frontier economics

Benchmarking analysis of Energex's and Ergon Energy's opex

A report prepared for Energy Queensland Limited | 18 January 2024

Contents

Figures

1 Executive summary

1.1 Background

1. The next regulatory control period for Energex and Ergon Energy is due to commence on 1 July 2025 and conclude on 30 June 2030 (the 2025-30 regulatory period). Energex and Ergon Energy are currently preparing their initial proposals to the Australian Energy Regulator (AER) for the 2025-30 regulatory period.

2. The AER uses benchmarking analysis to assess the efficiency of a Distribution Network Service Provider's (DNSP's) revealed base year operating expenditure (opex). If the AER's estimate of efficient base year opex (derived using benchmarking analysis) is higher than the DNSP's revealed base year opex, the AER accepts the DNSP's revealed opex as an appropriate starting point from which to forecast an efficient opex requirement over the regulatory control period. However, if the DNSP's revealed opex exceeds the AER's estimate of efficient base year opex, the AER deems the revealed opex of the DNSP to be materially inefficient. In these circumstances, the AER may apply an efficiency adjustment to the DNSP's revealed base year opex, and then use the adjusted base year opex as the starting point for forecasting an efficient opex requirement over the regulatory control period.

3. Energy Queensland Limited (EQL) has asked Frontier Economics to:

   a Derive an estimate of efficient base year opex for Energex and for Ergon Energy:

   b Consider how the AER should account for statistical uncertainty when deriving an estimate of efficient base year opex; and

   c Review the reliability of the AER's econometric benchmarking models and explain the implications for the assessment of the efficiency of Energex's and Ergon Energy's revealed base year opex.

1.2 Key findings

1.2.1 Accounting for statistical uncertainty around the estimate of efficient base year opex

4. The AER's methodology for determining an estimate of efficient base year opex relies on its econometric benchmarking models. The parameters in these models are estimated from data and are subject to statistical uncertainty. Consequently, the AER's base year opex target is also subject to statistical uncertainty.

5. However, the AER does not account for this statistical uncertainty when assessing the efficiency of a DNSP's actual base year opex because the AER effectively treats its point estimate of efficient base year opex as certain. The AER does not allow for any range of uncertainty around its point estimate of efficient base year opex. Rather, if its point estimate of efficient base year opex is lower than the DNSP's actual base year opex, then the AER concludes that the latter is materially inefficient.

6. In our view, this is a serious shortcoming in the AER's approach. The statistical uncertainty involved in estimating a DNSP's efficiency and the elasticities and other parameters specified in the econometric benchmarking models can be very material. This uncertainty means that the AER does not *know* the *true* level of efficient base year opex for a particular DNSP with certainty. Instead, the true level of efficient base year opex lies within a range of uncertainty that is defined by (amongst other factors) the statistical error involved in estimating:

   a the true level of average efficiency of a DNSP over the historical benchmarking period;

   b the true relationship between a DNSP's opex and outputs; and

   c the true values of other parameters specified in the AER's econometric models.

7. In our view, the AER should quantify formally the statistical uncertainty around its point estimate of efficient base year opex, by constructing confidence intervals around that estimate, and then use those confidence intervals to make a probabilistic assessment about the evidence for material inefficiency.

8. The Stata output files that accompany the Annual Benchmarking Reports provide information on the statistical uncertainty associated with:

   a AER's estimates of the period average efficiency; and

   b other relevant parameter estimates used to roll forward the period average estimate of efficient opex to the base year.

9. Using this information, and a well-accepted statistical technique known as the 'delta method', we have constructed confidence intervals around the estimates of efficient base year opex derived using each of the AER's econometric benchmarking models.

10. The AER's approach is to compare a DNSP's *actual* base year opex to an *estimate* of efficient base year opex, where that estimate is derived using statistical analysis. If the former is greater than the latter, then the AER concludes that the DNSP's actual base year opex is materially inefficient.

11. However, if the DNSP's actual base year opex lies within the confidence interval, then the AER cannot reject the possibility that there is no difference between a DNSP's revealed level of actual base year opex and the efficient level of base year opex—because the latter can only be estimated with statistical uncertainty, and the former lies within the range of statistical uncertainty.

12. In these circumstances, one could not conclude that a DNSP's revealed base year opex is efficient. But one could conclude that there is no evidence of material inefficiency. This approach to using confidence intervals is entirely consistent with standard hypothesis testing.

1.2.2 Shortcomings associated with the econometric benchmarking models

13. The AER uses four econometric opex cost function models to estimate the average efficiency of DNSPs' historical opex. The four models reflect two different specifications of the cost function (Cobb-Douglas and Translog) and two different estimation methods (Least Squares Econometrics (LSE) and Stochastic Frontier Analysis (SFA)).

14. We have identified several serious problems with these benchmarking models—the details of which are presented in Appendix A. In summary, we found that:

   a Statistical test results presented by Quantonomics indicate that the Cobb-Douglas model is seriously misspecified and that the Translog model, which allows for more flexibility in the specification of the output elasticities, fits the data significantly better than the Cobb-

Douglas model. In view of this, it is difficult to find a statistical justification for including estimates derived from the Cobb-Douglas models in the assessment of the efficiency of the DNSPs.

b The Translog models (particularly those estimated using the short sample) have exhibited monotonicity violations for a number of DNSPs and in a number of years. As the AER itself acknowledges, these monotonicity violations are becoming more prevalent over time.[1]

c These monotonicity violations are likely to be a symptom of a more fundamental model misspecification problem. Quantonomics' approach of restricting the flexibility of the Translog functional form to reduce the number of monotonicity violations simply treats the symptom rather than the root cause of the problem.

d Our analysis indicates that the Translog models are also misspecified. For example, plots of the residuals of the Translog models for the Australian DNSPs make it abundantly clear that the residuals of the models for the Australian DNSPs are not random with respect to time, and that there is a time-related factor that is not accounted for properly in the AER's models.

e There is convincing evidence that the DNSP industry as a whole in Australia has become more efficient over time—an observation that the AER itself has made.[2] However, all of the AER's benchmarking models assume that efficiency remains constant over time. This assumption of constant efficiencies over time is hard-wired into the specification of the models. Consequently, the AER's models are incapable, due to their specification, of accounting for the fact that some DNSPs have improved their level of efficiency considerably over time. Since the models cannot account for these changes in efficiency over time directly, they will tend to overfit the data to other time-varying variables in the model.

   i Given the highly flexible functional form of the Translog models, this response to the lack of time variation in the efficiencies in the model is likely to be more pronounced for the Translog models. This could be a key reason why the Translog models are prone to monotonicity violations—particularly when estimated using the short sample, which overlaps almost perfectly with the period over which the AER has been conducting benchmarking analysis.

   ii Given the significant changes in DNSP efficiencies since 2014, the assumption of constant efficiencies is likely to cause a serious misspecification problem for the Cobb-Douglas models as well as the Translog.[3] However, the consequences of this misspecification problem are harder to detect (e.g., as monotonicity violations) for the Cobb-Douglas models due to their more restrictive functional form. Nonetheless, both classes of models suffer from the same underlying issue.

f In short, there is compelling evidence that the Cobb-Douglas and Translog models are misspecified and therefore should not be relied upon by the AER. Misspecification of the benchmarking models will result in biased estimates of efficiency for individual DNSPs (and other model parameters). This means that the resulting estimates of efficient base

[1] AER, *Annual Benchmarking Report, Electricity distribution network service providers*, November 2022, p. 58.

[2] For example: AER, *Annual Benchmarking Report, Electricity distribution network service providers*, November 2023, p. v.

[3] As shown in Appendix A, the residual plots for the Cobb-Douglas models also exhibit a clear negative trend over time for the Australian DNSPs.

year opex derived using those models will be unreliable, and unsafe for the purposes of setting opex allowances.

15. Given the seriousness of the statistical problems we have identified, what is required is a fundamental review of the AER's econometric benchmarking models to ensure that they are capable of fitting the salient features of data well. Such a review should be done carefully and in proper consultation with stakeholders. Therefore, it should not be rushed.

16. Until this work can be completed properly, the AER should exercise extreme caution when interpreting the results derived from its existing models. The AER should not use those models mechanistically, as it has done in recent determinations, when assessing whether a DNSP's actual base year opex is materially inefficient.

1.2.3 Estimate of efficient base year opex

17. EQL has instructed us to assume, for the purposes of modelling Energex's and Ergon Energy's efficient base year opex, that the relevant base year for both DNSPs will be 2023-24.

18. In addition, EQL has asked us to derive estimates of efficient base year opex for Energex and Ergon Energy:

    a Using 2023-24 as the base year for forecasting Energex's and Ergon Energy's opex requirement for the 2025-30 regulatory period;

    b Employing the AER's existing benchmarking method, as applied in the most recent determinations (including for the NSW and ACT DNSPs);

    c Historical data used by the AER in the 2023 Annual Benchmarking Report (the latest dataset available at the time of preparation of this report);

    d Historical opex including all capitalised corporate overheads and defined using Energex's and Ergon Energy's 2022 Cost Allocation Methodology (consistent with AER's new approach to accounting for differences in capitalisation practices);

    e The latest data on backcast capitalised corporate overheads submitted by Ergon Energy and Energex to the AER; and

    f Using the latest Operating Environment Factor (OEF) adjustments employed by the AER and relevant to Energex and Ergon Energy.

19. We follow the AER's method for estimating an overall efficient level of base year opex, which involves:

    a Estimating an efficient level of opex over the relevant historical benchmarking period, using each statistical model that is not rejected due to monotonicity violations (i.e., the 'valid models');

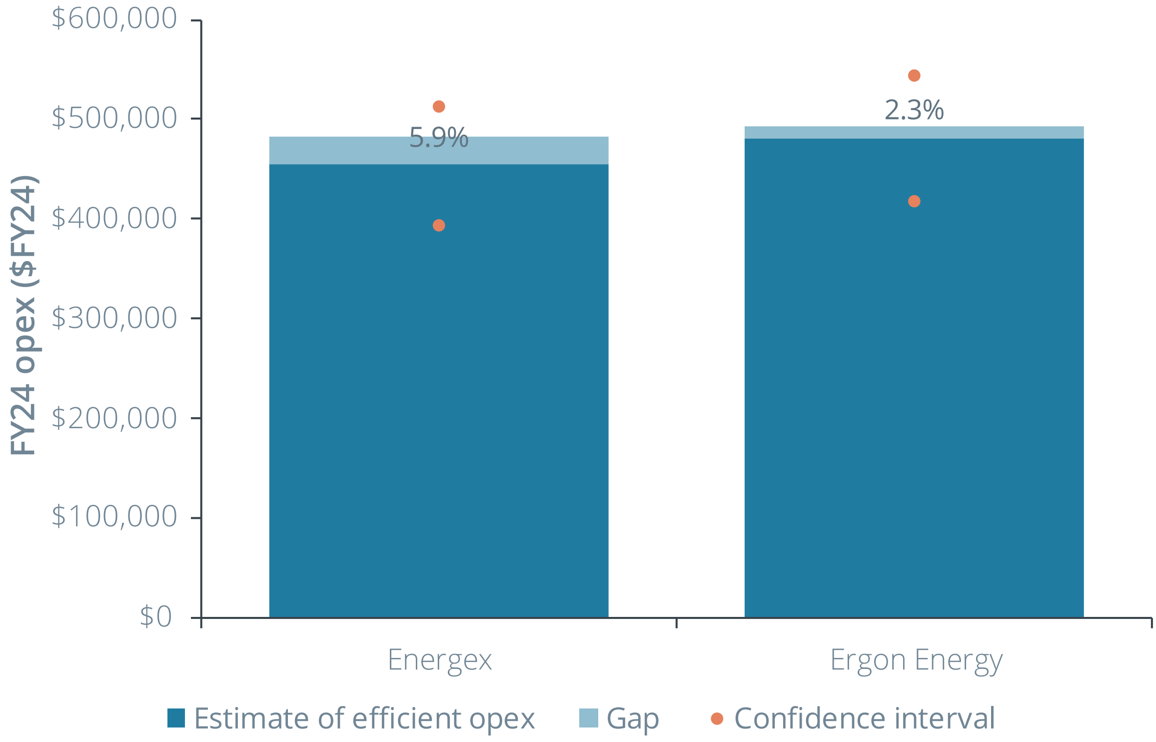
    b Rolling forward each of those estimates to the base year (using the AER's roll-forward procedure); and

    c Averaging the estimates across all of the valid models.

20. We have also constructed confidence intervals around the estimates of efficient base year opex for Energex and for Ergon Energy, as described in section 1.2.1 (with further detail in Appendix B).

21. The resulting estimates are presented in Figure 1 below.

Figure 1: Estimates of efficient base year opex ($FY2024)



Source: Frontier Economics analysis of EQL and Quantonomics data.

22. We make the following observations in relation to the results presented in Figure 1:

   a The current forecast of revealed opex for Energex and for Ergon Energy in 2023-24 is higher than the estimate of efficient opex obtained from the AER's benchmarking method:

      i The implied efficiency adjustment for Energex is 5.9%; and

      ii The implied efficiency adjustment for Ergon Energy is 2.3%.

   b The AER would typically conclude, on this basis, that the revealed base year opex for both DNSPs is materially inefficient and therefore should be adjusted downwards. The adjusted estimate of efficient opex for:

      i Energex would be $443.7 million ($FY2024) excluding capitalised corporate overheads (i.e., in current CAM terms); and

      ii Ergon Energy would be $461.9 million ($FY2024) excluding capitalised corporate overheads (i.e., in current CAM terms).

   c However, the estimate of efficient opex for both DNSPs lie comfortably within the 95% confidence interval around each estimate. Consequently, we conclude that there is no evidence that the forecast revealed base year opex for Energex or Ergon Energy is materially inefficient.

1.3 Structure of this report

23. The remainder of this report is structured as follows:

a. Section 2 provides a brief overview of the AER's approach to benchmarking DNSP opex;

b. Section 3 applies the AER's benchmarking method to assess the efficiency of Energex's and Ergon Energy's base year opex;

c. Appendix A discusses the statistical shortcomings associated with the AER's econometric benchmarking models, and the implications for the AER's assessment of the efficiency of Energex's and Ergon Energy's revealed base year opex for the 2024-29 regulatory control period; and

d. Appendix B explains how we take into account the statistical uncertainty associated with the estimates of efficient base year opex.

2 The AER's benchmarking approach

2.1 Overview of the AER's approach

24. The AER uses a 'base-step-trend' approach to derive a forecast of efficient opex over each regulatory control period. Under this approach, the AER uses an efficient level of opex in a nominated base year as the starting point for the forecast.

25. The AER has explained that its preferred approach is to use the DNSP's revealed opex in the base year as the starting point for its opex forecast over the regulatory control period, because opex tends to be recurrent and stable at a total opex level. However, the AER does not assume that the DNSP's revealed opex is efficient. Instead, the AER assesses whether there is evidence that the DNSP's base year revealed opex is materially inefficient.[4] The primary tool the AER uses to make such an assessment is economic benchmarking analysis.

26. Of the various benchmarking techniques used by the AER, the main technique the AER relies on to assess the efficiency of a DNSP's base year opex is the estimation of four different types of econometric benchmarking models:

    a Cobb-Douglas Stochastic Frontier Analysis (SFA-CD);

    b Cobb-Douglas Least Squares Econometrics (LSE-CD);

    c Translog Stochastic Frontier Analysis (SFA-TLG); and

    d Translog Least Squares Econometrics (LSE-TLG).

27. These four models are estimated using data over two historical time periods:

    a The long sample (using all the data available from 2006 onwards); and

    b The short sample (using all the data available from 2012 onwards).

28. Currently, the AER excludes any models for which a mathematical condition known as 'monotonicity' is violated for more than half the observations in the sample for the DNSP in question.[5] All surviving models are treated as 'valid' models.

29. Using each model, the AER estimates (for each historical time period), a 'period average' efficiency score (ranging from 0 to 1) for each DNSP. Under the AER's current approach, if the estimated efficiency score 0.75 (the 'benchmark comparison point') or higher, the DNSP is deemed to be 'reference' DNSP, and that DNSP's revealed base year opex is deemed to be an efficient point from which to forecast opex over the forthcoming regulatory control period.

[4] See, for example, AER, *Draft decision – Evoenergy electricity distribution determination 2024 – 29, Attachment 6*, September 2023, p. 13.

[5] Monotonicity is a property related to a DNSP's cost function that says that an increase in output can only be achieved with an increase in inputs (opex), all else remaining equal. If the estimated coefficient for a particular output is negative, that would imply that an increase in output could be achieved by the DNSP with a reduction in opex (all else remaining equal), which is not economically meaningful. Quantonomics has recently proposed a more stringent threshold for a monotonicity violation than is currently applied by the AER. In particular, Quantonomics has proposed that in order make a finding of a 'significant monotonicity violation', the estimated output coefficient must be both negative and statistically significant.

30. If the estimated period average efficiency score is less than 0.75, the AER calculates the gap between the DNSP's estimated efficiency score and the benchmark comparison point adjusted for the OEFs relevant to that particular DNSP. This estimated efficiency gap is then used to adjust the DNSP's average real revealed opex over the relevant benchmarking period to derive a 'period average' level of efficient opex.

31. The AER then rolls forward the period average estimate of efficient opex to the base year, accounting for the growth in the DNSP's outputs, technical progress and changes in business conditions.

32. In order to arrive at an overall estimate of efficient base year opex, the AER averages the estimates of efficient opex derived using all of the valid benchmarking models.

33. If the AER's overall estimate of efficient opex is higher than the DNSP's revealed base year opex, then the AER will accept the DNSP's revealed opex in the base year as an appropriate starting point from which to forecast an efficient level of opex over the regulatory control period.

34. If, on the other hand, the DNSP's revealed base year opex is higher than the AER's overall estimate of efficient opex, the AER will conclude that the DNSP's revealed opex is materially inefficient. In these circumstances, the AER may make an efficiency adjustment to the DNSP's revealed base year opex before using it as a starting point for forecasting efficient opex over the regulatory control period.

2.2 Serious problems with the econometric models

35. We have identified several serious problems with the econometric benchmarking models the AER uses to assess the efficiency of a DNSP's efficient opex. Appendix A explains in detail what these problems are, and the supporting evidence. In summary, we found that:

    a Statistical test results presented by Quantonomics indicate that the Cobb-Douglas model is seriously misspecified and that the Translog model, which allows for more flexibility in the specification of the output elasticities, fits the data significantly better than the Cobb-Douglas model. In view of this, it is difficult to find a statistical justification for including estimates derived from the Cobb-Douglas models in the assessment of the efficiency of the DNSPs.

    b The Translog models (particularly those estimated using the short sample) have exhibited monotonicity violations for a number of DNSPs and in a number of years. As the AER itself acknowledges, these monotonicity violations are becoming more prevalent over time.[6]

    c These monotonicity violations are likely to be a symptom of a more fundamental model misspecification problem. Quantonomics' approach of restricting the flexibility of the Translog functional form to reduce the number of monotonicity violations simply treats the symptom rather than the root cause of the problem.

    d Our analysis indicates that the Translog models are also misspecified. For example, plots of the residuals of the Translog models for the Australian DNSPs make it abundantly clear that the residuals of the models for the Australian DNSPs are not random with respect to time, and that there is a time-related factor that is not accounted for properly in the AER's models.

[6] AER, *Annual Benchmarking Report, Electricity distribution network service providers*, November 2022, p. 58.

e There is convincing evidence that the DNSP industry as a whole in Australia has become more efficient over time—an observation that the AER itself has made.[7] However, all of the AER's benchmarking models assume that efficiency remains constant over time. This assumption of constant efficiencies over time is hard-wired into the specification of the models. Consequently, the AER's models are incapable, due to their specification, of accounting for the fact that some DNSPs have improved their level of efficiency considerably over time. Since the models cannot account for these changes in efficiency over time directly, they will tend to overfit the data to other time-varying variables in the model.

i Given the highly flexible functional form of the Translog models, this response to the lack of time variation in the efficiencies in the model is likely to be more pronounced for the Translog models. This could be a key reason why the Translog models are prone to monotonicity violations—particularly when estimated using the short sample, which overlaps almost perfectly with the period over which the AER has been conducting benchmarking analysis.

ii Given the significant changes in DNSP efficiencies since 2014, the assumption of constant efficiencies is likely to cause a serious misspecification problem for the Cobb-Douglas models as well as the Translog.[8] However, the consequences of this misspecification problem are harder to detect (e.g., as monotonicity violations) for the Cobb-Douglas models due to their more restrictive functional form. Nonetheless, both classes of models suffer from the same underlying issue.

f In short, there is compelling evidence that the Cobb-Douglas and Translog models are misspecified and therefore should not be relied upon by the AER. Misspecification of the benchmarking models will result in biased estimates of efficiency for individual DNSPs (and other model parameters). This means that the resulting estimates of efficient base year opex derived using those models will be unreliable, and unsafe for the purposes of setting opex allowances.

36. Given the seriousness of the statistical problems we have identified, what is required is a fundamental review of the AER's econometric benchmarking models to ensure that they are capable of fitting the salient features of data well. Such a review should be done carefully and in proper consultation with stakeholders. Therefore, it should not be rushed.

37. Until this work can be completed properly, the AER should exercise extreme caution when interpreting the results derived from its existing models. The AER should not use those models mechanistically, as it has done in recent determinations, when assessing whether a DNSP's actual base year opex is materially inefficient.

2.3 Recent analysis by Quantonomics

38. A recent report by Quantonomics presents new modelling and analysis, some of which is related to issues we raise in this report (and in a recent report we prepared for ActewAGL) on the shortcomings of the AER's econometric benchmarking models.[9]

[7] For example: AER, *Annual Benchmarking Report, Electricity distribution network service providers*, November 2023, p. v.

[8] As shown in Appendix A, the residual plots for the Cobb-Douglas models also exhibit a clear negative trend over time for the Australian DNSPs.

[9] Quantonomics, *Opex Cost Function–Options to Address Performance Issues of Translog Models*, 25 October 2023.

39. There are four main issues raised in that recent Quantonomics report:

a Quantonomics proposes that the AER consider a lower standard for monotonicity violations;

b Quantonomics proposes that the AER consider the use of a new specification of the Cobb-Douglas models (for the short and long samples) and the Translog models (for the long sample) that introduces a time trend variable for Australian DNSPs to account for time-related factors that are not captured well by the existing models;

c Quantonomics presents a new restricted specification of the Translog models that reduces significantly the number of monotonicity violations associated with those models; and

d Quantonomics responds to the concerns we expressed in recent advice we provided to Ausgrid that the short sample SFA-TLG model used in the Draft 2023 Annual Benchmarking Report had been mis-estimated.

40. We address each of these issues in turn below.

Redefinition of the standard for a monotonicity violation

41. At present, the AER considers an observation to suffer from a monotonicity violation if one or more of the estimated elasticities of real opex with respect to each output variable is negative.

42. Quantonomics proposes a lower standard, whereby the AER would only concern itself with "significant" monotonicity violations. According to Quantonomics' recommendation, a significant monotonicity violation would be defined as a situation where an estimated elasticity of real opex with respect to any output is both negative and significantly different from zero.[10]

43. Our view is that, to the extent possible, the AER should address the root causes of the monotonicity violations rather than seek to define away the problem by lowering the standard for what is considered to be a monotonicity violation.

44. Quantonomics explains that monotonicity violations can arise when estimating Translog models due to "complex issues such as multicollinearity between the output variables, the effects of influential observations on the nonlinear shape of the function, and inadequate sample size."[11] As we argue in this report, monotonicity violations can also arise if the cost functions used to estimate efficiency are misspecified—for instance, by failing to account for the fact that some Australian DNSPs have achieved significant catch-up efficiency over time.

45. As the AER and Quantonomics have both acknowledged, monotonicity violations have become more prevalent over time. This has occurred despite the sample size growing over time as more data have become available with each passing year.

46. For the reasons explained in the previous section and in Appendix A, we think the increased prevalence of monotonicity violations is likely to be driven by the inability of the models to account for the catch-up efficiency exhibited by some Australian DNSPs over time. Our view is that a better way to address the issue of monotonicity violations would be to investigate alternative model specifications that are capable of accounting for time-varying efficiency.

[10] Quantonomics, *Opex Cost Function–Options to Address Performance Issues of Translog Models*, 25 October 2023, p. 2.

[11] Quantonomics, *Opex Cost Function–Options to Address Performance Issues of Translog Models*, 25 October 2023, p. 25.

47. We also note that models that suffer from a multicollinearity problem will produce elasticity estimates with large standard errors. This is because the individual effects of highly correlated explanatory variables on the dependent variable cannot be disentangled easily, and this lack of precision in individual elasticity estimates will be reflected in large standard errors. Large standard errors make it harder to conclude that a particular elasticity estimate is significantly different from zero. This means that the presence of multicollinearity (which Quantonomics acknowledges the AER's models suffer from) reduces the likelihood of finding significant monotonicity violations. In fact, the stronger the multicollinearity problem, the greater the likelihood and severity of actual monotonicity violations occurring, but the large standard errors associated with strong multicollinearity will make it harder to conclude those violations are significant.

Introduction of a time trend for Australian DNSPs

48. Quantonomics proposes that the AER consider the use of a new specification of the Cobb-Douglas models (for the short and long samples) and the Translog models (for the long sample) that introduces a time trend variable for Australian DNSPs to account for time-related factors that are not captured well by the existing models. Quantonomics does not recommend the inclusion of an additional Australian DNSP time trend variable in the short sample Translog models because the estimated coefficient on the ratcheted maximum demand variable is negative for Australian DNSPs on average.[12]

49. Quantonomics' analysis supports our own conclusions that there is a strong time-related factor for the Australian DNSPs that is not accounted for properly in the AER's models. As we explain in this report, we think this time-related factor is the significant catch-up efficiency that some Australian DNSPs have achieved since 2014.

50. Whilst we think it is positive that Quantonomics has undertaken some initial work to investigate this issue, we think it would be premature to adopt the additional time trend specification proposed by Quantonomics. We suggest that a range of alternative specifications be investigated, including models that allow for time varying efficiency over the period when catch-up efficiency has occurred.

51. Before Quantonomics' proposed alternative specification can be adopted, the AER would need to resolve how the estimated time trend term is to be interpreted and used in its opex roll-forward calculation. As Quantonomics explains, the time trend in the models is intended to capture the effects of technical change on opex over time. However, the estimated time trend invariably captures other factors that vary over time that are not accounted for explicitly in the model (e.g., changes in technical performance standards, regulatory frameworks and obligations, and/or environmental conditions).

52. If the time-related factor missing from the models is the catch-up efficiency achieved by some Australian DNSPs, and if this is reflected in the additional time trend estimate, then it would be inappropriate to include that as part of the calculation that rolls forward the estimate of efficient opex to the base year. This is because the estimated time trend would reflect the average catch-up efficiency achieved by Australian DNSPs. It would be inappropriate to impose this average rate of catch up on all Australian DNSPs. Furthermore, the scope for catch-up efficiency declines over time as the DNSP in question converges to the efficient frontier. Imposing a constant rate of catch-up on DNSPs when rolling forward their efficient opex to the base year would fail to recognise that the opportunities for such improvement decline and are eventually exhausted as the DNSP becomes more efficient.

[12] Quantonomics, *Opex Cost Function–Options to Address Performance Issues of Translog Models*, 25 October 2023, pp. 23-24.

Restricted Translog models

53. Quantonomics also proposes the consideration of a new restricted specification of the Translog models, whereby the coefficient on the higher order and interaction terms in relation to the customer number output variable is restricted to zero. This makes the models less flexible in their ability to fit the data, and results in significantly fewer monotonicity violations.

54. Our view is that a key cause of the monotonicity violations is misspecification of the AER's benchmarking models—in particular, the inability of the models to account for the fact that a number of Australian DNSPs have become significantly more efficient over time. We suggest that the AER address the root cause of the problem directly (by improving the specification of its models to account for changing efficiency over time), rather than simply treating the symptoms of the problem by limiting the flexibility of the Translog models.

Mis-estimation of the short sample SFA-TLG model

55. Our analysis of the short sample SFA-TLG model showed that there are parameter values for the model that produce a higher value of the likelihood function than the likelihood function value associated with Quantonomics' estimates for this model. This means that there are other SFA-TLG models that fit the data better than Quantonomics' model. It also means that Quantonomics' estimates are not maximum likelihood estimates, as is claimed, and hence their statistical properties are unknown. In particular, it has not been established that Quantonomics' estimates will tend to the true parameter values for the model as the sample size increases.

56. In its comments on our work,[13] Quantonomics chose to focus on the fact that the example of alternative estimates we provided had not converged. However, this mis-represents the point we were making, namely, that there are parameter values for which the likelihood value is higher than the likelihood value associated with Quantonomics' estimates. Despite the fact that the algorithm had not converged, the last parameter values reported by the algorithm correspond to a higher value of the likelihood function than the likelihood function value for Quantonomics' estimates. This established the key point we were making, namely, that the Quantonomics estimates do not correspond to the maximum value of the likelihood function, and hence are not maximum likelihood estimates.

57. New analysis that we present in Appendix A demonstrates that there are other examples of parameter values where the algorithm does converge and the corresponding likelihood value is higher than the likelihood value for Quantonomics' estimates.

58. Further investigation of the likelihood function shows that for this model the likelihood function has an unconventional shape, which we believe is the cause of the mis-estimation. The unusual likelihood function is likely due to the fact that the SFA-TLG model is a poor description of the true statistical model that underlies the data used to estimate the model. In other words, we believe that the model suffers from serious misspecification issues. While the case of the short sample SFA-TLG model provides a stark example of the statistical issues that can arise as a result of model misspecification, we believe that the same misspecification issues apply to *all* of the econometric benchmarking models used by the AER. This is because we consider that the main source of misspecification is the models' inability to recognise that some of the Australian DNSPs' have achieved significant catch-up efficiency since 2014. This is a shortcoming of all of the benchmarking models used by the AER; it is not confined just to the short sample SFA-TLG model.

[13] Quantonomics, *Opex Cost Function–Options to Address Performance Issues of Translog Models*, 25 October 2023, p. 39.

59. While the AER's approach of excluding models that exhibit monotonicity violations would, in this case, result in the problematic short sample SFA-TLG model being dropped, it does not deal with the more fundamental issue of correcting the misspecifications inherent in all of the AER's models. Hence, the current models continue to be misspecified and, therefore, are not fit-for-purpose.

3 Benchmarking outcomes for Energex and Ergon Energy

60. We set out below benchmarking outcomes for Energex and Ergon Energy by comparing the current forecast of base year opex to estimates of efficient base year opex for each DNSP.

3.1 Estimates of efficient base year opex

61. EQL has instructed us to assume, for the purposes of modelling Energex's and Ergon Energy's efficient base year opex, that the relevant base year for both DNSPs will be 2023-24.

62. In addition, EQL has asked us to derive estimates of efficient base year opex for Energex and Ergon Energy:

a Using 2023-24 as the base year for forecasting Energex's and Ergon Energy's opex requirement for the 2025-30 regulatory period;

b Employing the AER's existing benchmarking method, as applied in the most recent determinations (including for the NSW and ACT DNSPs);

c Historical data used by the AER in the 2023 Annual Benchmarking Report (the latest dataset available at the time of preparation of this report);

d Historical opex including all capitalised corporate overheads and defined using Energex's and Ergon Energy's 2022 Cost Allocation Methodology (consistent with AER's new approach to accounting for differences in capitalisation practices);

e The latest data on backcast capitalised corporate overheads submitted by Ergon Energy and Energex to the AER; and

f Using the latest OEF adjustments employed by the AER and relevant to Energex and Ergon Energy. The OEF adjustments we have applied to each DNSP, for the purposes of estimating an efficient level of base year opex using the AER's benchmarking method, are summarised in Table 1 below, and are consistent with the OEF adjustments adopted by the AER in the September 2023 Draft Decisions for the NSW and ACT DNSPs.

Table 1: OEF adjustments applied to Energex and Ergon Energy

OEF adjustment	Energex		Ergon Energy	
	2006-2022	2012-2022	2006-2022	2012-2022
Cyclones	n/a	n/a	5.04%	4.71%
Sub-transmission	1.53%	1.06%	5.34%	5.02%
Taxes and levies	1.53%	1.22%	0.31%	0.28%
Termite exposure	0.37%	0.32%	1.05%	0.98%
Worker's compensation	-0.22%	-0.22%	-0.22%	-0.22%
Division of responsibility	2.12%	1.98%	2.17%	1.97%
Bushfire risk	-3.65%	-5.17%	-3.65%	-5.17%
Network accessibility	n/a	n/a	1.42%	1.42%
Total	1.69%	-0.81%	11.46%	8.99%

Source: Frontier Economics analysis using AER data and OEF model.

63. We follow the AER's method for estimating an overall efficient level of base year opex, which involves:

    a Estimating an efficient level of opex over the relevant historical benchmarking period, using each statistical model that is not rejected due to monotonicity violations (i.e., the 'valid models');

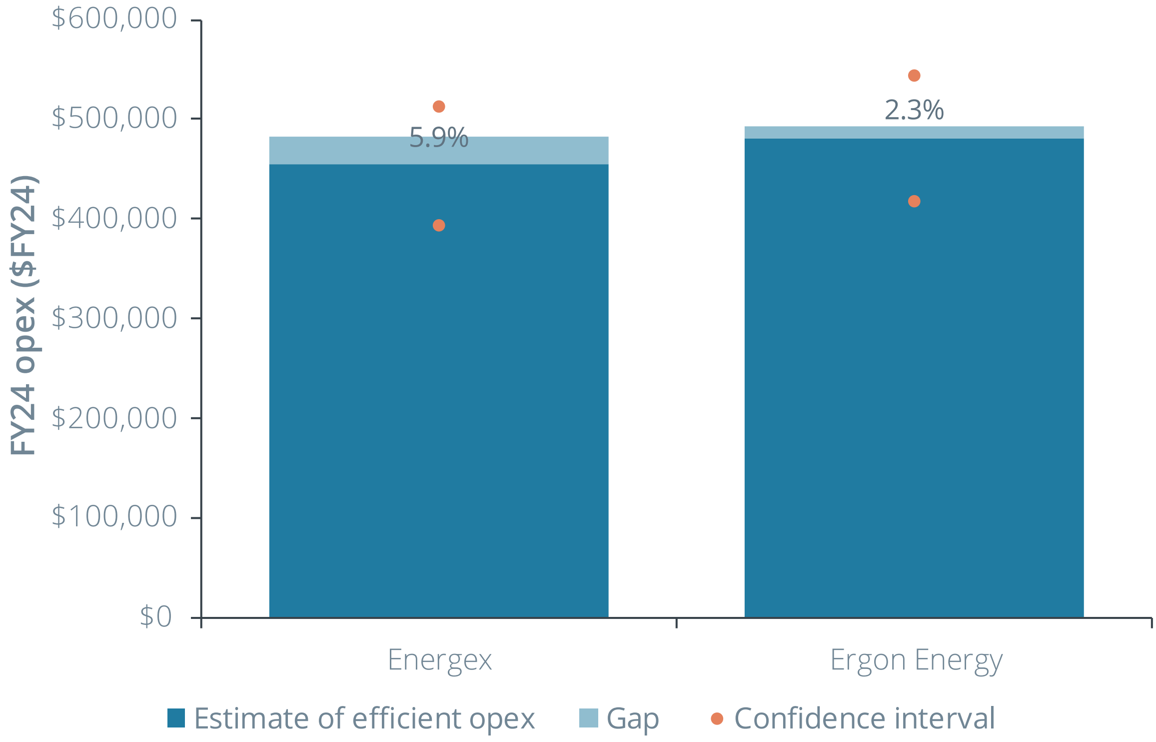
    b Rolling forward each of those estimates to the base year (using the AER's roll-forward procedure); and

    c Averaging the estimates across all of the valid models.

64. We also construct 95% confidence intervals around the point estimate of efficient base year opex for each DNSP using the method explained in Appendix B. We define the lower bound of the confidence interval for the short sample by averaging the lower bounds of the 95% confidence intervals for the valid models. We adopt a similar approach for the long sample models. We then average the short and long samples. We repeat this by taking upper bounds of the 95% confidence intervals to obtain the upper bound.

65. The resulting estimates are presented in Figure 2 below.

Figure 2: Estimates of efficient base year opex ($FY2024)



Source: Frontier Economics analysis of EQL and Quantonomics data.

66. We make the following observations in relation to the results presented in Figure 2:

   a The current forecast of revealed opex for Energex and for Ergon Energy in 2023-24 is higher than the estimate of efficient opex obtained from the AER's benchmarking method:

      i The implied efficiency adjustment for Energex is 5.9%; and

      ii The implied efficiency adjustment for Ergon Energy is 2.3%.

   b The AER would typically conclude, on this basis, that the revealed base year opex for both DNSPs is materially inefficient and therefore should be adjusted downwards. The adjusted estimate of efficient opex for:

      i Energex would be $443.7 million ($FY2024) excluding capitalised corporate overheads (i.e. in current CAM terms); and

      ii Ergon Energy would be $461.9 ($FY2024) excluding capitalised corporate overheads (i.e. in current CAM terms).

   c However, the estimate of efficient opex for both DNSPs lies comfortably within the 95% confidence interval around each estimate. Consequently, we conclude that there is no evidence that the forecast revealed base year opex for Energex or Ergon Energy is materially inefficient.

A Problems associated with the econometric benchmarking models

67. The AER uses four econometric opex cost function models to estimate the average efficiency of DNSPs' historical opex. The four models reflect two different specifications of the cost function (Cobb-Douglas and Translog) and two different estimation methods (Least Squares Econometrics (LSE) and Stochastic Frontier Analysis (SFA)), resulting in the following four models:

    a Cobb-Douglas Stochastic Frontier Analysis (SFA-CD);

    b Cobb-Douglas Least Squares Econometrics (LSE-CD);

    c Translog Stochastic Frontier Analysis (SFA-TLG); and

    d Translog Least Squares Econometrics (LSE-TLG).

68. These four models are estimated using data over two historical time periods:

    a The long sample (using all the data available from 2006 onwards); and

    b The short sample (using all the data available from 2012 onwards).

69. This Appendix discusses a number of statistical problems associated with the econometric benchmarking models relied upon by the AER to conduct its benchmarking analysis.

Monotonicity violations

70. For several years now, the AER has expressed concerns that some of the estimated opex cost functions fail to satisfy a mathematical property known as 'monotonicity.' As the AER has explained, monotonicity implies that an increase in output can only be achieved with an increase in inputs, holding other things constant. Monotonicity violations occur if the model predicts that an increase in any particular output leads to a decrease in opex. Such an outcome is inconsistent with economic theory.

71. The Cobb-Douglas models estimated by the AER do not exhibit monotonicity violations. Hence, the AER has typically assumed that these models are statistically sound and, therefore, has not expressed any concerns about its reliance on those models on statistical grounds.

72. However, the Translog models have exhibited monotonicity violations for a number of DNSPs and in a number of years. These violations tend to occur more often in the Translog models estimated using the short sample. The AER has attributed these violations to the more flexible functional form of the Translog models.[14]

73. These monotonicity violations are not becoming less prevalent over time as more data becomes available. To the contrary, the AER acknowledged that:

[14] AER, *Annual Benchmarking Report, Electricity distribution network service providers*, November 2023, p. 80.

> *this issue has generally become more prevalent since 2018.*[15]

74. In the 2023 Annual Benchmarking Report, the AER noted that the number of monotonicity violations had increased since the 2022 Annual Benchmarking Report:

> *For the current report, the number of instances where this property does not hold in the Translog models is prevalent again and has increased since last year. This year, for the 2006 to 2022 period, we observe monotonicity violations in the Translog LSE model for three DNSPs and in the Translog SFA model for a separate group of three DNSPs. In the 2022 Annual Benchmarking Report, we observed no monotonicity violations for all of the Australian DNSPs in both Translog models over the long period.*[16]

75. The 2023 Annual Benchmarking Report notes that for the short sample, and when using the AER's 'Option 5' definition of opex (to control for differences in capitalisation practices):

    a the SFA-TLG model exhibited monotonicity violations for 10 out of 13 DNSPs; and

    b the LSE-TLG model exhibited monotonicity violations for seven out of 13 DNSPs.

76. The AER deals with this problem by excluding from its process for deriving an estimate of efficient base year opex for a particular DNSP any models for which monotonicity is violated for more than half the observations in the sample for that particular DNSP. If, according to this criterion, a model is excluded for more than half the DNSPs, the model is excluded for all DNSPs.[17]

77. In 2022, the AER asked its adviser Quantonomics to investigate ways to overcome the problem of monotonicity violations. Quantonomics explored three models that were a 'hybrid' of the more restrictive Cobb-Douglas and the more flexible Translog functional forms. Quantonomics reasoned that if the cause of the monotonicity violations is the flexible nature of the Translog models, then making the Translog models less flexible (e.g., by excluding some of the second-order terms in the Translog models) might ameliorate the problem.

78. The AER concluded from Quantonomics' work that, whilst the hybrid models showed some promise (in terms of reducing the instances of monotonicity violations), they also suffered from statistical limitations, which meant that those models could not be adopted at the present time.[18]

79. The 2023 Annual Benchmarking Report notes that several DNSPs—including Evoenergy, Ausgrid, Jemena, Ergon Energy and Energex—have raised concerns about the issue of monotonicity violations.

[15] AER, *Annual Benchmarking Report, Electricity distribution network service providers*, November 2022, p. 58.

[16] AER, *Annual Benchmarking Report, Electricity distribution network service providers*, November 2023, p. 48.

[17] AER, *Draft decision – Evoenergy electricity distribution determination 2024 – 29, Attachment 6*, September 2023, p. 19.

[18] AER, *Annual Benchmarking Report, Electricity distribution network service providers*, November 2022, p. 58.

80. In our view, the monotonicity violations are likely to be a symptom of a more fundamental problem with the AER's econometric models. Quantonomics' approach of restricting the flexibility of the Translog functional form to reduce the number of monotonicity violations is an attempt to treat the symptom rather than the root cause of the problem.

81. As we explain in the remainder of this Appendix, there is mounting evidence that all the AER's econometric benchmarking models are misspecified and, therefore, are incapable of fitting the data well. That is likely to be the root cause of the monotonicity violation problem.

82. This has several important implications:

a The AER's solution of excluding the models that exhibit monotonicity violations is not a proper solution because it simply removes the cases where the symptoms associated with the underlying problem have manifested. That approach does not address the fundamental misspecification problem, which also affects those models that do not exhibit monotonicity violations.

b We also note that when the AER excludes a Translog model for some DNSPs but not for others, the calculation of efficient opex for the different DNSPs is no longer done on a like-with-like basis.

c Because the Quantonomics approach of seeking to make the Translog models less flexible does not address the root cause of the problem, it too is not a proper solution. Therefore, we see little value in the AER pursuing that approach in future.

d What is required is a fundamental review of the AER's econometric benchmarking models to ensure that they are capable of fitting the salient features of data well. We show below that the models do not capture one important feature, namely the time trends in the data. There may be other variables that are omitted. Misspecification of the benchmarking models is likely to result in biased estimates of the DNSPs' efficiencies, making them unreliable for the purposes of setting regulatory allowances.

e Such a review should be done carefully and in proper consultation with stakeholders. Therefore, it should not be rushed. Until this work can be completed properly, the AER should exercise extreme caution when interpreting the results derived from its existing models. The AER should not use those models mechanistically (as it has done in recent determinations) when assessing whether a DNSP's actual base year opex is materially inefficient.

Misspecification of the Cobb-Douglas models

83. Quantonomics undertakes statistical tests of the Cobb-Douglas specifications versus the Translog model specifications. The Cobb-Douglas specification is a special case of the Translog specification with a less flexible functional form. The null hypothesis for this test is that the restrictions imposed on the Translog model to obtain the Cobb-Douglas are consistent with the data. For the LSE models, Quantonomics conducts the Wald test to test this hypothesis, whereas, for the SFA models, Quantonomics conducts both the Wald test and the likelihood ratio test.

84. Quantonomics presents the results of the Wald tests for the Standard approach to opex in Appendices C.1.4 and C.2.4 of the 2023 report,[19] and notes that the Cobb-Douglas simplification of

[19] Quantonomics, *Economic Benchmarking Results for the Australian Energy Regulator's 2023 DNSP Annual Benchmarking Report*, November 2023.

the Translog model is soundly rejected in all cases. The likelihood ratio test for the SFA models (provided in the model output though not reported in the report), also soundly rejects the Cobb-Douglas specifications of the SFA models.

85. Quantonomics presents the results of the Wald tests for the Option 5 approach to opex in Appendices C.3.4 and C.4.4 of the 2023 report,[20] and notes that the Cobb-Douglas simplification of the Translog model is soundly rejected in all cases. The likelihood ratio tests for the SFA models, (provided in the model output though not reported in the report), also soundly rejects the Cobb-Douglas specifications of the SFA models. Evoenergy made a similar submission to the AER in its regulatory proposal.[21]

86. We summarise the results of all these statistical tests in Table 2.

Table 2: Adequacy of the Cobb-Douglas model vs the Translog model – probability values

	Standard approach opex Long sample	Standard approach opex Short sample	Option 5 opex Long sample	Option 5 opex Short sample
LSE CD vs TLG (Wald test)	0.0000	0.0000	0.0000	0.0000
SFA CD v TLG (Wald test)	0.0008	0.0011	0.0004	0.0003
SFA CD v TLG (Likelihood ratio test)	0.0047	0.0001	0.0005	0.0000

Source: Frontier Economics analysis of results in Quantonomics' supporting files for 2023 Annual Benchmarking Report dataset

Note: The probability value (p-value) is the probability that the estimated parameters in the Translog model are consistent with a Cobb-Douglas cost function. The null hypothesis that the data is consistent with the Cobb-Douglas simplification of the Translog specification is rejected if the p-value is smaller than the chosen significance level, which is usually taken to be 0.05. The p-values in this table are far smaller than 0.05.

87. The table shows that the hypothesis that the data is consistent with the Cobb-Douglas simplification of the Translog opex cost function is rejected soundly in all cases since the probability values are far smaller than the usual significance level of 0.05. This indicates that the Cobb-Douglas model is seriously misspecified and that the Translog model, which allows for more flexibility in the specification of the output elasticities, fits the data significantly better than the Cobb-Douglas model. In view of this, it is difficult to find a statistical justification for including estimates derived from the Cobb-Douglas models in the assessment of the efficiency of the DNSPs. However, Quantonomics always includes the results of the Cobb-Douglas models in its assessment of DNSPs' efficiencies despite the models being seriously misspecified from a statistical point of view.

[20] Quantonomics, *Economic Benchmarking Results for the Australian Energy Regulator's 2023 DNSP Annual Benchmarking Report*, November 2023.

[21] Evoenergy, *Regulatory proposal for the ACT electricity distribution network 2024–29, Appendix 2.1: Operating expenditure – base year efficiency*, p. 11.

88. In a recent report prepared for the AER, Quantonomics disagrees that the Translog models should be preferred over the Cobb-Douglas models on the basis of the Wald test because:

> *There are other criteria of model selection to be considered, including goodness-of-fit. Because goodness-of-fit measures penalise loss of degrees of freedom (ie, reward parsimony) the higher order terms can be jointly significant while at the same time, the fit is not improved. This has been shown to be the case in relation to the TLG and CD models*[22]

89. We note that contrary to the AER's assertion, statistical tests do take into account the loss in degrees of freedom when using more flexible models by requiring that the more flexible model fit the data not just better than the simpler model, but significantly better. Table 2 shows that the Translog models fit the data significantly better than the Cobb-Douglas models.

90. We also note that the AER has not presented any goodness-of-fit results to support the statement that the Cobb-Douglas model has a better goodness-of-fit than the Translog when parsimony is taken into account.

91. Table 3 presents the commonly used R-squared and adjusted R-squared measures for the LSE Cobb-Douglas and Translog models estimated by Quantonomics.[23] The R-squared measure does not penalise extra terms in the model, and a more flexible model will always have a higher R-squared value than the simpler version of the model. The adjusted R-squared modifies the R-squared by penalising an increase in the number of explanatory variables included in the model and hence rewards parsimony.

92. Table 3 shows that, as expected, the R-squared value of the Translog model is always larger than the Cobb-Douglas model for the same dataset. However, the adjusted R-squared values for the Translog models are also larger than for the Cobb-Douglas models. This implies that, even after allowing for a decrease in parsimony, the Translog models have the superior goodness-of-fit.

Table 3**:** Goodness-of-fit measures for the LSE models

	R-squared	R^2 difference	Adjusted R^2	Adj. R^2 difference
LSE-CD Long	99.168%		99.15%	
LSE-TLG Long	99.204%	0.04%	99.19%	0.03%
LSE-CD Short	99.505%		99.49%	
LSE-TLG Short	99.532%	0.03%	99.51%	0.02%

Source: Frontier Economics analysis

[22] Quantonomics, *Benchmarking limitations*, September 2023, p. 5.

[23] The Prais-Winsten regression using the xtpcse command in Stata.

93. For models like the SFA model, the commonly used R-squared measures of goodness-of-fit cannot be calculated, and alternative measures are used to select a preferred model; the most commonly used measures being the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). These goodness-of-fit measures both penalise the loss of degrees of freedom (i.e., reward parsimony), with the BIC penalising the inclusion of additional terms more heavily than the AIC. For both criteria, the specification with the lower value is considered to have the better fit.

94. Table 4 presents the values of these criteria for the SFA Cobb-Douglas and Translog models. The table shows that, for the short sample, both the AIC and the BIC select the Translog model as the preferred model, since the values for these criteria are lower (more negative) for the Translog model than for the Cobb-Douglas model. For the long sample, the Translog SFA is selected by the AIC criterion as the preferred model. However, the BIC criterion selects the Cobb-Douglas model as the preferred model – in this case, the improvement in the model's fit is considered to be outweighed by the decrease in parsimony.

Table 4**:** Alternative goodness of fit measure – SFA models

	AIC	AIC difference	BIC	BIC difference
SFA-CD Long	-1272.305		-1216.907	
SFA-TLG Long	-1284.286	-11.981	-1198.671	18.236
SFA-CD Short	-914.527		-864.019	
SFA-TLG Short	-969.578	-55.050	-896.111	-32.092

Source: Frontier Economics analysis

95. The above results for the goodness-of-fit of the Translog vs the Cobb-Douglas models indicate that, after allowing for the loss in degrees of freedom in the more flexible models, the Translog model, overall, fits the data better than the Cobb-Douglas model in all cases, except for one case, where the evidence is mixed. It is hard to reconcile these results with the AER's statement that the fit of the Translog models is not improved compared to the Cobb-Douglas model.

96. Quantonomics further observes that:

> *It is difficult to reconcile Evoenergy's apparent argument that the TLG model is to be preferred over the CD model … with its view that the varying rates of monotonicity violations in the TLG models when applied to different periods casts doubt on the reliability of all of the TLG opex cost function models, not just those with monotonicity violations.*[24]

97. This misinterprets Evoenergy's position as being in favour of the current Translog model specification. Evoenergy's main point (which we agree with) is that there is clear evidence that the

[24] Quantonomics, *Benchmarking limitations*, September 2023, p.5.

Cobb-Douglas models are misspecified and are, therefore, unreliable for setting revenue allowances.

98. As we explain below, the Translog models are also misspecified and are therefore also unreliable. We suspect that the misspecification problem affecting the Translog models also applies to the Cobb-Douglas models.

Misspecification of the Translog models

Residual plots

99. A standard technique used to check if an econometric model has been misspecified is to plot the residuals from the model (i.e., the differences between the fitted/predicted opex from the model and actual opex). If the model is well-specified, there should be no discernible pattern in the residual plot (i.e., the plotted residuals would be distributed randomly).

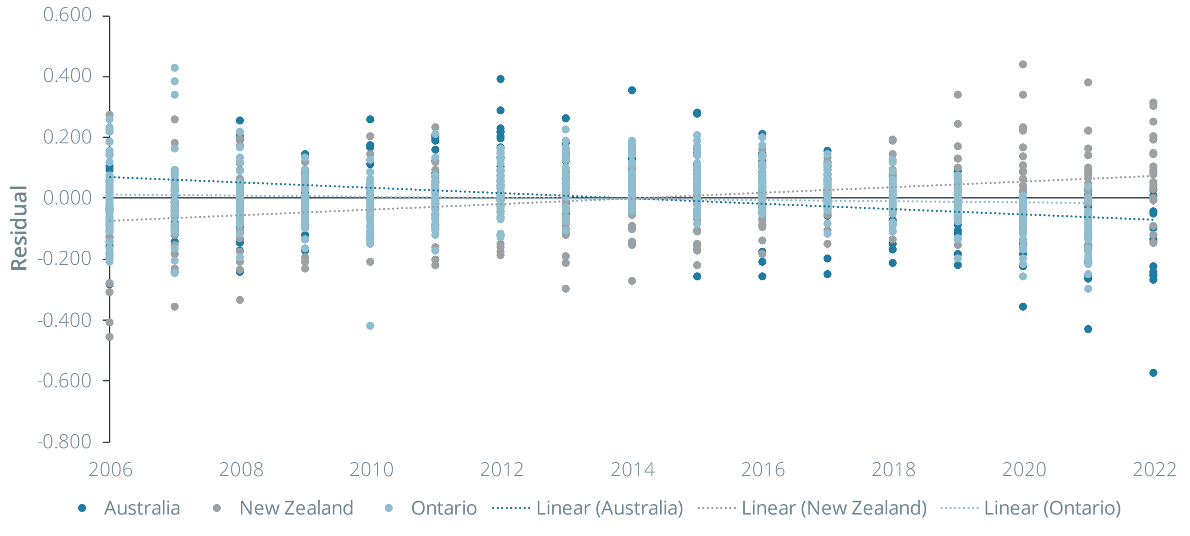
100. The plots show that for the Australian DNSPs there is a clear declining trend in the residuals over time. We found this to be true for all Cobb-Douglas and Translog models.

101. We have plotted the residuals from all the models against time. Figure 3 through Figure 6 plot the residuals for the Translog models. The figures also include a simple linear trend for each jurisdiction included in the figure. A downward-sloping trend line implies that residuals are decreasing over time (i.e., efficiencies are increasing), while an upward-sloping trend line shows that efficiencies are decreasing).

102. The figures show a negative trend for Australian DNSPs (i.e., improving efficiency), while New Zealand DNSPs appear to have an increasing trend (declining efficiency). The Ontario DNSPs residuals are relatively flat. The trend for Australian DNSPs is more visible in Figure 4 and Figure 6. The decrease in the residuals for the Australian DNSPs is particularly noticeable from about 2014 onwards, which corresponds to the start of the AER's current approach of using benchmarking to guide the setting of opex allowances.

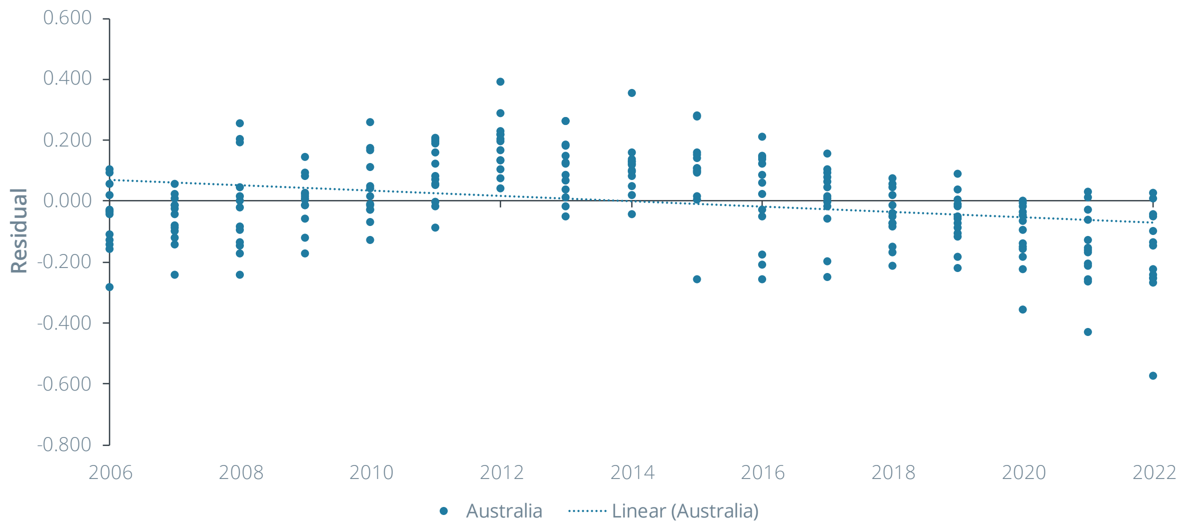
103. Analogous figures for the Cobb-Douglas models are presented in Figure 7 through Figure 10

Figure 3: Residuals – SFA-TLG long model



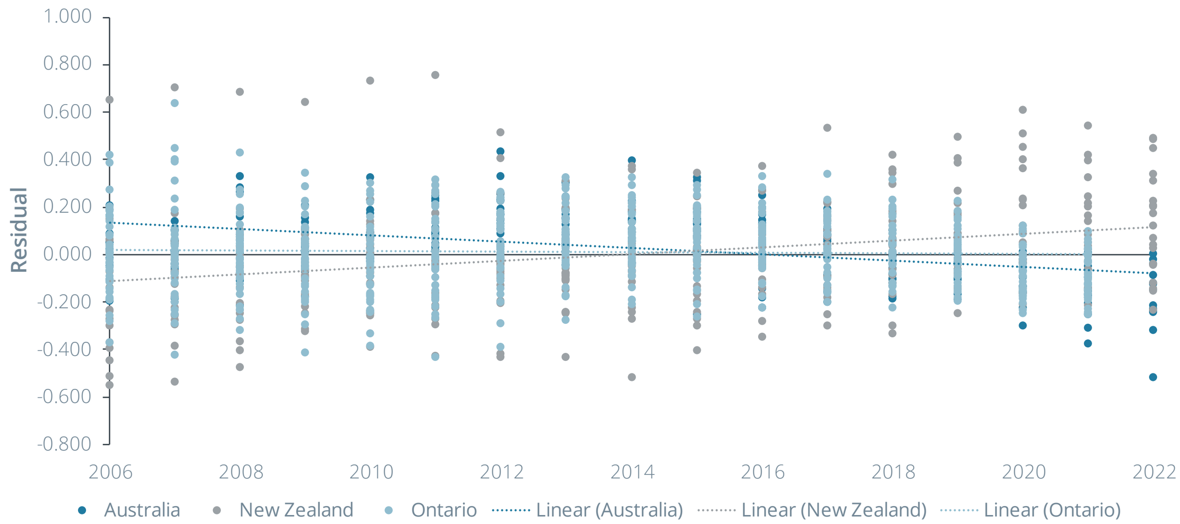
Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

Figure 4: Residuals for Australian DNSPs – SFA-TLG long model



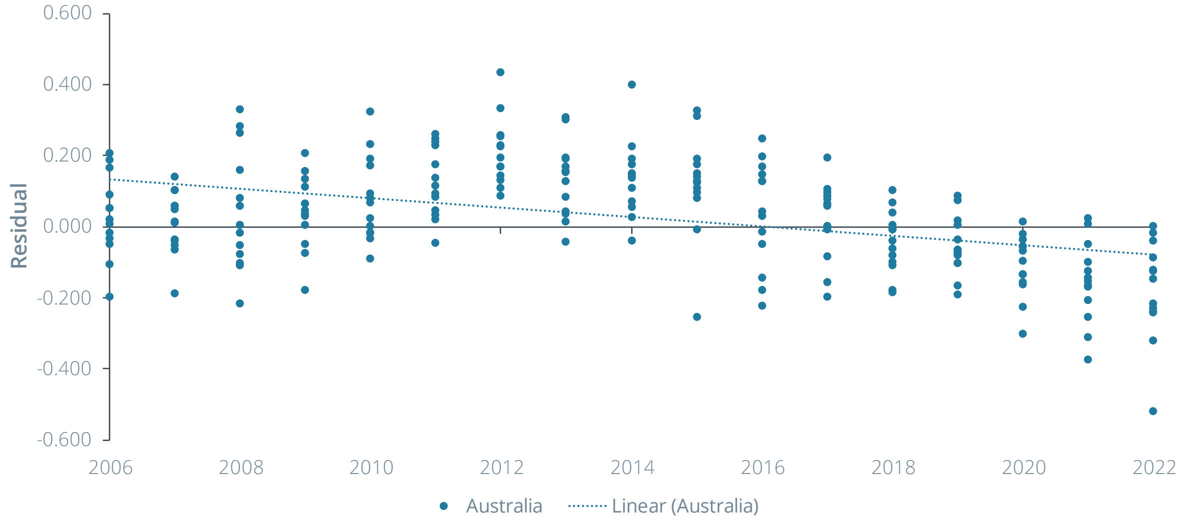
Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

Figure 5: Residuals – LSE-TLG long model



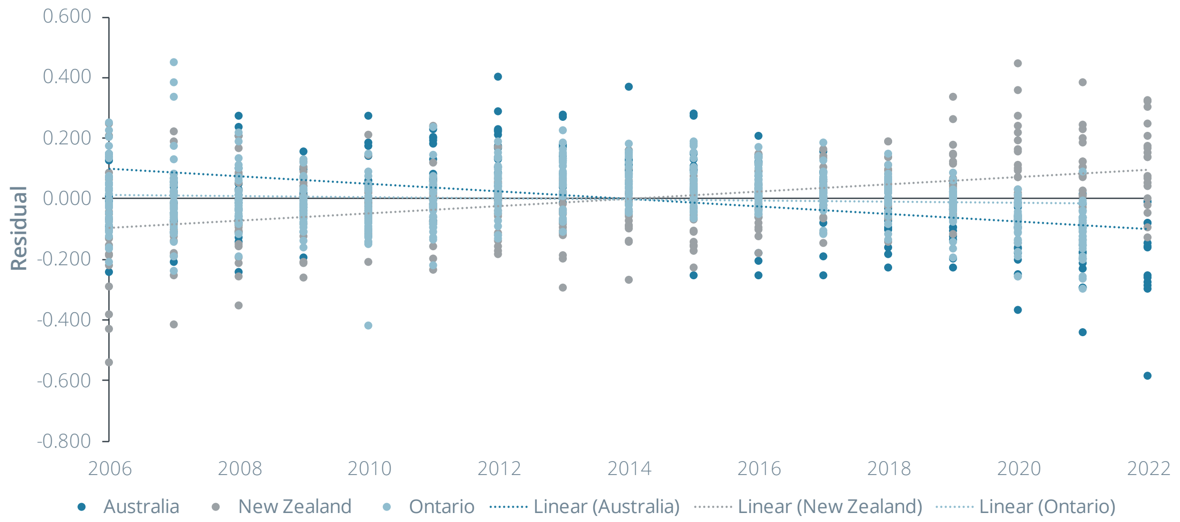
Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

Figure 6: Residuals for Australian DNSPs – LSE-TLG long model



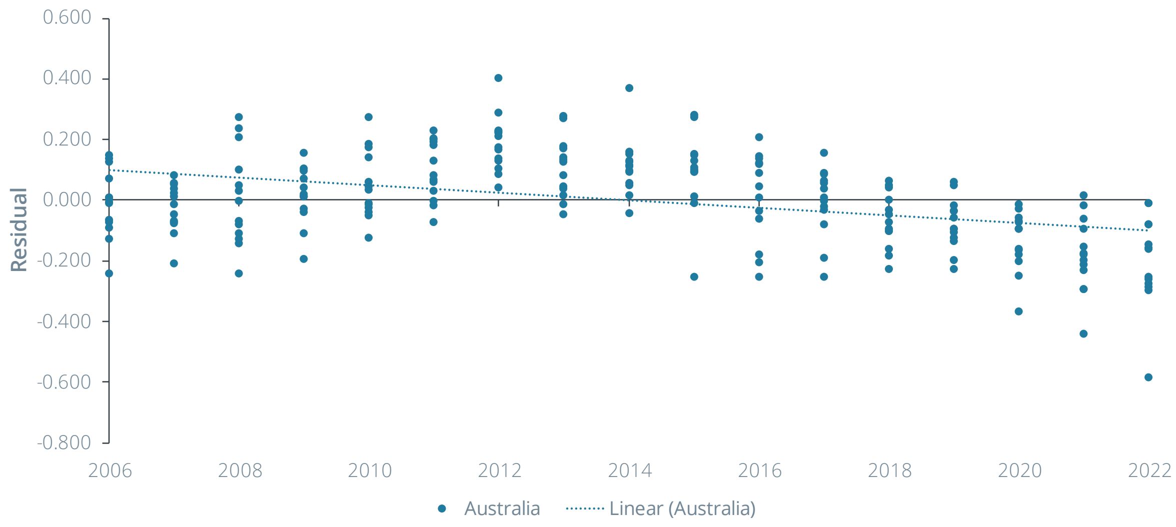
Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

Figure 7: Residuals – SFA-CD long model



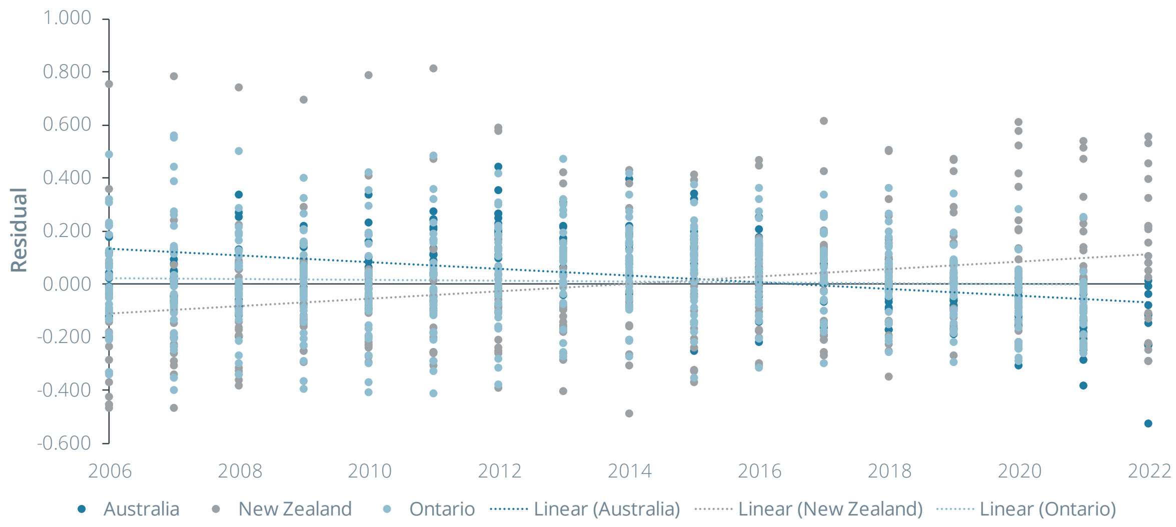
Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

Figure 8: Residuals for Australian DNSPs– SFA-CD long model



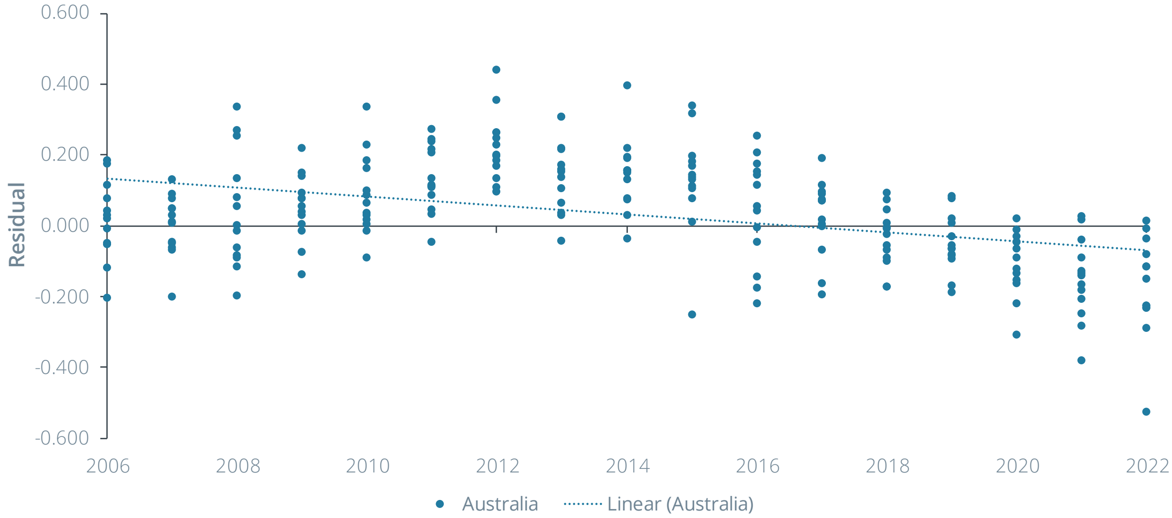
Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

Figure 9: Residuals – LSE-CD long model



Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

Figure 10: Residuals for Australian DNSPs– LSE-CD long model

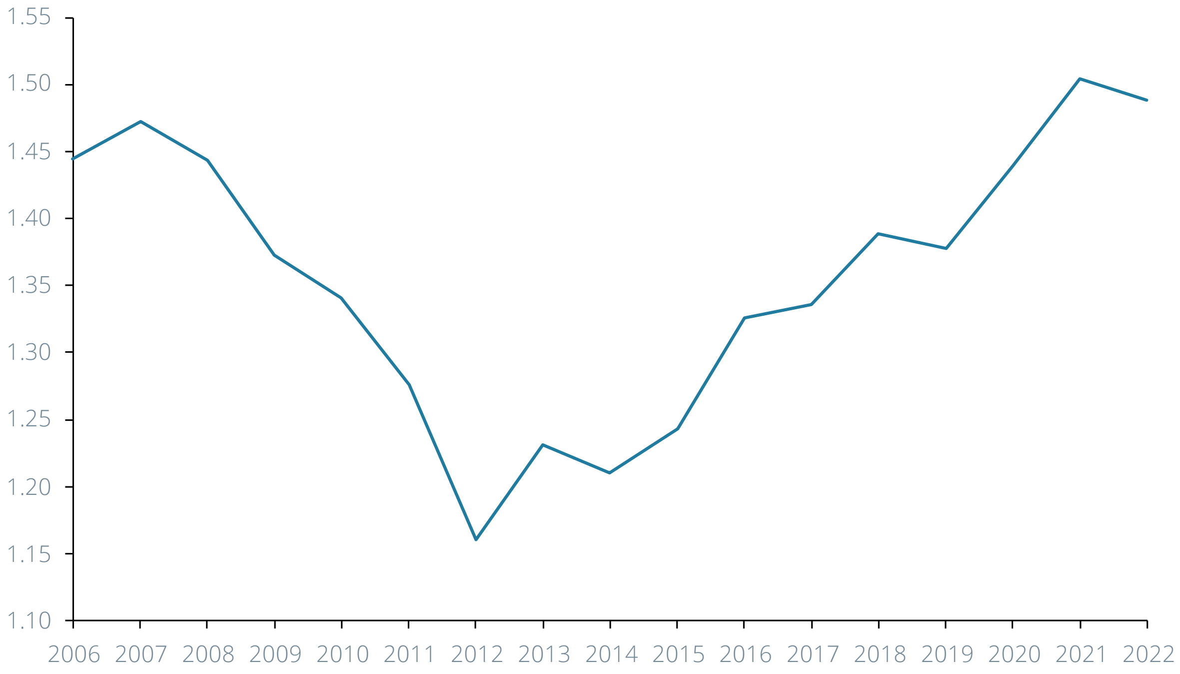


Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

104. The above figures make it abundantly clear that the residuals of the models for the Australian DNSPs are not random with respect to time, and that there is a time-related factor that is not accounted for properly in the AER's models when estimating the average efficiency of the Australian DNSPs. The downward trend in the residuals is consistent with the observation that the efficiency of the DNSP industry as a whole in Australia has improved significantly over time—particularly since the AER began using benchmarking analysis in 2014 as part of its revenue determinations, and since the publication of its Annual Benchmarking Reports.

105. This pattern is also observed in plots of the opex multilateral partial factor productivity (MPFP) for the Australian industry, as can be seen in Figure 11, which shows the opex MPFP for the Option 5 definition of opex. In this figure, a positive trend indicates an improvement in efficiency.

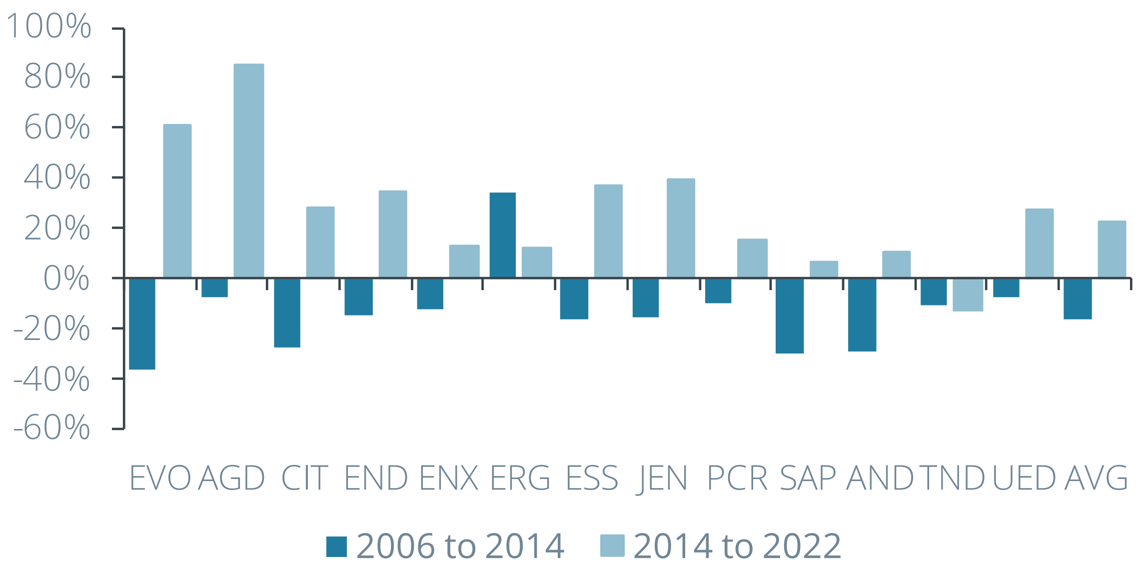
Figure 11: Opex MPFP for Australian DNSPs in aggregate over time



Source: Quantonomics analysis – DNSP23-MTFPtables-charts_Op5_12Oct2023.xlsx

106. The change in the trend in efficiencies of the Australian DNSPs is revealed even more starkly in Figure 12, which plots the change in the opex MPFP index for each DNSP, as well as the change in the weighted average opex MPFP for the industry as a whole, for the period before 2014 (dark blue bars) and after 2014 (light blue bars), i.e., the year when AER began using benchmarking analysis. All but one of the light blue bars are positive (indicating an improvement in efficiency since 2014), while all but one of the dark blue bars are negative (indicating declining efficiency before 2014).

Figure 12: Change in opex MPFP before and after 2014



Source: Quantonomics analysis – DNSP23-MTFPtables-charts_Op5_12Oct2023.xlsx

107. More specifically, Figure 12 shows that between 2006 and 2014 the opex MPFP index fell for the industry (i.e., on average) by approximately 16%. By sharp contrast, since 2014, the opex MPFP increased for the industry by approximately 23%. Evoenergy and Ausgrid have been the standout performers over that period.

108. The AER has recognised that its application of benchmarking analysis has contributed to an improvement in the efficiency of DNSPs. For example, the 2023 Annual Benchmark Report states:

> *Since 2014, the AER has used benchmarking in various ways to inform our assessments of network expenditure proposals. Our economic benchmarking analysis has been one contributor to the reductions in network costs and revenues for DNSPs and minimising retail prices, and retail price increases, faced by consumers.*[25]

109. Commenting on the opex partial factor productivity (PFP) and total factor productivity (TFP) indices for the industry, the AER observes in the 2023 Annual Benchmarking Report that:

> *…since 2012, opex reductions have been the most significant contributor to TFP growth, with opex PFP increasing on average by 2.9% each year.*[26]

110. The 2023 Annual Benchmarking Report also notes that:

> *Those DNSPs which have been the least productive over time have been improving their performance since 2012. In particular, Ausgrid and Evoenergy have increased their overall productivity, largely as a result of improvements in opex efficiency, noting Evoenergy's slight decline since 2016.*
>
> *Several middle-ranked DNSPs have also improved their relative MTFP performance to be closer to the top-ranked DNSPs. In recent years this includes United Energy, Jemena, Endeavour Energy and Essential Energy, again reflecting improved opex efficiency.*[27]

111. It is important to recognise that all the AER's current Cobb-Douglas and Translog models *assume* constant efficiencies over time. This is inconsistent with the AER's recognition that significant efficiency improvements have been achieved by many individual DNSPs and by the industry as a whole. This assumption of constant efficiencies over time is hard-wired into the specification of the models. Consequently, the AER's models are incapable, due to their specification, of accounting for

[25] AER, *Annual Benchmarking Report, Electricity distribution network service providers*, November 2023, p. 19.

[26] AER, *Annual Benchmarking Report, Electricity distribution network service providers*, November 2023, p. 25.

[27] AER, *Annual Benchmarking Report, Electricity distribution network service providers*, November 2023, p. v.

the fact that some DNSPs have improved their level of efficiency considerably over time. Since the models cannot account for these changes in efficiency over time directly, they will tend to overfit the data to other time-varying variables in the model. Given the highly flexible functional form of the Translog models, this response lack of time variation in the efficiencies in the model is likely to be more pronounced for the Translog models. This could be a key reason why the Translog models are prone to monotonicity violations— particularly when estimated using the short sample, which overlaps almost perfectly with the period over which the AER has been conducting benchmarking analysis.

112. Given the significant changes in DNSP efficiencies since 2014, the assumption of constant efficiencies is likely to cause a serious misspecification problem for the Cobb-Douglas models as well as the Translog.[28] However, the consequences of this misspecification problem are less easy to detect (e.g., as monotonicity violations) for the Cobb-Douglas models due to their more restrictive functional form. Nonetheless, both classes of models suffer from the same underlying issue.

Implausibly low efficiency estimates from SFA-TLG models

Detection of issue

113. Another telltale sign of a mis-specification problem is that some of the translog models produce implausibly low estimates of efficiency for some DNSPs. When analysing the preliminary results provided by Quantonomics,[29] we noted that the SFA-TLG models were producing very low estimates of efficiency for Ausgrid, for both the long and short models, and for both the standard opex and Option 5 opex approaches.[30] These scores, ranging from 26.1% to 37.9%, stood out as being low compared to the estimates of efficiency obtained from the other benchmarking models considered by the AER. This is most noticeable in the case of the short sample SFA-TLG.

114. When examining the long sample SFA-TLG model using Option 5 opex, we noted that the estimate of the *mu* parameter (the mean of the distribution of the inefficiency term) was negative at -0.825, while for the other models it was positive, ranging from 0.305 to 0.398.[31, 32]

115. In the SFA models used by Quantonomics, the inefficiency term u_i is distributed as a truncated normal distribution. Underlying this is a normal distribution with mean *mu* and some variance, but only the positive portion of the distribution is used to derive estimates of efficiency. When *mu* is positive, the positive part of the truncated normal distribution has a peak strictly greater than zero, but when *mu* is negative the positive part of the distribution is downward sloping.

116. These two possibilities are illustrated in Figure 13 using the efficiency distributions for the SFA-TL models presented by Quantonomics for Option 5 opex. The shape of the efficiency distribution for the long sample SFA-TLG model is very different to that of the short sample SFA-TLG model and implies that DNSPs with very low efficiency are far more common than would be feasible under the distributions resulting from models with positive *mu*. The negative *mu* also implies that 100%

[28] As noted above, the residual plots for the Cobb-Douglas models also exhibit a clear negative trend over time for the Australian DNSPs.

[29] Quantonomics, *Economic Benchmarking Results for the Australian Energy Regulator's 2023 DNSP Annual Benchmarking Report*, 17 August 2023.

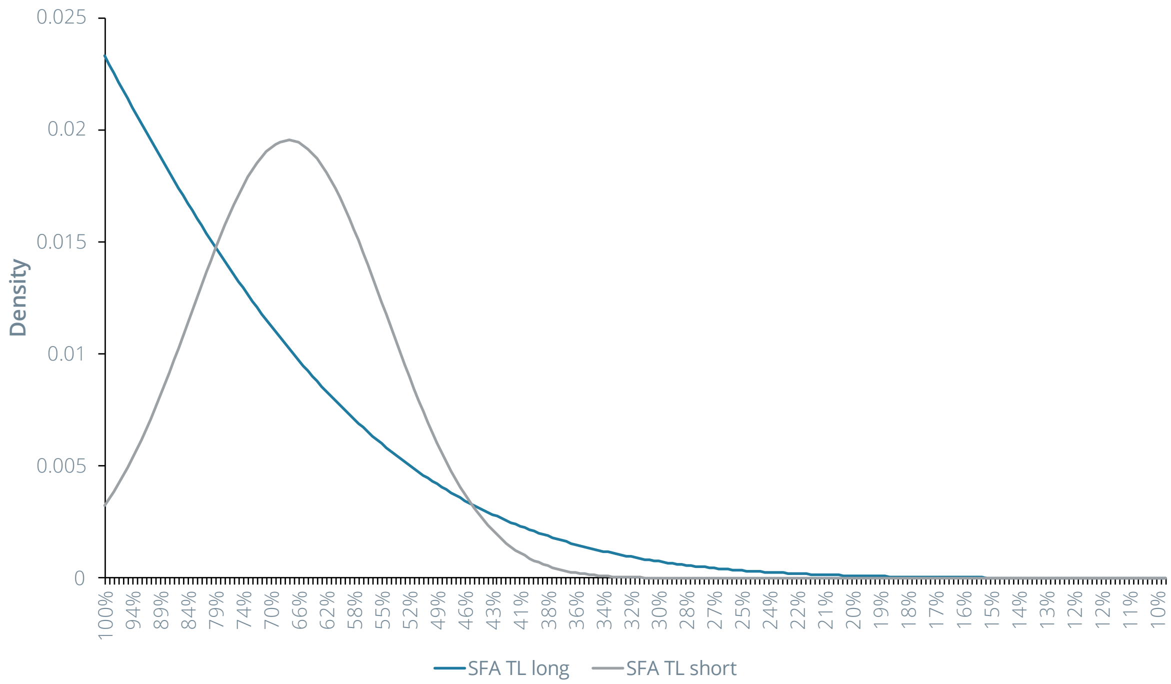
[30] See Quantonomics Tables 3.4, 3.6, 3.7, and 3.9.

[31] See Quantonomics Tables C.3, C.4, C.11, C.12, C.19, C.20, C.27, C.28.

[32] Similar observations apply to results from the standard opex approach.

efficiency is the most likely outcome, which is not the case for the models that the AER has historically relied upon.

Figure 13: Distribution of efficiency for Option 5 opex models estimated by Quantonomics



Source: Frontier Economics analysis

Note: We have used a non-evenly specified scale on the horizontal axis to better highlight the differences between these two types of efficiency distributions

Mis-estimation of short sample SFA-TLG model

117. Another issue of concern with the SFA models is whether the estimation algorithm for the SFA models has identified the best parameter estimates for the model. Unlike the LSE models, the SFA models cannot be estimated using a closed form solution. Instead, the SFA models are estimated iteratively using a technique known as maximum likelihood estimation (MLE). MLE involves searching iteratively for the model that best fits the data, which is achieved when the log-likelihood function associated with the observed data is maximised.

118. The statistical software for estimating the SFA models iteratively changes the values for the key parameters that define the log-likelihood function until the maximum of the log likelihood function is found. To commence the iterative process, it is necessary to specify starting values for *mu* and the other parameters in the model. The outcome of this process may be sensitive to the starting values. For example, the log-likelihood function could have several modes, i.e., several maximum turning points (referred to as local maxima). In these circumstances, it is possible that a *local* maximum for the log-likelihood function is identified rather than a *global* maximum (the global maximum is the largest of the local maxima). In such instances, because the true maximum of the log-likelihood function is not identified, the model will have been mis-estimated. That appears to have occurred in the case of the short sample SFA-TLG model.

119. One way to check whether a global maximum of the likelihood function has been found is to commence the iterative process using different starting values for the parameters and check

whether the iterative process leads to the same estimates as those found by Quantonomics. When we undertook this exercise for the long sample SFA-TLG model, we did indeed obtain the same parameter estimates for the long sample model SFA-TLG as those obtained by Quantonomics. This suggests that the long sample SFA-TLG model had not been mis-estimated.[33]

120. We then examined the short sample SFA-TLG model. In this case, to obtain alternative starting values for the parameters we set *mu* equal to 0 as a starting point and estimated the other parameters. This is, in fact, a popular simpler version of the truncated normal SFA model known as the half-normal SFA model.[34] When we examined the computer output for this initial step, we found that the log likelihood value for the half-normal SFA model is 491.7. This is larger than the log-likelihood value of 485.6 for the truncated normal SFA model estimated by Quantonomics.[35] This implies that the half-normal SFA model is, in fact, a better fitting model than the model estimated by Quantonomics.[36]

121. However, since the half-normal SFA model is a restricted version of the truncated normal SFA model, it is theoretically impossible for the half-normal SFA model to have a higher likelihood value than the likelihood function of the more general truncated normal model when this model is evaluated at the maximum likelihood estimates of the parameters. The fact that the truncated normal SFA model estimated by Quantonomics has a lower likelihood value than the half-normal SFA model implies that the estimates presented by Quantonomics are not associated with the maximum of the likelihood function. In other words, there are other parameter values for the truncated normal SFA model that produce higher values of the likelihood function than the parameter estimates obtained by Quantonomics. The Quantonomics estimates are likely produced by identifying a local maximum of the relevant log-likelihood function rather than the global maximum.

122. Since Quantonomics' parameter estimates for the short sample truncated normal SFA model are not associated with the global maximum of the likelihood function, they are not maximum likelihood estimates. Hence, we cannot infer that the desirable statistical properties that generally apply to maximum likelihood estimators apply to Quantonomics' estimates for the truncated normal SFA-TLG short model. In particular, we cannot assume that these estimates are statistically consistent estimates of the model's parameters. Hence, we do not know whether these estimates tend to the correct parameter values even in very large samples.

123. Table 5 below presents a comparison of the efficiency scores obtained from the short sample half-normal SFA-TLG model with the scores obtained from the Quantonomics model. The table shows that the short sample SFA-TLG estimates derived by imposing *mu* = 0 (i.e., half-normal) are materially different from those presented by Quantonomics. However, the half-normal estimates are clearly implausible and inconsistent with the estimates produced by the other models used by the AER. For instance, the efficiency estimates for Ausgrid and Energex from this model are below 10%, and the estimate for United Energy, a reference DNSP, is around 21%.

[33] To obtain alternative starting values for the iterative process, we set *mu* equal to 0.35 and estimated the other parameters in the model. We then used these estimates as the starting point for the iterative process used by the Stata package. Stata is a well-known econometric software package that has been used by Quantonomics to estimate the econometric benchmarking models.

[34] That is to say, the half normal model is *nested* within the truncated normal model.

[35] See Table C.28 of Quantonomics, *Economic Benchmarking Results for the Australian Energy Regulator's 2023 DNSP Annual Benchmarking Report*, November 2023.

[36] While the revised parameter estimates are a substantially better fit, the Quantonomics parameter estimates should be disregarded by virtue of not maximising the log-likelihood.

Table 5: Efficiency estimates for the short SFA-TLG model using Option 5 opex

DNSP	Quantonomics model	Model with *mu* = 0 (half-normal)
Evoenergy	51.8%	46.5%
Ausgrid	37.9%	4.0%
CitiPower	76.7%	37.7%
Endeavour Energy	58.7%	16.2%
Energex	48.6%	9.2%
Ergon Energy	72.0%	81.2%
Essential Energy	80.9%	94.9%
Jemena	54.8%	33.3%
Powercor	93.4%	62.7%
SA Power Networks	90.2%	62.4%
AusNet	64.4%	40.7%
TasNetworks	94.3%	96.7%
United Energy	66.8%	21.1%

Source: Frontier Economics analysis

124. To summarise, for the short sample SFA-TLG model, Quantonomics' estimates for the truncated normal model are not maximum likelihood estimates and the statistical properties of these estimates are unknown. An alternative model, the half-normal SFA model, which is widely used in SFA modelling, fits the data considerably better (i.e., it has a higher likelihood function). But the efficiency estimates produced by this model are unrealistic.

125. For the avoidance of doubt, we do not suggest that the AER should adopt the half-normal SFA model instead of the truncated normal SFA model. Indeed, we have found other parameter estimates for the truncated normal model for which the optimisation algorithm converges and for which the log-likelihood value is even higher than for the half-normal model. Rather, our investigation of the short sample SFA-TLG model likelihood function indicates that this likelihood function has an unconventional shape because the likelihood function keeps increasing as *mu* tends to minus infinity. This raises fundamental statistical issues in trying to obtain maximum likelihood estimates for this model. These issues are most likely due to the fact that the SFA-TLG model is a poor description of the true data generating process underlying the data used to estimate the model; that is, the model suffers from serious misspecification issues.

126. We have previously raised the fact that Quantonomics' estimates for the SFA-TLG short model are not maximum likelihood estimates, and provided an example of other parameter estimates that

result in a higher likelihood value. In its comments on our work,[37] Quantonomics chose to focus on the fact that the example of alternative estimates we provided had not converged. However, this mis-represents the point we were making, namely, that there are parameter values for which the likelihood value is higher than the likelihood value associated with Quantonomics' estimates. However, it is irrelevant whether the optimisation algorithm has converged; what matters is that our example established that the Quantonomics estimates do not correspond to the maximum value of the likelihood function since the parameter values in our example had a higher value of the likelihood function. We have, in fact, found numerous other examples of parameter values where the algorithm does converge and the associated likelihood function value is higher than the likelihood function value for Quantonomics' estimates.[38]

Implications

127. Given the statistical issues raised above regarding the maximum likelihood estimation of the short sample SFA-TGL model, we strongly recommend that this model, in its current specification, should be excluded completely from the benchmarking analysis. In our view, serious misspecification issues with the model are the cause of the unusual shape of the likelihood function and the resulting mis-estimation of the model's parameters by Quantonomics. Once it is recognised that the likelihood function has an unusual shape and that Quantonomics has not identified the maximum likelihood estimates for this model, it becomes clear that this model is not fit-for-purpose unless modifications are made to address these misspecification issues.

128. We acknowledge that due to monotonicity violations, the SFA-TLG short model would not have been used by the AER to assess the efficiency of base year opex. However, this is beside the point. The monotonicity violations are merely a symptom of a more fundamental problem with this model, namely that it is seriously misspecified. We have identified some of these misspecification issues elsewhere in this report.

129. While the case of the SFA-TLG short model discussed above provides a stark example of the statistical issues that can arise as a result of model misspecification, we believe that the same misspecification issues apply to *all* of the econometric benchmarking models used by the AER. For example, the long sample SFA-TLG model also exhibits implausibly low efficiency estimates.[39] However, there are no monotonicity violations associated with that model for Energex and it *would* be included by the AER when assessing the efficiency of a Energex's base year opex.[40] In other words, the AER's approach of excluding models that exhibit monotonicity violations only deals with one of the possible consequences of model misspecification. It does not deal with the more fundamental issue of correcting the misspecifications inherent in all of the AER's models. Hence, the current models continue to be misspecified and, therefore, are not fit-for-purpose.

130. As summarised in Figure 14, misspecification of the benchmarking models results in biased estimates of efficiency. Therefore, efficiency estimates derived using misspecified models should not be relied on for regulatory purposes.

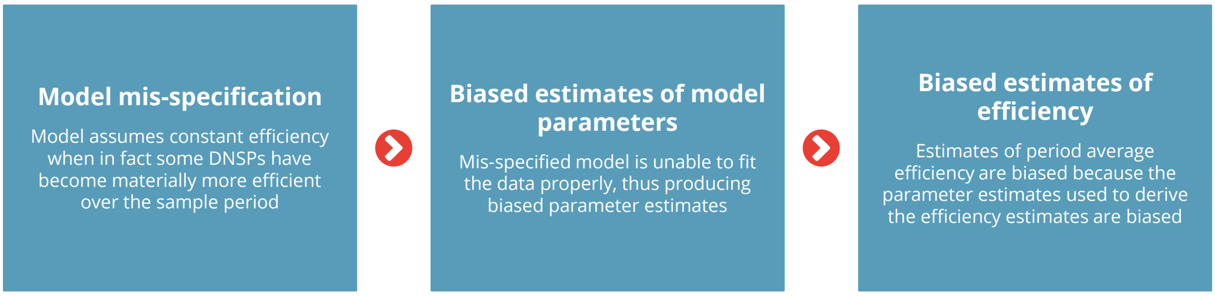
[37] Quantonomics, *Opex Cost Function–Options to Address Performance Issues of Translog Models*, 25 October 2023, p.39.

[38] We can provide the AER with the results of our investigations if requested.

[39] Quantonomics acknowledges in the October 2023 memo that this is an atypical *mu*, but point out that it is not statistically different from zero (p. 39). It is also not statistically different from 0.3, a typical *mu* value. If we re-estimate the model, imposing this value for *mu*, the estimate of efficient opex for Energex using the long sample SFA-TLG model increases from $303.3 million to $336.4 million.

[40] The long sample SFA-TLG is excluded for Ergon Energy as all observations have monotonicity violations.

Figure 14: The effect of model misspecification on period average efficiency estimates



Source: Frontier Economics

131. In our view, the proper course of action would be for the AER to review the specification of the models. That review cannot (and should not) be resolved without proper stakeholder consultation. Until the AER has had an opportunity to conduct such a review and demonstrate a material improvement in the reliability of the models, the AER should exercise extreme caution when interpreting the benchmarking results from its econometric models.

Failure of the models to account adequately for changes in efficiency over time of the Australian DNSPs

Simulated impact on estimated average efficiency

132. As explained above, we suspect that a key source of the statistical problems associated with the AER's models is that they are misspecified in the sense that they all assume that DNSPs maintain a constant level of efficiency over time. However, there is strong *prima facie* evidence (including evidence from the AER) that a number of the DNSPs have become more efficient over time. In these circumstances, the constant efficiency assumption built into the AER's benchmarking models will tend to produce biased estimates of efficiency.

133. We note that efficiency estimates are used in the AER's benchmarking roll-forward model to infer the efficient average opex of a DNSP over the sample period, by adjusting actual opex by the efficiency score and the efficiency target. In principle, it might not matter if a DNSP's efficiency changes over time, provided the estimated (constant) efficiency is a good estimate of the average efficiency. (This also assumes that the cost function elasticities are appropriate and not affected by the misspecification).

134. A key question, therefore, is whether the AER's models, although misspecified with respect to changing efficiencies over time, might still produce acceptable estimates of each DNSP's average efficiency over the sample period. One simple way to test this is via simulation analysis.

135. In summary, the simulation analysis we performed involved the following steps:

    a First, we assume a 'true' level of average efficiency for each DNSP over the long sample period;

    b Next, we assume starting values for efficiency in 2006 for that each DNSP, and a constant rate of efficiency improvement per annum that results in the assumed true average level of efficiency over the long sample period.

    c Then, we simulate the opex for each DNSP, for each year, using the assumed true (time-varying) efficiencies over the long sample period, and the AER's estimates of the parameters from the long sample SFA-TLG model.

d Finally, we fit the long sample SFA-TLG model to the simulated opex data and obtain efficiency estimates for each DNSP.

136. If the AER's long sample SFA-TLG model had accounted for the fact that the DNSPs had achieved constant efficiency improvements over time, then the estimated average efficiencies would match the assumed true level of efficiency that we had specified for each DNSP. In fact, what we found is that for most DNSPs, the higher the rate of efficiency improvement, the greater the extent to which the model underestimates the true level of average efficiency.

137. Since this simulation analysis is for illustrative purposes only, we focus on the long sample SFA-TLG model. Quantonomics' efficiency estimates from that model for the Australian DNSPs are presented in the second column of Table 6 below.

138. The third column of Table 6 specifies (for the purposes of the simulation analysis) the assumed true level of efficiency for each DNSP. The efficiencies in the third column are the same as those in the second column, except for Ausgrid, Endeavour Energy and Energex. For these DNSPs we specify more plausible assumed 'true' levels of efficiency than those presented in the second column, informed by the efficiency scores derived from the AER's other models.

Table 6: Efficiency estimates – long sample SFA-TLG, ABR dataset, Option 5 opex

DNSP	Quantonomics' estimated efficiencies	Assumed efficiencies in the simulation analysis
Evoenergy	54.19%	54.19%
Ausgrid	30.68%	55.00%
CitiPower	78.26%	78.26%
Endeavour Energy	48.80%	60.00%
Energex	44.97%	50.00%
Ergon Energy	74.41%	74.41%
Essential Energy	75.06%	75.06%
Jemena	76.14%	76.14%
Powercor	97.49%	97.49%
SA Power Network	90.99%	90.99%
AusNet	73.65%	73.65%
TasNetworks	96.06%	96.06%
United Energy	81.22%	81.22%

Source: Quantonomics estimates and Frontier Economics assumptions

139. We then allow the efficiency of each Australian DNSP to change over time. We specify an efficiency adjustment factor of x%, so that the DNSP's efficiency in 2006 is equal to the DNSP's specified average minus x% of the gap between their average and 100% and ending in 2022 at the average plus x% of the gap between their average and 100%. For example, if we set the adjustment factor to 25%, a DNSP with an average efficiency of 60% will have an efficiency of 60% in the middle of the sample period, increasing to 70%[41] at the end of the sample period (and starting at 50% at the beginning of the sample period).

140. We then simulate opex for each DNSP for each year by applying the specified true level of efficiency (time varying for Australian DNSPs) to the production function and applying random shocks (using the distribution of the random error term from the initial estimation). Thus, we obtain an alternative (simulated) opex for each of the 1,137 observations in the 2023 Annual Benchmarking Report dataset.

141. We then apply the SFA-TLG model estimation procedure, and obtain efficiency estimates for the Australian DNSPs, as well as the *mu* (mean of the distribution used for the inefficiency term).

142. The results of the simulation analysis are presented in Figure 15. The figure shows that the estimated average efficiency for Ausgrid falls dramatically as the efficiency time trend increases. By way of example, Ausgrid's true average efficiency is assumed to be 55%. In the scenario where the efficiency trend coefficient is 40%, Ausgrid's true efficiency rises from 37% in 2006 to 73% in 2022.[42] However, the estimated constant efficiency is only 27%, well below the assumed true average efficiency of 55%.

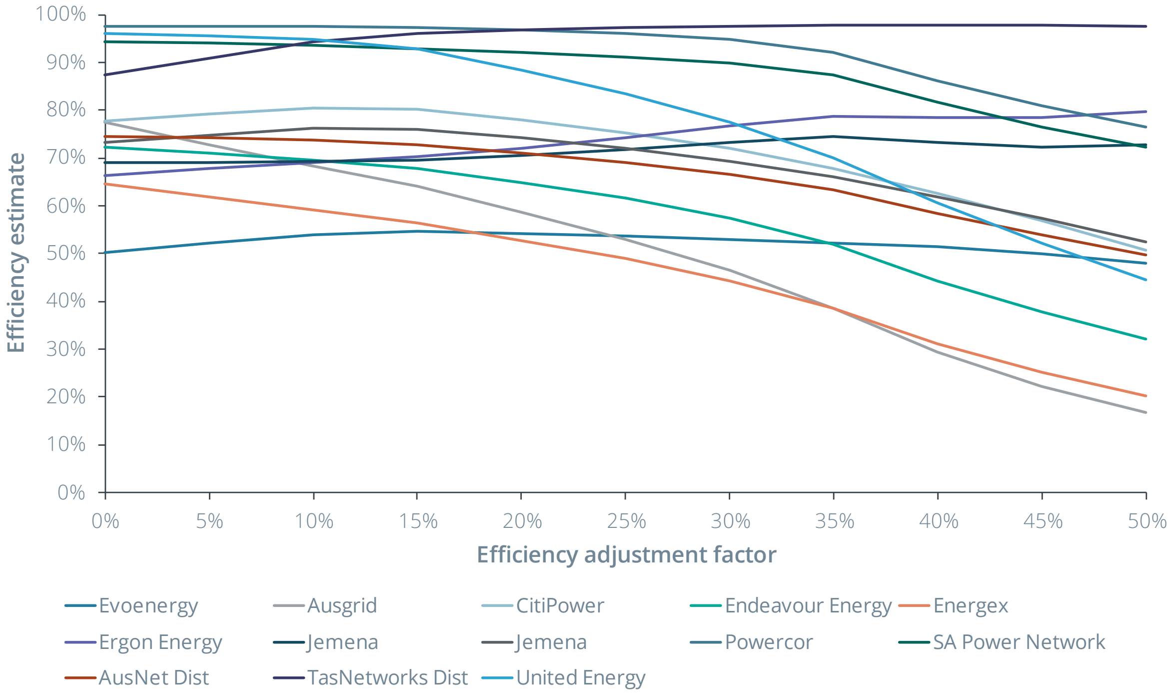
143. Furthermore, as Figure 16 shows, the estimated *mu* parameter becomes negative and large as the efficiency trend increases. As explained above, when this occurs, the SFA-TLG models will tend to produce excessively low efficiency estimates for DNSPs.

144. This simulation analysis suggests that the AER's models cannot estimate the average efficiency of DNSPs correctly when efficiency varies over time (because the models assume constant efficiency over time), and in fact may understate average efficiency by a considerable margin for firms that have improved their efficiency the most.

[41] Derived as 60% + 25% x (100% – 60%).

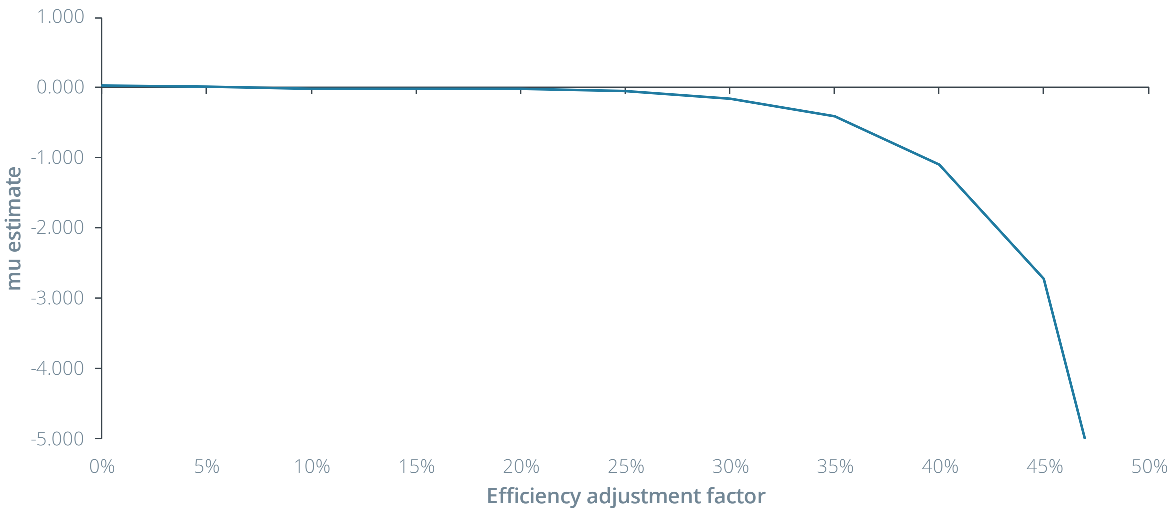
[42] Taking the gap between the actual and predicted opex as a proxy for efficiency, more specifically using the exponent of the negative of the residual, Ausgrid's estimated efficiency rises from 28% in 2006 to 54% in 2022. Thus, such a trend factor is not unreasonable and reflects the large improvements that some DNSPs have made.

Figure 15: Estimates of efficiency for different efficiency adjustment factors



Source: Frontier Economics analysis

Figure 16: Estimates of mu for different efficiency trend factors



Source: Frontier Economics analysis

Alternative LSE specification

145. As a further illustration of the potential for time-varying efficiencies to lead to biased estimates of sample average efficiency, we consider a simple modification to the LSE models to include a time varying efficiency trend for each Australian DNSP.

146. Instead of only including dummies for each Australian DNSP in the LSE models, we also include DNSP specific time trends for the Australian DNSPs. To achieve this, the DNSP dummy D_i is

multiplied by the sum of α_i and the time trend variable t multiplied by the DNSP's yearly change in the dummy variable coefficient, β_i, resulting in the following specification:

$$logRealOpex_{it} = X_t\beta + (\alpha_i + t\beta_i)D_i + \epsilon_{it}$$

147. These factors can be converted into efficiency scores. For each year, we find the lowest factor across DNSPs, and then average these factors over the sample period. Individual efficiencies are then calculated with respect to this factor. We then average the year specific efficiencies for each DNSP.

148. The results for the four LSE models are presented in Table 7 through Table 10. Some substantial differences can be observed between the estimates of average efficiency under the AER's existing LSE models and the modified LSE models that incorporate a DNSP-specific time trend. For example, Ausgrid's estimated efficiency for the short sample LSE-TLG model increases substantially (by more than seven percentage points).

Table 7**:** Efficiency score comparison, long LSE-CD

Efficiency	Quantonomics specification	Alternative model with time-varying efficiency
Evoenergy	48.0%	46.6%
Ausgrid	58.0%	57.3%
CitiPower	80.0%	77.1%
Endeavour Energy	66.0%	64.1%
Energex	68.7%	66.6%
Ergon Energy	55.6%	54.3%
Essential Energy	66.0%	64.0%
Jemena	72.9%	70.4%
Powercor	100.0%	96.9%
SA Power Networks	93.7%	90.7%
AusNet	82.2%	79.8%
TasNetworks	83.0%	80.4%
United Energy	94.2%	91.3%

Source: Frontier Economics analysis

Table 8: Efficiency score comparison, long LSE-TLG

Efficiency	Quantonomics	Alternative
Evoenergy	43.1%	41.0%
Ausgrid	55.5%	57.7%
CitiPower	73.0%	69.8%
Endeavour Energy	66.0%	64.6%
Energex	66.8%	67.0%
Ergon Energy	57.9%	52.4%
Essential Energy	73.1%	68.1%
Jemena	57.0%	55.2%
Powercor	100.0%	96.2%
SA Power Networks	97.0%	92.1%
AusNet	76.0%	75.1%
TasNetworks	78.2%	72.9%
United Energy	75.1%	74.8%

Source: Frontier Economics analysis

Table 9: Efficiency score comparison, short LSE-CD

Efficiency	Quantonomics	Alternative
Evoenergy	45.1%	46.2%
Ausgrid	58.7%	61.1%
CitiPower	71.5%	71.6%
Endeavour Energy	65.2%	66.2%
Energex	65.9%	66.8%
Ergon Energy	58.6%	59.4%
Essential Energy	66.3%	68.4%
Jemena	66.5%	66.5%
Powercor	100.0%	100.3%
SA Power Networks	87.5%	88.2%
AusNet	77.2%	77.4%
TasNetworks	80.5%	81.6%
United Energy	92.1%	92.8%

Source: Frontier Economics analysis

Table 10: Efficiency score comparison, short LSE-TLG

Efficiency	Quantonomics	Alternative
Evoenergy	43.9%	43.3%
Ausgrid	55.1%	62.5%
CitiPower	71.9%	70.9%
Endeavour Energy	67.9%	68.3%
Energex	63.4%	67.8%
Ergon Energy	69.1%	60.4%
Essential Energy	79.3%	75.4%
Jemena	52.2%	55.3%
Powercor	100.0%	100.3%
SA Power Networks	94.7%	91.1%
AusNet	68.5%	73.7%
TasNetworks	81.6%	77.4%
United Energy	72.2%	80.1%

Source: Frontier Economics analysis

149. We also find that modifying the Translog models to allow for different time trends for individual DNSPs reduces the number of monotonicity violations that occur. By way of example, Table 11 shows the effect of allowing separate time trends for the individual Australian DNSPs on the monotonicity violations associated with the LSE models. The table shows that:

   a Under the AER's 'standard' specification, the short sample LSE-CD model is not excluded due to monotonicity violations for any of the DNSPs. This does not change if individual time trends are allowed for the Australian DNSPs; and

   b Under the AER's standard specification, the short sample LSE-TLG model is excluded due to monotonicity violations for all of the DNSPs. However, the introduction of individual time trends for the Australian DNSPs results in the exclusion of the short sample LSE-TLG model for only two DNSPs.

Table 11: Effect of allowing individual Australian DNSP time trends on monotonicity violations

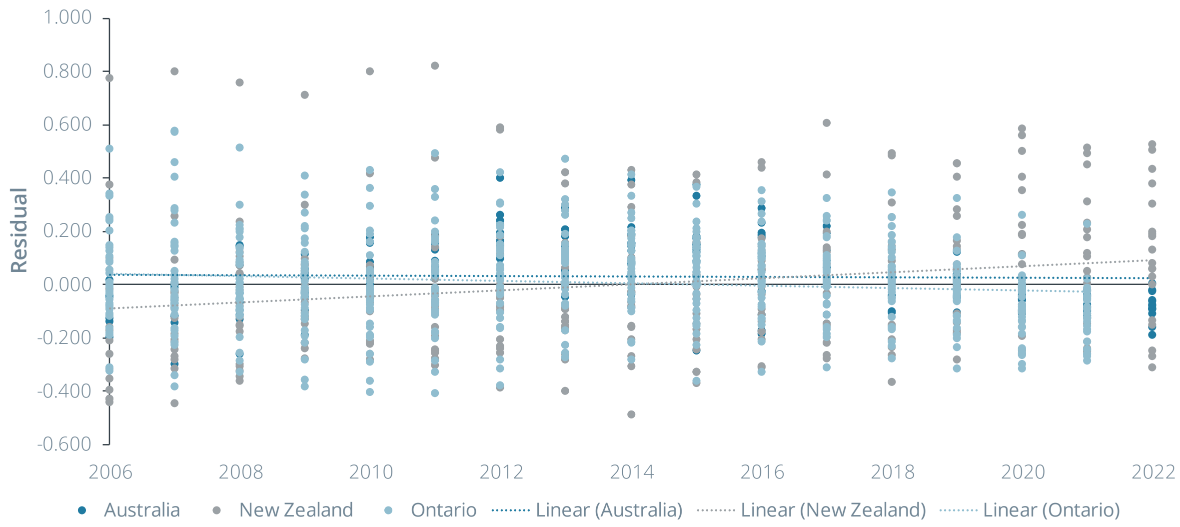
DNSP	LSE-CD short Standard	LSE-CD short Alternative	LSE-TLG short Standard	LSE-TLG short Alternative
Evoenergy	45%	46%	44%	43%
Ausgrid	59%	61%	55%	62%
CitiPower	71%	72%	72%	71%
Endeavour Energy	65%	66%	68%	68%
Energex	66%	67%	63%	68%
Ergon Energy	59%	59%	69%	60%
Essential Energy	66%	68%	79%	75%
Jemena	66%	67%	52%	55%
Powercor	100%	100%	100%	100%
SA Power Networks	88%	88%	95%	91%
AusNet Dist	77%	77%	69%	74%
TasNetworks Dist	80%	82%	82%	77%
United Energy	92%	93%	72%	80%

Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

Note: Green indicates that the model is included due to satisfying the monotonicity requirement

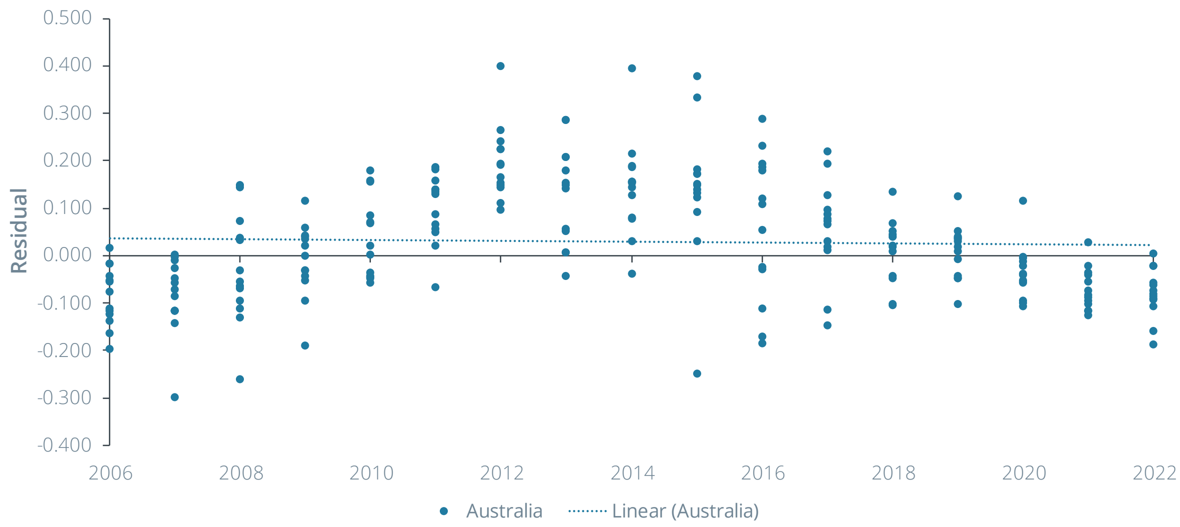
We also find that the residual plots for these alternative models are improved compared to the original specifications, with the residuals of Australian DNSPs having a flat trendline.

Figure 17: Residuals – LSE-CD long model – time varying efficiency



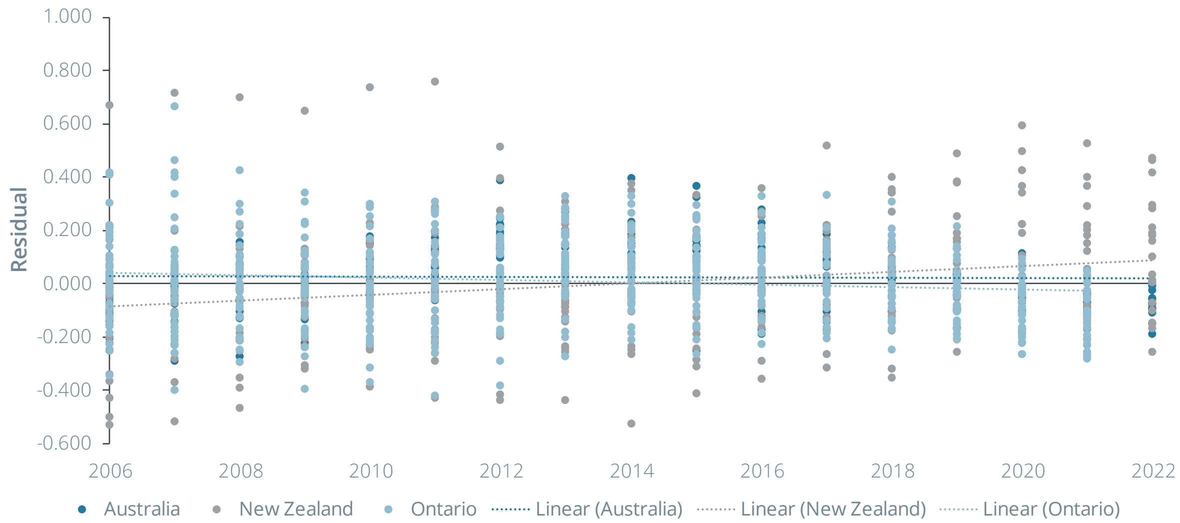
Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

Figure 18: Residuals for Australian DNSPs – LSE-CD long model – time varying efficiency



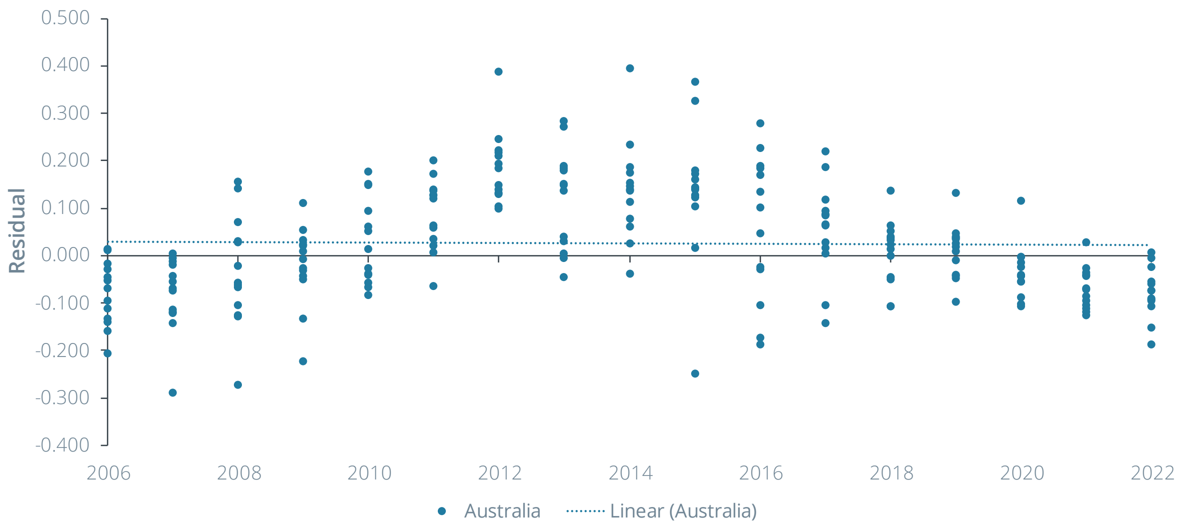
Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

Figure 19: Residuals – LSE-TLG long model – time varying efficiency



Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

Figure 20: Residuals for Australian DNSPs – LSE-TLG long model – time varying efficiency



Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

150. To avoid any doubt, we are not suggesting that the alternative model specification presented above (for illustrative purposes only) should be adopted by the AER. We simply show that making modest improvements to its models can reduce the misspecification problem and can also reduce the prevalence of monotonicity violations.

151. Our key recommendation is that the AER must rethink its econometric benchmarking models to reduce the misspecification problems that those models currently suffer from. A key focus of any such review should be how to allow for the fact that individual Australian DNSPs have made significant improvements in efficiency, particularly since 2014 — a change that the existing benchmarking models are unable to capture effectively.

Assumption of uniform time trend for all jurisdictions in the AER's sample

152. The AER's existing models do incorporate a time trend term. However, the time trend is assumed to be the same across all three jurisdictions. In other words, the AER's models cannot account for the possibility that the efficiency of the Australian DNSPs may have changed differently over time the efficiency of the New Zealand and Ontarian DNSPs.

153. As demonstrated in the residual plots presented above, the pattern of residuals over time differs between DNSPs in the different jurisdictions. That is, the residual plots indicate that:

    a There is a factor that is currently unaccounted for in the AER's models that is negatively related to time with respect to the Australian DNSPs; and

    b There is a factor that is currently unaccounted for in the AER's models that is positively related to time with respect to the New Zealand DNSPs.

154. A simple way to account for this is to add further time trend variables so that each jurisdiction has its own time trend. The 2023 Annual Benchmarking Report notes that the AER intends to explore and consult further on this issue.

155. As the AER notes in the 2023 Annual Benchmarking Report, the time trend in the models is intended to capture the effects of technical change on opex (i.e., 'frontier shift'/productivity improvements) over time. However, it is also likely to capture other factors that vary over time (e.g.,

improvements in efficiency/catch-up to the efficient frontier) but which are not accounted for explicitly in the model. For this reason, the time trend effect may differ between jurisdictions.

156. For illustrative purposes only, we have investigated an extension to the AER's standard model that allows for different time trends between jurisdictions. The resulting estimates of the time trends are presented in Table 12 below. The table also presents the estimated common time trend in the AER's current models. The resulting time trends estimates are consistent with what we would expect based on the residual plots in Figure 3 through Figure 6 above.

157. Table 12 also presents the results of a test of the null hypothesis that there is no difference in the time trends across the three jurisdictions. This null hypothesis is rejected convincingly for all eight models, with a p-value of 0.00% in every case. That is, the statistical evidence indicates that the efficiency of DNSPs in Australia are changing over time at a different rate to the DNSPs in New Zealand and Ontario—contrary to the uniform time trend assumed in the AER's existing models.

Table 12**:** Estimated time trends

	SFA CD short	**SFA TLG short**	**LSE CD short**	**LSE TLG short**	**SFA CD long**	**SFA TLG long**	**LSE CD long**	**LSE TLG long**
Standard model	0.33%	0.63%	0.30%	0.51%	1.19%	1.14%	1.03%	1.17%
Extended model - Australia	-3.34%	-3.38%	-3.02%	-2.90%	-0.28%	-0.52%	-0.18%	-0.25%
Extended model - NZ	2.89%	2.93%	2.79%	2.93%	2.45%	2.40%	2.58%	2.76%
Extended model - Ontario	-0.32%	-0.30%	0.08%	0.27%	0.76%	0.64%	0.61%	0.76%
Test								
ChiSq(2)	434.37	402.89	84.9	89.68	166.66	162.79	42.33	50.12
p-value	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Source: Frontier Economics analysis

158. Since there is *prima facie* evidence that the Australian DNSPs have become more efficient over time (particularly since 2014), the assumption of a uniform time trend could result in biased estimates of efficiency.

159. As shown in Table 13, the resulting efficiency estimates are more plausible for the short and long SFA-TLG model (results highlighted in red).

Table 13: Comparison of efficiency scores – standard model vs including separate time trends for each jurisdiction

	Short SFA-CD		Short SFA-TLG		Short LSE-CD		Short LSE-TLG		Long SFA-CD		Long SFA-TLG		Long LSE-CD		Long LSE-TLG	
DNSP	Standard	Alternative	Standard*	Alternative	Standard	Alternative	Standard	Alternative	Standard	Alternative	Standard	Alternative	Standard	Alternative	Standard	Alternative
Evoenergy	50%	48%	46%	47%	45%	45%	44%	43%	47%	46%	54%	44%	48%	47%	43%	43%
Ausgrid	59%	58%	3%	58%	59%	59%	55%	61%	55%	54%	31%	59%	58%	57%	55%	60%
CitiPower	77%	70%	37%	77%	71%	71%	72%	71%	83%	76%	78%	77%	80%	78%	73%	74%
Endeavour Energy	69%	65%	14%	63%	65%	65%	68%	68%	65%	63%	49%	62%	66%	65%	66%	67%
Energex	68%	67%	8%	65%	66%	66%	63%	68%	66%	66%	45%	69%	69%	68%	67%	71%
Ergon Energy	60%	52%	83%	59%	59%	58%	69%	60%	56%	52%	74%	59%	56%	55%	58%	53%
Essential Energy	66%	64%	96%	80%	66%	68%	79%	75%	60%	61%	75%	69%	66%	66%	73%	71%
Jemena	66%	69%	33%	64%	66%	67%	52%	55%	69%	71%	76%	71%	73%	73%	57%	59%
Powercor	96%	95%	58%	95%	100%	100%	100%	100%	96%	97%	97%	95%	100%	100%	100%	100%
SA Power Networks	89%	88%	57%	91%	88%	88%	95%	92%	88%	90%	91%	93%	94%	93%	97%	96%
AusNet Dist	73%	78%	38%	75%	77%	77%	69%	73%	74%	78%	74%	76%	82%	82%	76%	78%
TasNetworks Dist	87%	79%	97%	74%	80%	82%	82%	78%	85%	81%	96%	72%	83%	82%	78%	75%
United Energy	92%	94%	19%	88%	92%	93%	72%	81%	94%	94%	81%	96%	94%	95%	75%	80%

*Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset. Note: * These results presented here reflect the estimates presented in the final column in* ***Error! Reference source not found.*** *above r ather than the original estimates presented by Quantonomics.*

160. For example, under the short sample SFA-TLG:

   a Ausgrid's estimated efficiency increases from 3% to 58%;

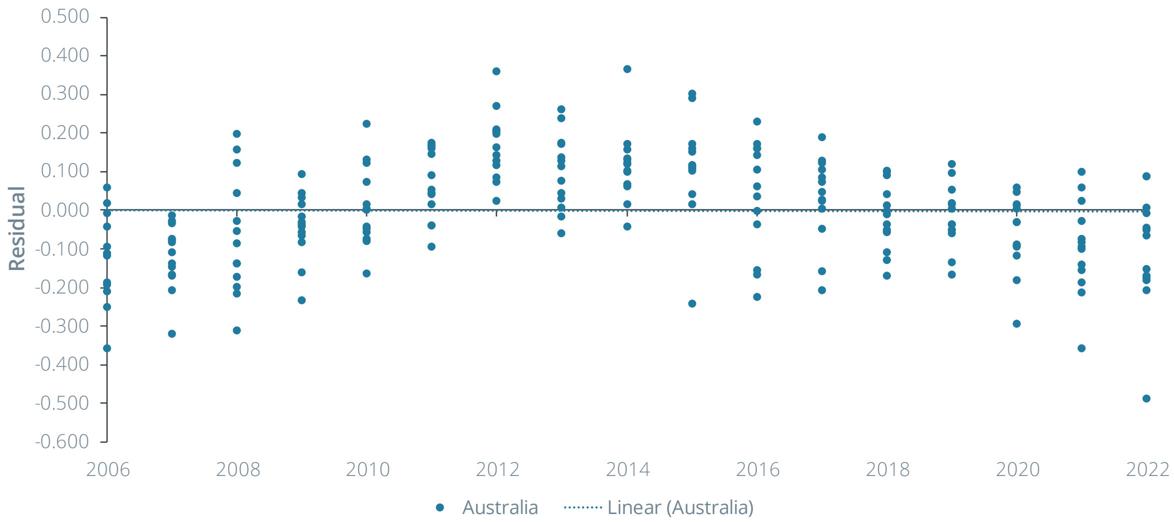
   b Energex's estimated efficiency increases from 8% to 65%; and

   c United Energy's estimated efficiency increases from 19% to 88%.

161. In every instance, the resulting efficiency estimates are more consistent with the efficiency estimates produced by the other models.

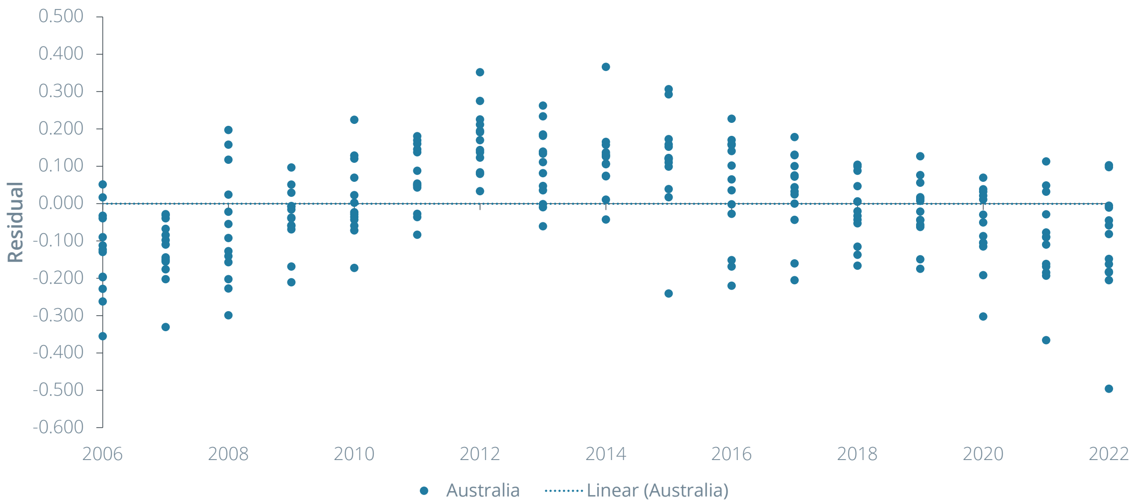
162. Analysis of the residual plots associated with the modified models, presented below in Figure 21 to Figure 24, illustrate that the residuals of Australian DNSPs no longer feature a time trend when differing time trends for New Zealand and Ontario DNSPs are allowed for.

Figure 21: Residuals for Australian DNSPs – SFA-CD model



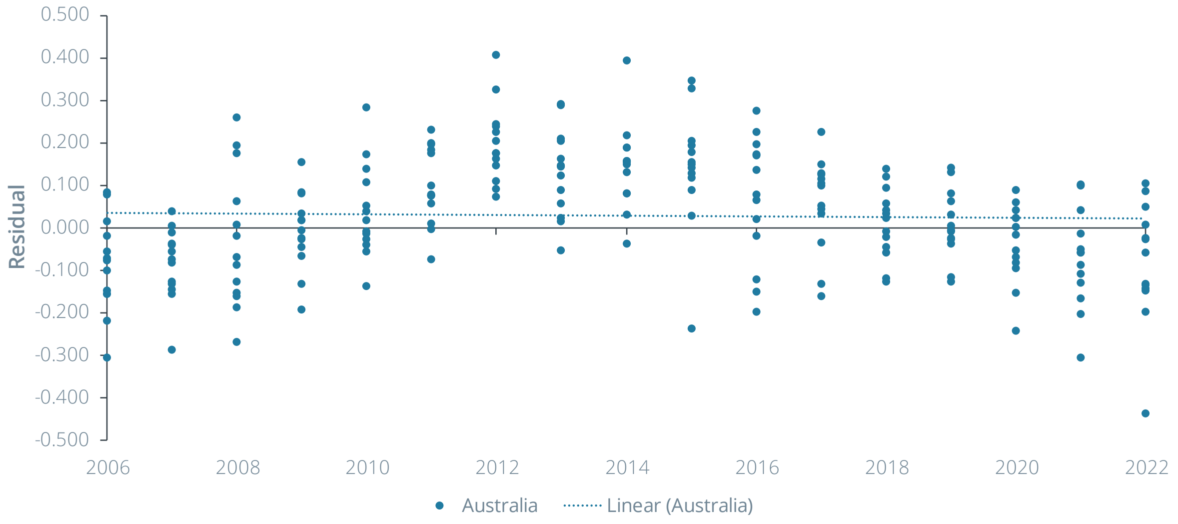
Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

Figure 22: Residuals for Australian DNSPs – SFA-TLG model



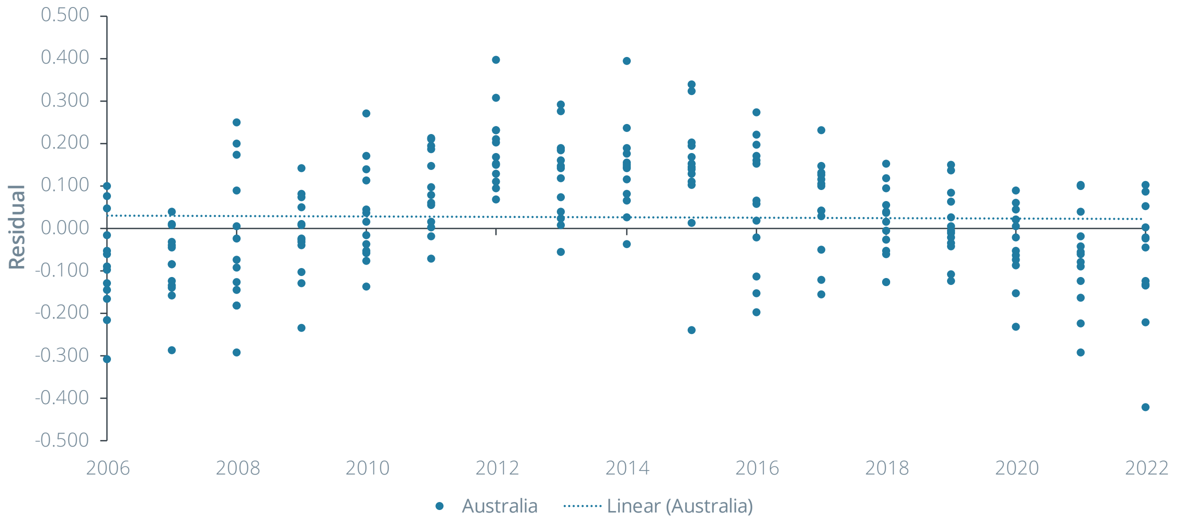
Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

Figure 23: Residuals for Australian DNSPs – LSE-CD model



Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

Figure 24: Residuals for Australian DNSPs – LSE-TLG model



Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

163. Table 14 shows that allowing different time trends for the Australian, New Zealand and Ontarian DNSPs results in fewer of the short sample SFA-TLG models being excluded due to monotonicity violations.

Table 14: Effect of allowing jurisdiction-specific time trends on monotonicity violations

	Short SFA-CD		Short SFA-TLG		Short LSE-CD		Short LSE-TLG	
DNSP	Standard	TimeTrends	Standard	TimeTrends	Standard	TimeTrends	Standard	TimeTrends
Evoenergy	50%	48%	46%	47%	45%	45%	44%	43%
Ausgrid	59%	58%	3%	58%	59%	59%	55%	61%
CitiPower	77%	70%	37%	77%	71%	71%	72%	71%
Endeavour Energy	69%	65%	14%	63%	65%	65%	68%	68%
Energex	68%	67%	8%	65%	66%	66%	63%	68%
Ergon Energy	60%	52%	83%	59%	59%	58%	69%	60%
Essential Energy	66%	64%	96%	80%	66%	68%	79%	75%
Jemena	66%	69%	33%	64%	66%	67%	52%	55%
Powercor	96%	95%	58%	95%	100%	100%	100%	100%
SA Power Networks	89%	88%	57%	91%	88%	88%	95%	92%
AusNet Dist	73%	78%	38%	75%	77%	77%	69%	73%
TasNetworks Dist	87%	79%	97%	74%	80%	82%	82%	78%
United Energy	92%	94%	19%	88%	92%	93%	72%	81%

Source: Frontier Economics analysis of 2023 Annual Benchmarking Report dataset

Note: Green indicates that the model is included due to satisfying the monotonicity requirement.

164. Again, we do not suggest that the AER should necessarily adopt the models specification presented above. The time trend term in the AER's models is intended to reflect an estimate of the rate of technical progress (i.e., frontier shift). However, it is likely that the estimated time trend in the modified model is capturing both the rate of frontier shift and catch-up efficiency achieved by the Australian DNSPs. If that is the case, the estimated time trend would not be suitable for rolling forward an estimate of efficient opex to the base year—since that would imply that DNSPs would need to achieve ongoing catch-up efficiency in addition to productivity improvements that reflect the shift in the efficient frontier over time.

165. Therefore, more work would need to be done to produce modified benchmarking model that disentangle the contributions of frontier shift and catch-up efficiency to the estimated time trend.

166. There are likely to be other issues that would need to be resolved before the modified model presented above could be considered for use to set regulatory allowances. The key point is that there is an imperative for the AER to give limited weight to the existing models, given the strong evidence that those models suffer from serious misspecification problems.

B Statistical uncertainty around the estimate of efficient base year opex

Sources of statistical uncertainty

Efficient base year opex is estimated with uncertainty

167. The AER's methodology for determining an estimate of efficient base year opex relies on its econometric benchmarking models. The parameters in these models are estimated from data and are subject to statistical uncertainty.[43] As a consequence, the AER's base year opex target is also subject to statistical uncertainty.

168. However, when testing the efficiency of a DNSP's actual base year opex, the AER does not account for this statistical uncertainty. Notwithstanding the statistical uncertainty around its estimates of each DNSP's efficiency score and other parameters estimated using its econometric benchmarking models, the AER effectively treats its point estimate of efficient base year opex as certain or deterministic. The AER has clarified recently that if the DNSP's actual base year opex is higher than its point estimate of efficient opex, then it concludes that the DNSPs' actual base year opex is materially inefficient:

> *We use results from our econometric opex cost function benchmarking and our benchmarking roll forward model to derive an estimate of efficient base year opex, and compare this to actual base year opex, in order to determine whether there is an efficiency "gap" and of what size. Where modelled efficient rolled-forward base year opex is below actual base year opex, we infer that the latter is materially inefficient.*[44]

169. The AER does not allow for any tolerance limits or range of uncertainty around its point estimate of efficient base year opex. Rather, as the AER explains in the excerpt above, if its point estimate of efficient base year opex is lower than the DNSP's actual base year opex, then the AER concludes that the latter is materially inefficient.

170. In our view, this is a serious shortcoming in the AER's approach. The statistical uncertainty involved in estimating a DNSP's efficiency and the elasticities and other parameters specified in the econometric benchmarking models can be very material. This uncertainty means that the AER does not *know* the *true* level of efficient base year opex for a particular DNSP with certainty. Instead, the true level of efficient base year opex lies within a range of uncertainty that is defined by (amongst other factors) the statistical error involved in estimating:

    a the true level of average efficiency of a DNSP over the historical benchmarking period;

[43] Statistical uncertainty refers to the 'spread' of estimates of a parameter around its true (unobserved) value. In statistics, statistical uncertainty is measured by the standard error of the estimates of the parameter in question. See Gujarati and Porter, *Basic Econometrics* (5th Edition), 2009, p. 69.

[44] AER, *Draft decision – Evoenergy electricity distribution determination 2024 – 29, Attachment 6*, September 2023, p. 23.

b the true relationship between a DNSP's opex and outputs; and

c the true values of other parameters specified in the AER's econometric models.

171. These uncertainties contribute to the overall uncertainty surrounding the AER's estimates of efficient base year opex. Failure to account for these uncertainties when assessing efficient base year opex could result in the AER concluding erroneously that the DNSP's actual base year opex is materially inefficient simply because it lies above the AER's *point estimate* of efficient opex—even if the DNSP's actual base year opex lies comfortably within a range of statistical uncertainty.

172. The AER has previously explained that it accounts for "uncertainties" and other limitations associated with its model by selecting a "conservative" benchmark comparison point of 0.75. For the reasons explained below, we disagree that this is an appropriate or adequate way to account for the statistical uncertainties described above.

173. In our view, an appropriate approach would be to quantify formally the statistical uncertainty around the AER's point estimate of efficient base year opex, by constructing confidence intervals around that estimate, and then using those confidence intervals to make a probabilistic assessment of the evidence for material inefficiency.

174. The remainder of this section presents a standard and well-accepted methodology for doing this.

Statistical uncertainty around estimates of efficiency scores

175. Information on the statistical uncertainty in the AER's estimates of the efficiency scores is provided in the Stata output files that accompany the Annual Benchmarking Reports.[45] The AER's previous adviser on benchmarking issues, Economic Insights, has noted that for the SFA models, information on the uncertainty is provided by confidence intervals around the estimated efficiency scores, and for the LSE models, it is provided by the asymptotic standard errors[46] for the coefficients of the dummy variables for the Australian DNSPs.[47]

176. We have extended these measures of the statistical uncertainty in the estimates of the DNSPs' efficiency scores to obtain confidence intervals around the estimate of efficient base year opex for Energex and Ergon Energy produced by each of the AER's econometric benchmarking models and opex roll-forward model.

177. The measures of statistical uncertainty provided in the AER's supplementary files only capture part of the statistical uncertainty of the estimated efficiency scores. For both the SFA and the LSE models, there are important additional sources of uncertainty of the efficiency score estimates that are not included in the measures provided by the AER.

[45] For example, the supplementary files for the 2022 Annual Benchmarking Report can be found in the Quantonomics folder for Distribution at: https://www.aer.gov.au/networks-pipelines/guidelines-schemes-models-reviews/annual-benchmarking-reports-2022/aer-position. The relevant Stata output files are "anOpexReg1-half.log" and "anOpexReg1-full.log".

[46] Quantonomics' LSE models are not standard linear regression models. For non-standard econometric models, it is sometimes hard to calculate the exact standard errors of some of the estimated parameters for finite sample sizes since the distribution of the uncertainty about the estimated parameter is complex. However, an estimate of the standard error can be obtained by assuming that the sample size becomes infinitely large, in which case the distribution of the uncertainty about the estimated parameter usually converges to the well-known normal distribution for which the standard error is easy to calculate. Standard errors estimated in this way are known as 'asymptotic standard errors.'

[47] Economic Insights, *Comments on 2019 Frontier Economics Benchmarking Reports for EQ*, Memorandum to the AER Opex Team, 11 March 2020, p.17.

178. For the SFA models, Quantonomics uses the Stata module "frontier_teci" to produce confidence intervals for the efficiency scores. These confidence intervals are calculated on the assumption that the estimates of the parameters in the truncated normal distribution for the efficiency scores are the true values of these parameters rather than estimates. Treating these values as estimates rather than true values adds to the uncertainty of the estimated efficiency scores.

179. We have been able to replicate results produced by the "frontier_teci" command using the Stata command "nlcom".[48] The "nlcom" command uses the same asymptotic approach as "frontier_teci", but it can be applied to more general algebraic expressions. In particular, "nlcom" can be used to produce asymptotic standard errors and confidence intervals for the estimates of the efficiency scores that take into account the additional source of uncertainty discussed above. Taking this additional source of statistical uncertainty into account can have a large impact on the width of the confidence intervals.

180. For the LSE models, the standard errors for the coefficients of the dummy variable referred to by Economic Insights, which are now produced by the AER's current consultant Quantonomics, do not take into account the fact that the estimated efficiency scores for this model are a function of the difference between the estimated coefficient of a DNSP's dummy variable in the model and the estimated coefficient of the most efficient DNSP.

181. When calculating the uncertainty around the difference between these two coefficients, it is not only the uncertainty in the estimated coefficient for a given DNSP's dummy variable that needs to be taken into account but also the uncertainty in the estimated coefficient of the most efficient DNSP. The Stata command "nlcom" can be used to take this additional uncertainty into account when calculating asymptotic standard errors and confidence intervals for the estimates of the efficiency scores. Taking this additional source of statistical uncertainty into account can have a large effect on the width of the confidence intervals.

Additional sources of statistical uncertainty around estimates of efficient base year opex

182. The AER's procedure for estimating base year efficient opex for each of the econometric benchmarking models involves:

    a estimating an efficient level of average opex over the relevant historical benchmarking period (i.e., the actual level of average opex over the period less the AER's estimate of any material inefficiency); and

    b rolling that efficient level of average opex forward to the base year using an annual rate of change.

183. The annual rate of change described in 182.b depends on the estimated elasticities (i.e., the coefficient on each of the output variables, the share of underground assets and the time-trend variable) in the Cobb-Douglas and Translog cost functions. All of these estimated elasticities are also subject to statistical uncertainty.

[48] Stata's "nlcom" command computes point estimates, standard errors, test statistics, significance levels, and confidence intervals for (possibly) nonlinear combinations of parameter estimates after any Stata estimation command using the delta method. The delta method is a standard statistical approach for obtaining estimates of the standard errors of non-linear combinations of parameters. See, for example Cramér, H. (1946), *Mathematical methods of statistics,* Princeton University Press.

184. When constructing confidence intervals for base year efficient opex, the Stata command "nlcom" can again be used to take into account the uncertainty of the estimates of these parameters.

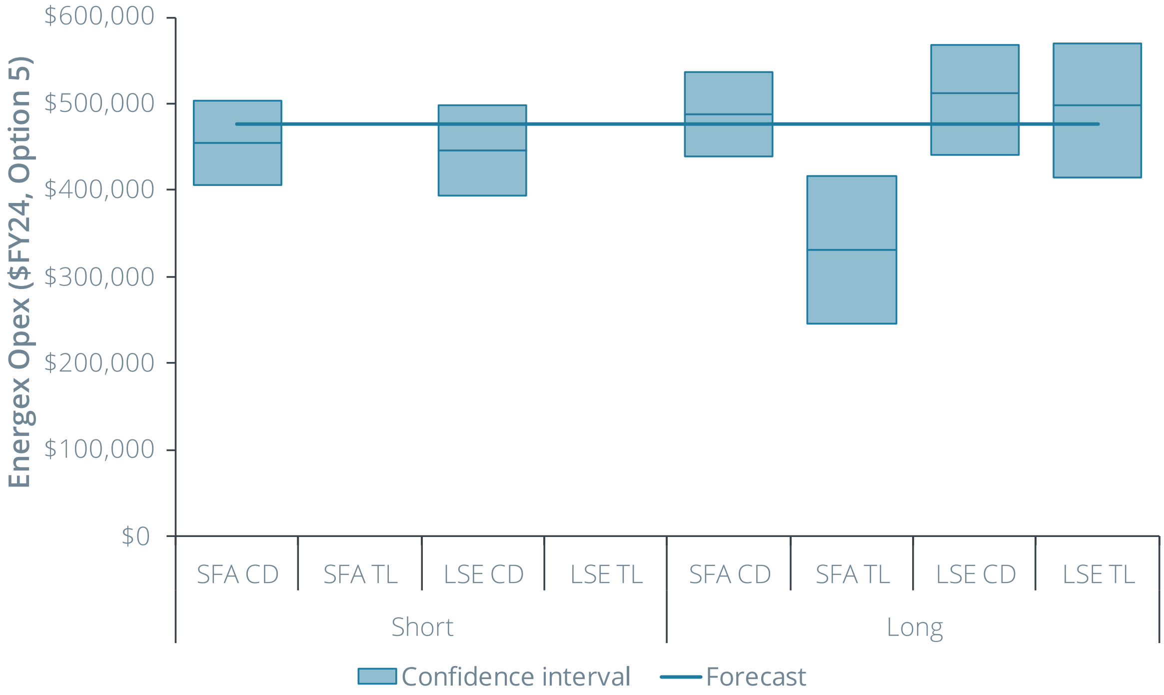
Construction of confidence intervals

185. We have converted the sources of statistical uncertainty discussed above into asymptotic standard errors. We then used those asymptotic standard errors to construct confidence intervals around the point estimate of efficient base year opex, for Energex and for Ergon Energy, using each econometric benchmarking model that is not excluded due to monotonicity violations.

186. We do this by specifying the steps in the AER's opex Excel files as algebraic equations and applying Stata's "nlcom" command. In doing so, we have taken into account constraints such as the fact that the AER restricts the target sample average opex never to be larger than the actual sample average opex.[49]

187. Figure 25 presents the 95% confidence intervals around the estimate of efficient base year opex for Energex, calculated using the approach described above.

Figure 25: Confidence intervals for efficient base year opex (including capitalised corporate overheads) – Energex

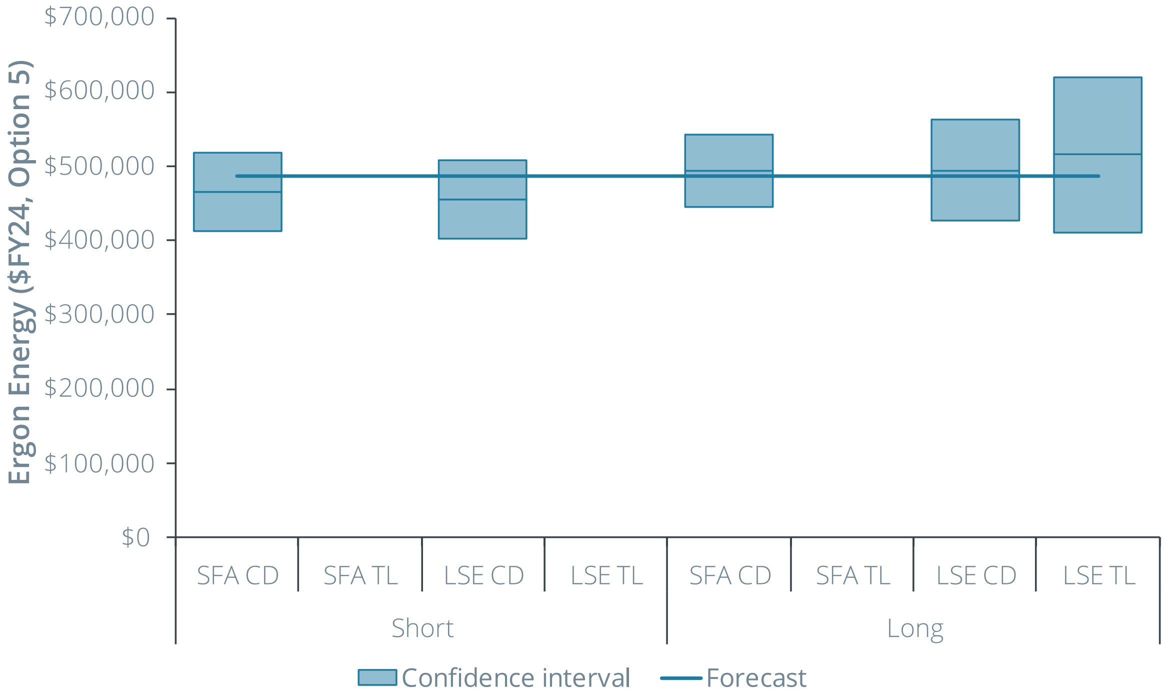


Source: Frontier Economics analysis using benchmarking data and models published along with 2023 Annual Benchmarking Report. Note: the central estimate is represented by the blue line at the centre of the confidence interval.

[49] For the LSE models we note that DNSP dummies, used to derive efficiency scores are random variables but the statistical uncertainty of these estimates is correlated with the error term for the relevant DNSP. To avoid 'double counting', we reconstruct the sample average opex for Energex and for Ergon Energy in the derivation of efficient sample average opex with respect to the regression estimates rather than as a fixed quantity. This consideration does not apply to the SFA models.

188. Figure 26 presents 95% confidence intervals around the estimate of efficient opex derived from each valid benchmarking model for Ergon Energy.

Figure 26: Confidence intervals for efficient base year opex (including capitalised corporate overheads) – Ergon Energy



Source: Source: Frontier Economics analysis using benchmarking data and models published along with 2023 Annual Benchmarking Report. Note: the central estimate is represented by the blue line at the centre of the confidence interval.

189. There are two types of uncertainty captured by the confidence intervals computed for each of the valid benchmarking models. The first type represents the uncertainty in the estimates of the efficiency scores discussed above, while the second type represents the additional uncertainty that arises due to the uncertainty in the estimated parameters of the Cobb-Douglas and Translog cost functions. Figure 25 and Figure 31 present 95% confidence intervals around the estimate of efficient opex derived from each valid benchmarking model for Ergon Energy, representing the full range of statistical uncertainty around the point estimate of efficient opex.

190. Figure 25 shows that for five of the six valid models, Energex's revealed opex lies within the 95% confidence surrounding the estimate of efficient base year opex. Figure 26 presents 95% confidence intervals around the estimate of efficient opex derived from each valid benchmarking model for Ergon Energy.

191. Figure 26 shows that for all five of the valid models, Ergon Energy's revealed opex lies within the 95% confidence surrounding the estimate of efficient base year opex.

192. This suggests that once the statistical uncertainty over the AER's estimates of:

   a the efficiency scores for the two DNSPs; and

   b the elasticities and the other parameters in the Cobb-Douglas and Translog cost functions are taken into account, one must conclude that there is no evidence that Energex's or Ergon Energy's revealed base year opex is materially inefficient.

Interpretation of confidence intervals

193. When we have previously proposed the use of confidence intervals to assess the efficiency of a DNSP's base year opex, the AER's adviser at the time, Economic Insights, argued that:

   a Regulators do not use confidence intervals to determine a range of efficient costs; and

   b Even if a regulator were to construct confidence estimates around a point estimate of efficient opex, it does not follow that all values within the confidence interval should be interpreted as being efficient.

194. Specifically, Economic Insights stated that:

> *Finally, FE (2019b) appears to argue that where the proposed opex sits within the confidence interval, then there is no evidence of material inefficiency. However, in regulatory applications, the confidence interval has not been used to set range of possible efficient values. Rather, it is a statistical construct used to estimate precision of the point estimate (eg the width of the confidence interval and the precision of the point estimate will generally be negatively related to the sample size). The point estimate provides the best estimate about the unknown true efficient value, while none of the other values within the confidence interval do. Confidence intervals may be useful in informing the degree of confidence in the point estimate, and thus the weights to apply to the estimate when multiple estimates from different sources/methods are available. They do not mean that all values within the confidence interval can be viewed as being efficient.*[50]

195. We address each of the two points above by Economic Insights in turn.

196. The first point Economic Insights makes is that regulators do not use confidence intervals to determine a range of possible efficient values. This is a misunderstanding of how we proposed that the confidence intervals should be used in the AER's particular context. As Economic Insights notes correctly, a confidence interval is a statistical construct used to assess the precision of a point estimate, and "may be useful in informing the degree of confidence in the point estimate." That is exactly how we propose confidence intervals should be used and interpreted by the AER.

197. The AER derives, with statistical uncertainty, a point *estimate* for the efficient level of DNSP opex in a base year. It would be wrong to simply assume away that statistical uncertainty and proceed as though the point estimate were the true level of efficient opex. Placing confidence intervals around a point estimate simply makes transparent the range of statistical uncertainty around that point estimate. That is precisely what we suggest the AER should do.

198. However, Economic Insights hints at another point—namely that regulators do not typically put confidence intervals around forecasts of efficient opex and, therefore, the novelty of doing so, in this case, should rule it out as a valid approach for the AER to take. If the novelty of a regulatory

[50] Economic Insights, *Comments on 2019 Frontier Economics Benchmarking Reports for EQ*, Memorandum to the AER Opex Team, 11 March 2020, p.19.

approach is sufficient to invalidate it, then the AER's entire approach to economic benchmarking should be discarded. No other regulator in the world performs economic benchmarking in the way the AER does. For example, no other regulator:

a uses the same econometric models employed by the AER; or

b accounts for OEFs in the way the AER does; or

c rolls forward an estimate of efficient opex to a base year in the way the AER does;

d and so on.

199. In our view, it is wrong to suggest that the AER should eschew an approach simply because it is not common regulatory practice. The usefulness of a particular approach should be judged on its own merits.

200. It is also important to recognise that the AER's use of econometric benchmarking models provides the statistical information required to construct confidence intervals in this particular case. Economic Insights itself acknowledges this.[51] Such information is not always available to other regulators. We do not see why the AER should discard such information if it can be useful in making a more informed decision about the efficiency of a DNSP's base year opex.

201. The second major point that Economic Insights makes is that the fact a DNSP's actual base year opex lies within a confidence interval does not mean that level of opex is efficient. This is a misrepresentation of how we say the AER should use confidence intervals.

202. The AER's approach is to compare a DNSP's *actual* base year opex to an *estimate* of efficient base year opex, where that estimate is derived using statistical analysis. If the former is greater than the latter, then the AER concludes that the DNSP's actual base year opex is materially inefficient.

203. We say that if the DNSP's actual base year opex lies within the confidence interval, then the AER cannot reject the possibility that there is no difference between a DNSP's revealed level of actual base year opex and the efficient level of base year opex—because the latter can only be estimated with statistical uncertainty, and the former lies within the range of statistical uncertainty.

204. We cannot conclude from such evidence that a DNSP's revealed base year opex is efficient. However, it would be legitimate to conclude that there is no evidence of material inefficiency.

205. The way we have suggested that confidence intervals be used in this context is entirely consistent with standard hypothesis testing.

Use of a conservative comparison point to deal with "uncertainties"

206. The AER has suggested that it accounts for general limitations associated with its econometric benchmarking models by selecting a conservative benchmark comparison point (75% before any adjustments for OEFs) rather than comparing each DNSP to (what the AER estimates to be) the most efficient DNSP. For example, the 2023 Annual Benchmarking Report states that:

[51] Economic Insights, *Comments on 2019 Frontier Economics Benchmarking Reports for EQ*, Memorandum to the AER Opex Team, 11 March 2020, p.17.

> *we consider our benchmarking comparison point is conservative and provides a margin for general limitations of the models. This includes with respect to the specification of outputs and inputs, data imperfections, other uncertainties when forecasting efficient opex and quantification of OEFs.*[52]

207. Neither the AER nor its advisers have been explicit (beyond statements similar to the one above) whether the margin between the efficiency estimate of the most efficient DNSP and the benchmark comparison point of 75% is designed to account for statistical uncertainty. The 75% comparison point was selected by the AER using regulatory judgment and on the advice of its previous adviser Economic Insights, who recommended it as a cutoff point for identifying the reference DNSPs.[53]

208. If the benchmark comparison point did account properly for the statistical uncertainty associated with estimating the efficient level of base year opex, then it would be useful to know exactly how much of the margin between the 75% benchmark comparison point and the estimated efficiency score of the most efficient DNSP (the 'margin for uncertainty') accounts for:

    a The statistical uncertainty involved in estimating the efficient level of base year opex; and

    b All the other general limitations associated with the benchmarking models that are distinct from and unrelated to statistical uncertainty, including (but not necessarily limited to):

        i uncertainty around the true form of the opex cost function (sometimes referred to as 'model uncertainty' in the economic literature)—i.e., whether the true functional form is something other than the Cobb-Douglas or Translog specifications;

        ii uncertainty about whether the true outputs of the DNSP have been identified and included properly in the models;

        iii limitations and imperfections in the data used to perform the benchmarking analysis;

        iv the scope for important OEFs that have not been accounted for at all, or not quantified and incorporated properly into the analysis; and

        v shortcomings in the process for rolling forward the estimate of efficient opex to the base year.

209. For the purposes of the remaining discussion, we refer to the examples of the uncertainties listed in paragraph 208.b as 'other uncertainties', to distinguish them from statistical uncertainty.

210. If the benchmark comparison point is indeed intended to account for statistical uncertainty, then, as we show below, it is possible to calculate how much of the margin for uncertainty allows for statistical uncertainty. Whatever is left over, therefore, must account for all other uncertainties. We can then consider whether the portion of the margin for uncertainty that does not account for statistical uncertainty would plausibly be sufficient to account for the other uncertainties.

[52] AER, *Annual Benchmarking Report, Electricity distribution network service providers*, November 2023, p. 83.

[53] Economic Insights, *Economic Benchmarking Assessment of Operating Expenditure for NSW and ACT Electricity DNSPs*, November 2014, p. 47.

211. To calculate how much of the margin for uncertainty must allow for statistical uncertainty, we first derive the 95% confidence interval for efficient base year opex, assuming a benchmark comparison point of 75%. We then calculate the estimate of efficient base year opex that would allow the AER to be 95% confident that it had not underestimated the true (unobservable) level of efficient base year opex, given the statistical uncertainty associated with its benchmarking models. That estimate of efficient base year opex is simply the upper bound of the 95% confidence interval.

212. If we were to perform this calculation Energex and for Ergon Energy, then the resulting estimate of efficient base year opex (i.e., that accounts fully for the statistical uncertainty associated with the benchmarking models applicable to the two DNSPs, as measured by the 95% confidence interval around the AER's point estimate of efficient base year opex) would be:

   a $512.0 million ($FY2024) for Energex; and

   b $544.0 million ($FY2024) for Ergon Energy.

213. These values are well above the estimates of efficient base year opex produced using the AER approach: $454.2 million and $481.0 million for Energex and Ergon Energy, respectively.

214. In order for the upper bounds of the confidence intervals to equal the point estimate for efficient base year opex as per the AER process, an efficiency target of 85.1% would be required for Energex and an efficiency target of 84.9% would be required for Ergon Energy.[54]

215. According to the benchmarking analysis presented in the 2023 Annual Benchmarking Report, the most efficient DNSP (Powercor) had an average efficiency estimate of 98.1%. That is:

   a The total margin for uncertainty implied from the 2023 Annual Benchmark Report was 23.1% (i.e., 98.1% – 75.0% = 23.1%).

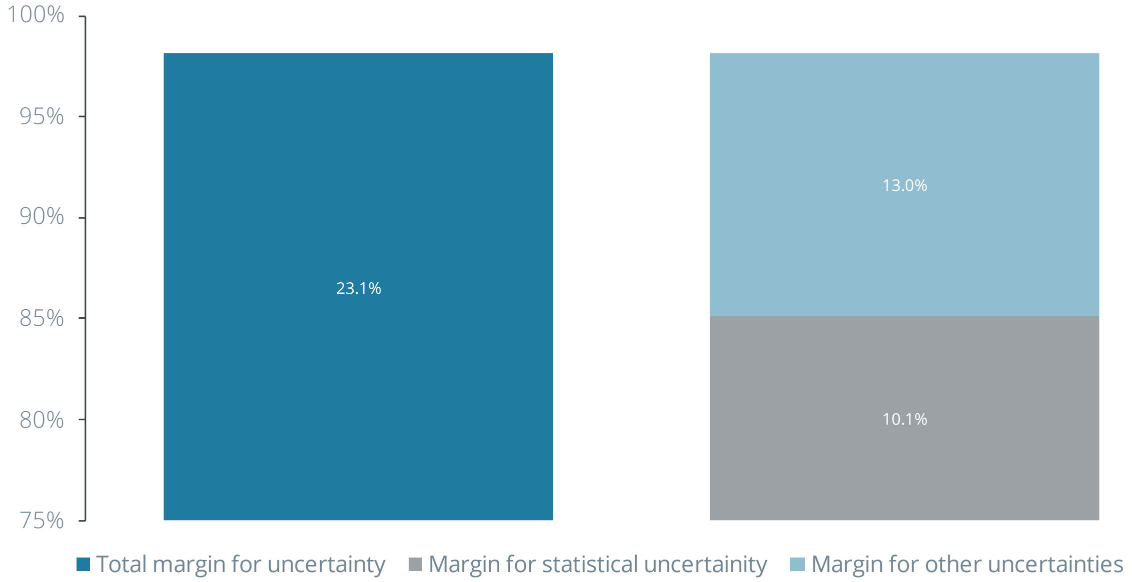
   b Of the total margin for uncertainty, 10.1% (i.e., 85.1% – 75.0% = 10.1%) would be required in order to allow properly for the statistical uncertainty around the estimate of efficient base year opex (9.9% for Ergon Energy).

   c That means that only 13.0% (i.e., 98.1% – 85.1% = 13.0%) would be left to account for all of the other uncertainties for Energex (13.3% for Ergon Energy).

216. This is illustrated in Figure 27 and Figure 28 below.

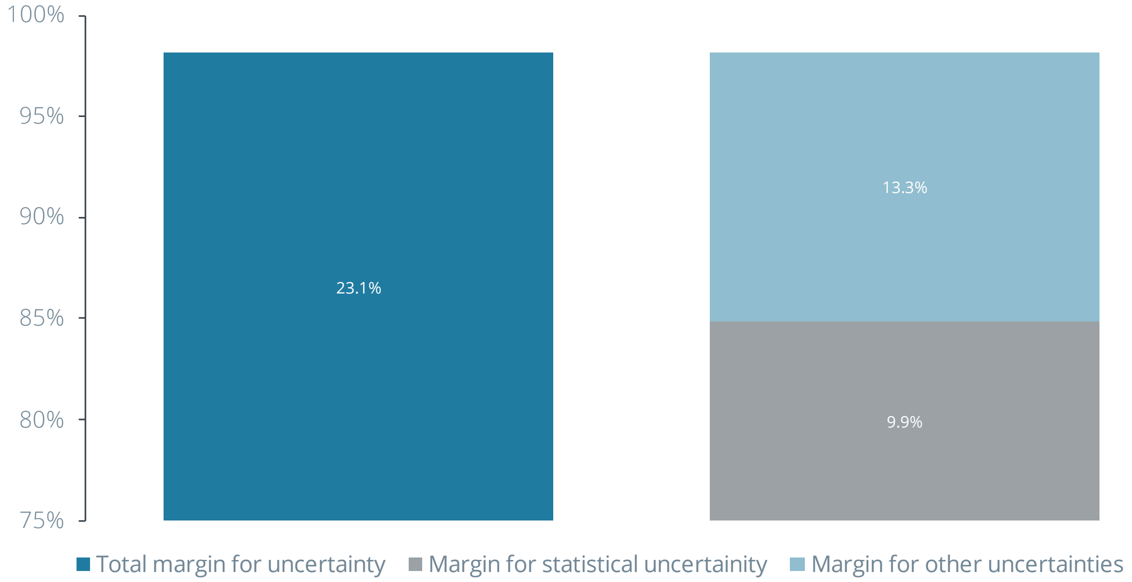
[54] Setting a target of 85.1% (rather than 75%) yields a confidence interval for efficient base year opex with an upper bound of $454.2 million for Energex; setting a target of 84.9% (rather than 75%) yields a confidence interval for efficient base year opex with an upper bound of $481.0 million for Ergon Energy.

Figure 27: Decomposition of margin for uncertainty into allowance for statistical uncertainty and other uncertainties for Energex



Source: Frontier Economics analysis

Figure 28: Decomposition of margin for uncertainty into allowance for statistical uncertainty and other uncertainties for Ergon Energy



Source: Frontier Economics analysis

217. Whilst the AER describes the benchmark comparison point it has selected (and, therefore, the margin for uncertainty) as "conservative", in our view, the existing comparison point of 75% does not allow properly for the significant statistical uncertainty associated with the estimate of efficient base year opex. This is because it is implausible that, once statistical uncertainty has been

accounted for properly, the remainder of the margin for uncertainty—just 13.0%—would be adequate to account for all of the other uncertainties, including model uncertainty, uncertainty over the true outputs of the DNSP, data limitations and imperfections, OEFs that have not been accounted for properly, shortcomings in the roll-forward process and other modelling limitations.

218. We conclude from this that the AER's benchmark comparison point of 75% does not account for statistical uncertainty properly; the allowed margin of uncertainty is simply too narrow for that to be so. Therefore, the AER should explicitly allow for statistical uncertainty associated with its estimate of efficient base year opex by quantifying confidence intervals around its point estimate of efficient base year opex. We have developed and applied a simple, standard procedure for doing so. Our method makes use of the information about the degree of statistical uncertainty around key estimated parameters, obtained directly from the AER's benchmarking models. Such an approach would be a more reasonable and transparent way to account for statistical uncertainty than via the benchmark comparison point.

Frontier Economics

Brisbane | Melbourne | Singapore | Sydney

Frontier Economics Pty Ltd
395 Collins Street Melbourne Victoria 3000

Tel: +61 3 9620 4488
https://www.frontier-economics.com.au

ACN: 087 553 124 ABN: 13 087 553 124