34	45.65	-0.19	-4.62	-0.63
3	42412	339400	4550	1478	0.50	2274	74.60	0.12	9.32	28.70	-0.31	-5.87	-0.63

REFERENCES

Australian Energy Regulator (AER) (2018), *Forecasting Productivity Growth for Electricity Distributors*, Draft Decision Paper, Melbourne, November.

Economic Insights (2014), *Economic Benchmarking Assessment of Operating Expenditure for NSW and ACT Electricity DNSPs*, Report prepared by Denis Lawrence, Tim Coelli and John Kain for the Australian Energy Regulator, Eden, 17 November.

Economic Insights (2018), *Economic Benchmarking Results for the Australian Energy Regulator's 2018 DNSP Benchmarking Report*, Report prepared by Denis Lawrence, Tim Coelli and John Kain for the Australian Energy Regulator, Eden, 9 November.

Endeavour Energy (2018), *RE: AER Draft Decision Paper – Forecasting productivity growth for electricity distributors (November 2018)*, Submission to the AER, Sydney, 21 December.

Kennedy, P. (1998), *A Guide to Econometrics*, 4th Ed., Blackwell.

Maddala, G.S. (1988), *Introduction to Econometrics*, MacMillan.

NERA Economic Consulting (NERA) (2018), *Assessment of the AER's Proposed Productivity Assumptions*, Report prepared for CitiPower, Powercor, SA Power Networks and United Energy, Sydney, 20 December.

StataCorp (2013), *Stata Statistical Software: Release 13*, College Station, Texas.