Electricity Distribution Benchmarking Opex Model Development

Report prepared for Australian Energy Regulator

26 November 2025

Michael Cunningham, Valentin Zelenyuk, Alice Giovani and Joe Hirschberg

Quantonomics Pty Ltd
www.quantonomics.com.au
ABN: 23 165 107 844
michael@quantonomics.com.au

Executive Summary

This report builds on the AER's earlier consultation on the memorandum *Electricity Distribution Opex Cost Function: Potential Misspecification Issues* (Quantonomics, 2024b). This consultation addresses the performance of econometric opex cost function Translog models, particularly the high frequency of monotonicity violations observed in recent annual benchmarking reports.

The previous memo (hereafter 'Phase 1') examined jurisdiction-specific time trends and found that while their inclusion improves model performance, it does not fully resolve the issue of excessive monotonicity violations. This indicates that such trends may only partially address the underlying limitations of the models. Two other issues remain potentially significant: the omission of relevant cost drivers, and the assumption, highlighted in Phase 1, that each DNSP's inefficiency is constant over time.

This report ('Phase 2') conducts more comprehensive analysis of these issues. It considers refinements to model specification and estimation techniques that could improve performance of opex cost function models; and approaches to incorporating time-varying inefficiency, and decomposing time trends into technical change and shifts in DNSP opex cost efficiency.

Opportunities for refining the current models are identified by considering best practices from benchmarking literature and applications of various frontier methods. Methods for introducing time-varying inefficiency are identified in the context of a broad review of stochastic frontier analysis (SFA) methods. This review identifies a number of potential approaches, and the criteria for evaluating the performance of alternative econometric models are clearly set out.

The review of options to refine current models addresses the following avenues:

- Findings in the literature suggest replacing the assumption in the standard SFA models that inefficiencies have a truncated-normal distribution, with an assumed half-normal distribution.[1] The latter has computational advantages and should remedy problems of non-convergence in the maximum likelihood procedure, which often arise for the SFA models with the more general truncated-normal distribution that often suffer from identification problems. It is found that using the half-normal distribution of inefficiencies does not substantially improve the MV problem, but this is not its purpose. It offers a valuable improvement by enhancing the computational stability of

[1] Assumptions about the distribution of inefficiency are necessary to estimate the standard SFA model. The most common distributional assumptions are the truncated-normal and the half-normal distributions. The truncated-normal assumes that inefficiency values are normally distributed about the pre-truncation mean, and restricted to non-negative values. The half-normal assumes inefficiency values are normally distributed about the pre-truncation mean of zero, and again only non-negative values are allowed. That is, the half-normal is a special case of truncated-normal where the pre-truncation mean is set to zero.

SFA models, without materially affecting the estimated output weights or efficiency scores.

- The standard LSE models are closely related to the frequently used Corrected Ordinary Least Squares (COLS) frontier technique. In the panel data context, the COLS approach uses fixed effects panel data regression,[2] which produces an efficiency estimate for all firms in the sample, unlike the standard LSE model, which includes fixed effects-based efficiency measures only for the Australian DNSPs. When the COLS fixed effects method is tested as an alternative, it is found that the precision and reliability of the main parameter estimates is detrimentally affected, and efficiency scores are unreliable. This suggests that much of the sample data variation relied on to identify the main parameters of the model comes from variation between firms, rather than from within-firm variation over time. This is due to the large degree of variation in the scales of DNSPs. When the fixed effects are introduced, much of the cross-sectional variation is absorbed by these fixed effects, potentially jeopardizing the reliability of the estimates. This approach does not represent an improvement over the standard LSE method.
- Whereas the TFP index analysis includes five outputs, the standard opex cost function includes only three. This raises the question whether the other two outputs, energy delivered and customer minutes off-supply (CMOS), are omitted variables. The inclusion of these variables is tested. Models that include energy delivered perform worse than the standard models in several respects. This tends to suggest that its inclusion results in excessive multicollinearity with the other outputs. CMOS is found to be statistically significant, but it has little effect on the model overall and does not mitigate any of the modelling difficulties. More importantly, it raises concerns about endogeneity because higher levels of routine (preventative) maintenance should reduce the frequency of outages. Therefore, it appears to be impractical to include either of these other two outputs into the SFA models with the available data.
- Cost functions typically include input prices. In the AER's standard opex cost function, they are dealt with by dividing opex by an index of opex input prices. This leaves open the possibility that there may be substitutability between capital and opex inputs that remains unaddressed. If so, the relative price of capital and opex inputs may be a relevant explanatory variable. This variable is tested and it is found that the log ratio of input prices has a statistically significant, if small, effect. Its inclusion may yield some marginal improvements. However, there remain measurement issues in relation to the price of capital inputs, for example relating to cost of capital parameters and

[2] The fixed effects panel data approach controls for unobserved characteristics that are constant over time for each firm, thereby reducing the risk of bias in the estimates caused by omitted firm-specific time-invariant factors.

depreciation rates, which would need to be adequately explored and resolved before reliably using it in the SFA models.

The central focus and findings of the report relate to the application of time-varying models. A wide variety of different approaches are tested, from relatively simple to more complex. The results confirm the challenges associated with maximum likelihood estimation of more complex SFA models. Often, while the more sophisticated SFA models are theoretically appealing, they often present difficulties in terms of estimation stability and optimization. This proved to be also the case for the most flexible of the models.

Among the models that were successful, the main alternative methods of allowing for time-varying inefficiency are by (a) using different jurisdictional time trends, with the same trends applying to all Australian DNSPs, and (b) using separate trends for each Australian DNSP plus each overseas jurisdiction.[3] The latter are generally more difficult to estimate than the former.

In the SFA models explored in this study, the time-varying inefficiency effects are introduced either:

- as determinants of the mean of the inefficiency distribution, specifically of the 'pre-truncation mean' of the stochastic distribution of inefficiencies in models based on the Battese and Coelli (1995) SFA specification;
- as shifts in the variance of the inefficiency distribution via a scaling function applied to a cross-sectionally stochastic distribution of inefficiencies, in the Kumbhakar (1990) SFA approach; or
- as additional explanatory variables directly in the opex cost function, specifically interactions between time trends and jurisdictional or DNSP-specific dummy variables, using the least squares econometrics (LSE) approach.

It is found that only one model variant using the Battese-Coelli (1995) method was successful, and at least two of the Kumbhakar (1990) and the LSE models were successful. Most of these models show meaningful improvements compared to standard applications, including a substantial reduction in monotonicity violations and producing time-profiles of efficiency scores that are well correlated with efficiency scores calculated from the multilateral Opex partial factor productivity (PFP) index.

However, the time-varying stochastic frontier models are more complex, with computational challenges in estimation, particularly when applied to shorter samples. They require sufficiently large and relatively rich (in variation) dataset so that short-period analyses are less suitable in this context. The reliability of the models also tends to decrease as the models

[3] Note that these approaches differ from those presented in the Phase 1 Memorandum. In the present analysis, inefficiency is allowed to vary over time, whereas in the Phase 1 models, although the time-trend term was permitted to differ between jurisdictions, the cost efficiency of DNSPs was assumed to remain constant across periods.

become more complex. This means there is a trade-off between the flexibility of estimating efficiency trends for each Australian DNSP, which increases risk of computational problems in estimation, versus models with more restrictive assumptions about the efficiency time profiles, which are computationally more robust. Some suitable simplifications, such as assuming a half-normal instead of truncated normal inefficiency distribution, assist computationally, including with convergence. Unlike the SFA models, the LSE time-varying models proved to be more robust in estimation and can reliably include DNSP-specific efficiency trends.

A comparison of the estimated efficiency trends produced by several feasible time-varying inefficiency models confirms that, as expected, a degree of uncertainty exists in the efficiency scores estimated by these models at any point in time. This is due to differences in the time-patterns of efficiency estimated using different models. We suggest that extending the current approach to include two different SFA models alongside an LSE model (rather than only one SFA and one LSE model, as currently applied) may improve the reliability of the estimated efficiency scores.

Unlike the standard current models, the time-varying models presented in this report allow for the decomposition of productivity changes into changes in efficiency, technical change, and other factors. They can also enable the efficient base-year opex to be calculated directly if the base year is the same as the last year of the sample period. This can remove the need to rely on the benchmarking roll-forward model (BRFM) framework or reduce its role to bridging between the last year of the sample and the base year.

We also show that the BRFM provides a means to mitigate such uncertainties and strengthen the robustness of the results. The BRFM can still be applied to the period-average efficiency scores produced by time-varying inefficiency models, and an estimate of the efficiency score at the end of the sample period can be derived from it. This provides additional estimates of the period-end efficiency score, which are arguably not constrained by restrictive assumptions relating to the efficiency time profile. Hence, the period-end efficiency score of each DNSP can be obtained by averaging over both the model period-end estimates and the BRFM period-estimates.

Contents

1 Introduction

1.1 Purpose

The AER's econometric opex cost function Translog models have exhibited somewhat declining performance in terms of the frequency of monotonicity violations (MVs) in recent annual benchmarking reports (ABRs). On 26 November 2024, the AER commenced a consultation on improving the econometric opex model used in benchmarking and released Quantonomics' memorandum *Electricity Distribution Opex Cost Function: Potential Misspecification Issues* ("Phase 1 memo").

The Phase 1 memo examined potential causes of MVs in the current models, with a particular focus on the treatment of time trends. In the current AER models, the time-trend coefficient – notionally included to capture technical change over time – is assumed to be identical across all three jurisdictions. In practice, the trend may capture a mix of technical change (frontier shift), average efficiency "catch-up", and any systematic movements linked to omitted operating-environment factors (OEFs). Forcing the same trend on every jurisdiction could be a likely source of misspecification. In particular, while it may be reasonable to assume that technical change is the same or very similar across jurisdictions, trends in efficiency "catch-up" and OEFs can be expected to differ somewhat.[4] The fact that model performance has declined as more years of data have been added is at least partly because differences between jurisdictions have been accumulating over time.

To test this hypothesis, the Phase 1 memo explored the use of jurisdiction-specific time trends and found that these specifications improved MV performance but did not completely resolve the problem of excessive MVs. This suggests that the inclusion of jurisdiction-specific trends may have only partially addressed the underlying problems in the models. One potential issue is the existence of omitted relevant cost drivers. When important cost drivers are excluded, the effects may be absorbed by other included variables, leading to biased parameter estimates and, ultimately, could contribute to monotonicity violations.

Another possible limitation of the current models, emphasised in the Phase 1 memo, that may be contributing to omitted variable issues is the assumption that each DNSP's inefficiency remains constant over time.[5] As discussed, *implicitly* the generic time trend includes the average rate of efficiency catch-up across all DNSPs in all jurisdictions in the sample. Time-invariant inefficiency and/or uniformity of the rate of efficiency catch-up is an important restriction. While this assumption may have been reasonable at the start of the AER's

[4] Technical change is expected to be similar across jurisdictions, as the electricity sectors in Australia, Ontario, and New Zealand generally use comparable technologies. However, catch-up rates may differ, reflecting variations in regulatory or other cost pressures to improve efficiency and in OEFs arising from jurisdiction-specific conditions.

[5] Changes in the efficiencies of Australian DNSPs are currently accommodated within the benchmarking roll-forward model, discussed in section 6.

benchmarking program, evidence of an upward trend in the opex partial factor productivity indexes of most Australian DNSPs since 2015 suggests that their opex inefficiency has changed, and specifically, reduced in the latter half of the sample period. A model that allows for jurisdiction-specific time-varying inefficiency is likely to provide a more accurate and robust assessment of DNSP performance.

As shown in the Phase 1 memo, jurisdictional time trend models which do not decompose the effects of technical change, efficiency changes or the effects of changes in OEFs not modelled, appear sufficient for measuring the opex cost efficiencies of Australian DNSPs at the mid-point of the sample. This is because they achieve a substantial reduction in the Australian monotonicity violations (MVs). However, our preference, and as discussed in section 1.2, that of stakeholders is for an approach that estimates separate trends in efficiencies, with the expectation that it will further improve MVs. This allows efficiency to be measured for each Australian DNSP in each sample year. Hence, models that can decompose, for Australian DNSPs, efficiency change from the other effects are needed.

The present Phase 2 analysis aims to address this issue. In particular, this report considers:

- aspects of specification or estimation technique, with the potential to improve the performance of the opex cost function models; and
- methods for decomposing time trends into technical change and the effects of changes in the opex cost efficiency of DNSPs.

This report does not explore available data and methods for incorporating a wider range of OEFs into the opex cost function model. That would be a separate exercise, which may be conducted at a later date.

1.2 Summary of submissions to Phase 1 memo and our response

In response to the Phase 1 memo, seven submissions were received, all from DNSPs: Ausgrid, AusNet, Endeavour Energy, Energy Queensland (Ergon and Energex), Essential Energy, Evoenergy, and Jemena. This section discusses the main points made in the submissions.

Most stakeholders welcomed the consultation and saw it as necessary to address specification issues with the current model, particularly by allowing for time-varying inefficiency. Five of the seven submissions (Energy Queensland, Ausgrid, Evoenergy, Jemena and Endeavour) emphasised that the review process should not be rushed. Endeavour also suggested that the benchmarking models be reviewed periodically via a consultation process. On the other hand, Essential Energy was concerned about the number of changes in benchmarking methodologies over recent years, which have impacted rankings.

The Phase 1 memo presented models with separate jurisdictional time trends (JTT) and with a separate Australian time trend (ATT). The following views were expressed about these models:

- AusNet supported the inclusion of jurisdictional time trend variables and regarded the JTT model as superior to the ATT and standard models. However, AusNet considers it an incomplete improvement due to its assumption of time-invariant inefficiency. AusNet suggested that methodologies like Battese & Coelli (1992) or Cuesta (2000) should be explored, but it recognises there will be methodological and identification challenges, especially when testing DNSP-specific time trends.
- Ausgrid, Endeavour, Energy Queensland and Evoenergy considered the ATT and JTT models useful in highlighting misspecification issues, but opposed their adoption. In their view, the assumption of time-invariant inefficiencies is a key misspecification and models addressing the issue of time-varying efficiency should be explored, rather than adopting the JTT models.
- Jemena warns that the JTT and ATT models misinterpret catch-up efficiency as technological progress, potentially overstating true frontier shift. Jemena recommends exploring ways to separate catch-up efficiency from genuine efficiency frontier shifts before modifying the current model specification. Likewise, Energy Queensland states that if the benchmarking models are properly specified, the time trend variable should only reflect the rate of technological progress.
- Endeavour also did not support the JTT model because "networks previously regarded as efficient are now materially inefficient, where the difference in outcomes may be due to model-specific shortcomings". However, in our view, this outcome is not unique to the JTT model and may be true of any change to model specification. We discuss this point further in Section 2.3.6.

In summary, submitters prefer models that separate changes in efficiency from technical change. The common theme was that rates of change in efficiency should be allowed to vary between jurisdictions, and if feasible, allowed to vary between Australian DNSPs. On behalf of Evoenergy, Frontier Economics suggested a specific model for separating technical change and technical efficiency change, which it implemented using the Stata ado-program *sfpanel* (Belotti et al., 2012). Frontier Economics emphasised that this model was only an example, and it was not advocating this specific model. In this report, we investigate models with time-varying inefficiency and the separation of technical change from changes in efficiency, including a model along the lines of Frontier Economics' suggestion.

Frontier Economics also recommended reducing the data sample period, citing concerns about a structural break in the data. In our view, structural breaks can be accommodated within the model using appropriately specified time trends. Whether or not it is necessary or convenient to reduce the sample period is a question that is considered in this review. Related issues that are also addressed include whether there may be merit in considering aligning the short sample period with the Australian structural break point, and whether or not to continue using the short sample period if the econometric models incorporate time-varying inefficiency.

Frontier Economics also expressed a concern with the inclusion of New Zealand and Ontario data. Economic Insights (2015) emphasised that the benefit of using the overseas data is to improve the precision of the parameter estimates, a view we maintain. The jurisdictional dummies ensure that the efficiency of Australian DNSPs is only measured against the most efficient Australian DNSPs. We are not measuring the efficiency of overseas DNSPs. Provided we adequately control for systematic differences of the two overseas jurisdictions, they will not have a significant effect on the comparative efficiency of Australian DNSPs. However, it is important to acknowledge that while jurisdictional dummies can control for unobserved time-invariant heterogeneity, they may not fully capture differences in the rates of efficiency change over time. To the extent that these trends differ systematically across jurisdictions, they may introduce residual bias not addressed by the dummy variables alone. This issue is considered in this report.

Some of the other views expressed in submissions are as follows.

- On the question of how monotonicity violations (MVs) should be measured, Jemena rejected an approach of focusing on statistically significant MVs. Jemena proposed that a more stringent definition be developed to include not only MVs, but also other data points close to an MV. We consider that since the theoretical requirement is that the elasticities of opex with respect to each output should all be non-negative, it is not clear why a more stringent criterion should be adopted and what that criterion would be. We have not seen Jemena's method in the literature, but we have seen less stringent monotonicity tests (compared to the AER method) applied at a subset of the sample space (Ryan and Wales, 2000).
- Essential Energy, Endeavour and Ausgrid all expressed the need for clarity in how a model with jurisdictional time trends or time-varying inefficiency would be used in the benchmarking roll-forward model (BRFM). Endeavour suggested that the BRFM be reviewed in this consultation process. This is discussed in section 6.
- Some submitters, including Endeavour and AusNet, suggested further investigation into operating environment factors (OEFs) that could be measured and included in the model. Suggested OEFs include the relative importance of bushfire prone areas, terrain and storm-related variables. We agree that OEFs can play a role in the econometric models. This issue will be addressed in future development work, as it is likely to require the development of new metrics (e.g., using geographic information system (GIS) data) which is beyond the scope of this report.
- The Phase 1 memo discussed Ontario circuit length data. AusNet agreed that there is no systematic evidence of measurement bias and supported ongoing data validation in future benchmarking studies to ensure circuit length remains a reliable variable in cost modelling. We agree that the benchmarking data must be thoroughly validated each time it is updated.

1.3 Outline of the Report

The remainder of this report is organised as follows:

- Section 2 outlines the directions of investigation adopted in this study. Section 2.1 identifies opportunities for refining the current models and explains the rationale for each proposed refinement. Section 2.2 reviews the literature on time-varying model options, discussing their challenges and potential benefits. Section 2.3 sets out the criteria used to evaluate econometric models.
- Section 3 presents the results of testing the refinements proposed in Section 2.1 and assesses the extent to which they can improve the current models.
- Section 4 examines the feasibility of the most promising time-varying models surveyed in Section 2.2, along with the specification presented as an example by Frontier Economics.
- Section 5 evaluates the time-varying models against the criteria outlined in Section 2.3, comparing their performance with one another and with the standard models. This assessment considers whether the time-varying specifications deliver meaningful improvements over the current framework and identifies any approaches that may warrant adoption based on the established criteria.
- Section 6 demonstrates the application of time-varying models in a regulatory context, focusing on their integration into the Benchmarking Roll Forward Model (BRFM).
- Appendix A provides background on the development of the current models, including past decisions on model specifications. Appendices B and C present additional analysis of the time-varying models. Appendix B examines the Pearson correlations between total output elasticities and DNSP size, providing further insights into models' performance. Appendix C examines the Pearson correlations between the efficiency scores of the standard models and the time-varying models, showing the extent to which the time-varying models align with the ABR24 efficiency scores of the DNSPs.

Separate from this report, Attachments A and B present modelling results in more detail.

2 Directions of Investigation

This chapter discusses stochastic frontier analysis (SFA) methodologies that could be applied for the purpose of improving the current model specifications. In this section the term 'SFA' is used more broadly to include the AER's LSE model. To properly consider improvements to the current model, an understanding of the development of the current approach is important. Appendix A summarises Economic Insights' reasoning behind the current modelling approach.

Section 2.1 presents a number of observations on potential improvements or refinements to the current opex cost function models. Section 2.2 provides a survey of SFA methods including more advanced time varying inefficiency econometric frontier models.

As outlined in the Section 1, the time-invariant inefficiency specification may be too restrictive, given the extended time period covered in the more recent annual benchmarking reports. The available evidence seems to indicate that there has been considerable improvement in opex efficiency since the AER's current benchmarking policies were introduced. This hints that models which allow inefficiency to vary over time might be more adequate. Section 2.2 therefore examines methodological approaches for incorporating more flexibility in the time patterns of inefficiency in SFA. It then focusses on methods that can be used to separately estimate technical change and time-varying inefficiency, especially those allowing time-varying inefficiency to differ between jurisdictions, DNSPs or periods. We consider whether they can be applied using Stata, including community contributed Stata programs.[6]

The methods reviewed include the suggestions of Frontier Economics relating to SFA, and several other SFA methods. From among the alternatives considered in section 2.2, several methods are selected based on their assessed:

- suitability for addressing the problems observed in the currently used methods;
- flexibility to separately estimate technical change and time-varying inefficiency, with different efficiency trends between jurisdictions, DNSPs, or periods; and
- stability and reliability, so that they can be readily applied in each annual benchmarking update without undue risk of problems or difficulties of application requiring troubleshooting or workarounds.

Once models are selected based on their theoretical attributes, their empirical performance is assessed in chapters 4 and 5. The criteria used for this assessment are detailed in Section 2.3.

[6] Stata is used here due to its reliability and consistency in estimating econometric models. In our experience, it has performed more robustly than alternatives software.

2.1 Opportunities for Refining Current Models

Before considering alternative models related to time varying inefficiency, we reanalyse the issues identified in the previously used models (i.e. the base models and those from the Phase 1 memo) to better understand the likely causes and to recommend potential improvements or refinements. This investigation is helpful to identify any other sources of misspecification which, ultimately, may be contributing to the excessive monotonicity violations observed in the TLG models. This step also helps to avoid or mitigate similar problems in alternative, more sophisticated models, such as time varying inefficiency models. Based on this analysis, several observations can be made:

(1) *Half-Normal Distribution*: In the SFA model, we recommend considering setting $\mu = 0$ (ie, the normal-half normal model). A truncated-normal inefficiency distribution is theoretically attractive because it allows the average level of inefficiency to vary, giving more flexibility than the half-normal (which concentrates inefficiency near zero). However, this flexibility comes at a cost of risk of convergence or identification problems, difficulties already evident in the standard SFATLG results. As described in Ritter and Simar (1994), the half-normal assumption avoids the computational identification problem of estimating both the intercept of the frontier (β_0) and the 'intercept' of the inefficiency term (μ).[7] This approach can reduce computational convergence issues, although it may not eliminate them entirely. All the time-varying models tested in this paper are based on the assumption that the intercept of the inefficiency term is equal to zero.

(2) *Fixed Effects:* For the current LSE estimation method, which includes fixed effects only for Australian DNSPs and uses the *xtpcse* command, we have also considered implementing it via the fixed effects model that allow for clustered standard errors (i.e., using *xtreg, fe vce*(*cl* <clustvar>)).

(3) *Omitted Variables:* Some of the outputs used in the index analysis are not included in the econometric model. This raises the possibility of omitted variable bias (which in turn may cause other issues, e.g., violation of monotonicity). We have explored the use of all five outputs in the econometric opex cost function (ie, including also energy throughput and reliability, as measured by Customer Numbers × SAIDI).

(4) *Input Prices:* In principle, according to the economic theory of production, the model for a cost function should also include input prices (in addition to outputs). Alternatively, other major inputs (e.g., capital input) should also be included if one models via the input distance function. Their exclusion may cause omitted variable

[7] An identification problem occurs when the model's parameters cannot be uniquely determined from the available data. This can result in a flat or poorly defined likelihood surface, making it difficult for optimisation algorithms to converge or for standard errors to be reliably estimated.

bias (which can in turn lead to other issues, e.g., violation of monotonicity). The opex model could be interpreted in different ways:

(a) If interpreted as an input demand function, it should include input prices among the explanatory variables and satisfy homogeneity of degree zero in input prices.[8]

(b) If interpreted as a conditional (short-run) cost function then it should also include the other inputs (or sub-costs) that are conditioned on, e.g., log capital input quantity, as an explanatory variable.[9]

(c) It can also be interpreted as a distance function, for example the Shephard's input distance function (which is dual to the cost function). This formulation would require the quantities of the other inputs, normalised by the quantity of opex inputs, among the explanatory variables (in addition to the outputs and other conditioning variables). E.g., for two inputs, it would require the log ratio of capital to opex inputs.

Interpretations (b) and (c) would require excluding Ontario DNSPs from the sample, because a consistent measure of capital input is unavailable. We test interpretation (a) in section 3.4 by including the ratio of prices for opex and capital inputs (although some assumptions are required to construct a consistent capital price index across all jurisdictions).

The proposed refinement methods based on time-invariant inefficiency approach are tested in Attachment A, with results summarised in Section 3.

2.2 Incorporating Time-Varying Inefficiency: Review of SFA methods

In the last four decades a wide variety of Stochastic Frontier Analysis models for panel data have been developed. Here, we provide a survey that focuses on those approaches we consider having the most relevance and potential for the present project.[10] We organise this section according to the chronological development of SFA methods, grouping models into distinct "waves" based on when they were introduced and the specific issues they aimed to address.

[8] Homogeneity of degree zero in input prices means that if all input prices change by the same proportion, *ceteris paribus*, the demand for inputs remains unchanged. In other words, input demand depends on relative prices, not absolute price levels, *ceteris paribus* (i.e., technology and outputs to be produced remain the same).

[9] This is the current interpretation of the models. However, capital input quantity was not included in the opex cost function modelling due to the lack of comparable capital data for Ontario DNSPs.

[10] Recent relevant academic surveys include Sickles (2005), Sickles and Zelenyuk (2019 ch.12-16), Kumbhakar, Parmeter and Zelenyuk (2020a, 2020b), Sickles, Song and Zelenyuk (2020), and Nguyen, Sickles and Zelenyuk (2022). Here we follow these works, briefly describing a selected set of SFA models. The reader can refer to these surveys for more details.

2.2.1 The first wave of panel data SFA models

The first panel data stochastic frontier analysis (SFA) models were introduced by Pitt and Lee (1981) and Schmidt and Sickles (1984) and are basically those used in the AER's opex benchmarking:[11]

- The AER's "SFA" approach is the Pitt and Lee approach via maximum likelihood estimation (MLE), with a truncated-normal distribution of inefficiencies rather than the half-normal distribution, although the half-normal may be as appropriate and potentially preferred.[12]
- What the AER refers to as the "LSE" method is essentially the fixed-effects approach of Schmidt and Sickles (with the exception that the fixed effects in the AER models are only for Australian DNSPs, not all DNSPs in the sample).

The general form of these panel data opex cost function models can be stated as follows:

$$c_{it} = \alpha_0 + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \epsilon_{it}, \tag{2.1}$$

where c_{it} represents the log of the measure of cost ($\ln C_{it}$) for firm i in period t. The term α_0 is the intercept, and $\boldsymbol{x}_{it}$ is a $J \times 1$ vector containing the log of the jth cost driver ($\ln X_{it}^{J}$).[13] The vector $\boldsymbol{\beta}$ is the slope-parameters (elasticities) to be estimated, while ϵ_{it} is the error term.

In Pitt and Lee's (1981) approach (from which the AER's SFA models are derived), the error term consists of two unobserved components, statistical noise (v_{it}) and cost inefficiency (u_i), i.e.,

$$\epsilon_{it} = v_{it} + u_i, \tag{2.2}$$

with a restriction that $u_i \geq 0$, (reflecting inefficiency can only be greater than or equal to zero) with finite mean and variance, and that v_{it} has zero mean and finite variance.[14] The cost

[11] Panel data SFA models emerged shortly after SFA was introduced in cross-sectional form by Aigner et al. (1977) and Meeusen and van den Broeck (1977).

[12] The assumption regarding the stochastic distribution of inefficiencies is used in formulating the log likelihood function, which is to be maximised.

[13] In this report, a reference to "cost" implies "opex cost" unless clarified otherwise. "Output" refers to the main output variables used in the models, such as customer numbers, circuit length, and ratcheted maximum demand. The term "cost driver" is used more broadly to include all variables that influence costs in the model, including outputs, time trends, jurisdictional dummies, and operating environment factors.

[14] This application is the cost function, where the main explanatory variables are output(s) and input prices, and homogeneity of degree one in the input prices is imposed via normalizing the total cost and all input prices by the price of one of the inputs. In the case of the short-run, or variable cost function, if there is a single input price for variable cost, this normalisation will eliminate input prices from the explanatory variables and cost will be divided by the variable input price. The variable cost function will normally include the fixed input(s) as explanatory variables. This has similarities to the input distance function. In the case of estimating an efficiency frontier based on the input distance function, the response variable will be an observed input of interest while the explanatory variables will be all other inputs (normalized by the input of interest) and all outputs.

inefficiency term (u_i) has no time subscript because in this model inefficiency is assumed to be time-invariant.

To disentangle the noise from inefficiency, Pitt and Lee (1981) makes parametric assumptions about their distributions and estimates the model via MLE as a random effects model. For example, assuming normal and half-normal distributions for the noise and inefficiency respectively:

$$v_{it} \sim \mathcal{N}(0, \sigma_v^2),$$
$$u_i \sim \mathcal{N}^+(0, \sigma_u^2). \tag{2.3}$$

The values of u_i are positive, and hence their distribution is truncated at zero (ie, half-normal). If the truncated-normal assumption is used then: $u_i \sim \mathcal{N}^+(\mu, \sigma_u^2)$, where μ is the mode or 'pre-truncation mean' of the distribution. This is the AER's current SFA approach.

The estimated *inefficiency* can be obtained from: [15]

$$\hat{u}_i = E(u_i | \epsilon_{it}). \tag{2.4}$$

The estimated cost efficiency scores (CE) can be obtained from:[16]

$$\widehat{CE}_i = E\{\exp(u_i | \epsilon_{it})\}. \tag{2.5}$$

The Pitt and Lee approach can be implemented in Stata using: **xtfrontier *y xlist*, cost ti**,[17] as implemented by AER. The **constraints(#)** option can be used to impose the half-normal distribution, as follows: **constraint # _b[mu:_cons] = 0**. Alternatively, the Pitt and Lee approach can be estimated via: **sfpanel *y xlist*, model(pl81)**, using the sfpanel command contributed by Belotti et al (2012).

Differently from the Pitt and Lee (1981) (which, as we have seen, is the basis of the AER's SFA method), Schmidt and Sickles (1984) do not decompose the error into inefficiency and noise. The approach to deriving u_i avoids distributional assumptions and estimates the model as a fixed effects panel data model and then recovers the estimated *inefficiency* from the estimated fixed effects (which is the basis of the AER's LSE method). The model in (2.1) is first transformed as follows:

$$c_{it} = \alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + v_{it}, \tag{2.6}$$

where $\alpha_i = \alpha_0 + u_i$ and v_{it} is the normally distributed noise. Firm k's inefficiency can be estimated (or predicted) via:

[15] As described in Jondrow et al. (1982)
[16] As described in Battese and Coelli (1988)
[17] Here *y* refers to the dependent variable, and *xlist* refers to the set of explanatory variables.

$$\hat{u}_k = \hat{\alpha}_k - \min_i\{\hat{\alpha}_i\}, \qquad k = 1, \dots, N \tag{2.7}$$

One can then obtain the estimated efficiency scores:

$$\widehat{CE}_k = \exp(-\hat{u}_k). \tag{2.8}$$

Again, the cost inefficiency term $\hat{u}_k$ has no time subscript because in this model inefficiency is time-invariant.

The AER implements this (LSE) model using firm-specific dummy variables to estimate the fixed effects, but only for Australian DNSPs. The Stata command **xtpcse** is used which produces panel-corrected standard errors, and a degree of autocorrelation within the panels is accommodated. The Schmidt and Sickles approach (strictly speaking, including fixed effects for all firms in the sample) can be implemented in Stata via **xtreg *y xlist*, fe**, and obtaining the individual fixed effects via **predict alpha, u**. Alternatively, one can do it as: **sfpanel *y xlist*, model(fe)**.[18]

2.2.2 The second wave: relaxing time-invariance

Recognizing the assumption of time invariant inefficiency as a key drawback of these models, the second wave of panel SFA models attempted to address this limitation. The most prominent of these are the approaches suggested by Cornwell et al. (1990), Kumbhakar (1990) and Battese and Coelli (1992).

Cornwell et al. (1990) modified the fixed effects approach of equation (2.6), replacing the time-invariant α_i with a time varying function α_{it}. They chose a quadratic function of the time trend t that allows for the parameters to vary across firms $i \in \{1, \dots, N\}$, i.e.,

$$\alpha_{it} = \theta_{0i} + \theta_{1i}t + \theta_{2i}t^2, \quad i = 1, \dots, N; \;\; t = 1, \dots, T. \tag{2.9}$$

The technical inefficiency scores for firm k at time t can then be estimated (predicted):

$$\hat{u}_{kt} = \hat{\alpha}_{kt} - \min_i\{\hat{\alpha}_{it}\}, \qquad k = 1, \dots, N; \;\; t = 1, \dots, T. \tag{2.10}$$

And the cost efficiency scores in each period can then be obtained as $\widehat{CE}_{kt} = \exp(-\hat{u}_{kt})$.

In Stata, this approach can be implemented using: **sfpanel *y xlist*, model(fecss)**.[19] A more cumbersome way of implementing the model would be to interact firm-specific dummy variables with the time trend and squared time trend variables. For example, this approach may be convenient for implementing a more restricted version of the model, since equation (2.9) has a great many parameters.

---

[18] This approach is tested in Attachment A, and the results are summarised in section 3.1.
[19] The Cornwell et al (1990) approach is tested in Attachment B.

The approaches of Kumbhakar (1990) and Battese and Coelli (1992) modify the Pitt and Lee (1981) approach by assuming a time varying inefficiency term that is a parametric function of time $f(t)$ and the time invariant inefficiency u_i:

$$u_{it} = f(t)u_i, \quad i = 1, \dots, N; \;\; t = 1, \dots, T. \tag{2.11}$$

Kumbhakar (1990) assumes that: $f(t) = (1 + \exp(at + bt^2))^{-1}$ and half-normal distribution for u_i. Substituting into equation (2.11), the inefficiency term becomes:

$$u_{it} = (1 + \exp(at + bt^2))^{-1} \times u_i\,, \qquad u_i \sim_{iid} \mathcal{N}^+(0, \sigma_u^2). \tag{2.12}$$

The function in $f(t)$ is a scaling function common to all firms, and u_i is the one-sided firm-specific inefficiency term, which has a half-normal distribution. The scaling function does not vary by firm, it adjusts each firm's inefficiency over time in the same way, imposing a common time trend across all firms. That is, inefficiency levels u_{it} evolve over time in parallel, with the shape of the time path determined by $f(t)$, and u_i is firm *i*'s efficiency when t = 0.

The *sfpanel* implementation command for Kumbhakar (1990) is a generalisation of this approach where, in place of $f(t)$ there is:

$$f(\boldsymbol{z_{it}}) = [1 + \exp(\delta_0 + \boldsymbol{z'_{it}\delta})]^{-1} \tag{2.13}$$

where $\boldsymbol{z_{it}}$ is a vector of exogenous variables. Note that in this model, the constant term δ_0 only affects the scaling of the inefficiency term and does not affect the truncation point of the random inefficiency term, which remains at zero. Hence, it does not give rise to partial identification issue, and does not require setting δ_0 to zero for estimation. Also note that the explanatory variables, $\boldsymbol{z_{it}}$, may vary across firms and across time, and hence in this generalised Kumbhakar (1990) formulation, each firm's inefficiency can have a different time-pattern.

Battese and Coelli (1992) assume a different form for $f(t)$ and a truncated-normal distribution for u_i:

$$u_{it} = \exp(-\xi(t - T)) \times u_i, \qquad u_i \sim_{iid} \mathcal{N}^+(\mu, \sigma_u^2) \tag{2.14}$$

where T is the last period in the sample. It can be applied to unbalanced panels, in which case $T = T_i$. In this model, u_i is firm *i*'s inefficiency when t = T, the last year of the sample period. Again, the scaling function imposes the same time trend to each u_i, so each firm's inefficiency moves in parallel over time.

Despite the advances in incorporating time variation into the models, it is important to note that including assumed firm-specific efficiency change as a quadratic function may lead to empirical estimation issues. Also, while offering improvements by allowing for time varying inefficiency, these models impose quite restrictive structures on how inefficiency varies over time, especially the Kumbhakar (1990) and the Battese and Coelli (1992) which assume that the rate of inefficiency change is the same for all firms. In contrast, the *generalised* version of the Kumbhakar (1990) formulation allows each firm's inefficiency to follow a different time pattern.

2.2.3 The third wave: decomposing the unobserved heterogeneity

The models developed during the first and second waves of panel data SFA have a key drawback in how they handle unobserved individual heterogeneity in the data. They are unable to identify which part of this heterogeneity is related to inefficiency and which part is not. Recognising this inability and other drawbacks, the third wave of panel data SFA approaches introduced further improvements, albeit with greater complexity. One of them was Cuesta (2000), who elaborated on the model of Battese and Coelli (1992) in equation (2.14), proposing a more general form for u_{it}, namely:

$$u_{it} = \exp(-\xi_i(t-T)) \times u_i, \qquad u_i \sim_{iid} \mathcal{N}^+(\mu, \sigma_u^2). \tag{2.15}$$

This model also assumes independence between the regressors (the x's) and inefficiency. Furthermore, as with Battese and Coelli (1992), this model imposes a monotonic relationship between the inefficiency and time, that is, it can only increase or decrease consistently throughout the period, which might be viewed as too restrictive.

However, a subtle yet important difference to the Battese and Coelli (1992) formulation is that the function that describes the change of efficiency over time is allowed to vary across firms due to the parameter ξ_i being firm-specific rather than constant (ξ). Thus, it does not restrict the efficiency change to be the same across all firms, allowing each firm to have its own inefficiency time profile.

This, however, comes at considerable sacrifice. Besides requiring much more complicated computations (with increased risk of non-convergence, reaching local optima, or intractability for large datasets), it significantly increases the number of parameters (reducing the degrees of freedom), which can lead to overfitting, especially in small samples. All these in turn may reduce the precision and robustness of inefficiency estimates.[20] More specifically, the number of parameters increases with the sample size, which is usually referred to as the 'incidental parameter problem' in the panel data econometrics literature. Hence, while theoretically appealing, this approach is likely to be not very practical.[21]

Among these and other prominent approaches that attempted to address the above-mentioned drawbacks are the works of Greene (2005a, 2005b), Chen et al. (2014), Colombi et al. (2014), Kumbhakar et al. (2014), and Belotti and Ilardi (2017).

[20] This problem was also noted in Cuesta (2000), who acknowledged that "further work is necessary to evaluate the model".

[21] These limitations explain why the model has seen limited adoption in empirical applications. Nevertheless, it served as an important stepping stone toward more flexible panel data SFA models that have since received greater attention.

In his two fundamental contributions to SFA literature, Greene (2005a, 2005b) argued that technical efficiency should be modelled separately from, or in addition to, the unobserved individual heterogeneity. In these models, efficiency is allowed to change over time (sometimes referred to as "transitory technical inefficiency"), but without imposing a specific trend on its evolution.[22] In contrast, time invariant differences between firms are assumed to represent unobserved heterogeneity. The models proposed by Greene (2005a, 2005b) were dubbed as 'true fixed effects' (TFE) and 'true random effects' (TRE) stochastic frontier models, depending on how the unobserved individual heterogeneity is being modelled and estimated. Using our previous notation, these models can be expressed as:

$$c_{it} = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \epsilon_{it}, \tag{2.16}$$

$$\epsilon_{it} = v_{it} + u_{it}, \quad i = 1, \dots, N; \; t = 1, \dots, T. \tag{2.17}$$

In the TFE model, α_i is a firm-specific fixed effect that captures the unobserved individual heterogeneity independent from the inefficiency (u_i) and noise (v_{it}), which are the two elements of the composite error ϵ_{it}. To disentangle these two unobserved error components, one needs to impose parametric assumptions on the distributions of these terms, e.g., normal and half-normal, respectively, as in most SFA models; such as the assumptions shown in equation (2.3). In the TRE model, $\alpha_i = \alpha_0 + w_i$, where w_i is a symmetric random variable with finite variance, which has a different value for each firm. Again, parametric assumptions on this distribution are required.

However, even for the normal-half-normal case, these approaches are computationally challenging to estimate, especially in the fixed effects framework. This is because, in general, there is no closed-form likelihood of the 'within transformation' of the composite error that can be done through standard procedures in panel data software.[23] Moreover, there is also the incidental parameters problem analogous to that in Cuesta (2000), potentially causing inconsistency of the estimates.

To mitigate these challenges, Greene (2005a) suggested using the dummy variable ML estimator for the fixed effects approach, supporting it with Monte Carlo evidence suggesting that the problem of incidental parameters is, while still present, not substantial if T is relatively large. To implement it in Stata, one can use command **sfpanel *y xlist*, model(tfe)** for the TFE approach, or the option **model(tre)** in the TRE approach. The additional option **distribution(h)** can be used to select the half-normal distribution for the inefficiency term (the truncated-normal and exponential distributions are also available).

[22] This means that while the model allows for inefficiency to go up and down, it does not assume (or impose) a trend. Inefficiency may vary from year to year and there is no systematic direction to the changes.

[23] The 'within transformation' (or 'fixed effects transformation') refers to the estimation procedure whereby the data is firstly de-meaned by subtracting the means for each panel, and then the model is estimated using this transformed data. The fixed effects can then be recovered by applying the estimated coefficients to the panel means that were previously removed.

Previous experience with these models suggests that the estimation is usually challenging, even for relatively large samples, especially for complex models (like the Translog, especially with many variables). Some of these challenges can be mitigated by following the approach of Chen et al. (2014). Leveraging on the properties of closed-skew normal distribution, Chen et al. established a closed-form solution for the likelihood of the within and first-difference transformation. This helps simplifying the computational burden to some extent, allowing consistent estimation in the fixed effects framework via MLE.

The approach of Chen et al. was further improved by Belotti and Ilardi (2017) who proposed using the simulated marginal MLE. Both approaches can be done in Stata using the community-contributed command **sftfe** (from Belotti and Ilardi (2017)). The option **estimator(within)** is used for the Chen et al. (2014) estimator. The option **estimator(mmsle)** is to be used for Belotti and Ilardi (2018) estimator. Again, the option **distribution(h)** can be used to select the half-normal distribution for the inefficiency term.

Summarizing, the models of Greene (2005a, 2005b) (and their improved estimators by Chen et al. (2014) and Belotti and Ilardi (2017)) made an important improvement relative to the earlier panel data SFA models. They allow the technical inefficiency to vary over both individual firms and over time (in a random fashion, according to some assumed distribution, like half-normal) and to disentangle it from the unobserved individual heterogeneity. However:

- These models can be difficult to estimate;
- They do not allow for trend-like changes in inefficiency and the separation of these trends from technical change (which is our principal aim here); and
- One may question whether all the inefficiency is 'fully' random (across individuals *and* across time), as modelled in Greene (2005a, 2005b), or if there is also some inefficiency associated with each individual that persists over time (often referred to as 'persistent inefficiency'). In an attempt to address this question, Colombi at el (2014) proposed what is usually referred to as "four-component" SFA model, which decomposes the inefficiency into persistent (over time) component and transitory component, as well as decomposes the remaining error into a persistent (over time) component and transitory component (statistical noise). The next section briefly explains this model.

2.2.4 The four-component SFA model

The basic four-component SFA model, first proposed by Colombi et al. (2014) and elaborated in a few other recent papers, allows capturing more nuanced structures (systemic vs. situational) of the unobserved part that hides in the residual of a regression analysis. Specifically, it aims to capture separately four different components of the composite error:

(a) the usual random error outside the control of the firm;

(b) the differences between units that are not explained by observed variables, usually called unobserved heterogeneity;

(c) the long-term inefficiency, e.g., due to structural or managerial issues, usually named as persistent inefficiency;

(d) the short-term inefficiency that can change over time, called time-varying inefficiency.

In the context of DNSP efficiency analysis, unobserved heterogeneity refers to differences between firms that are not necessarily due to inefficiency—such as variations in operating environments (which at present are partially accounted for through post-estimation OEF adjustments). Persistent inefficiency could reflect established operational practices. Time-varying inefficiency, in turn, may capture shorter-term fluctuations in performance across years.

Consider the initial SFA model (2.1), reproduced as (2.18):

$$c_{it} = \alpha_0 + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \epsilon_{it}, \quad i = 1, \dots, N; \; t = 1, \dots, T. \tag{2.18}$$

Here ϵ_{it}, the composite error, consists of *four* unobserved components:

$$\epsilon_{it} = \alpha_i + v_{it} + (\eta_i + u_{it}), \tag{2.19}$$

where $\eta_i \geq 0$ is a one-sided random inefficiency effect with a single value for each DNSP, and $u_{it} \geq 0$ a one-sided random inefficiency effect with a different value for each observation. These respectively represent persistent and transitory technical inefficiency (both with finite means and variances), α represents the usual fixed or random effect to capture time-invariant unobserved heterogeneity across firms, and v_{it} is the usual statistical noise with zero mean and finite variance.

As usual for SFA, to disentangle these four components one needs to impose parametric assumptions on the distributions of these terms, e.g., normal for the statistical noise and half-normal for the inefficiency, as in most SFA models. That is, for $i = 1, \dots, N$ and $t = 1, \dots, T$:

$$\begin{aligned} &\alpha_i \sim_{iid} N(0, \sigma_\alpha^2), \\ &v_{it} \sim_{iid} N(0, \sigma_v^2), \\ &\eta_i \sim_{iid} N^+\left(0, \sigma_\eta^2\right), \\ &u_{it} \sim_{iid} N^+(0, \sigma_u^2). \end{aligned} \tag{2.20}$$

In practice, estimation may often run into computational challenges, especially for relatively large dimensions (e.g., like the Translog with many variables). To mitigate some (though likely not all) of the difficulties, the *multi-step procedure* of Kumbhakar et al. (2014) can be deployed. For example, in Stata it can be done as follows:

(a) Predict combined transitory error (v_{it}) and transitory inefficiency (u_{it}) via random effects regression: **xtreg *y xlist*, re**; then use postestimation commands: **predict esl, e** (to predict the composite transitory elements) and: **predict lam, u** (to predict lambda = $\alpha_i + \eta_i$ + E(η_i));

(b) Predict persistent (in)efficiency via cross-sectional SFA without regressors: **frontier lam, distribution(hnormal)**. Then use: predict **ineff_persistent, u** (to predict the persistent inefficiency, E($\eta_i \mid \epsilon_{it}$); and: predict **eff_persistent, te** (to predict persistent efficiency, E(exp($-\eta_i$) | ϵ_{it}))

(c) Predict transitory (in)efficiency via cross-sectional SFA without regressors: frontier **esl, distribution(hnormal)**. Then: **predict ineff_transitory, u** (to predict transitory inefficiency, E($u_{it} \mid \epsilon_{it}$)); and: **predict eff_transitory, te** (to predict transitory efficiency, E(exp($-u_{it}$) | ϵ_{it})).

In summary, by decomposing the unobserved variation ("error") into more components than the other models, this approach is expected to have higher distinguishing power between the noise and the inefficiency and its two distinct parts (persistent or long term and transitory or short-term), which is particularly relevant in a panel data context. If applied adequately, this can provide an improved policy insight, e.g., if persistent inefficiency can be addressed differently (with different policy actions) from time-varying inefficiency. This approach also provides better statistical "controls" or "conditioning" for the unobserved heterogeneity, reducing the chances of potential misattributing of some unobserved firm heterogeneity (e.g., sub-jurisdictional effects not controlled by dummies) to inefficiency.

As usual, the additional benefits this model offers comes at some additional cost. This cost comes through the additional assumptions needed to identify the model, potentially higher sensitivity of results with respect to admitted assumptions and relatively higher computational complexity. In turn, may put additional pressure on the quality of the data needed for this model to provide robust results.[24]

2.2.5 The third wave (continued): Modelling the determinants of inefficiency

Another important stream of SFA literature allows (in)efficiency to be determined by some explanatory variables, such as operating environment factors (OEFs), instead of either being constant (in the Pitt and Lee approach) or time-varying according to a particular time pattern, such as the time-varying decay model of Battese-Coelli 1992, or the quadratic time trend

[24] Ideally, all variables should consistently reflect the same output definitions across all DNSPs included in the analysis. In practice, however, this is rarely the case. The output variables used are often proxies for the actual services delivered, and incorporating new data can increase model complexity by introducing additional heterogeneity.

model of Kumbhakar 1990.[25] The generalised form of the Kumbhakar 1990 model, as it is implemented in the *sfpanel* package and shown in equation (2.13) supports this approach.

Kumbhakar, Ghosh and McGuckin (1991), hereafter KGM, proposed characterizing the 'pre-truncated mean' of the inefficiency term as a function of potential determinants of inefficiency. In particular, they focused on the standard model in equation (2.1), with the idiosyncratic error component being normally distributed as shown in equation (2.3). However, the distribution of inefficiencies is given by:

$$u_i \sim N^+(\mu_i, \sigma_u^2) \qquad \text{with: } \mu_i = \delta_0 + \boldsymbol{z}_i'\boldsymbol{\delta}, \qquad i = 1, \ldots, N \qquad (2.21)$$

where $\boldsymbol{z}_i$ is a vector of variables (assumed to be exogenous in the model) for observation i that are expected to 'determine' or explain the inefficiency of the observation i via some vector of parameters (common to all observations) $\boldsymbol{\delta}$ that is unknown and is to be estimated along with other parameters of the model.

The KGM model was later adapted to the panel data context by Battese and Coelli (1995b), in which case the 'pre-truncated mean' of the inefficiency distribution can vary over time as well as between firms:

$$u_{it} \sim N^+(\mu_{it}, \sigma_u^2) \qquad \text{with: } \mu_{it} = \delta_0 + \boldsymbol{z}_{it}'\boldsymbol{\delta}, \qquad (2.22)$$
$$i = 1, \ldots, N; \;\; t = 1, \ldots, T.$$

Battese and Coelli explain that an equivalent way of expressing the inefficiency term u_{it}, is: $u_{it} = \delta_0 + \boldsymbol{z}_{it}'\boldsymbol{\delta} + \varpi_{it}$, where ϖ_{it} is a random variable distributed as a truncation of a normal distribution with zero mean and constant variance, and with the truncation point being: $\varpi_{it} \geq -\delta_0 - \boldsymbol{z}_{it}'\boldsymbol{\delta}$ (Battese and Coelli, 1995: 327). This formulation makes clear the potential difficulties that can arise in adequately identifying both α_0, the intercept of the frontier model, and δ_0, the intercept of the function determining μ_{it}.

This model (referred to as the KGM-BC or BC95 model) can be estimated via MLE. These models can be implemented via Stata commands from Belotti et al. (2012). The panel data version can be implemented via: **sfpanel *y xlist*, model(bc95)**. The inefficiency and efficiency measures can be obtained from: **predict ineff, u**; and **predict eff, bc**. The post-estimation command **marginal** can be used to calculate the marginal effects of the *z*-variables.

The KGM-BC model is a promising model.[26] However, it is worth noting its major caveats (often overlooked in the applied literature):

[25] The work of Cornwell et al. (1990) mentioned in section 3.1.2 was the first attempt on this question, although it mainly focused on modelling inefficiency with respect to time

[26] It became very popular largely due to its relative simplicity and intuitive nature, as well as due to the software package *frontier,* programmed by Professor Tim Coelli, that was made freely available at that time.

- The first, and perhaps the most important, caveat is that the estimated relationship between $\boldsymbol{z_{it}}$ and inefficiency should not be interpreted as causal, rather as an empirically identified statistical dependency. Indeed, one must be careful when selecting the set of $\boldsymbol{z_{it}}$ variables, having solid justifications for each, to avoid identifying various 'spurious' relationships, and keeping in mind the possibility of endogeneity or reverse causality problems (pertinent to many econometrics models).
- Another important caveat is related to the computational convergence problems that often arise for this type of model. Some of these problems are related to inability to accurately identify β_0 and δ_0 together, which can be mitigated by setting $\delta_0 = 0$. Similar problems may occur for some slope coefficients in $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ if the corresponding variables (or their combinations) are highly correlated, resulting in multicollinearity. As a result, thorough diagnostics of robustness of the estimated results (e.g., with respect to the starting values, the definition of variables) is highly advised for this model.

Alternatively, some of the computational challenges can be mitigated by modelling the determinants not via the mean but via the variance (or skedastic function) of inefficiency, e.g., as in Caudill et al. (1995) where:

$$u_{it} \sim N^{+}\left(0, \sigma_u^2(\boldsymbol{z_{it}})\right) \qquad \text{with: } \sigma_u^2(\boldsymbol{z_{it}}) = \exp(\delta_0 + \boldsymbol{z}_{it}'\boldsymbol{\delta}). \tag{2.23}$$

In Stata, this approach can be implemented in a cross-sectional setting via commands: **sfcross y xlist , distribution(hnormal) usigma(zlist)**.

A more general approach involves assuming the so-called scaling property for the inefficiency term, as proposed by Wang and Schmidt (2002) and Alvarez et al. (2006):

$$u_{it} = g(\boldsymbol{z_{it}}|\boldsymbol{\delta})u_{it}^{*}, \qquad i = 1, \dots, N; \;\; t = 1, \dots, T \tag{2.24}$$

where $g(\boldsymbol{z_{it}}|\boldsymbol{\delta})$ is some non-negative function characterizing the explainable part of inefficiency via the *z*-variables (e.g., $g(\boldsymbol{z_{it}}|\delta) = \exp(\delta_0 + \boldsymbol{z_{it}}'\boldsymbol{\delta})$), while u_{it}^{*} is a non-negative random term, e.g., half-normal.[27]

The Wang and Schmidt (2002) model can be implemented in a cross-sectional context via commands from Kumbhakar et al. (2015) as follows. To set up the log-likelihood function: **sfmodel *y* , cost dist(t) frontier(*xlist*) scaling hscale(*zlist*) tau cu**. To execute the model: **ml max**. The inefficiency and efficiency measures can be obtained from: **sf_predict , jlms(ineff)**; and **sf_predict , bc(eff)**. A specific variant of the Wang and Schmidt model, which may be of interest here, is where $g(\boldsymbol{z_{it}}|\boldsymbol{\delta}) = [1 + \exp(\delta_0 + \boldsymbol{z_{it}}'\boldsymbol{\delta})]^{-1}$, which can be implemented in

[27] Note that this can be viewed as a generalization of Kumbhakar (1990) and Battese and Coelli (1992), where some of the z-variables can include time variables.

sfpanel using: **sfpanel *y xlist*, model(kumb90) bt(*zlist*)**. This is *sfpanel*'s generalised form of Kumbhakar 1990 which we discussed previously.

It is also possible to model inefficiency both through the mean (as in KGM-BC) and through the variance (as in Caudill et al., 1995) together, e.g., as done by Wang (2002). In a panel data context it can be implemented by: **sfpanel *y xlist*, model(bc95) emean(*zlist_m*) usigma(*zlist_u*)**. In both the emean(.) and usigma(.) "**, nocons**" can be specified to set $\delta_0 = 0$.

It is worth noting that for a case where the model contains a time variable (e.g., to estimate the time-trend of technological change), an inclusion of the same time variable among the *z*-variables may lead to identification problems in all these models, especially in KGM-BC models. A more feasible alternative could be to include the time-trend in the frontier part of the model (to estimate average annual technological change), and a time-dummy variable can be included among the z-variables (e.g. to estimate the coefficient of change in efficiency after some time threshold(s)).

Finally, the same type of caveats as discussed above for the KGM-BC model apply for these models as well. Indeed, depending on the data, some of these models may encounter more or less of computational issues, yet the most important caveat—the risk of identifying a reverse causality or even a spurious relationship—remain for all of them. Accurately establishing a causal relationship would require different estimators (e.g., with a well-justified instrumental variables strategy, difference-in-difference methods adapted to SFA or other methods) and with appropriate data needed for such estimators to perform well.

2.2.6 Summary

In summary, the four-component SFA model of Colombi et al. (2014) is of particular interest for the present application, in principle improving on the early SFA models and the Greene (2005a,b) models. In particular, under certain assumptions (mentioned above), it would allow to disentangle the total error into potentially four distinct components: persistent unobserved heterogeneity, transitory unobserved heterogeneity (statistical noise), persistent inefficiency and transitory inefficiency. Indeed, the transitory inefficiency there is of the same nature as the one in Greene (2005a,b) models and is a generalization of the original inefficiency term in the first SFA models (Aigner et al (1977)). Meanwhile, the persistent inefficiency can be understood as the unobserved heterogeneity associated with inefficiency pertinent (in potentially different degrees) for each specific i that persist across all the time periods.[28]

In parallel to these models, the SFA models that try to explain the inefficiency through some expected 'z-variables' are also recommended for the purpose of this review, however keeping in mind the caveats of such models mentioned above. In particular, the KGM-BC (bc95) model would be a natural starting point for this task, with a restriction that $\delta_0 = 0$ and

[28] Before trying the four-component model it might still be worthwhile exploring the "true fixed effects" model from Greene (2005a) through either the estimator of Chen et al. (2014) or that of Belotti and Ilardi (2018), after sorting out the issues identified for the simpler models used in the earlier reports.

thorough diagnostics of robustness of the estimated results (e.g., with respect to the starting values, the definition of variables, etc.).

The models which rely on the scaling property, can also be explored. For these models, a time-trend can be included in the frontier part of the model (to estimate average annual technological change), and time variable-dummy variable interactions can be included among the z-variables to estimate the rates of change in efficiency. It would also be possible to include some time-threshold(s), e.g., the time of policy interventions, time of reforms by the regulator, etc.

To conclude this survey, it is worth clarifying that, as usual in economics, there is no panacea or a single model (or estimator) that is *a priori* 'the best': each comes at some costs of various assumptions or/and additional computational challenges, often with substantially increased complexity for some aspects while oversimplifying other aspects. Moreover, even models that are very appealing theoretically may still perform poorly in practice, largely depending on the quality and quantity of data at hand.

The empirical results of the selected time-varying inefficiency models are presented in Attachment B and summarised in Section 4.

2.3 Criteria for Econometric Models Evaluation

The evaluation of alternative opex cost function models presented in this report is based on the criteria discussed in this section.

2.3.1 Consistency with Economic Theory or Industry Knowledge

This involves verifying that the estimated models adhere to key theoretical expectations regarding cost behaviour:

- *Sign and Significance of (Primary) Output Coefficients:* Output variables are expected to have a positive marginal effect on costs. When an estimated (primary) output coefficient is negative, this violates the monotonicity condition (in the Cobb-Douglas (CD) case, at every point in the sample, and in the Translog (TLG) case, at the sample mean).[29] If an output coefficient is statistically insignificant, it does not necessarily imply that the output has little effect on costs. It may imply the relationship is not precisely estimated. If the output variable is expected to be an important cost driver, insignificance suggests a problem with the definition or measurement of that variable or other problems in the model leading to such results.

[29] The monotonicity requirement states that, all else being equal, an increase in output should not result in a decrease in cost. That is, the marginal cost of outputs should be non-negative. Because the output variables have been centred, the primary output coefficients on the Translog model equal the output elasticities at the sample mean.

- *Frequency of Monotonicity Violations (MVs) in the TLG models:* For the TLG models, we also assess the frequency of monotonicity violations at the observation level.[30] Models with fewer MVs are preferred to those with more, having regard to the total sample, and particular regard to the Australian sample, *ceteris paribus*. For example, for a given Australian DNSP, a TLG model is, under recent practice by the AER, considered invalid if 50% or more of the observations for that DNSP are MVs, *ceteris paribus*. Models that are considered invalid for some Australian DNSPs are considered inferior to those that are considered valid for all Australian DNSPs, *ceteris paribus*.

- *Reasonableness of Output Weights:* The output weights derived from the opex cost function models can be compared to other relevant comparators. For example, they can be compared with those used in the index analysis (which are estimated for Australian DNSPs only), and those derived in opex cost function analysis previously using different data samples or methods. In Translog models, the relationship between the total output elasticity and total output can also shed light on whether smaller DNSPs face economies of scale and whether there is an optimal scale.

- *Coefficients of other variables have the correct sign:* The *a priori* sign of the effect on opex of some variables is known. For example, the log share of undergrounding in circuit length has a negative sign because undergrounding has a higher capital cost but lowers opex, as these cables are not exposed to damage or vegetation encroachment. However, a coefficient value that is negative but not statistically different from zero is also considered acceptable as it may imply the relationship cannot be precisely estimated. This criterion is reasonable given that the undergrounding share is not a primary output in the cost function but rather an operating environment factor. On average across the sample, this factor may not exert a statistically significant effect on costs.

2.3.2 Convergence

Particularly for models estimated using maximum likelihood, convergence of the estimation algorithm is a critical requirement. Models that do not converge are not considered reliable. Other computational problems can arise, such as inability to calculate some or all of the parameter standard errors. While failure to calculate one or two parameter standard errors may not be fatal, if many or all of the standard errors are missing, this will typically impact on calculations that can be made from the model. Such models cannot be used.

Optimisation will not necessarily always reach a global optimum, instead it may reach some local optimum. Different optimisation algorithms are available in Stata which can assist to

[30] Unlike Cobb-Douglas models, Translog models offer greater flexibility by allowing the marginal effect of outputs on costs to vary across observations.

address such problems. Further, parameters that do not fall within a domain of reasonableness may also flag that a computational problem has occurred.

2.3.3 Goodness-of-fit

We evaluate the explanatory power of the models using measures such as the Bayesian Information Criterion (BIC) when it is available for both models being compared that are relatively similar (although not necessarily nested). We also report the adjusted R-squared for the models estimated with least-squares method, and consider the pseudo-adjusted-R^2, which is available for all models. Models with substantially higher goodness-of-fit statistics are preferred, provided they also satisfy theoretical and specification criteria.

2.3.4 Performance on Specification Tests

Statistical tests and plots are used to assess whether the model is appropriately specified, including:

- Statistical tests and plots are used to investigate the normality and homoskedasticity of residuals. Similarly, significant correlations of residuals with efficiency scores may imply inaccurate efficiency estimates if such correlation is not part of the model.
- Specification tests (e.g., the *linktest* in Stata)
- Multicollinearity and its impact on parameter standard errors (e.g., variance inflation factors).
- Hypothesis tests of the single and joint significance of parameters (e.g., the second-order terms in the TLG model, the jurisdictional dummies, etc.).

2.3.5 Stability to Sample Changes

The stability of models can be tested by varying the sample length. For example:

- Beginning with the period 2006–2018, and sequentially adding one more year at the end to reach 2006–2023. Then evaluate the stability of each parameter or statistic of interest (eg, output weights, efficiency scores, goodness-of-fit, per cent of MVs) over these six sample periods.
- Beginning with the period 2012–2023, and sequentially adding one more year at the beginning to reach 2007–2023. Then, evaluate the stability of the parameters of interest over these sample periods.

In total, this would represent 12 different sample periods over which the statistics can be calculated. The standard deviations of the parameters and statistics of interest will provide information on the stability of the models.

2.3.6 Consistent Efficiency Scores

We are not seeking any predetermined result. As noted in Section 1.2, changes in model specification will inevitably alter efficiency scores because each specification captures factors that earlier models omit. Some variation is therefore both expected and acceptable, provided the model meets the agreed selection criteria. All models carry some uncertainty, which the AER already moderates through its 75 per cent efficiency threshold, post-adjustment factors, and by avoiding fully mechanical price resets.

Experience shows that models with excessive monotonicity violations produce unreliable efficiency scores; such models are excluded from the DNSP efficiency average. Seeking models that eliminate excessive MVs should therefore yield different—yet more credible—efficiency estimates.

If a new model delivers efficiency scores that differ markedly from previous benchmarking studies, historic Opex MPFP results, or established DNSP rankings, the discrepancy should be investigated and explained. Consistency checks remain essential.

Consistency across specifications using the same modelling approach (such as Cobb-Douglas and Translog functional forms) is also important. Where appropriate, efficiency scores should be compared across these specifications. While some variation is expected, the results should be broadly consistent within a reasonable margin.

2.3.7 Parsimony

If two models perform similarly, the more parsimonious model is usually preferred. A preferred model should also be straightforward to apply, e.g., with reliable software packages and minimal troubleshooting required.

3 Preliminary Assessment: Options to Refine Current Models

This chapter summarises the findings of testing the various opportunities for refining the current models discussed in section 2.1, excluding the time-varying inefficiency approaches, which are addressed in chapters 4 and 5. The details of the modelling summarised in this chapter is set out in Attachment A.

Here, we first present the results for the Half-Normal SFA model, followed by an alternative approach to estimating the LSE model. This is followed by a model incorporating two additional outputs to address potential omitted variables, and then a model including an input price component. The section concludes with a summary of findings.

3.1 Standard SFA with Half-normal inefficiency distribution

As discussed in section 2.1, although the truncated normal distribution of inefficiencies has greater flexibility than the more restricted half-normal distribution, estimation of the additional parameter can give rise to computational identification problems, given the constant term in the frontier function. In this section, we summarise the results of testing the half-normal distribution (with details in Attachment A).

In the half-normal SFA model, the main output variables are statistically significant and display the expected positive signs across all models except the SFATLG short-period model shows negative coefficient with customer numbers. The share of undergrounding variable is of the expected sign (negative) and statistically significant in the long-period models but not significant in the short-period models. The time trend variable is positive in all models, and statistically significant in all but the short-period CD model. These results are consistent with those obtained for the Truncated-Normal specification. However, the SFATLG short-period model with Truncated-Normal inefficiency distribution did not converge, whereas the half-normal SFATLG models converge in both periods.

Table 3.1 presents the output elasticities for the Half-Normal models in both the long and short periods, compared to the standard truncated-normal models. The results when using the half-normal distribution are close to those of the standard model. The only exception is the short-period SFATLG model, the standard version of which did not converge, whereas the half-normal version, while convergent, is not satisfactory, having a statistically insignificant and negative elasticity on customer numbers at the sample mean.

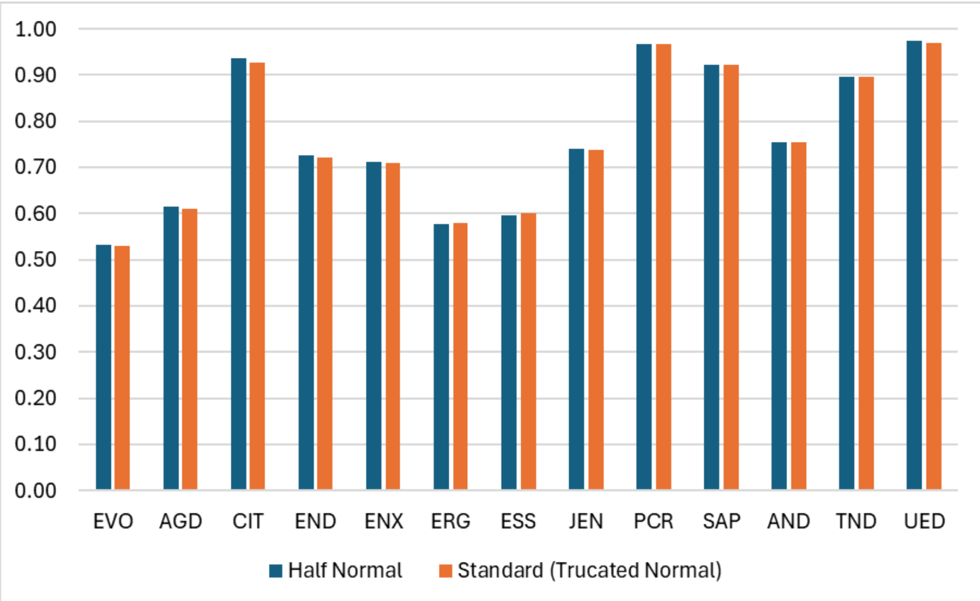
Both the half-normal SFATLG model exhibits a high proportion of MVs in both long and short periods. In the Half-Normal SFATLG model for the long period, 60.3 per cent of Australian DNSPs and 41.3 per cent of the total sample display MVs. These are slightly lower than the corresponding figures for the standard (truncated-normal) model; 79.5 per cent and 45.4 per cent respectively. In the short-period half-normal SFATLG model, MVs occur in 100 per cent of the Australian DNSPs and 67.9 per cent of the total sample. As previously indicated, although the half-normal SFATLG model does converge, the resulting model is not satisfactory.

Table 3.1 Output elasticities: Comparison with standard SFA model

	Half-normal SFA				*Standard (truncated normal) SFA*			
	Long Period		*Short Period*		*Long Period*		*Short Period*	
Sample	*CD*	*TLG*	*CD*	*TLG*	*CD*	*TLG*	*CD*	*TLG*
Cust.	0.280	0.318	0.283	-0.021	0.280	0.318	0.251	NA
CL	0.122	0.147	0.244	0.487	0.129	0.166	0.307	NA
RMD	0.560	0.483	0.428	0.410	0.553	0.443	0.392	NA
Total	0.963	0.948	0.955	0.876	0.962	0.927	0.950	NA

Figures 3.1 and 3.3 compare the efficiency scores of the half-normal and standard (truncated-normal) SFACD and SFATLG models respectively for the long sample and for the SFACD in the short period. The SFATLG short-period efficiency scores are not compared, as they are not available for the standard SFATLG model. The efficiency scores estimated by the half-normal distribution of inefficiencies are very similar to those estimated using the truncated normal distribution.

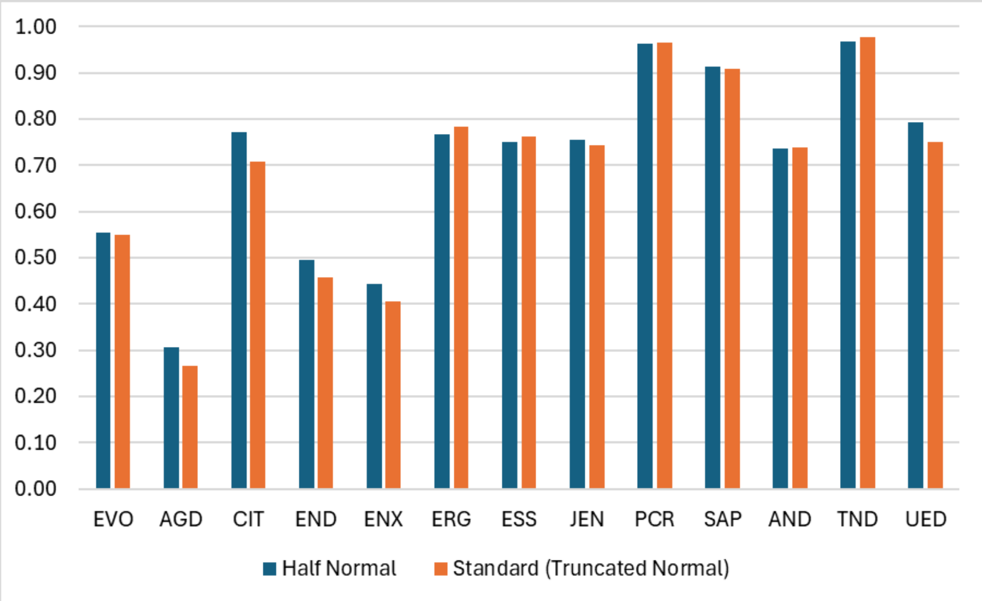
Figure 3.1 Efficiency Scores – SFACD models – Long Period*



* The correlation coefficient between the efficiency scores from the half normal and truncated normal models is equal to 1. On average, the half normal efficiency scores are 0.2% higher than those from the truncated normal model.

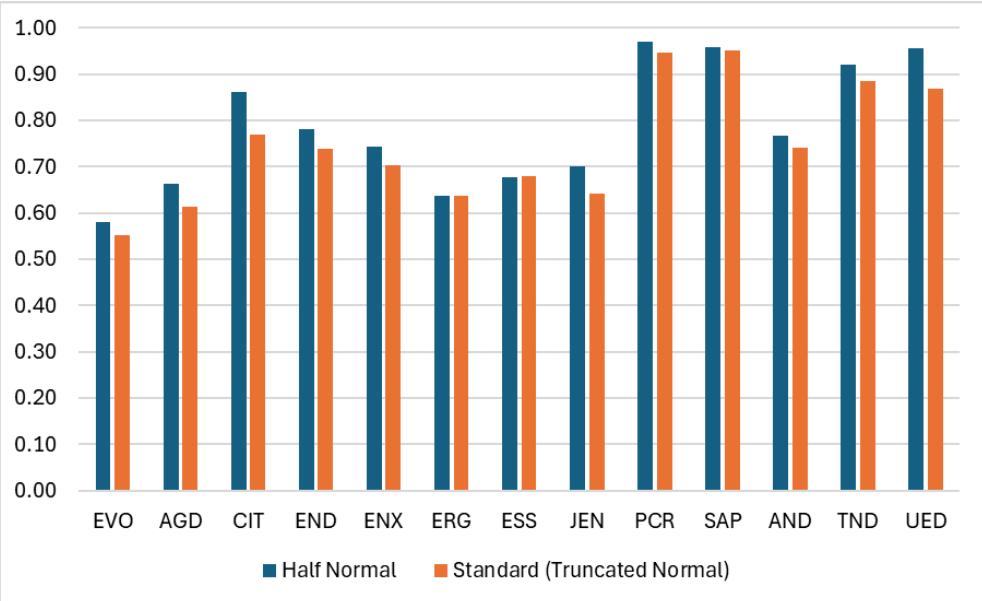
In conclusion, the Half-Normal specification resolves the convergence issue in the short-period SFATLG model, but remains unsatisfactory in other respects, including coefficients with incorrect signs, severe monotonicity violations, unreasonable overall output elasticities and abnormal efficiency scores in some instances. For the CD models for both periods and the long-period TLG model, the estimated efficiency scores and output weights are very similar to those estimated by the standard model using the truncated normal distribution of inefficiencies. This shows that the more flexible truncated-normal assumption does not translate into a significantly different distribution of efficiency score estimates. In the SFATLG long-period model, the frequency of MVs is slightly lower under the Half-Normal assumption than under the Truncated-Normal assumption.

Figure 3.2 Efficiency Scores – SFATLG models - Long Period*



* The correlation coefficient between the efficiency scores from the half-normal and truncated-normal models is equal to 0.994. On average, the half normal efficiency scores are 2.3% higher than those from the truncated normal model.

Figure 3.3 Efficiency Scores – SFACD models – Short Period*



* The correlation coefficient between the efficiency scores from the half-normal and truncated-normal models is equal to 0.976. On average, the half normal efficiency scores are 5.2% higher than those from the truncated normal model.

Although the Half-Normal specification does not address the MV problem— and is not intended to— it can offer improvement by enhancing the computational stability of SFA models, particularly their convergence. That said, whilst the half-normal assumption ensures convergence in cases where the truncated-normal model did not converge, the resulting models remained inadequate. Generally, the efficiency scores from the half-normal model are similar to those from the truncated normal model, with correlations exceeding 99 per cent in the two long period models, 97 per cent in the SFACD short period case. On average, the half-normal model produces efficiency scores that are 2.6 per cent higher than those of the truncated-normal models. The Half-Normal model slightly improved the MVs. Also, in the long-period models, the coefficient of μ in the Truncated-Normal specification is not statistically different from zero, suggesting that the truncated distribution does not significantly depart from the Half-Normal restriction when the ABR24 long-period data sample is used.

These results suggest that the Half-Normal model could enhance the computational stability of SFA models without materially affecting the estimated output weights or efficiency scores. One may argue in favour of the greater flexibility of the Truncated Normal compared with the Half Normal distribution. However, the results presented here suggest that the half-normal assumption is a marginal improvement against the truncated-normal specification. We believe that the Half-Normal specification offers computational advantages. Hence, the AER may want to further examine this assumption in the context of the standard model as the dataset expands over time. Although it does not resolve all modelling issues, especially in the short-period sample, it mitigates some and provides a workable pathway for addressing others. Indeed, the various unresolved issues may be more related to data limitations, omitted factors, or other empirical challenges rather than the inefficiency distribution itself. We consider that the Half-Normal specification represents the more parsimonious choice, which provides a suitable assumption for exploring more computationally challenging models such as time-varying inefficiency models.

3.2 Alternative methods of estimating the LSE models

The standard LSE employs the panel-corrected standard errors (PCSE) technique with inefficiencies measured by fixed effects only for Australian DNSPs. The Stata implementation is via **xtpcse *y xlist*, c(a) het**, where *xlist* includes fixed effects for Australian DNSPs, and the options specify panel-level heteroskedasticity with a first-order autocorrelation structure (AR1).

The standard LSE method is a particular form of the corrected ordinary least squares (COLS) method of estimating an efficiency frontier. This section briefly summarises the results of testing an alternative COLS method, which estimates an inefficiency measure for all DNSPs in the sample. This is a panel data fixed effects model, estimated using Stata's command **xtreg *y xlist*, fe**.

The results of the fixed effects analysis show that the coefficients obtained from this model differ substantially from those in the standard model presented in the 2024 benchmarking report (Quantonomics 2024: Appendix C). With respect to the first-order output coefficients,[31] the coefficient on customer numbers is not statistically significant in either the CD or TLG models in either the long or short period. In the short period models, it is negative. The coefficients on circuit length and RMD are all positive. However, circuit length is insignificant in the long period CD model, and RMD is insignificant in the short-period TLG model. These results suggest that the Fixed Effects model fails to produce reliable coefficients for the three main outputs, particularly due to lack of statistical significance and monotonicity issues.

The rho value is high in both models, indicating that the majority of the variation in log opex is due to systematic DNSP-specific factors. This is consistent with Economic Insights' previous observations that much of the sample data variation relied on to identify the main parameters of the model is cross-sectional variations.

Monotonicity violations are extremely high, representing 95.7 per cent of all observations on Australian DNSPs in the TLG long-period sample, and 64.5 per cent of all DNSPs. In the short sample, they represent 100.0 per cent of Australian DNSP observations and 73.1 per cent of all DNSPs.

The efficiency scores are generally very low, and the rankings differ markedly from those observed in the standard models. In the long sample period, the average Australian DNSP efficiency score is 0.337 in the CD model and 0.139 in the TLG model. In the short sample, the corresponding efficiency scores are 0.645 and -0.072 respectively.

These results underscore that expanding the fixed effects specification from only Australian DNSPs to all DNSPs, significantly affects the parameter estimates overall. Further, the model produces unreliable efficiency scores. It does not represent an improvement over the standard LSE method.

3.3 Including additional outputs

The standard econometric opex cost function used in the ABR incorporates three outputs: customer numbers, circuit length, and ratcheted maximum demand (RMD). In contrast, the TFP index analysis includes five outputs—the same three, plus energy delivered and customer minutes off supply (CMOS). The exclusion of the latter two variables from the econometric analysis was primarily due to concerns about data availability and reliability for international DNSPs (Economic Insights, 2014). Additionally, there were concerns about potential multicollinearity arising from their inclusion. We have been able to obtain for international DNSP data for energy delivered and, with some challenges, we have extracted measures of

[31] The first-order output coefficients represent the cost-output elasticities, in the TLG model only at the sample mean values of the outputs (since the log outputs are centred at the sample means).

the System Average Interruption Duration Index (SAIDI) on a similar basis to the measure used by the AER for calculating CMOS.

This section summarises the results when including energy delivered and CMOS. Details are provided in Attachment A.

In these models, the role of CMOS differs from that in the productivity index analysis. In the index analysis, CMOS is a negative output, because the costs to customers of outages represents the negative value of this output. In these econometric models, the coefficient on CMOS reflects the partial effect of customers interruption on opex, which will be associated with the cost of dealing with outages through reactive maintenance. More outages will require more work in rectifying the supply failures. However, it is also the case that higher levels of routine (preventative) maintenance should reduce the frequency of outages. The sign of the coefficient on CMOS will depend on which effect is more important. More importantly, there is an endogeneity issue with the inclusion of CMOS as an explanatory variable for opex, because it is not fully exogenous.

3.3.1 Models with 5 outputs

In the long period, all four models (SFACD, SFATLG, LSECD and LSETLG) yield positive coefficients for the five output variables, although not all are statistically significant at the 5 per cent level. In the LSECD and SFACD models, Energy Delivered is not significant, while in the LSETLG model, CMOS lacks significance. The SFATLG model shows two insignificant outputs: Circuit Length and Energy Delivered. In the short period, results are more mixed. Most outputs are of the expected sign (positive) and significant. However, in the LSECD model, RMD is negative and not significant. In the LSETLG model, CMOS is positive but not significant. The SFACD model shows Customer Numbers and RMD as positive but not significant. In the SFATLG model, Customer Numbers are negative and not significant, and RMD is positive but not significant. Unlike the standard models where the SFATLG specification for the short period failed to converge, in this extended specification it achieves convergence.

Overall, including the two additional outputs generally worsens model performance in terms of monotonicity violations when compared to the standard specification. In the long period, MVs for Australian DNSPs account for 72.2 per cent of observations under LSETLG and 77.8 per cent under SFATLG. For the full sample, the corresponding totals are 48.3 per cent for LSETLG and 71.2 per cent for SFATLG. In the short period MVs become even more prevalent. The total violation rate for Australian DNSPs reaches 98.7 per cent under both LSETLG and SFATLG models. For the full sample, violations rise to 70.0 per cent for LSETLG and 89.8 per cent for SFATLG.

Overall, the five-output specification produced weaker results compared to the standard ABR24 models. This is primarily due to the lack of statistical significance in some output coefficients and a substantial increase in monotonicity violations across all TLG models. The only apparent advantage is that the SFATLG model converged in the short-period estimation. However, this gain is limited, as the resulting estimates are not considered reliable.

3.3.2 Models with 4 outputs including CMOS

In this model the inclusion of CMOS is the only change to the standard model. In the long period, all four models (LSECD, LSETLG, SFACD, and SFATLG) produce positive and statistically significant coefficients for all four output variables. Similar to the standard models, the short period SFATLG model does not achieve convergence. In the three successfully estimated short period models, the estimated output coefficients are all positive and statistically significant.

The frequency of MVs for the LSETLG and SFATLG models in the long and short periods is broadly similar to the standard models. In the long period, the frequency of MVs among Australian DNSPs is 34.6 per cent for LSETLG (compared to 22.2 per cent for the corresponding standard model) and 71.4 per cent for SFATLG (compared to 79.5 per cent). In the short sample period, the LSETLG model has MVs in 65.4 per cent of observations on Australian DNSPs (compared to 48.7 per cent in the standard model). Hence, the inclusion of the additional output variable does not result in any improvement regarding MVs.

The efficiency scores produced by the models that include CMOS are, in most cases, similar to those produced by the standard models. This is especially so for the LSE models. For example, in the long-period LSECD and LSETLG models, the average absolute percentage differences in efficiency scores from the standard model are 0.4 percent and 0.9 per cent respectively, and the difference in the average DNSP's efficiency score is –0.2 per cent and 0.4 per cent respectively. In the SFACD and SFATLG models the efficiency scores tend to be slightly higher in the CMOS models. The average absolute percentage differences in efficiency scores from the standard model are 1.9 percent and 5.0 per cent respectively, and the difference in the average DNSP's efficiency score is 1.4 per cent and 3.2 per cent respectively.

In this model, the four output coefficients are all positive and statistically significant. However, there is no improvement and, if anything, a slight deterioration in MV performance compared to the standard specification. Like the standard model, the SFATLG model for the short period fails to converge. The efficiency scores are similar. It may be that this model represents a slight improvement over the standard model by including a variable that appears to be statistically significant. However, if so, it is a very small improvement. Furthermore, there is a concern that including the CMOS variable may introduce endogeneity.

3.3.3 Models with 4 outputs including Energy

In this model the inclusion of Energy Delivered is the only change to the standard model. Unlike in the standard SFATLG model, this extended specification achieves convergence in the short period. In the long period, all four models generally produce positive coefficients for the four output variables, and except for a few instances, they are statistically significant. In the LSECD model, RMD is not statistically significant at the 5 per cent level and in the SFATLG model, Circuit Length is not significant (although each is significant at the 10 per cent level). In the SFACD model, Energy Delivered lacks statistical significance.

In the short period, RMD is not statistically significant in any of the four models, and is negative in the LSECD and SFACD models. In the LSECD and LSETLG models, the other three outputs are all positive and statistically significant. However, in both the SFACD and SFATLG models, the Customer Numbers coefficient is not statistically significant, and in the SFATLG model it is negative.

The inclusion of the energy delivered output leads to a considerable increase in MVs compared to the standard models. In the long period, the rate of MVs among Australian DNSPs is 59.8 per cent for LSETLG and 78.6 per cent for SFATLG (compared to 22.2 per cent and 79.5 per cent for the corresponding standard models). For all DNSPs, the rate is 44.5 per cent for LSETLG and 67.3 per cent for SFATLG (compared to 21.9 per cent and 45.4 per cent for the standard models). In the short period, MVs for Australian DNSPs, reach 98.7 per cent for LSETLG and 100.0 per cent under SFATLG (compared to 48.7 per cent and 100.0 per cent for the standard models). For the full sample, violations rise from 68.6 per cent under LSETLG to 88.9 per cent under SFATLG (compared to 36.6 per cent and 71.4 per cent for the standard models).

As with the preceding models that include CMOS, in the models that includes Energy Delivered the efficiency scores produced by the LSECD and LSETLG models are similar to those produced by the standard models. However, the SFACD and SFATLG models produce efficiency scores that vary to a greater degree from those of the corresponding standard models. In the long-period SFACD model, the efficiency score of the average Australian DNSP is 2.0 per cent lower in the model that includes Energy compared to the standard model, and the average absolute percentage difference in efficiency scores is 2.4 per cent. In the long-period SFATLG model, the efficiency score of the average Australian DNSP is 3.5 per cent higher than in the standard model, and the average absolute percentage difference in efficiency scores is 7.8 per cent.

In summary, compared to the standard ABR24 models, this specification performs worse in several respects. Several output coefficients are not statistically significant, and the Translog models exhibit an increase in MVs. While the SFATLG model did achieve convergence in the short-period estimation, the resulting model is not reliable. These observations tend to support the previous concerns of Economic Insights that the inclusion of Energy Delivered results in excessive multicollinearity.

3.4 Input prices

As discussed in section 2.1, cost functions typically include input prices. In the AER's standard opex cost function, the dependent variable is nominal opex divided by a composite index of opex input prices. The resulting function could be interpreted as the demand function for non-capital inputs. In addition to being a function of output and OEFs, it may be a function of the prices of capital and non-capital inputs.[32] Input demand functions should be homogeneous of degree zero; ie, if all input prices increase by an equal proportion, then the demand for the input must remain unchanged.[33] Including input prices only in relative terms, such as a ratio or log ratio of capital and non-input prices, satisfies that requirement.

We have explored the inclusion of the log ratio of the prices of non-capital and capital inputs. The purpose is to determine whether these variables have any effect on log real opex, which depends on whether there is substitution between inputs. For this purpose, we have developed a price index for capital services input, based on the method of Jorgenson (1967), and consistent with the user cost of capital method used by the AER in calculating input index weights.

The weighted average cost of capital, which is one element of the price of capital inputs, is equal to the nominal risk-free rate of return and a weighted average risk premium for debt and equity. The risk-free rate of return is based on the yields of government long-term bonds and varies between jurisdictions. The risk premium is assumed to be common across jurisdictions. This is a simplifying assumption, sufficient for the present exploratory analysis, and given the potential limitations of confidently quantifying differences in risk premia between jurisdictions with available regulatory disclosure and other readily available public data.

The coefficient on log ratio of opex input price to capital input price is expected to be negative. It is equal to the elasticity of opex input demand with respect to the opex input price (which should be negative), and the negative of the cross-price elasticity of opex input demand with respect to the capital input price (which should be positive if the two inputs are substitutes). The estimation results show that the coefficient on the log input price ratio is negative and statistically significant at the 5 per cent level across all estimated models. Hence, the sign of the coefficient is consistent with economic theory in all models. The price elasticities are small, ranging from 0.06 to 0.15 (in absolute value), which suggests that there is not a great deal of substitutability between opex (non-capital) inputs and capital inputs.

The findings suggest that the input price ratio is potentially an omitted variable. That said, the inclusion of the input substitution variable did not result in any notable improvements over the standard models in terms of goodness-of-fit or convergence of the SFA-TLG model in the short period. Further, the introduction of the input price ratio has little effect on efficiency score estimates. The rankings of average efficiency scores for the input-substitution models show only minor differences compared to the ABR24 results.

[32] This is because, according microeconomic theory (as well as practical intuition) a firm may adjust its mix of capital and opex inputs in response to relative changes in their prices, seeking to minimise total costs for a given level of output given the technological substitution possibilities.

[33] When there is no change in relative input prices (for given output levels), there is no change in the optimal input mix, and hence the quantities of each input demanded remain unchanged.

In most cases there is a very modest improvement in the frequency of MVs relative to the standard TLG models. However, this not consistent across both the LSE and SFA models in both the short and long periods, and the improvement is only marginal. This suggests that the inclusion of the input price ratio does not substantially improve the problem of monotonicity violations in the TLG models.

The main implication of these findings is that including the log ratio of input prices into the model may yield some marginal improvements.

There remain measurement issues in relation to the price of capital inputs, for example relating to cost of capital parameters and depreciation rates. Assumptions were needed to construct some of these parameters, and there is room for improvement. For example, there remain questions about the calculation of the risk premium in the cost of capital, including the imputation of missing values for these parameters in jurisdictions and years where they are unavailable; whether there should be any averaging of parameters across jurisdictions; and whether all of these parameters should be treated as constants or be allowed to vary over time.

3.5 Concluding Comments

The analyses summarised in this chapter address the opportunities for refining the current models listed in section 2.1. The main points are:

- There are good reasons for preferring the assumption of half-normally distributed inefficiencies rather than truncated-normal inefficiencies. The literature, especially Ritter and Simar (1994), finds that the latter can more often have computational identification problems. Estimation of the half normal model using the ABR24 data sample shows that the half-normal model resolves the convergence issue in the short-period SFATLG model. However, some computational problems remain, such as irregularities in the efficiency scores. This indicates that while the half-normal specification is computationally more stable, it does not fully resolve all issues. Although it does not address the problem of monotonicity violations, it is not targeted at solving the MV problem. The estimated efficiency scores are very similar to those estimated by the standard model using the truncated-normal distribution of inefficiencies.
- The standard LSE model can be characterised as a form of the COLS method of estimating an efficiency frontier. A common way of estimating a COLS model is to use panel data fixed effects estimation. This implies a measure of inefficiency for every DNSP in the sample, not only Australian DNSPs. Testing this approach with the ABR24 data sample shows that the panel data fixed effects approach fails to produce reliable coefficient for the three main outputs, particularly due to lack of statistical significance. This is because much of the sample data variation relied on to identify the main parameters of the model is cross-sectional variation. It also produces extremely high frequencies of monotonicity violations the estimated and efficiency

scores are generally unrealistically low. These results show that expanding the fixed effects allowed for in the model yields unreliable results.

- Although the TFP index analysis includes five outputs, Economic Insights did not include two of them in the opex cost function analysis: energy delivered and CMOS (which is a 'bad' output, as it imposes a cost on consumers). We have tested the inclusion of these two outputs, and the results suggest that the inclusion of energy delivered introduces an increased degree of multicollinearity, making it more difficult to obtain sufficiently precise and reliable estimates of the output parameters. This reduces the performance of the models in several respects including increasing MVs. The inclusion of CMOS does introduce a statistically significant variable, although its interpretation differs from its use in the TFP index analysis. Although it appears to represent a small improvement, in all respects its effect on the model is marginal. On the other hand, there is a concern that the CMOS variable is not independent of opex, and its inclusion may introduce endogeneity. Hence, we find no benefits to including these other two outputs.

- The inclusion of the log ratio of capital and non-capital input prices would permit the opex cost function model to take account of input substitution. This is a long-run phenomenon which has not to date been the focus of opex efficiency modelling. We have tested this approach by constructing a price index of capital inputs; however, some assumptions need to be made in this exercise. It is found that the log ratio of input prices is a statistically significant determinant of real opex, and its sign is correct (negative). The implied substitution effect on opex is small. The inclusion of the log ratio of input prices into the model may yield some marginal improvements. However, there remain measurement issues in relation to the price of capital inputs to be resolved and further developed before it could be used in the benchmarking model.

4 Feasibility of Time Varying Inefficiency Approaches

In this section we examine several alternative time-varying efficiency models. Section 4.1 begins by replicating Frontier Economics' proposed model. The following sections examine selected approaches for testing alternative methodologies to incorporate time-varying inefficiency into the AER's benchmarking framework which were outlined in Section 3.

The following main approaches are discussed:

- *Section 4.2*: The theoretically attractive four-components method of Colombi et al (2014). In this approach, the residual, after excluding inefficiency, has a transitory component which is the idiosyncratic disturbance, and a non-transitory component, which is interpreted as the effect of unobserved heterogeneity. The one-sided random component associated with inefficiency, also has transitory and non-transitory components. This is intended to allow for a more refined understanding of the unobserved elements in a regression's residuals by distinguishing between systemic and situational factors.
- *Section 4.3*: Models relying on the *sfpanel* model option *bc95* (other than the Frontier Economics model discussed in section 4.1). This represents the Battese and Coelli 1995 approach which allows the pre-truncation mean of the inefficiency distribution (ie, μ) to be a linear function of explanatory variables. Here time trend variables are used. This is a highly flexible command, which also allows the variance of the inefficiency distribution and the variance of the idiosyncratic error to both be linear functions of other variables. Eight variants of this method were tested.
- *Section 4.4*: Models relying on the *sfpanel* model option *kum90*, which is a generalisation of the Kumbhakar (1990) model. As discussed in section 3.1.5, the time-varying inefficiency term $u_{it} = g(\boldsymbol{z}'_{it}\boldsymbol{\delta}).u_i$, where $u_i \sim \mathcal{N}^+(0, \sigma_u^2)$, and $g(\boldsymbol{z}'_{it}\boldsymbol{\delta}) = [1 + \exp(\boldsymbol{z_{it}}'\boldsymbol{\delta})]^{-1}$. Here the z-variables are time trend variables. Four variants of this method were tested.
- *Section 4.5*: Alternative modelling approaches using LSE models, which incorporate not only firm-specific fixed effects for Australian DNSPs to measure inefficiency, but also interactions of those effects with time trends to allow for time-varying inefficiency. Four variants of this method were tested.
- *Section 4.6*: The Cornwell et al (1990) method described in section 2.2. This approach is based on corrected ordinary least squares (COLS) and provides for a highly flexible relationship between individual firm inefficiencies and time, involving a great many parameters. Two variants of this method were tested.

Regarding the BC95 and Kumb90 sets of models, and to some extent with the LSE models, in each model we test some alternative methods of incorporating time trends to measure time-varying inefficiency. These are

(a) Jurisdictional time trends (JTT) for Australia, NZ and Ontario;

(b) Individual time trends for Australian DNSPs (used in the LSE models only); or

(c) Individual time trends for Australian DNSPs, and Jurisdictional time trends for NZ and Ontario.

Technical change is modelled via two alternative methods:

(i) The time trend variable used in the standard model. When jurisdiction-specific or firm-specific time-varying inefficiency effects are included in the models, this general time trend seeks to measure technical change.

(ii) The full sample period is divided into 6 periods:[34]

- 2006-2008
- 2009-2011
- 2012-2014
- 2015-2017
- 2018-2020
- 2021-2023

In the long-period, 5 of these periods are represented by dummy variables for each period. In the short period 3 are represented by dummies. This approach, referred to as 'general technical change' (GTC), follows Baltagi and Griffin (1988). Sets of years are used rather than individual years to conserve degrees of freedom.

In addition to these alternatives, in the bc95 models we have also tested whether the variance of the idiosyncratic disturbance is related to cross-sectional measures of DNSP scale. The mean values of log output variables for each DNSP were used. When a single measure of scale was used, such as mean log customer numbers per DNSP, or the weighted average of the mean values of all three outputs per DNSP, the scale effect was found to be statistically insignificant. Alternatively, when the three outputs were included separately in commands that could accommodate this, they were found to be significant, but with a mix of positive and negative signs. This raises a concern that the resulting correlations may be spurious, since the outputs are correlated with each other. For these reasons, we do not report the models where heteroscedasticity of the idiosyncratic disturbance is determined by DNSP scale.

For convenience, we have adopted a naming convention using suffixes to identify the different modelling approaches. These are:

[34] The choice of periods as three-year intervals is convenient because the short sample excludes the first two periods.

- *Pre-truncation mean of inefficiencies:* HN refers to either the half-normal distribution of inefficiencies (which means that the 'pre-truncation mean', μ, is zero), or where there is a linear function determining μ, such that there is no constant term.
- *Determinants of time-varying inefficiency:* JTT indicates that time-varying inefficiency is modelled using jurisdiction-specific time trends; ADTT means it is modelled using Australian DNSP-specific time trends; and AJTT mean there are time trends for each Australian DNSP and jurisdiction-specific time trends. These suffixes are not included for some SFA methods which have a specific approach to modelling time-varying inefficiency.
- *Technical change:* The GTC suffix indicates that technical change is modelled using dummy variables for distinct periods (5 for the long period and 3 for the short period). If GTC is absent, the standard method is applied (i.e. using the year variable to proxy technical change).

The discussion focuses primarily on the long-period specifications of the time-varying inefficiency models, as these account directly for changes in inefficiency and therefore reduce the reliance on the short-period results. This focus is also warranted because the short-period models generally performed poorly. Further details on both the long- and short-period time-varying models are provided in Attachment B.

4.1 Frontier Economics' Specification

As part of the Evoenergy submission to Quantonomics (2024), Frontier Economics (2025) provided several comments on approaches to improve the AER's econometric benchmarking estimates. In particular, it considers the assumption of time-invariant inefficiencies as serious misspecification. Frontier Economics also noted that the jurisdictional time-trend models explored by Quantonomics (2024) suggest likely differences across jurisdictions in the effects of changes in OEFs over the sample period. These differences, it argues, should be captured through separate time trends. Frontier Economics also conducted a residual analysis, suggesting that omitted variables may be captured in the residuals, indicating potential model misspecification.

Frontier Economics recommended using the Battese-Coelli 1995 (*bc95)* model, in which the 'pre-truncation mean' of the stochastic inefficiency terms is a linear function of *z*-variables, as a starting point for investigating time-varying inefficiency in SFA models. Using this method, Frontier Economics presented an example of an SFATLG time-varying inefficiency model using the short period. It emphasised that this model was intended solely to demonstrate the use of the *sfpanel* command in Stata and was not a proposed or preferred specification. Frontier Economics provided the Stata code and output files used in its analysis, enabling us to replicate its model. In doing so, we have also run the corresponding SFACD version.

In the Frontier Economics model, the linear function that determines the 'pre-truncation mean' of the inefficiency distribution includes an intercept, and the z-variables are dummy variables for 60 of the 61 DNSPs in the sample and jurisdiction-specific time trend variables.[35] There is no explanation for DNSP-specific fixed effects in determining u_{it} additional to the random effect in ϖ_{it}; recalling from section 2.2.5 that in this model the inefficiency term u_{it} can be expressed as: $u_{it} = \delta_0 + \boldsymbol{z}_{it}'\boldsymbol{\delta} + \varpi_{it}$, where ϖ_{it} is a random variable distributed as a truncation of a normal distribution with zero mean and constant variance, and with the truncation point being: $u_{it} \geq -\delta_0 - \boldsymbol{z}_{it}'\boldsymbol{\delta}$.

To produce starting values for the main TLG model, Frontier Economics first estimated simplified models. Unlike the AER's approach, which first estimates the TLG model using ordinary least squares (OLS), Frontier Economics has two steps using OLS and then SFA to estimate the Cobb-Douglas (CD) model. For estimating the final SFA model, initial values equal to zero were used for the second-order TLG terms, and the coefficients of the z-variables (the country-specific dummies and the jurisdiction-specific trends). The Davidon–Fletcher–Powell (DFP) algorithm was chosen for numerical optimisation.

Tables 4.1.1 and 4.1.2 present our replication of Frontier Economics Model in both the CD and TLG version for both the long and short sample periods. The short-period TLG results can be compared to Table 1.2.2 of Frontier Economics (2025). Frontier Economics' CD short-period results are available in its supporting files. We are able to replicate Frontier Economics' results to a reasonable degree of precision, although we are unable to obtain the exact parameters reported by Frontier Economics using the *sfpanel* command.[36] Most of these differences are under 1 per cent. Discrepancies of this kind are common when using complex MLE models.

Frontier Economics also included in its report a figure showing the efficiency scores produced by its SFATLG short-period model (Frontier Economics 2025, p.11). However, these are not the direct results of its model. The directly estimated efficiency scores are shown in Table 4.1.3 and Figure 4.1.1. As Frontier Economics noted in its Stata program, its SFATLG model did not produce efficiency scores for all DNSPs because of extremely large negative values of the inefficiency intercept for two DNSPs,[37] which led to a division-by-zero error in the calculation of the conditional expectation of the inefficiency term. To work around this problem, Frontier Economics "rearrang[ed] the BC1988 formula to use the lognormal functionality in Stata".[38]

[35] The time trend variables—aus_yr, nz_yr, and ont_yr—are each defined as yr – 2017.5 for their respective countries. The reference year, 2017.5, is the midpoint of the 2012–2023 period (or 2014.5 for long-period models), which Frontier Economics suggests facilitates both interpretation of parameter estimates and optimisation of the model.
[36] We ran Frontier Economics' Stata programs using both Stata version 14—which appears to be the version used by Frontier Economics—and Stata 19. In both cases, the results differed from Frontier Economics' outputs, but only for the *sfpanel* command. All other estimation commands were replicated without discrepancies in values.
[37] More specifically, 2 Australians DNSPs, PCR and UED.
[38] Coding comment in *2025-04-14 SF models.do*.

This was done by manually reconstructing the efficiency scores for Australian DNSPs in the SFATLG model.

The frequency of monotonicity violations (MVs) in the TLG versions of these models are shown in Table 4.1.4.

Table 4.1.1 Frontier Economics model results (long period)*

	SFACD			SFATLG		
Variable	*Coefficient*	*SE*	*t–ratio*	*Coefficient*	*SE*	*t–ratio*
Frontier						
ly1	0.475	0.047	10.13	0.639	0.061	10.54
ly2	0.136	0.022	6.29	0.087	0.022	3.89
ly3	0.365	0.047	7.78	0.226	0.061	3.730
ly11				-0.196	0.646	-0.300
ly12				-0.023	0.221	-0.100
ly13				0.266	0.455	0.580
ly22				-0.010	0.094	-0.100
ly23				0.083	0.143	0.580
ly33				-0.425	0.356	-1.200
lz1	-0.069	0.028	-2.49	-0.097	0.038	-2.580
yr	0.009	0.001	8.11	0.007	0.001	6.300
jur2	0.048	0.041	1.17	0.134	0.090	1.480
jur3	0.216	0.036	5.94	0.227	0.101	2.240
_cons	-7.654	2.177	-3.52	-3.675	2.124	-1.730
Mu						
aus_yr	-0.019	0.003	-7.15	-0.016	0.003	-5.440
nz_yr	0.022	0.002	10.44	0.022	0.002	10.700
ont_yr	-0.009	0.002	-3.77	-0.004	0.002	-2.320
_cons	0.361	0.035	10.18	0.312	0.033	9.340
Usigma						
_cons	-4.778	0.145	-32.85	-4.776	0.134	-35.540
Vsigma						
_cons	-4.916	0.083	-59.28	-5.064	0.086	-59.220
sigma_u	0.092	0.007	13.75	0.092	0.006	14.880
sigma_v	0.086	0.004	24.12	0.080	0.003	23.390
lambda	1.071	0.009	116.65	1.154	0.009	134.010
LLH	892.69			911.91		
Iterations #	752**			784**		
N	1098			1098		

* The coefficients on the DNSP dummies used as z-variables are omitted.

** Stata reports an error in MLE convergence.

Table 4.1.2 Frontier Economics models results (short period)*

	SFACD			SFATLG		
Variable	*Coefficient*	*SE*	*t–ratio*	*Coefficient*	*SE*	*t–ratio*
Frontier						
ly1	0.280	0.068	4.12	0.474	0.053	8.93
ly2	0.190	0.021	8.90	0.225	0.022	10.20
ly3	0.516	0.072	7.14	0.204	0.060	3.41
ly11				0.540	0.373	1.45
ly12				-0.114	0.074	-1.54
ly13				-0.384	0.351	-1.09
ly22				0.199	0.027	7.50
ly23				-0.045	0.075	-0.61
ly33				0.243	0.337	0.72
lz1	-0.101	0.022	-4.62	0.013	0.027	0.50
yr	0.004	0.002	1.59	-0.001	0.002	-0.33
jur2	0.015	0.034	0.43	-0.264	0.045	-5.88
jur3	0.158	0.036	4.42	-0.045	0.037	-1.21
_cons	2.749	4.490	0.61	11.379	3.574	3.18
Mu						
aus_yr	-0.039	0.004	-11.05	-0.040	0.003	-11.85
nz_yr	0.033	0.003	11.61	0.033	0.002	13.89
ont_yr	-0.006	0.003	-2.37	-0.005	0.002	-2.05
_cons	0.318	0.031	10.39	0.229	0.030	7.59
Usigma						
_cons	-6.917	1.218	-5.68	-6.773	0.679	-9.97
Vsigma						
_cons	-5.189	0.175	-29.60	-5.270	0.117	-45.02
sigma_u	0.031	0.019	1.64	0.034	0.011	2.95
sigma_v	0.075	0.007	11.41	0.072	0.004	17.09
lambda	0.421	0.025	16.58	0.472	0.015	30.95
LLH	811.92			835.024		
Iterations #	512**			617**		
N	732			732		

* The coefficients on the DNSP dummies used as z-variables are omitted.

** Stata reports an error in MLE convergence.

Table 4.1.3 Frontier Economics model average efficiency scores (long period)*

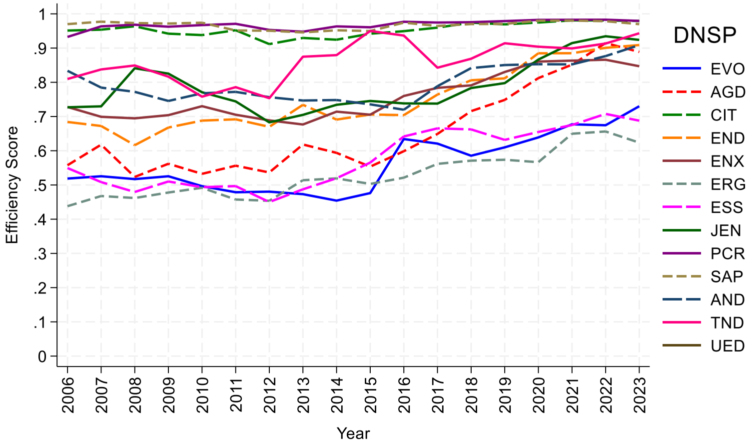
	Long Period		Short Period	
	SFACD	SFATLG	SFACD	SFATLG
EVO	0.562	0.515	0.565	0.676
AGD	0.657	0.605	0.710	0.465
CIT	0.954	0.828	0.899	0.946
END	0.749	0.680	0.813	0.610
ENX	0.758	0.730	0.775	0.570
ERG	0.528	0.545	0.672	0.690
ESS	0.577	0.637	0.712	0.874
JEN	0.789	0.770	0.717	0.845
PCR	0.968	.	.	.
SAP	0.967	0.970	0.980	0.983
AND	0.797	0.824	0.784	0.880
TND	0.863	0.813	0.974	0.981
UED	.	.	.	.
Australia	0.764	0.720	0.782	0.774

* These efficiency scores are raw outputs from the model, without applying the lognormal transformation used by Frontier Economics.

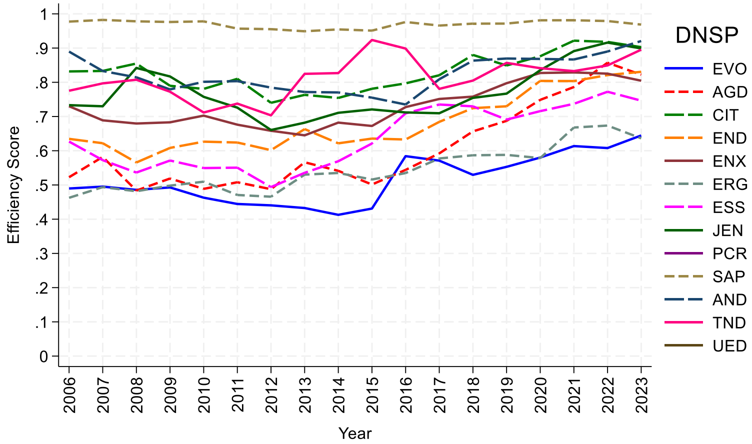
Table 4.1.4 Monotonicity violations in Frontier Economics models

	Long Period				*Short Period*			
Sample	*Cust.*	*CL*	*RMD*	*Total*	*Cust.*	*CL*	*RMD*	*Total*
By DNSP								
EVO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
AGD	0.00	0.00	72.20	72.20	0.00	0.00	100.00	100.00
CIT	0.00	0.00	22.20	22.20	0.00	0.00	0.00	0.00
END	0.00	0.00	5.60	5.60	0.00	0.00	100.00	100.00
ENX	0.00	0.00	0.00	0.00	0.00	0.00	100.00	100.00
ERG	0.00	0.00	0.00	0.00	0.00	0.00	100.00	100.00
ESS	0.00	0.00	0.00	0.00	0.00	0.00	100.00	100.00
JEN	0.00	0.00	0.00	0.00	0.00	0.00	100.00	100.00
PCR	0.00	0.00	0.00	0.00	0.00	0.00	100.00	100.00
SAP	0.00	0.00	0.00	0.00	0.00	0.00	100.00	100.00
AND	0.00	0.00	0.00	0.00	0.00	0.00	100.00	100.00
TND	0.00	0.00	0.00	0.00	0.00	0.00	100.00	100.00
UED	0.00	0.00	0.00	0.00	0.00	0.00	100.00	100.00
By jurisdiction								
Australia	0.00	0.00	7.70	7.70	0.00	0.00	84.60	84.60
New Zealand	0.00	36.80	0.00	36.80	0.00	0.00	10.50	10.50
Ontario	0.00	0.00	12.10	12.10	0.00	16.70	9.20	25.90
Full sample	0.00	11.50	7.40	18.90	0.00	7.90	25.70	33.60

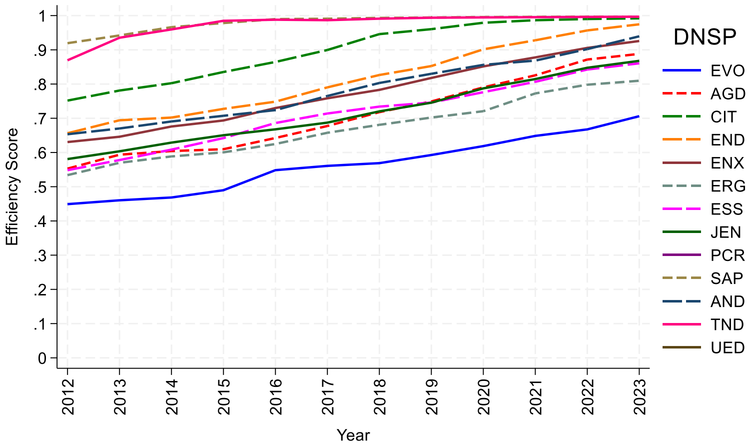
Figure 4.1.1 Efficiency Trends by DNSP- Frontier Economics



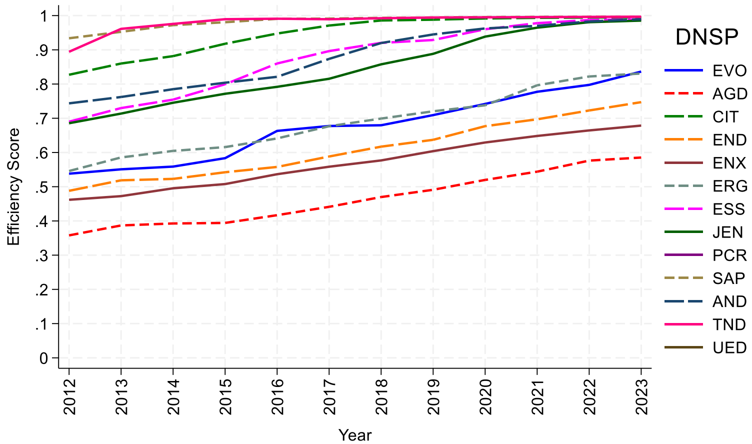
SFACD Long Period



SFATLG Long Period



SFACD Short Period



SFATLG Short Period

We recognise that Frontier Economics presented its approach merely as an example to demonstrate the potential application of the *sfpanel bc95* approach. However, in practice, it highlighted the challenges associated with implementing this method.

The problem of being unable to produce efficiency scores for all the Australian DNSPs is sufficient grounds for rejecting Frontier Economics' presented specification. However, there are at least two other major problems with it:

- For all four models (the CD and TLG models in short and long periods), Stata returned an error message in relation to MLE convergence, indicating failure to converge correctly at the specified convergence criteria.
- The frequency of monotonicity violations (MVs) remains high in the short-period model (see Table 4.1.4). Compared to the JTT models presented in Quantonomics (2024)—which can be used as a benchmark for comparing time-varying inefficiency models—the Frontier Economics model has much poorer MV performance in the short period, and in the long sample period it performs better for Australian DNSPs but worse for the full sample of all DNSPs.

We reiterate that Frontier Economics presented this model discussed in this section as a basis for further discussion and not as a proposed or preferred specification. In the following sections we test several different approaches to introducing time-varying inefficiency. Section 4.3 discusses the results of testing alternative specifications of the bc95 SFA method.

4.2 Colombi et al (2014) Model

This section briefly summarises the results of the Four Components approach, which we consider the most theoretically appealing. More detail is available in Attachment B. Recall that in this method, time varying inefficiency refers to year-to-year variations, but does not specifically include any trended component to inefficiency. The explanatory variables included in this model were the same as those in the standard model, except the jurisdictional dummy variables were excluded because the Four Components model includes a firm-specific effect to capture time-invariant unobserved heterogeneity across firms. There is a time trend variable in the model, which may capture various trended effects including technical change.

Over the long period, the coefficients on the main log output variables have the expected signs and are statistically significant in both the CD and TLG versions of the model. However, in the short period, the coefficient on the log of customer numbers is not statistically significant in either the CD or TLG models.

The model shows a frequency of monotonicity violations in the long period for the Australian DNSPs of 34.2 per cent. For all DNSPs, 13.0 per cent of observations exhibit MVs in the long period. The MV results for Australian DNSPs are intermediate between the standard LSETLG and SFATLG models, whereas the results over all DNSPs are an improvement (Quantonomics, 2024a: 148). In the short period, the MV issue becomes more severe, with

98.7 per cent of Australian observations and 51.8 per cent of the total sample affected. These results indicate that the model is not viable in the short period.

The efficiency scores produced by the Four Components model are not reliable,[39] as they are extremely high, with no DNSP scoring below 0.9. See Table 4.2.1.

Table 4.2.1 Four Components *Average* Efficiency Scores by Australian DNSP

	Long Period				*Short Period*			
Sample	*CD*	*Rank*	*TLG*	*Rank*	*CD*	*Rank*	*TLG*	*Rank*
EVO	0.927	11	0.929	11	0.997	1	0.997	4
AGD	0.921	13	0.925	13	0.997	3	0.997	1
CIT	0.932	4	0.934	4	0.997	9	0.997	10
END	0.931	8	0.934	5	0.997	8	0.997	5
ENX	0.934	1	0.937	1	0.997	6	0.997	2
ERG	0.931	5	0.934	6	0.997	2	0.997	8
ESS	0.926	12	0.929	12	0.997	4	0.997	9
JEN	0.930	10	0.932	10	0.997	5	0.997	3
PCR	0.933	3	0.934	2	0.997	13	0.997	11
SAP	0.931	7	0.933	9	0.997	12	0.997	12
AND	0.933	2	0.934	3	0.997	7	0.997	6
TND	0.931	6	0.933	7	0.997	10	0.997	13
UED	0.930	9	0.933	8	0.997	11	0.997	7
Australia	0.930		0.932		0.997		0.997	

The high efficiency levels may be because this model may attribute a substantial portion of the variation to firm-specific heterogeneity effects rather than inefficiency. This may reflect limitations in the model's ability—or that of its estimation algorithm—to separate inefficiency from other components of the error term in our data sample. This issue seems particularly pronounced in the short-period sample, where the number of observations may be insufficient relative to the complexity of the model. The use of a three-step estimation approach, while simpler from an optimisation perspective, is known to be less statistically efficient than a one-step method and generally requires larger samples to reliably identify each component of the error structure.

Overall, the results indicate that the Four Components model is not suitable for application within the AER's benchmarking framework. Despite its strong theoretical appeal, the model did not perform well in practice. The Four Components model lacks empirical credibility in this context and should not be retained for further analysis.

[39] In this model, the overall efficiency score is calculated as the product of persistent inefficiency and transitory efficiency scores. This is similar to adding together the persistent inefficiency and transitory inefficiency, before the transformation: $\exp(-ineff.)$

4.3 Battese and Coelli (1995) Model

The Battese-Coelli 1995 (*bc95)* method is implemented in Stata using **sfpanel *y xlist*, model(bc95)**. The methodology of this model is detailed in Section 3.1.5. See Attachment B for further details of the models discussed in this section.

In addition to replicating the Frontier Economics specification, presented in section 4.1, we tested eight different variants of the Battese–Coelli model. However, only one (referred to here as BC95-JTT-HN) successfully produced results consistently for both the CD and TLG forms in both the long and short sample periods.

The models we tested are characterised by the mean of the inefficiency term u_{it}, in the distribution of inefficiencies: $u_{it} \sim N^{+}(\mu_{it}, \sigma_u^2)$ being determined by jurisdictional or DNSP-specific time trends, with no constant term. In the successful model, the trends that affect μ are three jurisdiction-specific time trends, and the constant term is excluded. It is found that the *bc95* model is increasingly difficult to estimate successfully in this sample as complexity is increased (ie, more determinants of the pre-truncation mean or allowance for heterogeneity).

Section 4.3.1 briefly describes the results for the successful BC95-JTT-HN model. Section 4.3.2 overviews the unsuccessful models, which all involve estimating more parameters than the successful model.

4.3.1 BC95-JTT-HN: Jurisdiction time trends as determinants of μ

In this specification, SFACD and SFATLG models were estimated for both the long period and short period, converging after between 125 and 212 iterations. However, in both SFACD and SFATLG short-period estimation, the models returned the message "*cannot compute an improvement —flat region encountered*", indicating that the optimisation reached an area of the likelihood surface where further improvement was not possible. As a result, the estimates produced under these conditions may be unreliable or unstable.[40]

The results of estimating the Cobb-Douglas and Translog versions with the long sample are presented in Table 4.3.1. In both the SFACD and SFATLG long period models, the coefficients on the primary log output variables have the expected signs and are statistically significant. These represent the output elasticities, for the TLG model at the sample mean values of the outputs.

[40] See Section 5.1.1 for a further discussion of this convergence issue.

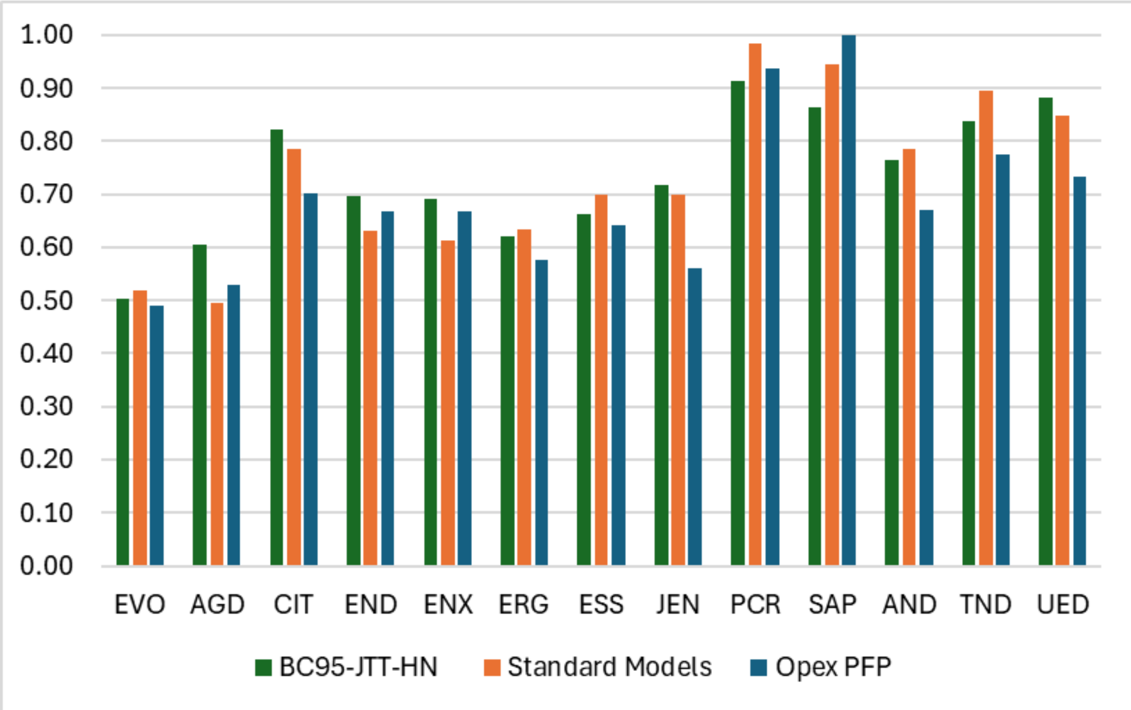
Table 4.3.1 BC95-JTT-HN Parameter Estimates (long sample period)

	CD			*TLG*		
Variable	*Coefficient*	*SE*	*t–ratio*	*Coefficient*	*SE*	*t–ratio*
Frontier						
ly1	0.496	0.036	13.62	0.448	0.038	11.75
ly2	0.113	0.016	7.17	0.103	0.016	6.54
ly3	0.381	0.034	11.18	0.415	0.034	12.07
ly11				0.334	0.289	1.16
ly12				0.026	0.065	0.40
ly13				-0.538	0.237	-2.27
ly22				-0.005	0.023	-0.23
ly23				0.016	0.054	0.29
ly33				0.680	0.194	3.51
lz1	-0.181	0.014	-13.17	-0.178	0.015	-12.25
yr	0.010	0.001	10.09	0.011	0.001	11.05
jur2	0.128	0.034	3.72	0.070	0.040	1.76
jur3	0.394	0.033	11.87	0.298	0.038	7.78
_cons	-9.849	1.925	-5.12	-12.368	1.992	-6.21
Mu						
aus_yr	0.000	0.000	0.26	0.000	0.000	0.57
nz_yr	0.000	0.000	0.07	0.000	0.000	0.28
ont_yr	-0.001	0.000	-2.89	-0.001	0.000	-2.53
Usigma						
_cons	-1.768	0.239	-7.41	-2.044	0.262	-7.79
Vsigma						
_cons	-4.521	0.132	-34.15	-4.457	0.135	-32.95
sigma_u	0.413	0.049	8.38	0.360	0.047	7.63
sigma_v	0.104	0.007	15.11	0.108	0.007	14.78
lambda	3.960	0.048	82.23	3.342	0.046	72.39
LLH	261.76			318.69		
Iterations #	125			164		
Pseudo Adj R^2	0.997			0.997		
BIC	-432.50			-504.35		
N	1,098			1,098		

Regarding MVs in the SFATLG long-period model, 9.8 per cent of all Australian observations are violations, and 14.8 per cent of all sample observations. One of the Australian DNSPs has excessive MVs in the long sample. This is a better result than the standard SFATLG and LSETLG models in the long sample period (which have MVs of 79.5 per cent and 22.2 per cent respectively for Australian DNSPs, and 45.4 per cent and 21.9. per cent respectively for all observations). Although it is an improvement, it does not entirely solve the problem of excessive MVs.

Figure 4.3.1 presents the average of the BC95-JTT-HN SFACD and SFATLG efficiency scores for each Australian DNSP in the long period and compares them to the efficiency scores from the standard models and the comparative Opex PFP. The efficiency scores and rankings of the BC95-JTT-HN models are broadly aligned with those from the standard models.

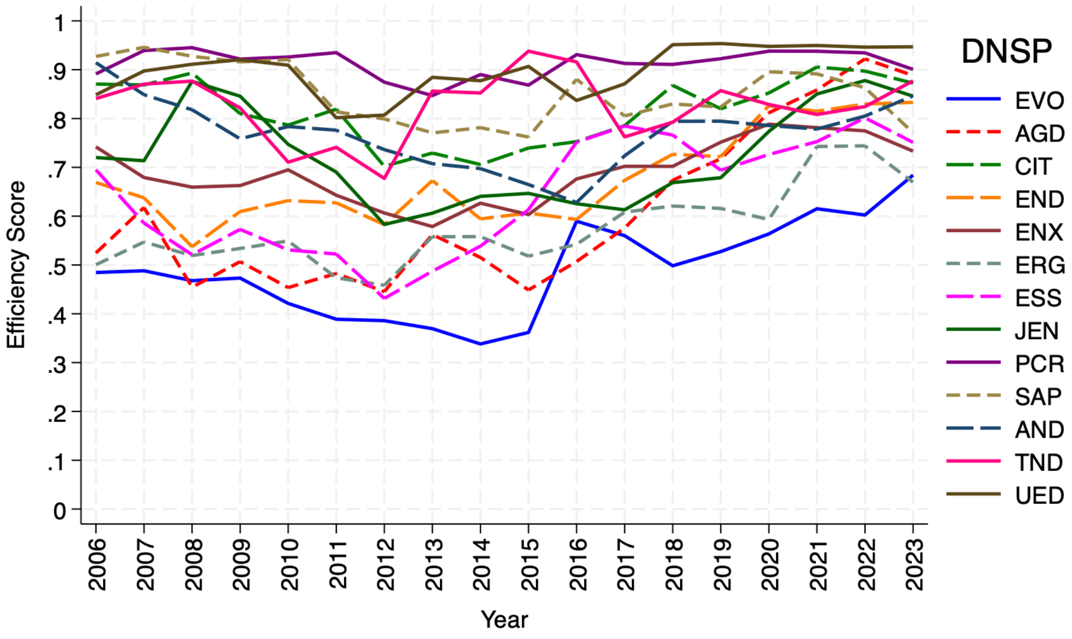
Figure 4.3.1 Average Efficiency Scores by DNSP (2006–2023)



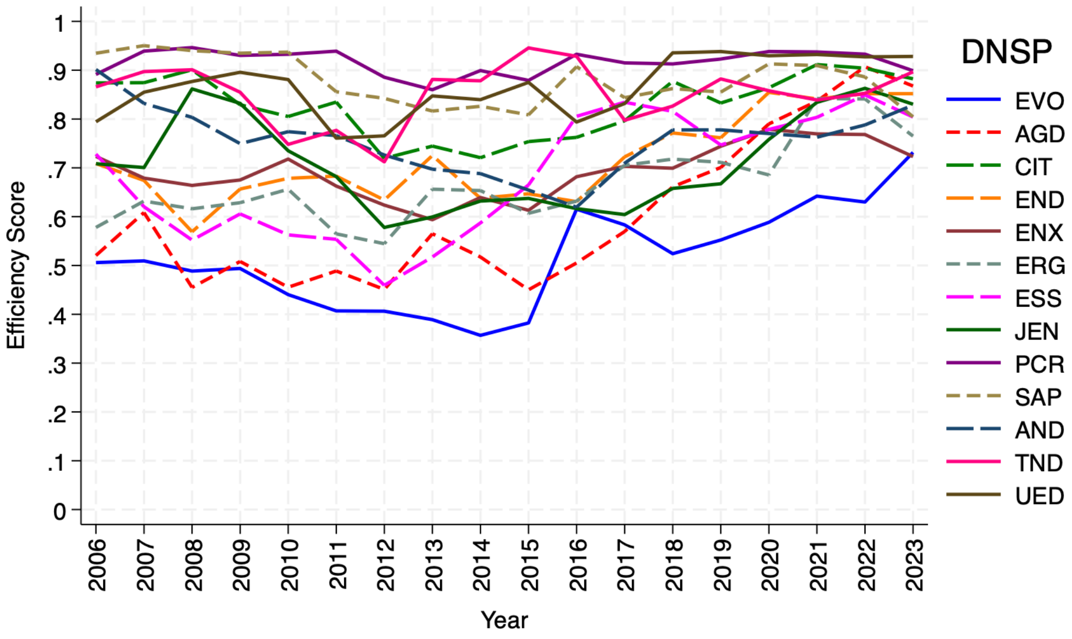
The correlation coefficients between the efficiency scores for Australian DNSPs from the long period SFACD and SFATLG models and the multilateral Opex PFP indexes are: 0.840 for SFACD and 0.876 for SFATLG. This is better than the Standard models (in which each DNSP's efficiency is constant), for which the correlation coefficients are 0.686 for SFACD and 0.734 for SFATLG.

Figure 4.3.2 illustrate the trends in efficiency scores over time for each model. Note that, unlike the other time trend models presented in the following sections, the BC95 model allows efficiency to vary flexibly over time, meaning it can increase in some years and decrease in others for the same DNSP. A general pattern of increasing efficiency over time is observed for most DNSPs, particularly from around 2015 onwards, which is aligned with the findings from the Opex multilateral partial factor productivity (MPFP) indexes analysis in ABR24.

Figure 4.2.2 BC95-JTT-HN Efficiency Trends by DNSP



SFACD Long Period



SFATLG Long Period

4.3.2 Unsuccessful BC95 variants

The seven BC95 specifications that were tested but found to be unsuccessful are as follows:

- *BC95-JTT-HN-GTC:* This model includes jurisdictional time trends as determinants of inefficiency (via μ) and uses the General Technical Change Index (ie, dummy variables for a set of periods). It failed to produce efficiency scores for the Australian DNSPs in the long sample period, although it was able to do so in the short sample period. In the long-period SFACD model, no standard errors were produced for any estimated coefficients, indicating computational issues. The long-period SFATLG specification also performed poorly, producing a negative and statistically significant coefficient for output circuit length (CL), failing to generate meaningful standard errors for any coefficients, and producing excessive monotonicity violations (100% for Australian DNSPs). These issues were not present in the short-period results, where the model performed reasonably well: standard errors were available, monotonicity violations were low, and efficiency scores were reasonable.
- *BC95-AJTT-HN:* This model generalises the *BC95-JTT-HN* model by including, as determinants of μ, DNSP-specific time trends for Australian businesses in addition to jurisdictional trends for New Zealand and Ontario. Technical change is reflected by a standard time trend. This model failed to produce efficiency scores for the Australian DNSPs in the long sample period. In both SFACD and SFATLG long-period specifications, no standard errors were produced, pointing again to computational issues. In the short-period models, although estimation was partially successful, some standard errors were still missing and not all DNSPs received efficiency scores.
- *BC95-AJTT-HN-GTC:* This model is similar to BC95-AJTT-HN but replaces the standard time trend with general technical change variables. It failed to estimate efficiency scores for the Australian DNSPs in both the long and short sample periods. In both the CD and TLG models in long and short-periods, standard errors were either missing for all or some coefficients.

This sequence of results shows that the BC95 model becomes more difficult to estimate as the models become more complex, and with the ABR data, estimation was only feasible in the simplest of the model specifications tested.

We also tested variants of these models which allowed for the variance of the idiosyncratic error term to be a function of a measure of scale. The scales variables were a weighted average of the mean values, for each DNSP, of each log output. The weights were given by the estimated coefficients of the ordinary least squares CD model also used in producing starting values for the SFACD model. This line of inquiry was not promising since this scale variable was typically not a statistically significant explanator of the squared residuals of the *BC95-JTT-HN* models.

4.4 Kumbhakar (1990) Model

This section presents the empirical results from the Kumbhakar 1990 (*Kumb90)* method, as generalised in the *sfpanel* program. The method is briefly described in sections 2.2.2 and 2.2.5. As presented in those sections, in this approach inefficiency, u_{it}, is defined such that it has a cross-sectional one-sided random element, and a deterministic element which can be time varying:

$$u_{it} = g(\boldsymbol{z_{it}}) \cdot u_i \qquad \text{where: } u_i \sim N^+(0, \sigma_u^2) \tag{4.4.1}$$

Here $g(\boldsymbol{z})$ is a scaling function, and u_i is the one-sided random cross-sectional inefficiency term, which has a half-normal distribution. Although in Kumbhakar (1990) has a specific function for $g(\boldsymbol{z})$ quadratic in the time trend, the *sfpanel* implementation is a generalisation of this approach with:

$$g(\boldsymbol{z_{it}}) = [1 + \exp(\delta_0 + \boldsymbol{z}'_{\boldsymbol{it}}\boldsymbol{\delta})]^{-1} \tag{4.4.2}$$

where $\boldsymbol{z_{it}}$ is a vector of exogenous variables. Note that in this model, the constant term δ_0 only affects the scaling of the inefficiency term and does not affect the truncation point of the random inefficiency term, which remains at zero (ie, it is half-normal). It does not give rise to the same partial identification issue as with the bc95 model because the inclusion of the constant term δ_0 in (4.4.2) does not mean that we are departing from the half-normal assumption.

Four models are presented here, focussing mainly on the long sample period results. Short-period results are available in Attachment B.

4.4.1 Kumb90-JTT-HN: Jurisdiction Time Trends

This specification incorporates jurisdiction-specific time trends as determinants of the inefficiency term. All models were estimated with convergence achieved between 100 and 225 iterations. However, in short period estimation, the message "*cannot compute an improvement — flat region encountered*" was returned, indicating the optimisation reached an area of the likelihood surface where further improvement was not possible. As a result, the estimates produced under these conditions may be unreliable or unstable.

The results of estimating the Cobb-Douglas and Translog versions with the long sample are presented in Table 4.4.1. The coefficients on the primary log output variables have the expected signs and are statistically significant. The undergrounding share variable is negative although not statistically significant in the SFATLG model.

Regarding MVs in the long-period model, there are no MVs in the whole sample. This result indicates that the Kumb90-JTT-HN model performs better than the standard LSETLG and SFATLG specifications in the long period in terms of reducing monotonicity violations.

Table 4.4.1 Kumb90-JTT-HN: Parameter Estimates

	SFACD Long Period			SFATLG Long Period		
Variable	*Coefficient*	*SE*	*t–ratio*	*Coefficient*	*SE*	*t–ratio*
Frontier						
ly1	0.469	0.065	7.25	0.487	0.066	7.38
ly2	0.213	0.029	7.39	0.223	0.033	6.84
ly3	0.292	0.057	5.12	0.276	0.060	4.60
ly11				0.593	0.371	1.60
ly12				-0.197	0.081	-2.43
ly13				-0.421	0.322	-1.31
ly22				0.121	0.037	3.22
ly23				0.093	0.067	1.38
ly33				0.313	0.291	1.07
lz1	-0.063	0.028	-2.21	-0.017	0.032	-0.55
yr	0.021	0.001	16.11	0.020	0.001	13.92
jur2	0.074	0.071	1.04	0.064	0.060	1.06
jur3	0.234	0.055	4.28	0.217	0.054	3.99
_cons	-31.699	2.598	-12.20	-29.814	2.867	-10.40
Bt						
t_aus	0.454	0.048	9.40	0.473	0.049	9.73
t_nz	-0.071	0.040	-1.78	-0.066	0.040	-1.67
t_ont	0.383	0.041	9.33	0.366	0.041	9.01
_cons	-2.262	0.212	-10.66	-2.277	0.215	-10.60
/sigmau_2	0.194	0.040	4.84	0.201	0.043	4.66
/sigmav_2	0.011	0.000	22.65	0.011	0.000	22.58
sigma_u	0.440	0.045	9.68	0.449	0.048	9.32
sigma_v	0.104	0.002	45.31	0.103	0.002	45.16
lambda	4.222	0.045	93.04	4.345	0.048	90.46
LLH	800.56			808.47		
Iterations #	100			183		
Pseudo Adj R^2	0.994			0.994		
BIC	-1503.1			-1478.69		
N	1098			1098		

Figure 4.4.1 presents the average of the Kumb90-JTT-HN SFACD and SFATLG efficiency scores for each Australian DNSP in the long period and compares them to the efficiency scores from the standard models and the comparative Opex PFP. The efficiency scores and rankings of the Kumb90-JTT-HN models are broadly aligned with those from the standard models.

Figure 4.4.1 Average Efficiency Scores by DNSP (2006–2023)

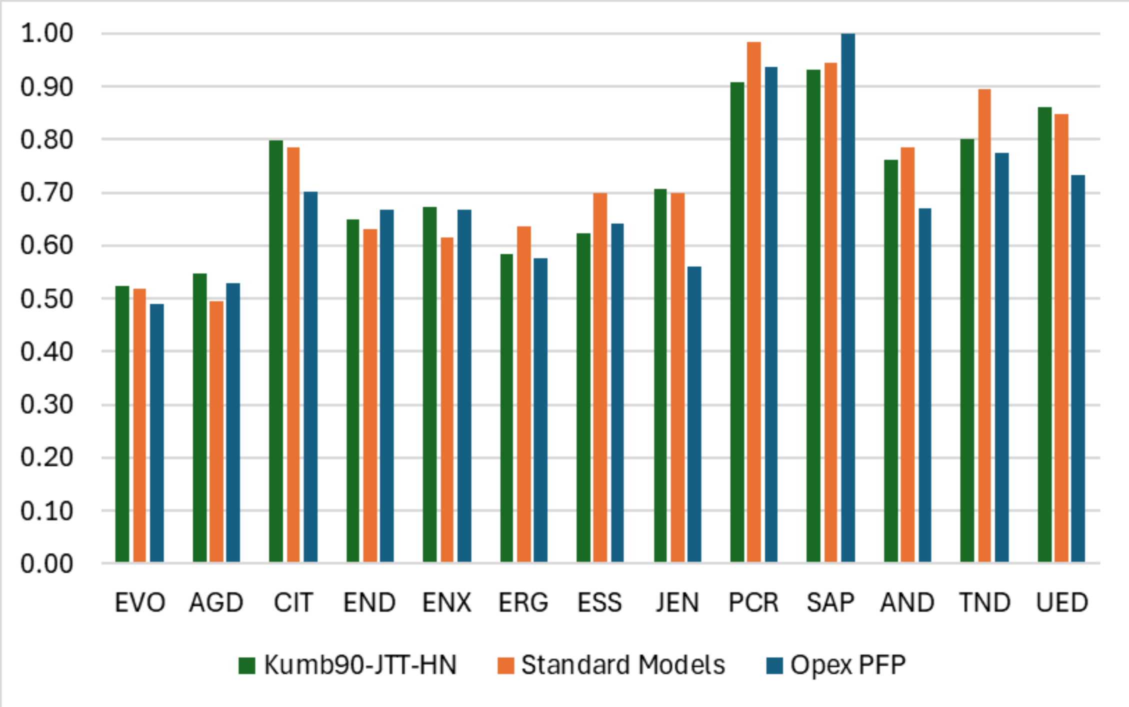


Figure 4.4.2 illustrates the trends in efficiency scores over time for all models.[41] The functional form for the scaling function is quite flexible. It is bounded by (0,1) and can be monotonically increasing, decreasing, concave or convex depending on the values of the parameters. A consistent pattern emerges: most DNSPs show a gradual but steady increase in efficiency over time. The upward trend is particularly evident from around 2014 onwards, with all DNSPs converging toward higher efficiency scores by the end of the sample period. In this case, the strong upward trends in efficiency scores in the latter half of the period indicates that most DNSPs have substantially improved their efficiency since the introduction of opex benchmarking.

This specification imposes the restriction that the inefficiencies of Australian DNSPs differ in size but follow the same time pattern.[42] This means that the time-pattern is an average for all Australian DNSPs. This presents a drawback in the specification, as the estimated trend does not reflect the diversity of time patterns observed across individual Australian DNSPs, and the model does not appear to capture performance deteriorations.

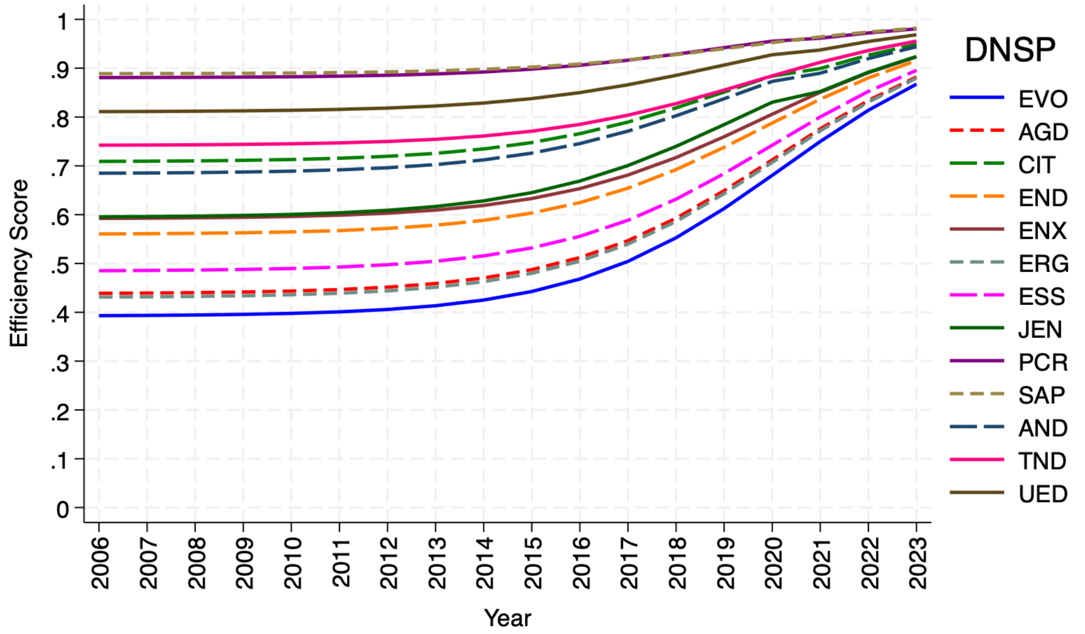
The efficiency scores for Australian DNSPs from the long-period SFACD and SFATLG models show strong correlations with the multilateral Opex PFP measures, at 0.722 and 0.721,

[41] Unlike the efficiency trends observed in the BC95 model, the efficiency scores in the Kumb models are smoother and do not exhibit the year-to-year variation seen in the BC95 results. This is because the Kumb models do not include the random component that is present in the BC95 specification.

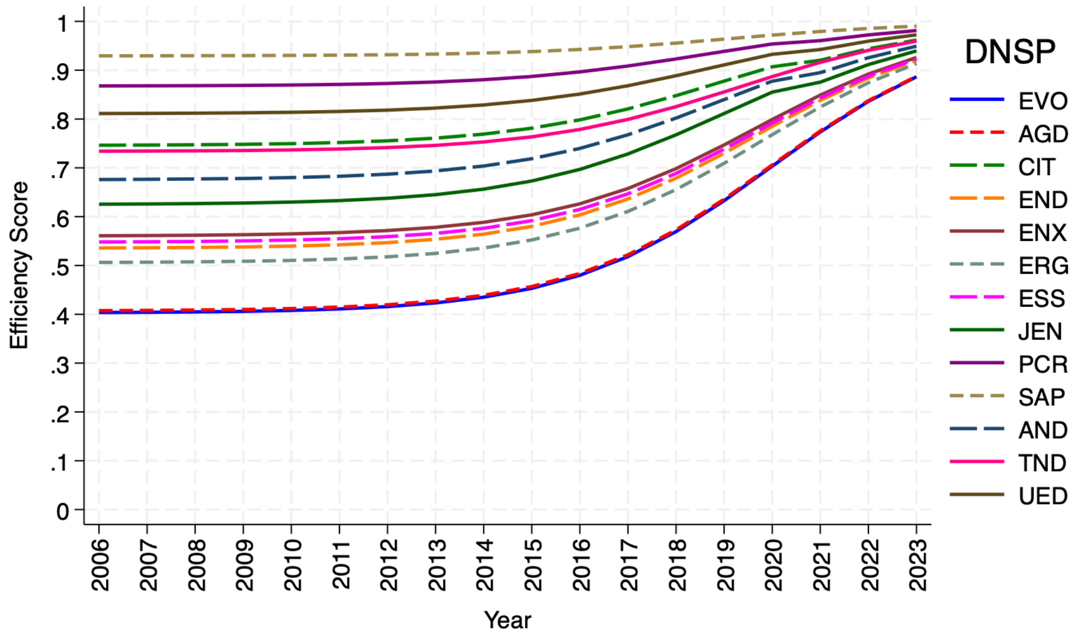
[42] The small kinks in some curves relate to the switch from calendar to financial years for some DNSPs.

respectively. These correlations are higher than those observed in the standard SFACD models (0.686) and comparable to the standard SFATLG model (0.734).

Figure 4.4.2 Kumb90-JTT-HN Efficiency Trends by DNSP



SFACD Long Period



SFATLG Long Period

4.4.2 Kumb90-JTT-HN-GTC: Jurisdiction Time Trends & General Technical Change

This specification is different from the previous Kumb90 specification by including the GTC variables to measure technical change. Both the SFACD and SFATLG models were estimated with convergence achieved after 57 to 334 iterations in the long sample. In the short sample, the SFACD and SFATLG models both converged successfully. The results of estimating the Cobb-Douglas and Translog versions with the long sample are presented in Table 4.4.2.

Table 4.4.2 Kumb90-JTT-HN-GTC: Parameter Estimates

	SFACD Long Period			*SFATLG Long Period*		
Variable	*Coefficient*	*SE*	*t–ratio*	*Coefficient*	*SE*	*t–ratio*
Frontier						
ly1	0.461	0.066	6.98	0.503	0.073	6.91
ly2	0.225	0.029	7.86	0.222	0.034	6.52
ly3	0.287	0.059	4.84	0.262	0.064	4.07
ly11				0.555	0.405	1.37
ly12				-0.224	0.095	-2.35
ly13				-0.351	0.339	-1.03
ly22				0.135	0.044	3.03
ly23				0.112	0.071	1.58
ly33				0.214	0.303	0.71
lz1	-0.038	0.029	-1.28	0.009	0.032	0.27
gtc2	0.042	0.012	3.62	0.039	0.012	3.37
gtc3	0.112	0.013	8.72	0.107	0.013	8.13
gtc4	0.164	0.015	11.00	0.155	0.016	9.94
gtc5	0.230	0.017	13.40	0.217	0.018	11.96
gtc6	0.266	0.020	13.06	0.249	0.021	11.73
jur2	0.018	0.069	0.26	0.037	0.054	0.69
jur3	0.200	0.052	3.88	0.192	0.051	3.74
_cons	9.683	0.071	136.22	9.719	0.064	152.01
Bt						
t_aus	0.443	0.056	7.95	0.458	0.054	8.43
t_nz	-0.101	0.037	-2.76	-0.094	0.035	-2.67
t_ont	0.343	0.051	6.73	0.322	0.048	6.74
_cons	-2.133	0.267	-8.00	-2.146	0.262	-8.19
/sigmau_2	0.184	0.038	4.88	0.199	0.043	4.65
/sigmav_2	0.011	0.000	22.67	0.011	0.000	22.57
sigma_u	0.429	0.044	9.76	0.446	0.048	9.30
sigma_v	0.106	0.002	45.35	0.105	0.002	45.14
lambda	4.046	0.044	92.08	4.255	0.048	88.88
LLH	783.65			793.03		
Iterations #	57			334		
Pseudo Adj R^2	0.994			0.994		
BIC	-1441.29			-1418.03		
N	1098			1098		

The SFACD and SFATLG models for the long period produce output coefficients with the expected positive signs, and most are statistically significant. The coefficient on the log share of underground cables has the expected negative sign only in the SFACD model but is not statistically significant. It has a small positive and statistically insignificant value in the SFATLG model, which goes against expectations and suggests a possible issue with this specification.

Regarding MVs in the long-period model, none were observed across the full sample. This result indicates that the Kumb90-JTT-HN-GTC model performs better than the standard SFATLG specification in the long sample period by eliminating all monotonicity violations.

Figure 4.4.3 presents the average of the Kumb90-JTT-HN-GTC SFACD and SFATLG efficiency scores for each Australian DNSP in the long period and compares them to the efficiency scores from the standard models and the comparative Opex PFP. The efficiency scores and rankings of the Kumb90-JTT-HN-GTC models are broadly aligned with those from the standard models.

Figure 4.4.3 Average Efficiency Scores by DNSP (2006–2023)

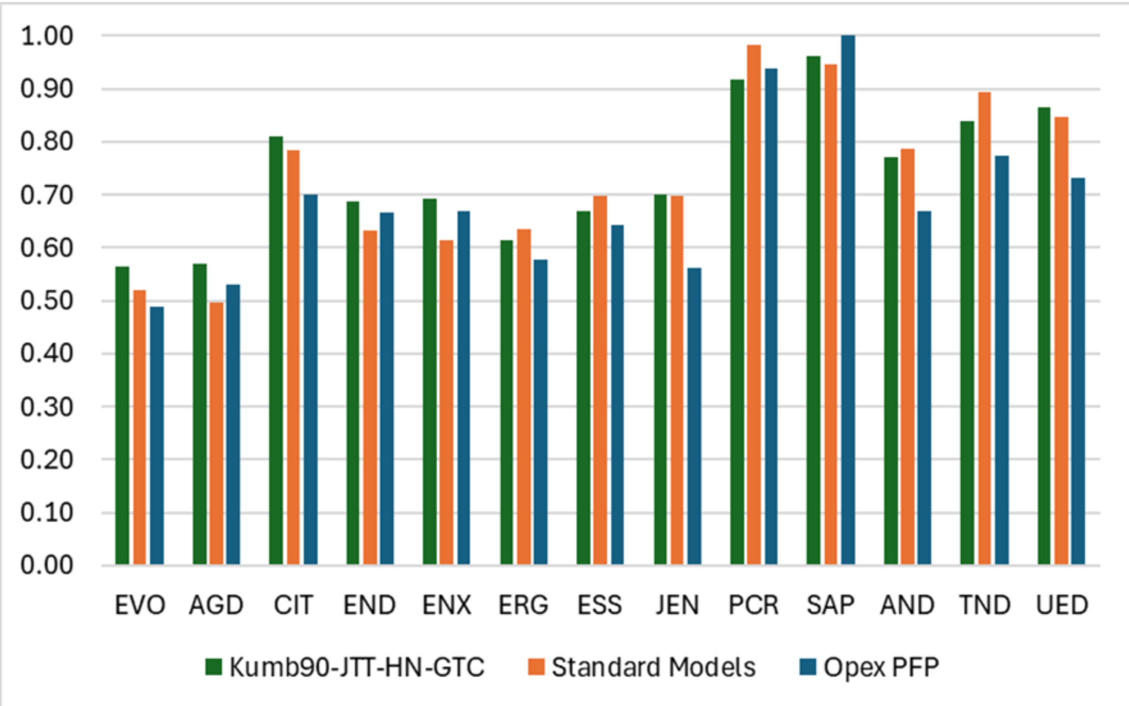
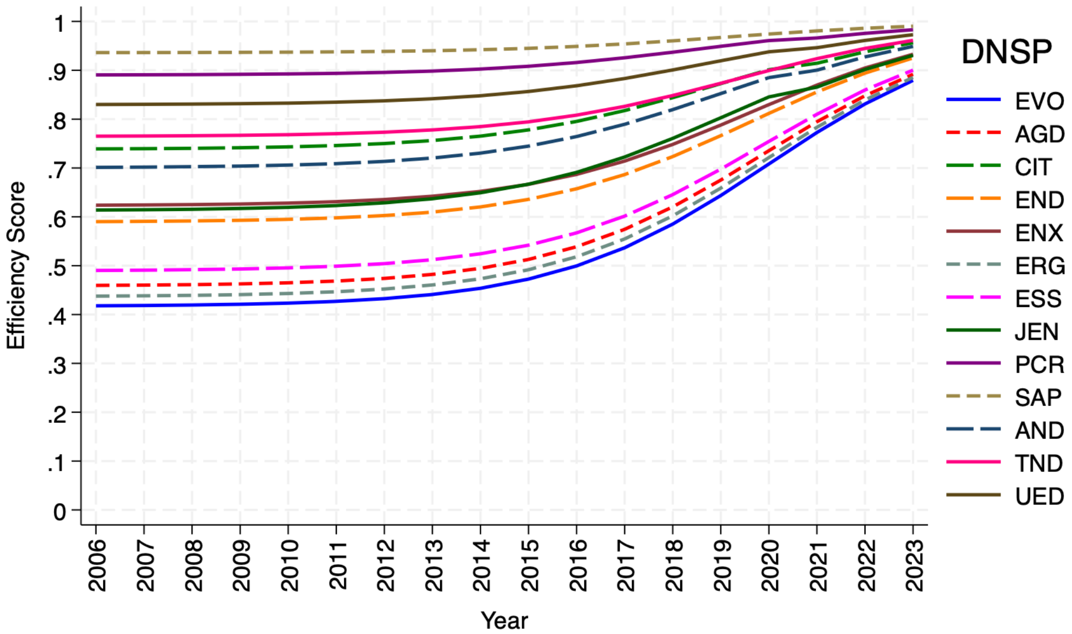
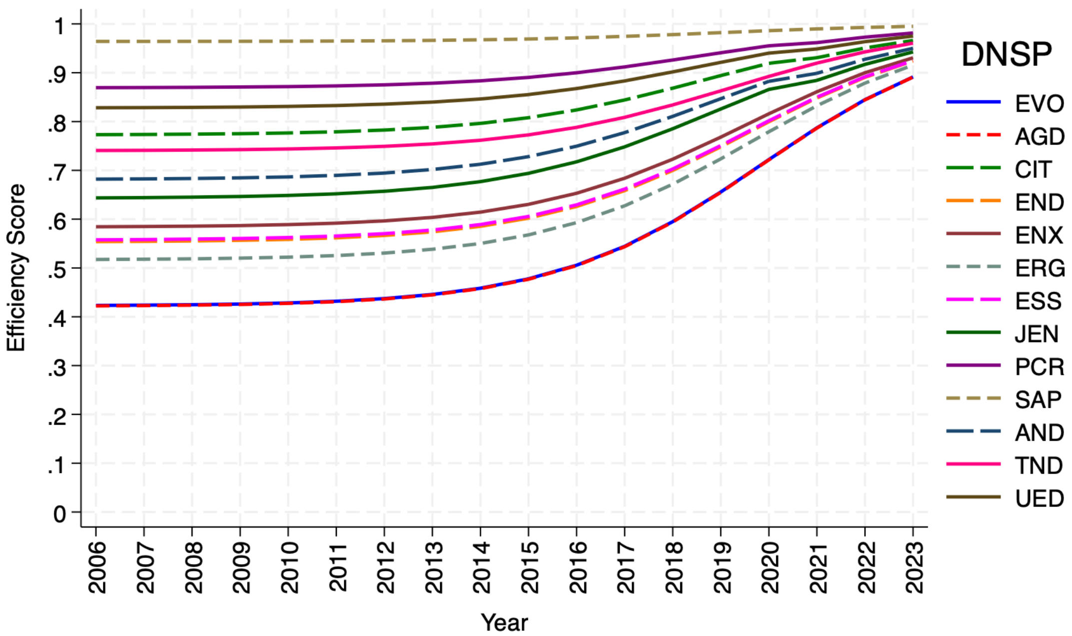


Figure 4.4.4 shows the efficiency score trends over time by DNSP for the Kumb90-JTT-HN-GTC specification. The efficiency scores for Australian DNSPs from the long-period SFACD and SFATLG models show strong correlations with the multilateral Opex PFP measures, at 0.732 and 0.727, respectively. These correlations are higher than those observed in the standard SFACD models (0.686) and comparable to the standard SFATLG model (0.734).

Figure 4.4.4 Kumb90-JTT-HN-GTC Efficiency Trends by DNSP



SFACD Long Period



SFATLG Long Period

4.4.3 Kumb90-AJTT-HN: Australian DNSP Specific & Jurisdiction Time Trends

This model differs from the one presented in Section 4.4.1 by replacing the jurisdiction-specific time trends with individual time trends for each Australian DNSP, along with time trends for New Zealand and Ontario. Both the SFACD and SFATLG models were estimated for the long and short periods, with convergence achieved between 172 and 220 in the long period. In the short sample the SFATLG model converged after 341 iterations, but the SFACD encountered a flat region. The results of estimating the Kumb90-AJTT-HN models with the long sample are presented in Table 4.4.3. The SFACD and SFATLG models for the long period produce output coefficients with the expected positive signs, and are statistically significant. The coefficient on the log share of underground cables has the expected negative sign but it is not statistically significant at 5 per cent in the SFATLG.

Regarding MVs in the long-period model, ten Australian DNSPs exhibit monotonicity violations in more than 50 per cent of observations. Overall, 55.1 per cent of the Australian sample and 27.0 per cent of the total sample are affected. These results do not support the usefulness of this specification. However, the Kumb90–AJTT–HN model still performs better than the standard SFATLG specifications in terms of reducing monotonicity violations.

The efficiency scores from the long sample SFACD specification under this model remain aligned to established results, however the SFATLG version diverges more significantly.

Table 4.4.3 Kumb90-AJTT-HN Parameter Estimates

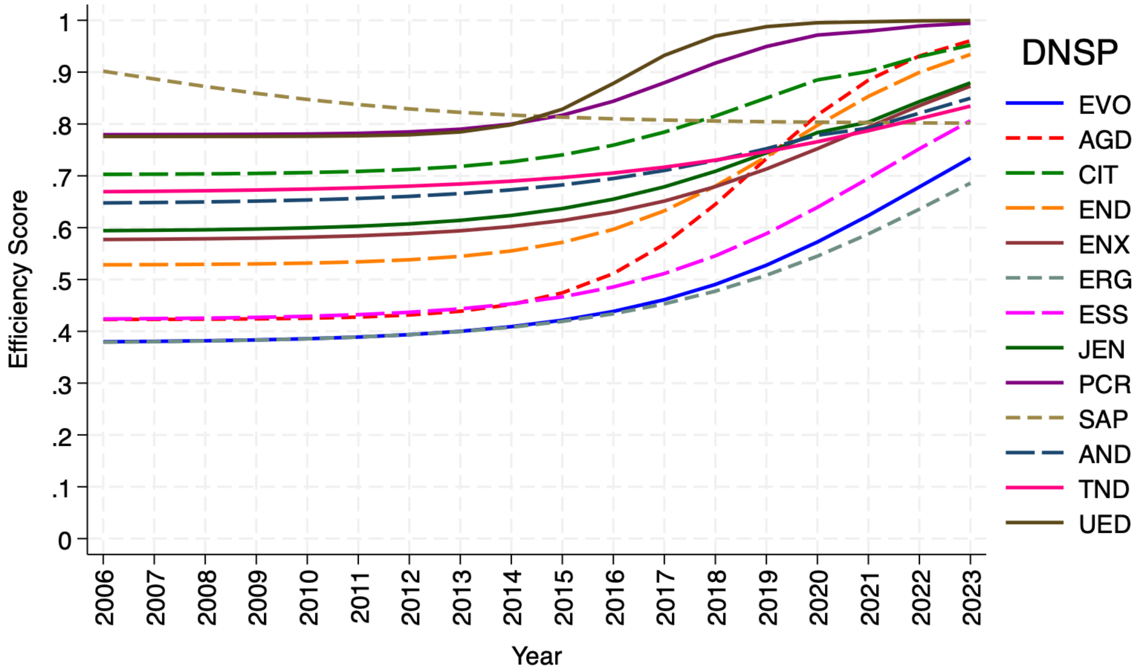
	SFACD Long Period			*SFATLG Long Period*		
Variable	*Coefficient*	*SE*	*t–ratio*	*Coefficient*	*SE*	*t–ratio*
Frontier						
ly1	0.582	0.073	7.95	0.616	0.077	7.99
ly2	0.154	0.033	4.63	0.221	0.033	6.76
ly3	0.256	0.063	4.10	0.172	0.072	2.39
ly11				1.000	0.466	2.15
ly12				-0.376	0.109	-3.46
ly13				-0.539	0.374	-1.44
ly22				0.246	0.046	5.33
ly23				0.163	0.084	1.93
ly33				0.275	0.313	0.88
lz1	-0.066	0.031	-2.16	-0.012	0.032	-0.38
yr	0.019	0.001	14.70	0.018	0.001	13.99
jur2	0.206	0.065	3.15	0.176	0.063	2.81
jur3	0.298	0.055	5.39	0.417	0.080	5.18
_cons	-29.497	2.689	-10.97	-27.384	2.640	-10.37

Table 4.4.3 (cont.)

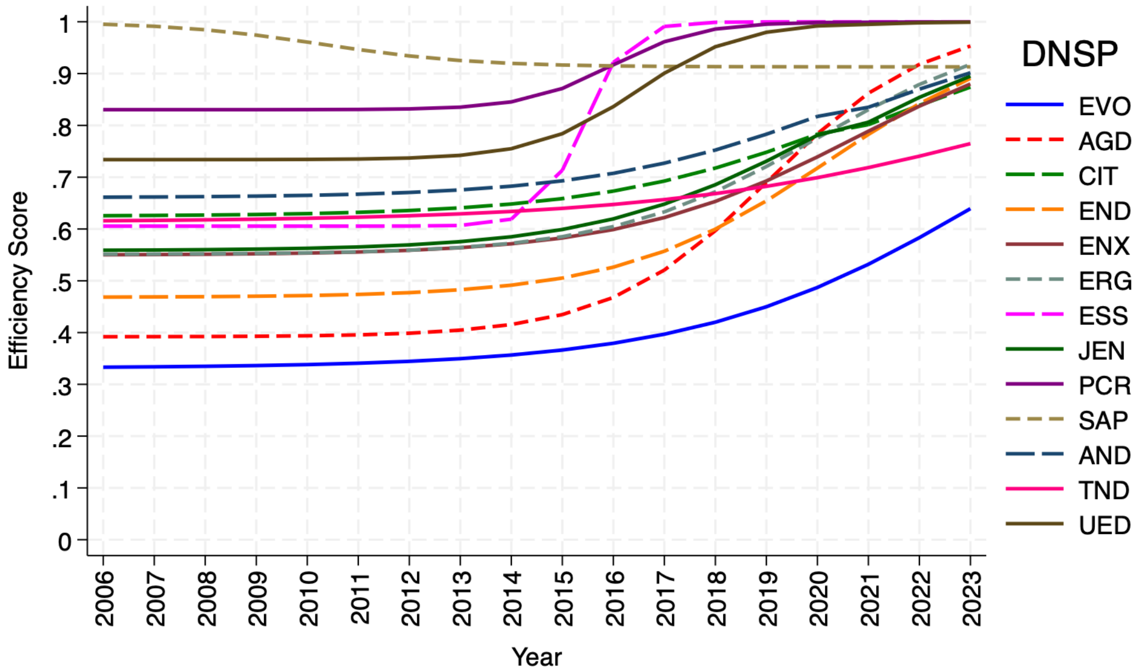
	SFACD Long Period			*SFATLG Long Period*		
Variable	*Coefficient*	*SE*	*t–ratio*	*Coefficient*	*SE*	*t–ratio*
Bt						
t_dnsp1	0.355	0.043	8.32	0.336	0.041	8.23
t_dnsp2	0.611	0.072	8.52	0.626	0.076	8.21
t_dnsp3	0.475	0.116	4.09	0.397	0.087	4.58
t_dnsp4	0.509	0.083	6.17	0.488	0.068	7.22
t_dnsp5	0.395	0.074	5.32	0.441	0.081	5.46
t_dnsp6	0.320	0.044	7.26	0.496	0.078	6.38
t_dnsp7	0.392	0.060	6.57	2.365	0.802	2.95
t_dnsp8	0.395	0.064	6.20	0.458	0.067	6.87
t_dnsp9	0.698	0.305	2.29	1.187	0.678	1.75
t_dnsp10	-0.303	0.181	-1.67	-0.669	0.336	-1.99
t_dnsp11	0.327	0.074	4.44	0.418	0.084	4.95
t_dnsp12	0.291	0.078	3.75	0.271	0.068	3.97
t_dnsp13	1.009	0.265	3.81	0.989	0.212	4.68
t_nz	-0.110	0.039	-2.80	-0.090	0.038	-2.37
t_ont	0.384	0.040	9.54	0.383	0.045	8.56
_cons	-2.346	0.207	-11.32	-2.570	0.219	-11.74
/sigmau_2	0.215	0.045	4.79	0.250	0.060	4.18
/sigmav_2	0.010	0.000	22.61	0.010	0.000	22.29
sigma_u	0.463	0.048	9.58	0.500	0.060	8.37
sigma_v	0.101	0.002	45.22	0.099	0.002	44.58
lambda	4.573	0.048	94.70	5.030	0.060	84.51
LLH	825.95			840.18		
Iterations #	172			220		
Pseudo Adj R^2	0.994			0.994		
BIC	-1469.86			-1456.32		
N	1098			1098		

Figure 4.4.5 shows the efficiency score trends over time by DNSP for the Kumb90-AJTT-HN specification. The efficiency scores for Australian DNSPs from the long-period SFACD and SFATLG models show strong correlations with the multilateral Opex PFP measures, at 0.688 and 0.733, respectively. These correlations are similar to those observed in the standard SFACD and SFATLG models (0.686 and 0.734 respectively).

Figure 4.4.5 Kumb90-AJTT-HN Efficiency Trends by DNSP



SFACD Long Period



SFATLG Long Period

4.4.4 Kumb90-AJTT-HN-GTC: Australian DNSP Specific & Jurisdiction Time Trends & General Technical Change

This model differs from the one presented in Section 4.4.3 by including the GTC variables to measure technical change. Both the SFACD and SFATLG models were estimated for the long and short periods, with convergence achieved between 59 and 342 iterations, although the SFATLG in the short period could not compute an improvement during estimation as it encountered a flat region in the likelihood surface. The results of estimating the Kumb90-AJTT-HN-GTC models with the long sample are presented in Table 4.4.4.

The SFACD and SFATLG models for the long period produce output coefficients with the expected positive signs and are statistically significant. The coefficient on the log share of underground cables has the expected negative sign but it is not statistically significant in the SFACD, but it is positive although not statistically significant in the SFATLG, suggesting a possible issue with this specification.

Regarding MVs in the long-period model, ten Australian DNSPs exhibit monotonicity violations in more than 50 per cent of observations. Overall, 54.7 per cent of the Australian sample and 29.8 per cent of the total sample are affected. These results do not support the usefulness of this specification. However, the Kumb90-AJTT-HN-GTC model still performs better than the standard SFATLG specifications in terms of reducing monotonicity violations.

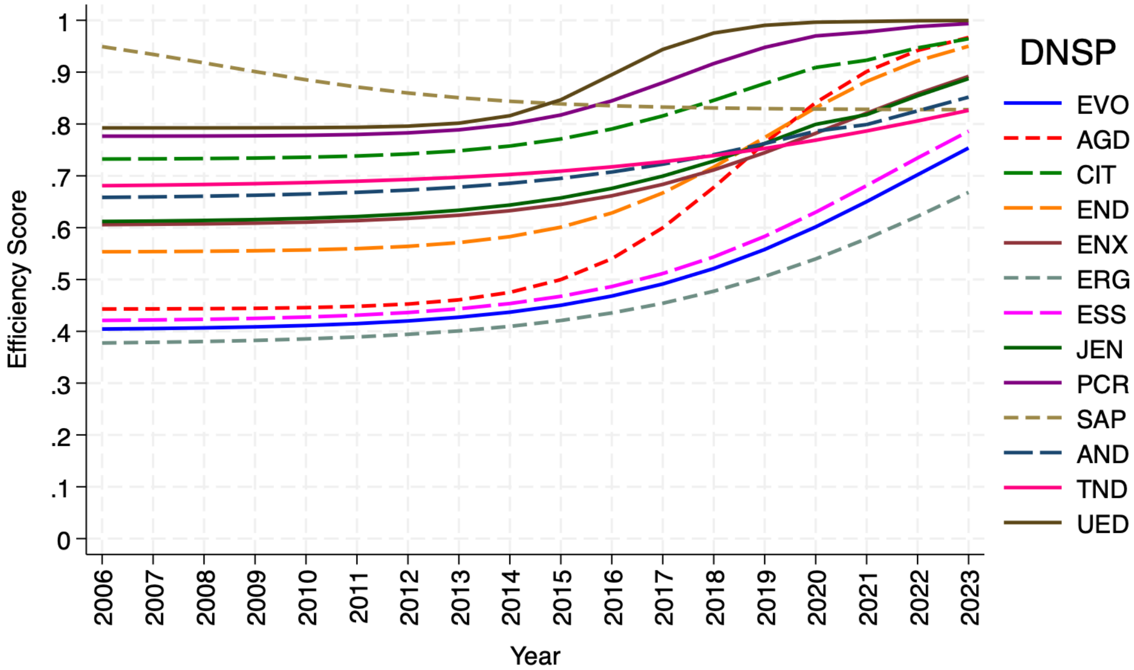
The efficiency scores from the long sample SFACD specification under this model remain aligned to established results, however the SFATLG version diverges more considerably.

Figure 4.4.6 shows the efficiency score trends over time by DNSP for the Kumb90-AJTT-HN specification. Unlike the Kumb90-JTT specification, the Kumb90-AJTT specification allows profile shapes to vary across DNSPs. The efficiency scores for Australian DNSPs from the long-period SFACD and SFATLG models show strong correlations with the OPFP measures, at 0.683 and 0.727, respectively. These correlations are similar to those observed in the standard SFACD models and SFATLG models (0.686 and 0.734 respectively).

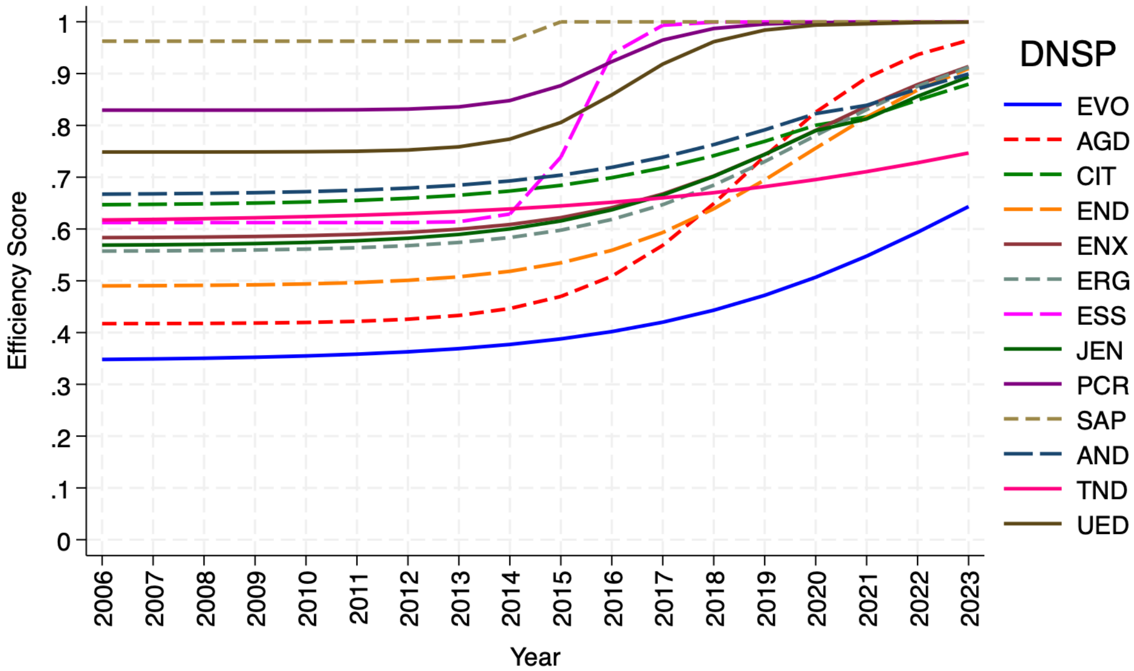
Table 4.4.4 Kumb90-AJTT-HN-GTC Parameter Estimates

	SFACD Long Period			*SFATLG Long Period*		
Variable	*Coefficient*	*SE*	*t–ratio*	*Coefficient*	*SE*	*t–ratio*
Frontier						
ly1	0.584	0.078	7.47	0.608	0.076	7.99
ly2	0.166	0.033	4.95	0.230	0.033	6.97
ly3	0.242	0.067	3.61	0.165	0.073	2.25
ly11				0.962	0.481	2.00
ly12				-0.395	0.109	-3.62
ly13				-0.462	0.388	-1.19
ly22				0.257	0.046	5.62
ly23				0.173	0.086	2.01
ly33				0.180	0.325	0.55
lz1	-0.031	0.032	-0.97	0.024	0.032	0.75
gtc2	0.038	0.011	3.29	0.038	0.011	3.36
gtc3	0.104	0.013	7.97	0.102	0.013	8.06
gtc4	0.153	0.015	9.93	0.150	0.015	9.93
gtc5	0.214	0.018	12.13	0.206	0.018	11.72
gtc6	0.247	0.021	12.02	0.232	0.021	10.98
jur2	0.146	0.064	2.29	0.152	0.057	2.68
jur3	0.261	0.057	4.59	0.400	0.078	5.15
_cons	9.605	0.083	115.60	9.487	0.079	119.45
Bt						
t_dnsp1	0.344	0.050	6.91	0.305	0.046	6.59
t_dnsp2	0.611	0.081	7.57	0.624	0.086	7.23
t_dnsp3	0.485	0.142	3.41	0.366	0.098	3.75
t_dnsp4	0.522	0.098	5.35	0.480	0.077	6.24
t_dnsp5	0.392	0.088	4.44	0.449	0.103	4.37
t_dnsp6	0.294	0.047	6.21	0.458	0.083	5.49
t_dnsp7	0.362	0.060	6.05	2.377	0.909	2.61
t_dnsp8	0.383	0.072	5.34	0.425	0.072	5.89
t_dnsp9	0.670	0.299	2.24	1.151	0.650	1.77
t_dnsp10	-0.389	0.202	-1.93	24.896	351.596	0.07
t_dnsp11	0.309	0.081	3.82	0.385	0.090	4.28
t_dnsp12	0.257	0.083	3.10	0.221	0.071	3.11
t_dnsp13	1.016	0.290	3.51	0.979	0.229	4.27
t_nz	-0.132	0.038	-3.51	-0.104	0.033	-3.11
t_ont	0.339	0.049	6.90	0.335	0.048	7.02
_cons	-2.221	0.262	-8.47	-2.335	0.260	-8.99
/sigmau_2	0.213	0.045	4.73	0.235	0.054	4.39
/sigmav_2	0.011	0.000	22.58	0.010	0.000	22.45
sigma_u	0.461	0.049	9.45	0.485	0.055	8.79
sigma_v	0.103	0.002	45.16	0.101	0.002	44.90
lambda	4.479	0.049	91.99	4.790	0.055	87.09
LLH	809.89			823.88		
Iterations #	80			342		
Pseudo Adj R^2	0.994			0.994		
BIC	-1409.73			-1395.72		
N	1098			1098		

Figure 4.4.6 Kumb90-AJTT-HN-GTC Efficiency Trends by DNSP



SFACD Long Period



SFATLG Long Period

4.5 LSE Time Varying Models

The standard LSE time invariant opex cost function has the specification:

$$c_{it} = \alpha + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \varphi t + \sum_{k=2}^{13} \gamma_k d_k + \varepsilon_{it} \tag{4.5.1}$$

where i refers to firm i; c_{it} is the log opex cost of firm i in period t; $\boldsymbol{x_{it}}$ is the vector of the values of the explanatory variables for firm i in period t;[43] the parameter φ represents the coefficient for the time trend variable, capturing the rate of technical change, trends in omitted OEFs and changes in efficiency over time, and d_k represent firm-specific dummy variables for each of the 13 Australian DNSPs;[44] and ε_{it} is a white noise disturbance.

The technical efficiency measures are based on the estimates of γ_k parameters. The technical efficiency of firm k is obtained using the formula:

$$\theta_k = \exp[\min(0, \gamma_2, \dots, \gamma_{13}) - \gamma_k] \tag{4.5.2}$$

The time varying LSE opex cost function extends Equation 4.5.1 to allow for time-varying inefficiency by introducing interaction terms between the time trend variable and the firm-specific dummy variables. In the linear case, this can be formulated as:

$$c_{it} = \alpha + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \varphi t + \sum_{k=2}^{13} d_k(\gamma_k + \gamma_{kt} t) + \gamma_{1t} d_1 t + \varepsilon_{it} \tag{4.5.3}$$

where the terms additional to equation (4.5.1) are time-trend effects specific to each Australian DNSP. For DNSP k this is: $\gamma_{kt} d_k t$, where d_k is the firm-specific dummy variable for DNSP k, and γ_{kt} is a parameter to be estimated. Note that although for $k = 1$, there is no DNSP-specific fixed effect, there is a time trend for DNSP 1.

It is feasible to have both the time-trend variable (t) and the time-interactions with the inefficiency effects ($d_k t$) because the 13 Australian DNSPs are a subset of the total sample of DNSPs. DNSPs. In this case, the cost efficiency of firm k in period t is obtained using the formula:

$$\theta_{kt} = \exp[\min(\gamma_{1t} t, \gamma_2 + \gamma_{2t} t, \dots, \gamma_{13} + \gamma_{13t} t) - (\gamma_k + \gamma_{kt} t)] \tag{4.5.4}$$

More specifically, by incorporating DNSP-specific time trends for each Australian DNSP, the estimated value of φ represents the average time trend for the 2 remaining jurisdictions (New Zealand and Ontario). This model allows for a specific time trend for each Australian DNSP, rather than for Australian DNSPs to be grouped together as a whole. However, the model

[43] The explanatory variables include the log of output variables—customer numbers, circuit length, and RMD—the log of the share of underground cables, which is a proxy for operational environmental factors (OEF), and jurisdictional dummy variables for the New Zealand and Ontario DNSPs.
[44] DNSP 1 has been arbitrarily chosen as the base firm so that $\gamma_1 = 0$. The choice of base has no effect because the inefficiency score is calculated relative to the minimum γ_j.

does not decompose each Australian DNSP's time trend further into separate estimates of technical change and efficiency change. To make such a decomposition would require the assumption that the time trend for Ontario and New Zealand, φ, is the rate of technical change applicable to Australian DNSPs.[45]

In this section, we present empirical results for four alternative specifications for the standard LSE model using time-varying approaches. These models do not have the convergence issues sometimes accompanying the previously considered models.

4.5.1 LSE-ADTT: LSE time-varying inefficiency through interaction of Australian DNSP-specific fixed effects and the time trend variable

This specification allows for time-varying inefficiency through interaction terms between the time trend variable and Australian DNSP-specific dummy variables. This specification is discussed in Quantonomics (2023). Technical change is proxied by a general time trend variable applying to all DNSPs (including Ontario and New Zealand).

The results of estimating the Cobb-Douglas and Translog versions with the long sample are presented in Table 4.5.1. The coefficients on the primary log output variables have the expected signs and are statistically significant. The undergrounding share variable is negative and statistically significant. Regarding MVs in the long-period model, one Australian DNSP exhibits monotonicity violations in more than 50 per cent of observations. Overall, 11.1 per cent of the Australian sample and 18.7 per cent of the total sample are affected. The LSE-ADTT model performs better than the standard LSETLG specifications in terms of reducing monotonicity violations.

[45] See footnote 47 for a qualification to this statement, in section 6.2: decomposing productivity changes.

Table 4.5.1 LSE-ADTT: Parameter Estimates

	LSECD - Long Period			*LSETLG- Long Period*		
Variable	*Coefficient*	*SE*	*t–ratio*	*Coefficient*	*SE*	*t–ratio*
ly1	0.544	0.074	7.32	0.380	0.078	4.85
ly2	0.227	0.034	6.59	0.230	0.034	6.84
ly3	0.194	0.064	3.03	0.345	0.066	5.20
ly11				-0.243	0.527	-0.46
ly12				0.295	0.123	2.41
ly13				-0.103	0.419	-0.25
ly22				-0.045	0.043	-1.04
ly23				-0.227	0.101	-2.25
ly33				0.388	0.335	1.16
lz1	-0.091	0.025	-3.56	-0.102	0.028	-3.62
yr	0.013	0.002	7.72	0.015	0.002	8.98
jur2	-48.711	32.863	-1.48	-55.968	32.252	-1.74
jur3	-48.479	32.864	-1.48	-55.798	32.253	-1.73
d2	36.923	46.606	0.79	30.354	45.203	0.67
d3	-38.963	37.320	-1.04	-43.465	36.702	-1.18
d4	-1.625	38.707	-0.04	-7.814	38.080	-0.21
d5	-32.882	35.771	-0.92	-36.390	34.886	-1.04
d6	4.873	39.458	0.12	5.434	38.705	0.14
d7	-12.406	42.188	-0.29	-15.756	41.415	-0.38
d8	-22.482	40.630	-0.55	-30.175	39.969	-0.75
d9	-33.363	37.184	-0.90	-37.392	36.503	-1.02
d10	-70.505	38.678	-1.82	-72.892	37.832	-1.93
d11	-51.399	37.115	-1.38	-56.650	36.444	-1.55
d12	-32.053	42.519	-0.75	-35.532	41.506	-0.86
d13	-9.403	38.906	-0.24	-12.922	38.081	-0.34
dt1	-0.024	0.016	-1.47	-0.028	0.016	-1.72
dt2	-0.042	0.017	-2.56	-0.043	0.016	-2.69
dt3	-0.005	0.009	-0.53	-0.006	0.009	-0.69
dt4	-0.023	0.010	-2.25	-0.024	0.010	-2.32
dt5	-0.008	0.007	-1.06	-0.010	0.007	-1.39
dt6	-0.027	0.011	-2.39	-0.030	0.011	-2.79
dt7	-0.018	0.013	-1.35	-0.020	0.013	-1.52
dt8	-0.013	0.012	-1.08	-0.013	0.012	-1.07
dt9	-0.008	0.009	-0.87	-0.009	0.009	-1.07
dt10	0.011	0.010	1.03	0.008	0.010	0.82
dt11	0.001	0.009	0.14	0.000	0.009	0.03
dt12	-0.008	0.014	-0.61	-0.010	0.013	-0.77
dt13	-0.020	0.011	-1.85	-0.021	0.010	-2.08
_cons	32.937	32.689	1.01	35.857	32.076	1.12
rho	0.766			0.753		
R2	0.992			0.992		
Adj. Pseudo R^2	0.980			0.982		
N	1,098			1,098		

Figure 4.5.1 presents the average of the LSE-ADTT LSECD and LSETLG efficiency scores for each Australian DNSP in the long period and compares them to the efficiency scores from the standard models and the comparative Opex PFP. The efficiency scores and rankings of the LSE-ADTT models are broadly aligned with those from the standard models. The efficiency scores for Australian DNSPs from the long-period LSECD and LSETLG models show strong correlations with the multilateral Opex PFP measures, at 0.760 and 0.821, respectively. These correlations are higher than those observed in the standard LSECD and SFATLG models (0.741 and 0.798 respectively).

Figure 4.5.1 Average Efficiency Scores by DNSP (2006–2023)

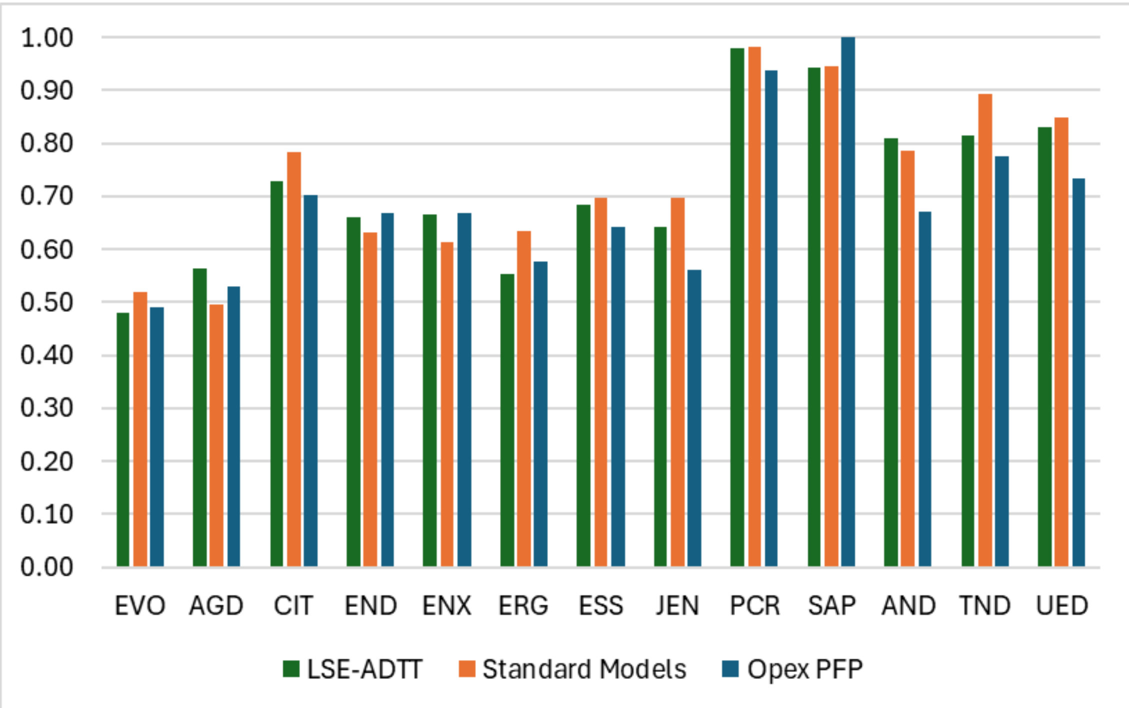
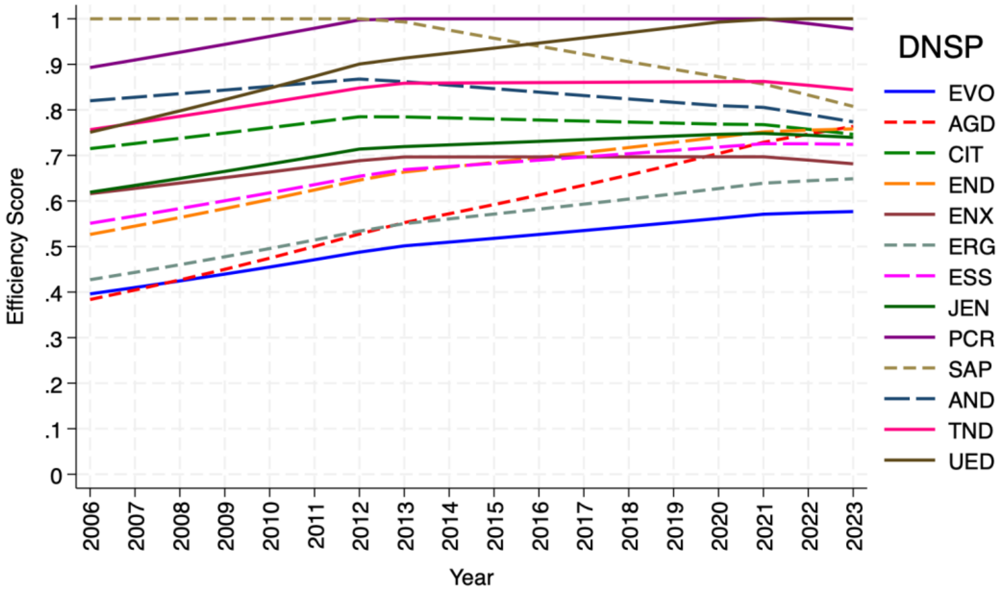
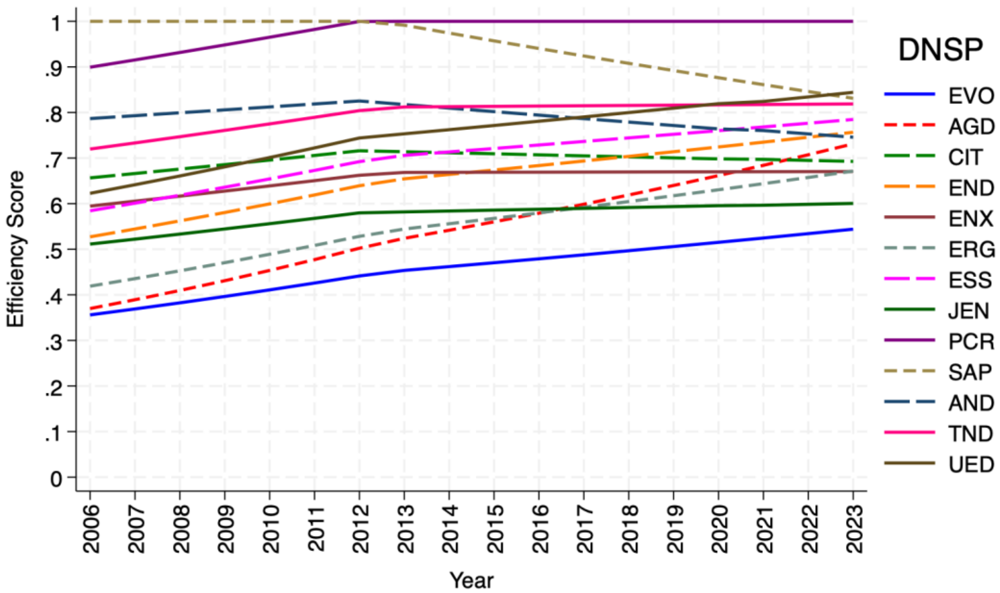


Figure 4.5.2 illustrates the trends in efficiency scores over time for both the LSECD and LSETLG long-sample models. These figures show that the LSE-ADTT models benefit from having separate time trends for each DNSP. These trends are not all the same, a restriction imposed by some of the previously considered models. The apparent nonlinearities in the efficiency trends of individual Australian DNSPs arise because efficiency is measured relative to the best performer in each year.

Figure 4.5.2 LSE-ADTT Efficiency Trends by DNSP



LSECD Long Period



LSETLG Long Period

4.5.2 LSE-ADTT–GTC: LSE time-varying inefficiency through interaction & General Technical Change

This specification differs from the specification in section 4.5.1 by including the GTC terms. The results of estimating the Cobb-Douglas and Translog versions with the long sample are presented in Table 4.5.2.

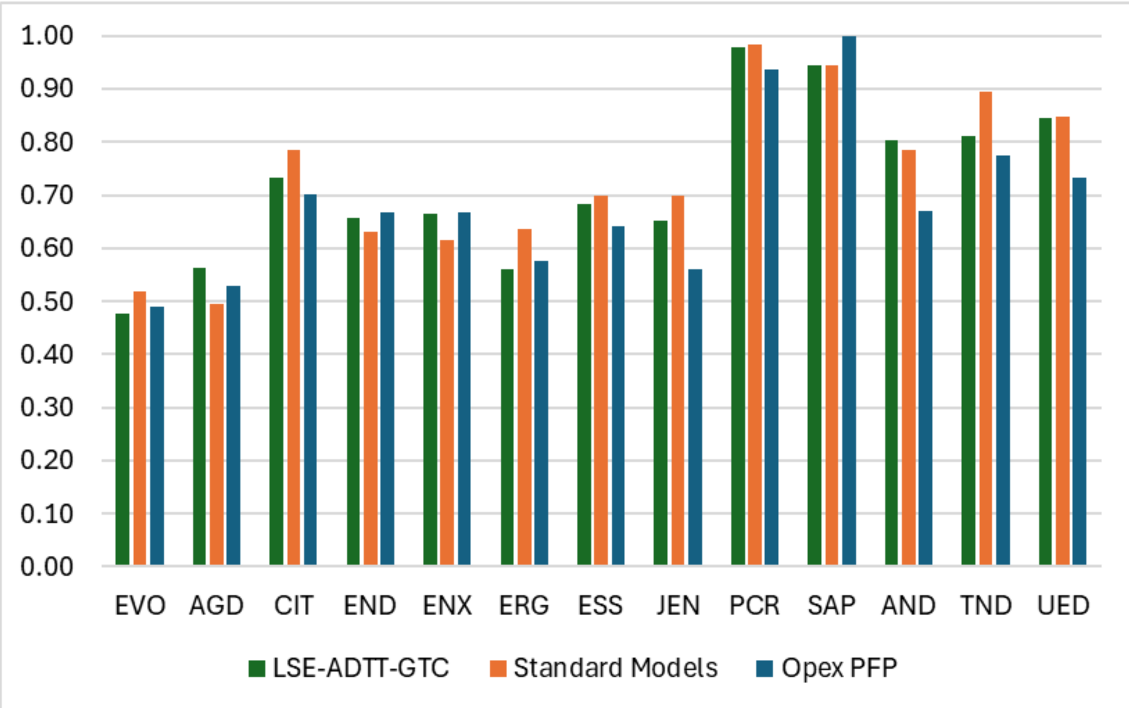
Table 4.5.2 LSE-ADTT-GTC: Parameter Estimates

	LSECD - Long Period			LSETLG- Long Period		
Variable	*Coefficient*	*SE*	*t–ratio*	*Coefficient*	*SE*	*t–ratio*
ly1	0.575	0.068	8.48	0.432	0.071	6.05
ly2	0.223	0.030	7.33	0.225	0.030	7.44
ly3	0.170	0.059	2.86	0.300	0.060	4.97
ly11				-0.058	0.518	-0.11
ly12				0.202	0.117	1.73
ly13				-0.204	0.412	-0.49
ly22				-0.008	0.039	-0.20
ly23				-0.171	0.098	-1.75
ly33				0.441	0.326	1.35
lz1	-0.092	0.022	-4.14	-0.093	0.025	-3.70
gtc2	0.034	0.013	2.65	0.037	0.013	2.95
gtc3	0.104	0.016	6.43	0.112	0.016	6.94
gtc4	0.129	0.018	7.08	0.141	0.018	7.80
gtc5	0.146	0.020	7.39	0.163	0.020	8.26
gtc6	0.121	0.021	5.71	0.142	0.021	6.70
jur2	-0.413	0.094	-4.38	-0.465	0.094	-4.93
jur3	-0.175	0.092	-1.89	-0.280	0.092	-3.03
d2	-0.128	0.129	-0.99	-0.168	0.141	-1.19
d3	-0.430	0.102	-4.20	-0.447	0.104	-4.31
d4	-0.285	0.107	-2.66	-0.358	0.110	-3.25
d5	-0.322	0.100	-3.22	-0.369	0.110	-3.36
d6	-0.104	0.116	-0.90	-0.217	0.137	-1.59
d7	-0.288	0.125	-2.31	-0.438	0.146	-3.01
d8	-0.358	0.112	-3.19	-0.265	0.124	-2.14
d9	-0.689	0.105	-6.57	-0.762	0.111	-6.88
d10	-0.655	0.110	-5.93	-0.743	0.117	-6.33
d11	-0.524	0.106	-4.94	-0.548	0.115	-4.76
d12	-0.526	0.119	-4.41	-0.555	0.120	-4.64
d13	-0.608	0.113	-5.37	-0.525	0.128	-4.10
dt1	-0.020	0.014	-1.41	-0.022	0.014	-1.60
dt2	-0.038	0.014	-2.76	-0.038	0.014	-2.80
dt3	-0.001	0.007	-0.17	-0.002	0.007	-0.26
dt4	-0.020	0.009	-2.19	-0.019	0.009	-2.15
dt5	-0.004	0.006	-0.69	-0.005	0.006	-0.88
dt6	-0.022	0.009	-2.48	-0.024	0.009	-2.70
dt7	-0.015	0.011	-1.34	-0.016	0.011	-1.42
dt8	-0.008	0.010	-0.87	-0.008	0.010	-0.82
dt9	-0.004	0.007	-0.53	-0.005	0.007	-0.66
dt10	0.015	0.009	1.64	0.013	0.009	1.50
dt11	0.005	0.007	0.67	0.005	0.008	0.59
dt12	-0.004	0.012	-0.32	-0.005	0.012	-0.40
dt13	-0.015	0.009	-1.63	-0.017	0.009	-1.75
_cons	10.341	0.094	110.27	10.339	0.094	110.33
rho	0.699			0.693		
R2	0.991			0.992		
Adj. Pseudo R^2	0.980			0.983		
N	1,098			1,098		

The coefficients on the primary log output variables have the expected signs and are statistically significant. The undergrounding share variable is negative and statistically significant in both models. Regarding MVs in the long-period model, overall, zero per cent of the Australian sample and 16.7 per cent of the total sample are affected. The LSE-ADTT-GTC model performs better than the standard LSETLG specifications in terms of reducing monotonicity violations.

Figure 4.5.3 presents the average of the LSE-ADTT-GTC Cobb-Douglas and Translog efficiency scores for each Australian DNSP in the long period and compares them to the average efficiency scores from the standard LSE models and the comparative Opex PFP. The efficiency scores and rankings of the LSE-ADTT-GTC models are broadly aligned with those from the standard models.

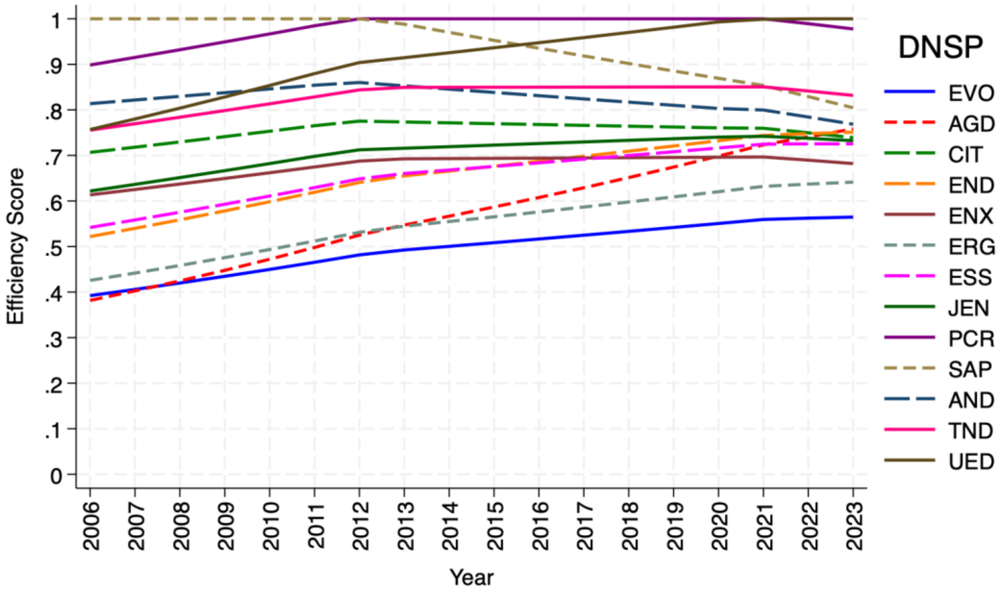
Figure 4.5.3 Average Efficiency Scores by DNSP (2006–2023)



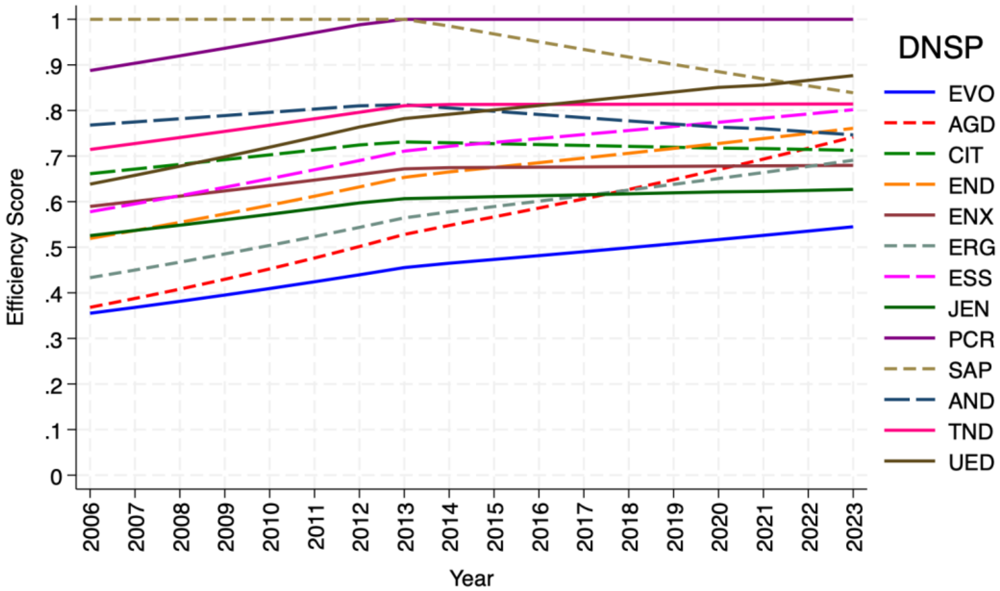
The efficiency scores for Australian DNSPs from the long-period LSECD and LSETLG models show strong correlations with the OPFP measures, at 0.763 and 0.818, respectively. These correlations are higher than those observed in the standard LSECD models (0.741) and comparable to the standard LSETLG model (0.798).

Figure 4.5.4 illustrates the trends in efficiency scores over time for the LSECD and LSETLG models.

Figure 4.5.4 LSE-ADTT-GTC Efficiency Trends by DNSP



LSECD Long Period



LSETLG Long Period

4.5.3 LSE-AJTT: LSE time-varying inefficiency through interaction & Jurisdictional Time Trend Variable

This specification differs from the specification in section 4.5.1 by including a jurisdictional time trend variable for NZ DNSPs. This allows the average rate of change of opex PFP to differ between Ontario and NZ. The general time trend variable now represents the average rate of change of opex PFP for Ontario. If we wish to decompose the productivity changes of Australian DNSPs into technical change and efficiency change, then we can assume that the general (ie, Ontario) time trend represents the underlying rate of technical change.

The results of estimating the Cobb-Douglas and Translog versions with the long sample are presented in Table 4.5.3. The coefficients on the primary log output variables have the expected signs and are statistically significant. The undergrounding share variable is negative and statistically significant in both models.

Regarding MVs in the long-period Translog model, 8.1 per cent of the Australian sample and 19.9 per cent of the total sample are affected. The LSE-AJTT model performs better than the standard LSETLG specifications in terms of reducing monotonicity violations. It also performs better than the LSE-ADTT model in terms of the frequency of MVs for Australian DNSPs, but is similar for all MVs.

Figure 4.5.5 presents the average of the LSE-AJTT Cobb-Douglas and Translog efficiency scores for each Australian DNSP in the long period and compares them to the efficiency scores from the standard LSE models and the comparative Opex PFP. The efficiency scores and rankings of the LSE-AJTT models are broadly aligned with those from the standard models. The efficiency scores for Australian DNSPs from the long-period LSECD and LSETLG models show strong correlations with the multilateral Opex PFP index, at 0.757 and 0.820, respectively. These correlations are marginally higher than those observed in the standard LSECD and LSETLG models (0.741 and 0.798 respectively).

Figure 4.5.5 Average Efficiency Scores by DNSP (2006–2023)

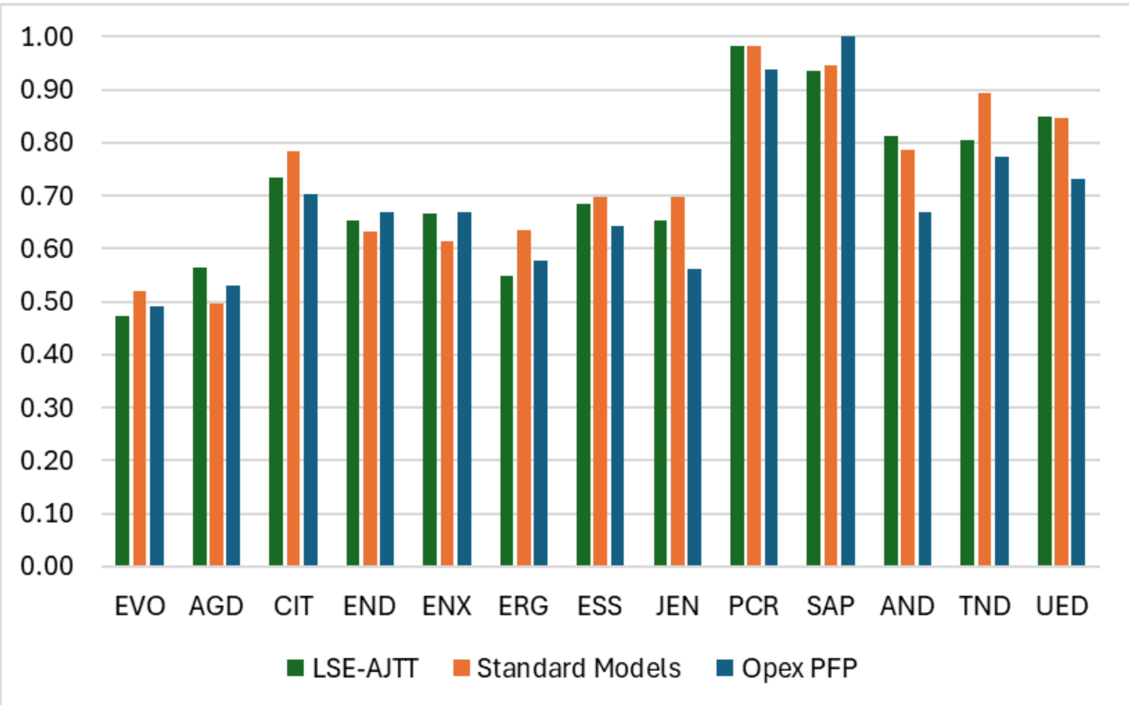
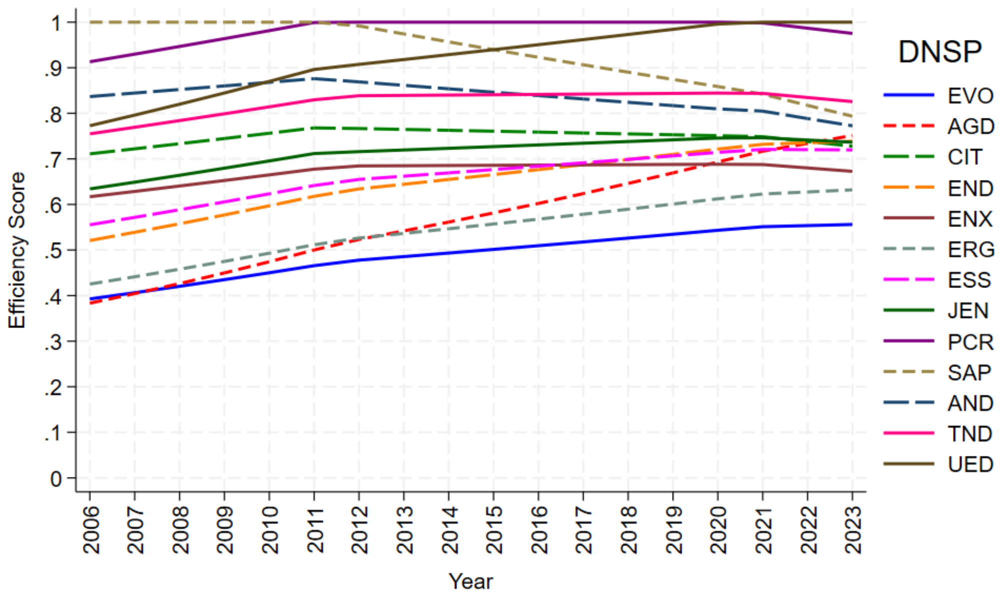


Table 4.5.3 LSE-AJTT: Parameter Estimates

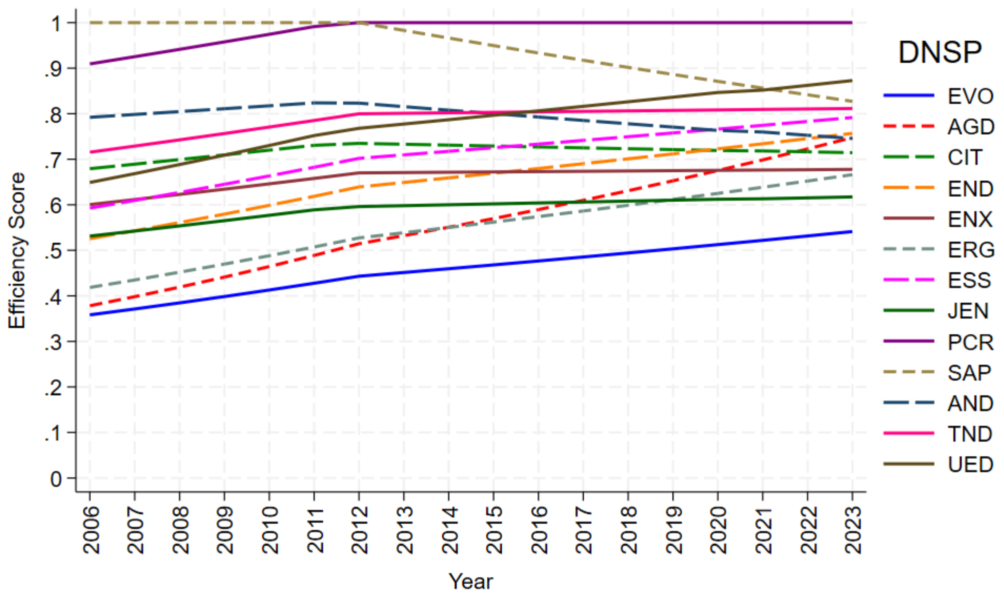
	LSECD - Long Period			LSETLG- Long Period		
Variable	*Coefficient*	*SE*	*t–ratio*	*Coefficient*	*SE*	*t–ratio*
ly1	0.616	0.068	9.01	0.471	0.076	6.23
ly2	0.217	0.032	6.81	0.216	0.032	6.86
ly3	0.134	0.059	2.26	0.269	0.064	4.23
ly11				-0.388	0.490	-0.79
ly12				0.270	0.114	2.37
ly13				0.066	0.389	0.17
ly22				-0.016	0.039	-0.40
ly23				-0.236	0.094	-2.53
ly33				0.237	0.309	0.77
lz1	-0.100	0.023	-4.39	-0.103	0.025	-4.10
yr	0.004	0.002	2.16	0.006	0.002	3.20
jur2	-0.416	0.106	-3.92	-0.471	0.105	-4.49
jur3	-0.166	0.104	-1.60	-0.273	0.103	-2.65
d2	-0.133	0.147	-0.91	-0.184	0.156	-1.18
d3	-0.433	0.118	-3.67	-0.461	0.118	-3.90
d4	-0.283	0.121	-2.35	-0.360	0.122	-2.96
d5	-0.325	0.113	-2.88	-0.375	0.121	-3.09
d6	-0.103	0.130	-0.79	-0.181	0.150	-1.20
d7	-0.304	0.140	-2.17	-0.444	0.158	-2.80
d8	-0.377	0.129	-2.93	-0.262	0.138	-1.90
d9	-0.703	0.120	-5.87	-0.774	0.124	-6.22
d10	-0.655	0.123	-5.30	-0.735	0.129	-5.70
d11	-0.547	0.120	-4.58	-0.563	0.127	-4.44
d12	-0.529	0.133	-3.97	-0.553	0.132	-4.18
d13	-0.626	0.125	-5.00	-0.530	0.138	-3.85
dt1	-0.015	0.015	-0.99	-0.019	0.015	-1.23
dt2	-0.034	0.015	-2.21	-0.034	0.015	-2.29
dt3	0.004	0.009	0.43	0.002	0.009	0.27
dt4	-0.015	0.010	-1.57	-0.016	0.010	-1.63
dt5	0.000	0.007	0.05	-0.001	0.007	-0.21
dt6	-0.018	0.010	-1.75	-0.022	0.010	-2.11
dt7	-0.010	0.013	-0.78	-0.011	0.013	-0.91
dt8	-0.004	0.011	-0.36	-0.004	0.011	-0.33
dt9	0.001	0.008	0.13	0.000	0.008	-0.04
dt10	0.019	0.010	1.97	0.017	0.010	1.77
dt11	0.010	0.009	1.16	0.009	0.008	1.08
dt12	0.000	0.013	0.01	-0.002	0.012	-0.14
dt13	-0.011	0.010	-1.08	-0.013	0.010	-1.30
t-nz	0.022	0.003	6.71	0.022	0.003	6.75
_cons	2.419	3.709	0.65	-1.528	3.733	-0.41
rho	0.732			0.725		
R2	0.992			0.992		
Adj. Pseudo R^2	0.981			0.984		
N	1,098			1,098		

Figure 4.5.6 illustrates the trends in efficiency scores over time for the LSECD and LSETLG long sample models.

Figure 4.5.6 LSE-AJTT Efficiency Trends by DNSP



LSECD Long Period



LSETLG Long Period

4.5.4 LSE-AJTT-GTC: LSE time-varying inefficiency through interaction & Jurisdictional Time Trend Variable & General Technical Change

This specification differs from the specification in section 4.5.3 by including the GTC terms instead of the year variable. The results of estimating the Cobb-Douglas and Translog versions with the long sample are presented in Table 4.5.4. The coefficients on the primary log output variables have the expected signs and are statistically significant. The undergrounding share variable is negative and statistically significant in both models.

Regarding MVs in the long-period model, overall, 8.1 per cent of the Australian sample and 22.9 per cent of the total sample are affected. The LSE-AJTT-GTC model performs better than the standard LSETLG specifications in terms of reducing monotonicity violations. However, it does not perform as well as the LSE-ADTT-GTC model in terms of frequency of MVs.

Figure 4.5.7 presents the average of the LSE-AJTT-GTC CD and TLG efficiency scores for each Australian DNSP in the long period and compares them to the average efficiency scores from the standard LSE models and the comparative Opex PFP. The efficiency scores and rankings of the LSE-AJTT-GTC models are broadly aligned with those from the standard models.

Figure 4.5.7 Average Efficiency Scores by DNSP (2006–2023)

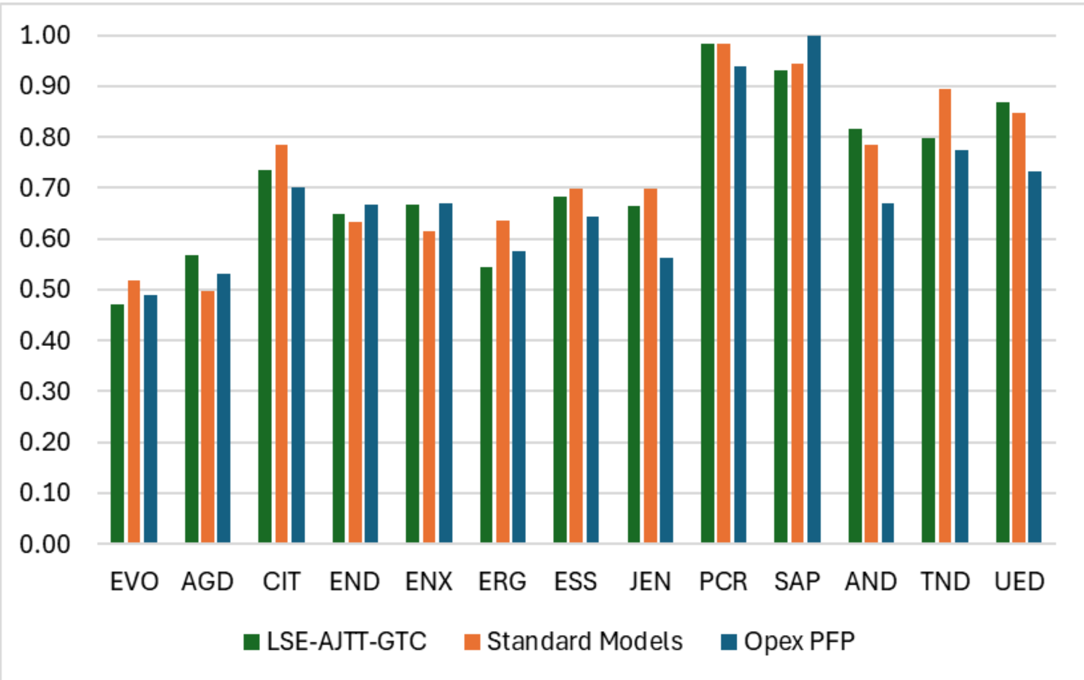
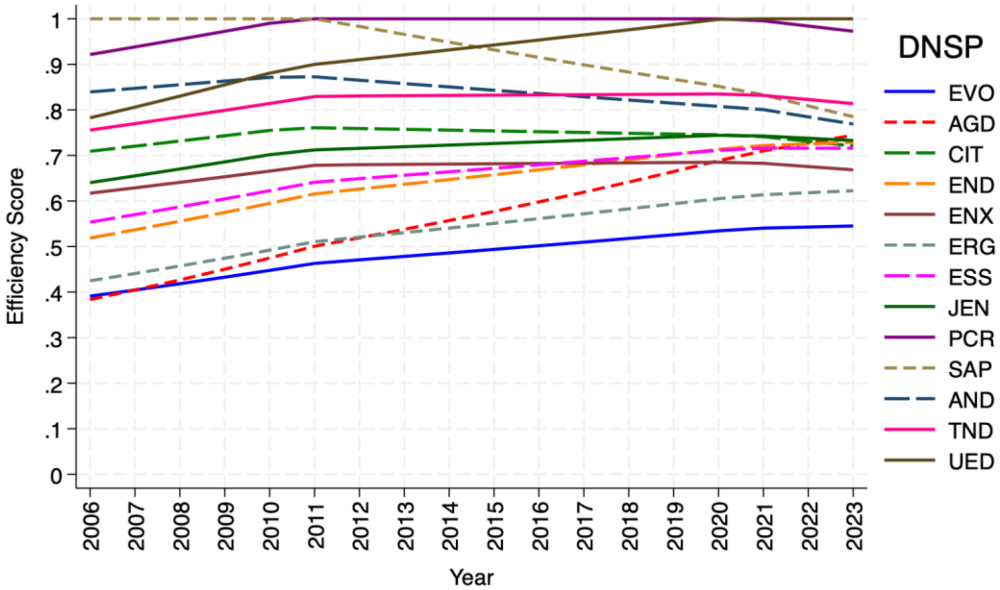


Figure 4.5.8 illustrates the trends in efficiency scores over time for the long sample models. The efficiency scores for Australian DNSPs from the long-period LSECD and LSETLG models show strong correlations with the multilateral Opex PFP measures, at 0.756 and 0.812, respectively. These correlations are marginally higher than those observed in the standard LSECD and LSETLG models (0.741 and 0.798 respectively).

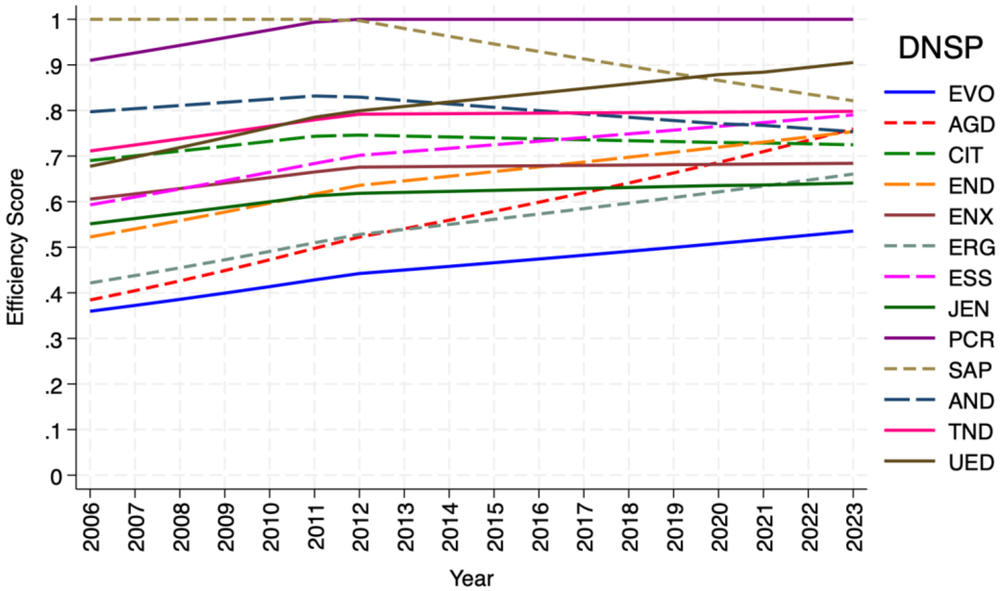
Table 4.5.4 LSE-AJTT-GTC: Parameter Estimates

	LSECD - Long Period			LSETLG- Long Period		
Variable	*Coefficient*	*SE*	*t–ratio*	*Coefficient*	*SE*	*t–ratio*
ly1	0.646	0.063	10.18	0.527	0.071	7.45
ly2	0.211	0.030	7.10	0.208	0.030	7.01
ly3	0.112	0.056	2.01	0.222	0.059	3.76
ly11				-0.231	0.475	-0.49
ly12				0.210	0.109	1.94
ly13				-0.019	0.376	-0.05
ly22				0.008	0.037	0.22
ly23				-0.204	0.090	-2.28
ly33				0.283	0.297	0.95
lz1	-0.104	0.021	-5.00	-0.101	0.023	-4.33
gtc2	0.013	0.012	1.07	0.016	0.012	1.28
gtc3	0.061	0.016	3.80	0.066	0.016	4.18
gtc4	0.061	0.018	3.31	0.070	0.018	3.82
gtc5	0.054	0.021	2.61	0.066	0.020	3.22
gtc6	0.007	0.022	0.33	0.023	0.022	1.02
jur2	-0.422	0.094	-4.50	-0.480	0.095	-5.08
jur3	-0.168	0.091	-1.84	-0.267	0.092	-2.90
d2	-0.141	0.128	-1.10	-0.203	0.141	-1.45
d3	-0.439	0.101	-4.34	-0.479	0.103	-4.65
d4	-0.286	0.106	-2.71	-0.358	0.109	-3.29
d5	-0.334	0.099	-3.37	-0.388	0.110	-3.54
d6	-0.107	0.113	-0.95	-0.184	0.134	-1.37
d7	-0.311	0.123	-2.54	-0.447	0.143	-3.13
d8	-0.392	0.111	-3.54	-0.302	0.123	-2.46
d9	-0.718	0.104	-6.89	-0.777	0.111	-7.03
d10	-0.661	0.109	-6.08	-0.735	0.116	-6.34
d11	-0.559	0.105	-5.33	-0.574	0.115	-5.01
d12	-0.534	0.117	-4.56	-0.545	0.119	-4.59
d13	-0.644	0.110	-5.84	-0.574	0.126	-4.56
dt1	-0.012	0.014	-0.88	-0.015	0.014	-1.05
dt2	-0.032	0.014	-2.28	-0.031	0.014	-2.28
dt3	0.006	0.007	0.90	0.006	0.007	0.82
dt4	-0.013	0.009	-1.45	-0.013	0.009	-1.43
dt5	0.003	0.006	0.43	0.002	0.006	0.28
dt6	-0.015	0.009	-1.72	-0.018	0.009	-1.97
dt7	-0.008	0.011	-0.69	-0.008	0.011	-0.71
dt8	-0.001	0.010	-0.12	-0.001	0.010	-0.06
dt9	0.004	0.007	0.53	0.003	0.007	0.39
dt10	0.022	0.009	2.46	0.020	0.009	2.31
dt11	0.012	0.008	1.64	0.012	0.008	1.57
dt12	0.003	0.011	0.27	0.002	0.011	0.18
dt13	-0.008	0.009	-0.88	-0.009	0.009	-1.01
t_nz	0.025	0.003	8.34	0.026	0.003	8.69
_cons	10.383	0.093	111.88	10.380	0.093	111.17
rho	0.696			0.695		
R2	0.992			0.993		
Adj. Pseudo R^2	0.981			0.984		
N	1,098			1,098		

Figure 4.5.8 LSE-AJTT-GTC Efficiency Trends by DNSP



LSECD Long Period



LSETLG Long Period

4.6 Cornwell et al (1990) models

In the FECSS model, inefficiency is captured through a fixed effect for each firm combined with a firm-specific quadratic time trend. We test two specifications of the FECSS model: the first uses the standard technical change variable, while the second incorporates a general index of technical change (GTC). Only a brief summary of the empirical results is presented here.[46]

4.6.1 FECSS Model

Considering both the Cobb–Douglas and Translog specifications across the long and short periods, none of the coefficients are statistically significant at the 5 per cent level, except for the coefficient on circuit length in the Translog model during the short period. As a result, the estimated output elasticities are inconsistent with expectations. Moreover, the incidence of monotonicity violations is extremely high, affecting 99.6 per cent of the Australian sample in the long period and 100 per cent in the short period.

4.6.2 FECSS-GTC Model

This specification extends the model presented in Section 4.6.1 by incorporating GTC variables. Across both the Cobb–Douglas and Translog forms, and for both the long and short periods, almost all coefficients fail to reach statistical significance at the 5 per cent level—apart from the circuit length coefficient in the short-period Translog model. In addition, the rate of monotonicity violations is exceptionally high, impacting 81.6 per cent of the Australian sample in the long period and the entirety of the sample in the short period.

The shortcomings of the FECSS models can be ascribed to two characteristics of these models. Firstly, they include a fixed effect for each DNSP. As seen in section 3.2.1, the introduction of fixed effects for all DNSPs makes it difficult to identify the output elasticities, because such of the sample data variation relied on to identify the main parameters of the model is cross-sectional variations, which in this model is captured by the fixed effects. Secondly, there are DNSP-specific coefficients on both the time trend and the squared time trend, and with 61 DNSPs in the sample, there are a great many parameters. This results in over-fitting.

4.7 Concluding Comments

By testing the models proposed in the theoretical section, we confirmed the challenges associated with maximum likelihood estimation of more complex SFA models. While the more sophisticated SFA models are theoretically appealing, they often present difficulties in terms of estimation stability and optimization. This proved to be the case for the most flexible of the models we tested, the four-component model. Among the bc95 models we tested, only one produced sufficiently reliable results that for it to be considered a candidate. Among the kumb90, the four models presented can be considered candidate models, having produced

[46] Details of these models can be found in Attachment B.

reasonably reliable results for the long sample period. The reliability of these models appears to decrease as the models become more complex. However, while the jurisdiction-specific model performs better in terms of monotonicity violations, the DNSP-specific approach offers greater flexibility in efficiency scores.

In contrast, the LSE time-varying models proved to be more robust and consistent across specifications and time periods. This reflects their relative simplicity compared with the SFA models, which are more complex as the stochastic term comprises two components, white noise and inefficiency, with the latter identified through the composed-error formulation. The LSE models have a simple white noise disturbance, permitting a more straight-forward computation. Hence, some of the LSE models that include DNSP-specific efficiency trends are reasonably reliable and can be considered candidate models. This is not the case for the highly flexible FECSS models which were unsuccessful in identifying the main parameters of the opex model.

Most of the candidate time varying models show meaningful improvements when estimated with the long-period sample, compared to standard applications, including a reduction in monotonicity violations and successful improved correlation of estimated efficiency scores with the multilateral Opex PFP index. In contrast, the models encountered difficulties when estimated with the short-period sample, suggesting that sample size may influence their performance. In Section 5, we assess all models against the established evaluation criteria and present a comparative analysis of their relative strengths and limitations.

5 Preliminary Assessment: Time-varying Inefficiency Modelling Alternatives

In this section, we evaluate the remaining model alternatives against selected statistical criteria (see section 2.3). This includes assessing the performance of the time-varying specifications to determine whether they offer meaningful improvements over the current framework. The goal is to identify whether any of the tested approaches merit adoption based on the established evaluation criteria.

As mentioned in section 5, we give greatest attention to the long sample period for time-varying inefficiency models, as they are capable of directly accounting for changes in inefficiency, which may reduce the need for, or weight given to, the short sample results. Further, particular emphasis should be placed on the performance of the Cobb–Douglas models, as they have proven the most reliable in previous applications.

5.1 Comparative assessment of Time Varying models alternatives

5.1.1 Convergence of Maximum Likelihood Estimation

As discussed, the SFA models which use maximum likelihood estimation (MLE) sometimes do not converge properly or fully, producing messages such as "convergence not achieved" or "cannot compute an improvement -- flat region encountered", or failing to estimate some or all parameter standard errors. Table 5.1 summarises the convergence results for the BC95-JTT-HN model presented in section 4.3 and the four Kumb90 models presented in section 4.4, in comparison to the standard models from ABR24.

Table 5.1 MLE Convergence

	Long Period		*Short Period*	
	SFACD	*SFATLG*	*SFACD*	*SFATLG*
Standard ABR24 Models	✓	✓	✓	X[(a)]
BC95-JTT-HN	✓	✓	(b)	(b)
Kumb90-JTT-HN	✓	✓	(b)	(b)
Kumb90-JTT-HN-GTC	✓	✓	✓	✓
Kumb90-AJTT-HN	✓	✓	(b)	(b)
Kumb90-AJTT-HN-GTC	✓	✓	✓	X[(c)]

Notes: (a) "convergence not achieved".
(b) "cannot compute an improvement -- flat region encountered".
(c) Failed to estimate some or all parameter standard errors.

The results in Table 5.1 show that most of the time-varying SFA models presented in sections 4.3 and 4.4 have some convergence problems in the short-term. The cells with crosses either failed to converge or failed or estimate some or all of the parameter standard errors. This means the short sample SFATLG models for ABR24 and Kumb90-AJTT-HN-GTC are precluded from further consideration.

Note that we do not consider the case of "cannot compute an improvement -- flat region encountered" necessarily means the model solution is unsatisfactory. It could be that it converged to a local optimum rather than the global optimum and the neighbourhood around that optimum happens to be flat hence no improvement is possible in that neighbourhood. This can happen even to other cases that do not display this specific message. Nonetheless, we highlight it as a point of caution.

5.1.2 Theory-Consistent Parameter Signs and Monotonicity Violations

Table 5.2 compares the consistency with economic theory performance of the time-varying SFA models with the standard SFA ABR24 specifications. Only the long sample period results are presented, for reasons given, and due to the poor convergence performance of the short-period models. In the standard models, all models apart from the short-sample SFATLG model (which failed to converge) deliver output and undergrounding coefficients with the correct signs and statistical significance. However, they perform poorly on monotonicity, especially for Australian DNSPs.

The BC95-JTT-HN models improve markedly over the standard models. Across both CD and TLG specifications and both sample periods, the output coefficients retain the expected signs and significance, and the undergrounding term remains consistently negative and significant. Monotonicity violations fall to 9.8 per cent in the long-period SFATLG and 34.6 per cent in the short-period version for Australian sample. The total sample shows a similar pattern of improvement.

For Kumb90-JTT-HN models, the output coefficients again display the correct signs and significance. Whilst the undergrounding coefficient is negative in long sample models, and in the SFATLG in the short period, it is positive (but insignificant) in the SFACD short period. Monotonicity violations are fully eliminated in the long-period model for both Australian DNSPs and the total sample. In the short period, there are just 3.8 per cent of monotonicity violations for Australian DNSPs and 22.1 for the total sample. In terms of MVs this model outperforms both the standard and BC95 variants.

In the Kumb90-JTT-HN-GTC models, all specifications produce significant positive output coefficients. The undergrounding coefficient is negative only in the SFACD long period model and positive in all other models (though it is significant only in the short-period SFATLG). Monotonicity violations are again eliminated in the long-period model across both Australian DNSPs and the total sample and are reduced to below 9.0 per cent in the short-period.

By contrast, the Kumb90-AJTT-HN models perform poorly in terms of monotonicity violations. In the long-period specification, more than 50 per cent of the Australian sample is affected by MVs. While the output coefficients have the correct signs and are statistically significant, the undergrounding coefficient fails to meet the expected coefficient criteria in the short-period SFATLG, where it is positive but not statistically significant.

Table 5.2 Consistency with Economic Theory – SFA Models Performance

		First-order Output Coefficients		*Undergrounding*	*MVs (%)*	
		Correct Signs	*Significance**	*negative***	*Australia*	*Total*
Standard ABR24 Models	SFACD - LP	✓	✓	✓	0.0	0.0
	SFATLG - LP	✓	✓	✓	79.5	45.4
	SFACD - SP	✓	✓	✓	0.0	0.0
	SFATLG - SP	NA	NA	NA	NA	NA
BC95 JTT-HN	SFACD - LP	✓	✓	✓	0.0	0.0
	SFATLG - LP	✓	✓	✓	9.8	14.8
	SFACD - SP	✓	✓	✓	0.0	0.0
	SFATLG - SP	✓	✓	✓	34.6	23.6
Kumb90 JTT-HN	SFACD - LP	✓	✓	✓	0.0	0.0
	SFATLG - LP	✓	✓	✓	0.0	0.0
	SFACD - SP	✓	✓	x	0.0	0.0
	SFATLG - SP	✓	✓	✓	3.8	22.1
Kumb90 JTT-HN-GTC	SFACD - LP	✓	✓	✓	0.0	0.0
	SFATLG - LP	✓	✓	x	0.0	0.0
	SFACD - SP	✓	✓	x	0.0	0.0
	SFATLG - SP	✓	✓	x	8.3	8.1
Kumb90 AJTT-HN	SFACD - LP	✓	✓	✓	0.0	0.0
	SFATLG - LP	✓	✓	✓	55.1	27.0
	SFACD - SP	✓	✓	✓	0.0	0.0
	SFATLG - SP	✓	✓	x	15.4	20.1
Kumb90 AJTT-HN-GTC	SFACD - LP	✓	✓	✓	0.0	0.0
	SFATLG - LP	✓	✓	x	54.7	29.8
	SFACD - SP	✓	#	x	0.0	0.0
	SFATLG - SP	NA	NA	NA	NA	NA

Notes: * ✓ indicates significance at 0.05 level; # indicates significance at 0.10 level.
** Irrespective of statistical significance.
NA means the model did not produce standard errors.

The Kumb90-AJTT-HN-GTC is successfully estimated only for the long-period sample. In the short-period SFATLG specification, the model fails to produce standard errors for the coefficients, while in the short-period SFACD model, one of the output variables (RMD) is not statistically significant at 5 per cent level, although it is significant at 10 per cent level. For the long period, the output coefficients display the expected signs and are statistically significant. The rate of monotonicity violations is high, approximately 55 per cent for Australian DNSPs and 30 per cent across the full sample, an improvement over the standard model, though still less favourable than other Kumb90 specifications.

In summary, the BC95-JTT-HN models have output and share of undergrounding coefficients with the expected signs and statistical significance and are the only models that use the *bc95* formulation that are successful. Among the kumb90 formulations, the Kumb90-JTT-HN and

Kumb90-JTT-HN-GTC models achieve the best results, either eliminating MVs entirely or reducing them to minimal levels.

Table 5.3 compares the consistency with economic theory performance of the time-varying LSE models with the standard ABR24 LSE specifications. The standard and time-varying LSE models do not present any issues with coefficient signs or statistical significance.

Table 5.3 Consistency with Economic Theory – LSE Models Performance

		First-order Output Coefficients		*Undergrounding negative***	*MVs (%)*	
		Correct Signs	*Significance**		*Australia*	*Total*
Standard ABR24 Models	LSECD - LP	✓	✓	✓	0.0	0.0
	LSETLG - LP	✓	✓	✓	22.2	21.9
	LSECD - SP	✓	✓	✓	0.0	0.0
	LSETLG - SP	✓	✓	✓	48.7	36.6
LSE-ADTT	LSECD - LP	✓	✓	✓	0.0	0.0
	LSETLG - LP	✓	✓	✓	11.1	18.7
	LSECD - SP	✓	✓	✓	0.0	0.0
	LSETLG - SP	✓	✓	✓	1.3	21.6
LSE-ADTT-GTC	LSECD - LP	✓	✓	✓	0.0	0.0
	LSETLG - LP	✓	✓	✓	0.0	16.7
	LSECD - SP	✓	✓	✓	0.0	0.0
	LSETLG - SP	✓	✓	✓	0.0	20.6
LSE-AJTT	LSECD - LP	✓	✓	✓	0.0	0.0
	LSETLG - LP	✓	✓	✓	8.1	19.9
	LSECD - SP	✓	✓	✓	0.0	0.0
	LSETLG - SP	✓	✓	✓	0.0	19.7
LSE-AJTT-GTC	LSECD - LP	✓	✓	✓	0.0	0.0
	LSETLG - LP	✓	✓	✓	8.1	22.9
	LSECD - SP	✓	✓	✓	0.0	0.0
	LSETLG - SP	✓	✓	✓	0.0	19.7

Notes: * a ✓ indicates significance at 0.05 level; a # indicates significance at 0.10 level.
** Irrespective of statistical significance.
NA means the model did not produce standard errors.

In terms of monotonicity violations, the standard specification shows notable issues, with 22.2 per cent of Australian observations affected in the SFATLG long period and 48.7 per cent in the short period. The LSE-ADTT improves on this, reducing MVs for Australian DNSPs to 11.1 per cent in the long period and 1.3 per cent in the short period. The LSE-ADTT-GTC model performs even better, fully eliminating violations in both periods for the Australian sample. The LSE-AJTT model shows good results in the long period (8.1 per cent) and eliminates the MVs in the short period for the Australian sample. The LSE-AJTT-GTC has MV results similar to the LSE-AJTT model.

In summary, all tested LSE time-varying models perform better than the standard LSE model in terms of monotonicity violations. The LSE-ADTT-GTC model shows the best MV performance, closely followed by the LSE-AJTT and LSE-AJTT-GTC specifications.

5.1.3 Reasonableness of Output Elasticities

This section evaluates the cost output elasticities derived from the models.[47] Unlike the CD specification which assumes constant elasticities, the TLG specification allows elasticities to vary across observations. Consequently, the output elasticities for the TLG models reported here have been calculated at the sample mean.

Figure 5.1 presents the cost output elasticities for each output variable (evaluated at the sample mean for the TLG models) and the total of the cost elasticities for the long-period SFA models. Across all specifications shown, the total output elasticity is close to 1, suggesting near-constant returns to scale on average. The standard ABR24 models exhibit the lowest total output elasticities (0.962 for SFACD and 0.927 for SFATLG), whereas the Kumb90-AJTT-HN and Kumb90-AJTT-HN-GTC Translog models have total output elasticities slightly greater than 1 (1.071 to 1.003), indicating diseconomies of scale.

In terms of individual outputs, in the standard models most weight is placed on RMD, followed by customer numbers and circuit length. The time-varying inefficiency models place most weight in customer numbers. In most cases, circuit length has a lower elasticity than RMD, with the exceptions of the Translog versions of Kumb90-AJTT-HN and Kumb90-AJTT-HN-GTC. The comparative sizes of the output elasticities are reasonably consistent between the CD and TLG versions these models. Kumb90-JTT-HN is the most consistent in this way.

Figure 5.2 illustrates the cost output elasticities for each output and the total elasticity for each specification of the long-period LSE models. In terms of individual outputs, all models place the greatest weight on customer numbers. When the specification is Cobb–Douglas, circuit length is typically the second most heavily weighted output, followed by RMD. In contrast, under the Translog specification, this order is changed, with RMD receiving more weight than circuit length. The only exception to this pattern is the standard LSETLG model, which places the greatest weight on RMD, followed closely by customer numbers, with circuit length receiving the least weight. As a general rule, the LSETLG models have higher elasticities for RMD and lower elasticities for customer numbers compared to the LSECD models.

[47] Cost output elasticity reflects the percentage change in cost associated with a 1 per cent change in output. The sum of the elasticities for each output is the total output elasticity. A total output elasticity equal to 1 indicates constant returns to scale, ie, costs increase proportionally with output. If the elasticity is less than 1, it suggests economies of scale, meaning a 1 per cent increase in output results in a less than 1 per cent increase in cost. Conversely, an elasticity greater than 1 implies diseconomies of scale, where costs increase more than proportionally with output. A negative elasticity for any individual output indicates a violation of monotonicity, the theoretic requirement that increasing output cannot reduce costs.

Figures 5.3 and 5.4 present the cost output elasticities for the short-period SFA and LSE models respectively.

Figure 5.1 SFA Cost Elasticity – Long Period

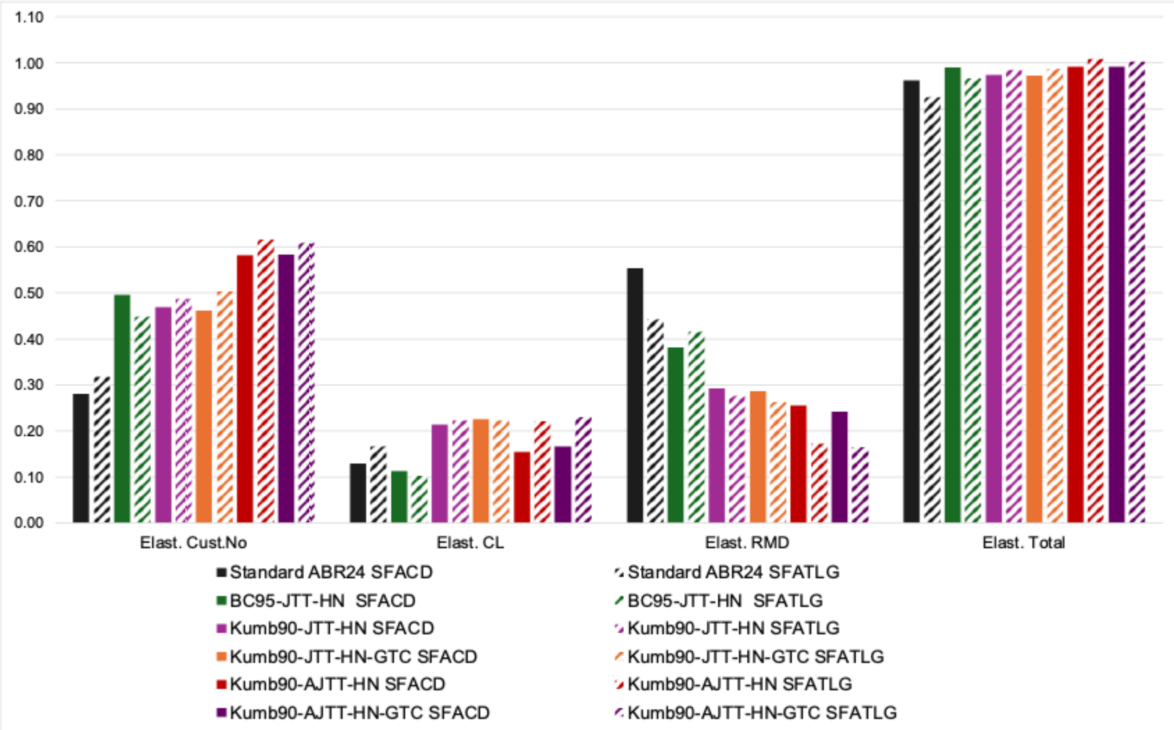


Figure 5.2 LSE Cost Elasticity – Long Period

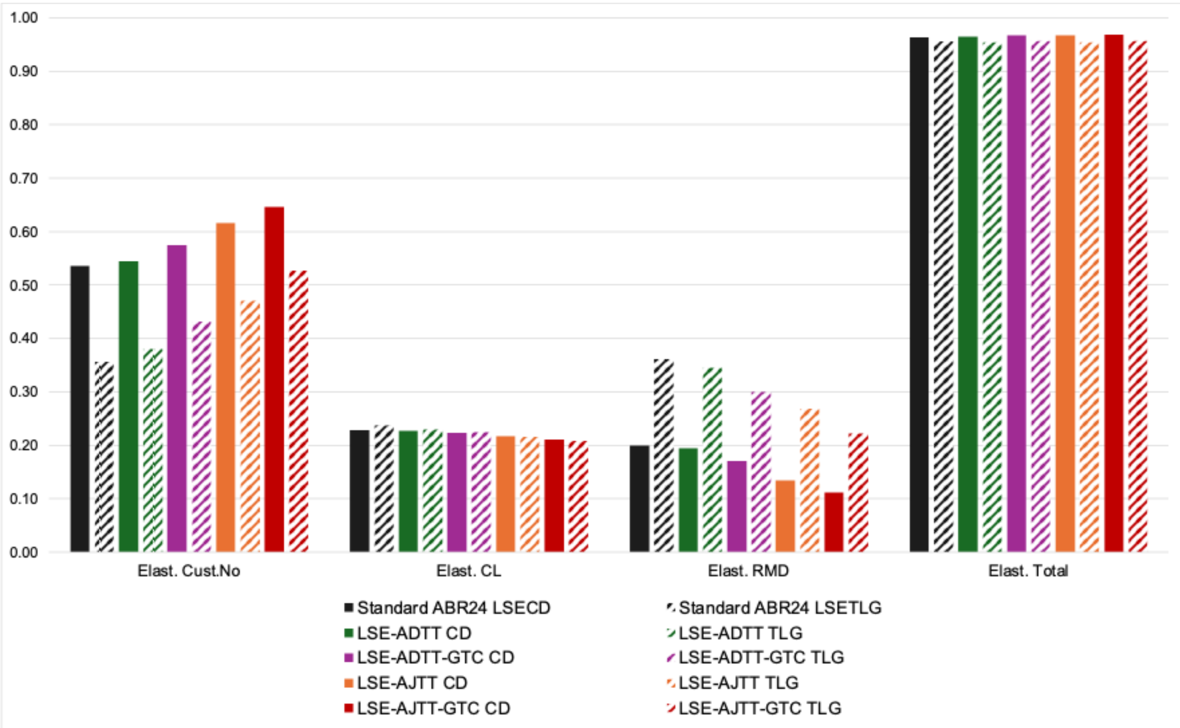


Figure 5.3 SFA Cost Elasticity – Short Period

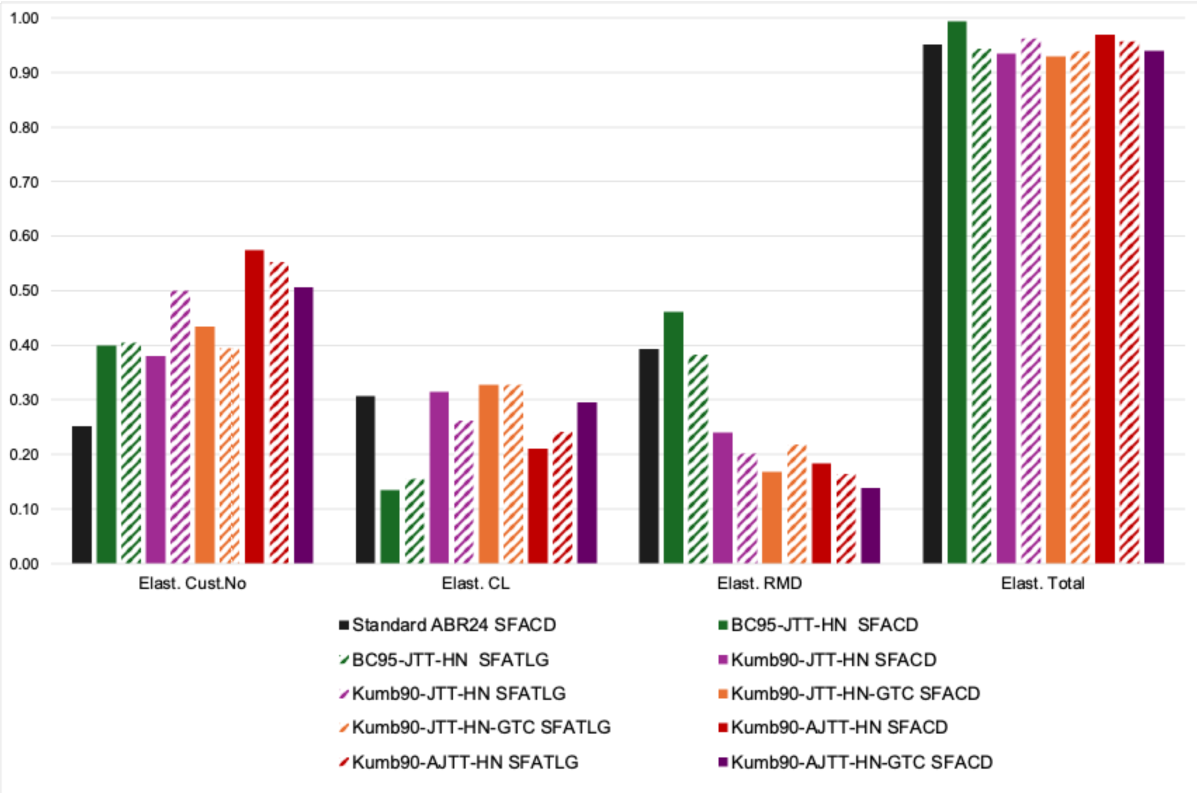
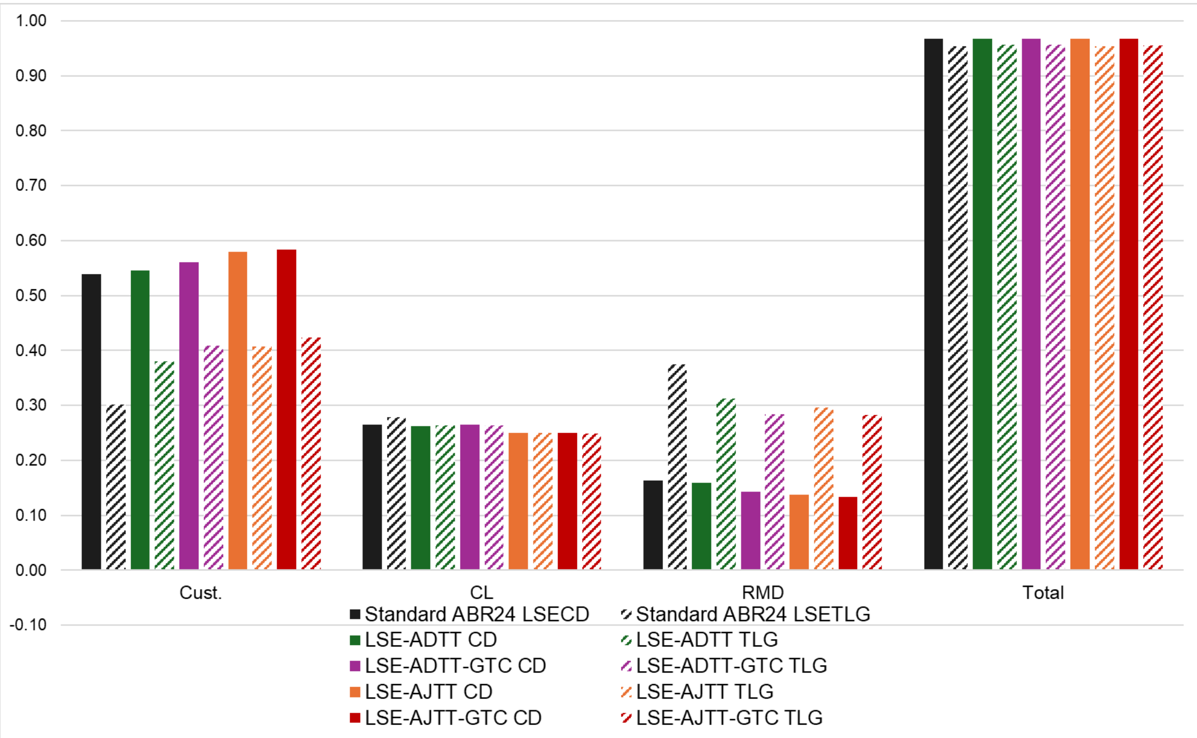


Figure 5.4 LSE Cost Elasticity – Short Period



5.1.4 Goodness of Fit Performance

Table 5.4 presents the Pseudo-Adjusted R^2 and Bayesian Information Criterion (BIC) results across the standard and time-varying SFA model specifications.[48] Lower values of BIC imply a "better fit". Both measures penalise the addition of explanatory variables, so variables with only marginal explanatory power could reduce the goodness-of-fit.

While these goodness-of-fit indicators are useful, caution is required in their application. BIC is suitable for comparing models with different functional forms or variables, but only when they belong to the same model family (in particular, the same likelihood framework, the same data and the same scale of the dependent variable). Strictly speaking, BIC is less reliable, and potentially questionable, when models come from different statistical families or assume different distributions for the same outcome. In this context, it is therefore only appropriate to compare the Kumb specifications against each other. The Pseudo-Adjusted R^2 is more flexible than BIC, though less theoretically grounded. It can be applied across different model families, provided that the dependent variable remains on the same scale and the same data are used.

Table 5.4 Goodness of Fit – SFA Models Performance

		Pseudo Adj. R^2		*BIC*	
		Long Period	*Short Period*	*Long Period*	*Short Period*
Standard ABR24 Models	SFACD	0.991	0.993	-1103	-833
	SFATLG	0.992	NA	-1108	NA
	avg	0.992	NA	-1105	NA
BC95-JTT-HN	SFACD	0.997	0.996	-432	-387
	SFATLG	0.997	0.997	-504	-485
	avg	0.997	0.997	-468	-436
Kumb90-JTT-HN	SFACD	0.994	0.996	-1503	-1245
	SFATLG	0.994	0.997	-1477	-1239
	avg	0.994	0.996	-1490	-1242
Kumb90-JTT-HN-GTC	SFACD	0.994	0.996	-1441	-1242
	SFATLG	0.994	0.997	-1418	-1229
	avg	0.994	0.997	-1430	-1236
Kumb90-AJTT-HN	SFACD	0.994	0.997	-1470	-1235
	SFATLG	0.994	0.997	-1456	-1214
	avg	0.994	0.997	-1463	-1225
Kumb90-AJTT-HN-GTC	SFACD	0.994	0.997	-1410	-1206
	SFATLG	0.994	NA	-1396	NA
	avg	0.994	NA	-1403	NA

[48] $BIC = -2 \cdot \ln(ll) + k \cdot ln(n)$, where $\ln(ll)$ is the is the log-likelihood of the fitted model, k is the number of estimated parameters and n is the number of observations. Pseudo-Adjusted $R^2 = 1 - (1-r)^2\,(n-1)/(n-k)$, where r is the Pearson correlation coefficient between the actual and predicted values of the dependent variable.

Some observations from Table 5.4:

- The standard ABR24 models have a substantially lower Pseudo-Adjusted R^2 than the other models shown. The average for the long-period standard models is 0.91, compared to 0.97 for the BC95-JTT-HN model, and 0.94 for all of the Kumb90 models. However, the pseudo-adjusted R^2 values are not sufficiently distinct across the time-varying models to allow any definitive conclusions.
- Using the BIC, the standard models' average value of –1105 in the long period is inferior to the Kumb90 models, which average from –1403 to –1490.
- The BC95-JTT-HN yields the least favourable BIC scores.[49] However, as noted above, BIC is not a reliable criterion for comparing models from different families, and a lower BIC for BC95-JTT-HN does not necessarily imply a poorer goodness-of-fit relative to the others.
- All the Kumb90 models have a similar average Pseudo-Adjusted R^2 (in the long period, 0.94). In terms of BIC, Kumb90-JTT-HN's average BIC value of –1490 in the long period outperforms the other Kumb90 models. The Kumb90-AJTT-HN model ranks second in the long period (BIC = –1463). It is important to note, that this is arguably only a small loss of fit given that the Kumb90-AJTT-HN model provides for a separate inefficiency time trend for each DNSP.

These findings suggest that the time-varying SFA models perform similarly in terms of Pseudo-Adjusted R^2. Within the Kumb90 specifications, the Kumb90-JTT-HN model offers the most favourable trade-off between model fit and parsimony, as indicated by the BIC.

Turning to the LSE models, Table 5.5 presents the results of Pseudo Adjusted R^2 across the standard and time varying LSE alternative models. BIC is not available for the panel-corrected standard error models used for LSE estimation. The LSE-AJTT and LSE-AJTT-GTC models exhibit the best overall fit among the LSE specifications, with average Pseudo-R^2 values of 0.983 for the long period and 0.986 for the short period. The other alternative time-trend LSE models follow closely, with only slightly lower fit measures. These results represent only marginal improvements over the standard ABR24 LSE models, which record the lowest Pseudo-R^2 values in the group, averaging 0.980 in the long period and 0.983 in the short. Given the small magnitude of these differences, the gains in model fit may not be particularly meaningful in practical terms.

[49] This discrepancy reflects the lower log-likelihood values of the BC95 specification. For example, the CD version of the BC95-JTT-HN model has a log likelihood of 261.8, compared to 800.6 for CD version of the Kumb90-JTT-HN model.

Table 5.5 Goodness of Fit – LSE Models Performance

		Pseudo Adj. R2	
		Long Period	*Short Period*
Standard ABR24 Models	LSECD	0.979	0.981
	LSETLG	0.981	0.984
	avg	0.980	0.983
LSE-ADTT	LSECD	0.980	0.982
	LSETLG	0.982	0.986
	avg	0.981	0.984
LSE-ADTT-GTC	LSECD	0.980	0.982
	LSETLG	0.983	0.986
	avg	0.982	0.984
LSE-AJTT	LSECD	0.981	0.984
	LSETLG	0.984	0.987
	avg	0.983	0.986
LSE-AJTT-GTC	LSECD	0.981	0.984
	LSETLG	0.984	0.987
	avg	0.983	0.986

5.1.5 Statistical Tests Performance

Table 5.6 presents key diagnostic statistics, assessing multicollinearity, normality of residuals, the presence of severe outliers, and the joint significance of Translog parameters of the different SFA model specifications. Some results are consistent across all models.

There is no evidence to support the assumption of normally distributed residuals for any of the models, as indicated by the Shapiro-Wilk normality of residuals test. An important reason for this will be relatively fat tails, as indicated by the percentage of severe outliers. Severe outliers comprise about 0.0002% of the normal population, so although the percentages of severe outliers appear to be small, they are inconsistent with normality. Averaging over the two long-sample models in each group, the standard models have the lowest rates of severe outliers (0.09%), followed by Kumb90-JTT-HN-GTC with the second lowest (0.23%), then Kumb90-JTT-HN (0.27%), Kumb90-AJTT-HN (0.32%), BC95-JTT-HN (0.37%) and Kumb90-AJTT-HN (0.41%).

The multicollinearity statistics reported in Table 5.6—the average variance inflation factor (VIF) and the Condition Number—are consistent across some groups of models because they only depend on the variables determining the efficiency frontier, and not the z-variables that are determinants of the inefficiency term. The Cobb-Douglas specifications, across all models and periods, exhibit much lower VIFs and Condition Numbers compared to the Translog models, which have severe multicollinearity. This is because the multiple interaction and squared terms lead to high correlation among regressors. This is an important reason why the Cobb-Douglas models continue to be important, even when the 2^{nd}-order terms in the Translog models are jointly statistically significant. Among the time-varying models, the Kumb90-JTT-

HN-GTC and Kumb90-AJTT-HN-GTC models have lower multicollinearity than the other models. This is because of the reduced collinearity between the measurement of technical change and time-varying inefficiency.

Table 5.6 Statistical Tests – SFA Models Performance

		Multicollinearity		*Normality of Residuals*	*% Severe Outlier*	*TLG Joint Parameters Significance*	*Specification Link Test*
		Avg VIF	*Cond. No*				
Standard ABR24 Models	SFACD - LP	25.0	1084.2	x	0.18		
	SFATLG - LP	714.1	1641.0	x	0.00	✓	
	SFACD - SP	25.9	1619.5	x	0.27		
	SFATLG - SP	NA	NA	NA	NA	NA	
BC95 JTT-HN	SFACD - LP	25.0	1084.2	x	0.64		x
	SFATLG - LP	714.1	1641.0	x	0.09	✓	x
	SFACD - SP	25.9	1619.5	x	0.27		✓
	SFATLG - SP	717.5	2427.5	x	0.27	✓	x
Kumb90 JTT-HN	SFACD - LP	25.0	1084.2	x	0.27		x
	SFATLG - LP	714.1	1641.0	x	0.27	✓	x
	SFACD - SP	25.9	1619.5	x	0.14		x
	SFATLG - SP	717.5	2427.5	x	0.28	✓	x
Kumb90 JTT-HN-GTC	SFACD - LP	16.6	22.2	x	0.27		x
	SFATLG - LP	547.5	367.0	x	0.18	✓	x
	SFACD - SP	20.5	22.6	x	0.41		x
	SFATLG - SP	621.9	366.7	x	0.27	✓	x
Kumb90 AJTT-HN	SFACD - LP	25.0	1084.2	x	0.27		x
	SFATLG - LP	714.1	1641.0	x	0.36	✓	✓
	SFACD - SP	25.9	1619.5	x	0.41		x
	SFATLG - SP	717.5	2427.5	x	0.41	✓	x
Kumb90 AJTT-HN-GTC	SFACD - LP	16.6	22.2	x	0.45		x
	SFATLG - LP	547.5	367.0	x	0.36	✓	✓
	SFACD - SP	20.5	22.6	x	0.55		x
	SFATLG - SP	NA	NA	NA	NA	NA	NA

The joint significance of the Translog parameters is confirmed in all SFATLG models. In every case, the additional parameters introduced in the Translog specification are statistically different from zero, supporting their inclusion and justifying the added model complexity.

Finally, the specification link test indicates potential misspecification in most models. Among the long-period models, only the Translog versions of Kumb90-AJTT-HN and Kumb90-AJTT-HN-GTC satisfy the link test. These models have separate trends for inefficiency for each Australian DNSP and the two overseas jurisdictions. For all other models where the test was conducted, the results suggest potential misspecification, implying that the model may be omitting relevant variables or misrepresenting the functional form. The link test is not available for the standard ABR24 models estimated using Stata's *xtfrontier* command.

Table 5.7 presents key diagnostic statistics for the LSE models, assessing multicollinearity, normality of residuals, the presence of severe outliers, and the joint significance of Translog parameters. Note that the link test for model specification is not available after Stata's *xtpcse* command.

Table 5.7 Statistical Tests – LSE Models Performance

		Multicollinearity		*Normality of Residuals*	*% Severe Outlier*	*TLG Joint Parameters Significance*
		Avg VIF	*Cond. No*			
Standard ABR24 Models	SFACD - LP	15	1,107	x	0.00	
	SFATLG - LP	714	1,641	x	0.00	✓
	SFACD - SP	15	1,650	x	0.00	
	SFATLG - SP	718	2,428	x	0.00	✓
LSE-ADTT	SFACD - LP	311,380	9,061	x	0.27	
	SFATLG - LP	264,608	13,100	x	0.18	✓
	SFACD - SP	710,855	13,637	x	0.00	
	SFATLG - SP	604,380	19,730	x	0.00	✓
LSE-ADTT-GTC	SFACD - LP	9	25	x	0.46	
	SFATLG - LP	390	487	x	0.36	✓
	SFACD - SP	9	27	x	0.00	
	SFATLG - SP	400	497	x	0.00	✓
LSE-AJTT	SFACD - LP	9	1,589	x	0.55	
	SFATLG - LP	401	2,463	x	0.46	✓
	SFACD - SP	9	2,372	x	0.14	
	SFATLG - SP	412	3,623	x	0.00	✓
LSE-AJTT-GTC	SFACD - LP	9	26	x	0.55	
	SFATLG - LP	381	487	x	0.55	✓
	SFACD - SP	9	27	x	0.27	
	SFATLG - SP	392	498	x	0.00	✓

The main observations from Table 5.7 are:

- None of the models pass the normality test for residuals, indicating that the residuals deviate from the normal distribution in all cases.
- Multicollinearity results vary considerably across specifications. The standard ABR24 LSE models have similar VIFs and Condition Numbers as the standard SFA model, except the VIF for the LSECD model is lower at 15. In contrast, the LSE-ADTT models show extremely high multicollinearity, particularly in the LSECD specifications. For instance, the long-period LSECD model reports an average VIF exceeding 311,000 and a condition number over 9,000. The short-period model is even more extreme, with VIFs above 710,000 and a condition number of 13,637. The

remaining time-varying inefficiency LSE models perform much better in this regard, returning low to moderate VIFs of around 9 for Cobb-Douglas, and VIFs of around 400 for TLG. The condition numbers suggest that the LSE-AJTT model also exhibits high multicollinearity. The LSE-ADTT-GTC and LSE-AJTT-GTC have condition numbers of around 25 for the CD versions and around 500 for TLG versions. This suggests that using the GTC terms rather than the time trend helps to resolve the multicollinearity issue.

5.1.6 Plots of Residuals versus Fitted Values

In a well-specified model, a residual versus fitted values plot should display residuals that are randomly scattered around zero, with no clear patterns or structure. This randomness suggests that the model is capturing the main systematic variation in the data, leaving only random noise in the residuals. Ideally, the spread of residuals should be consistent across the range of fitted values (homoskedasticity), and there should be no clustering by subgroups such as jurisdiction. If the residuals show systematic patterns, such as trends, banding, or group-specific clustering, it may indicate issues like omitted variables, model misspecification, or unaccounted heterogeneity in the data.

Figures 5.5 to 5.10 plot the residuals versus fitted values for the six SFA models, including the standard ABR24 SFA model, the BC95-JTT-HN, and the four Kumb90 specifications. Figures 5.11 to 5.15 plot the residuals versus fitted values for the five LSE models, including the standard LSE model, and the four time-varying LSE specifications. In these charts, Australian data points are shown in blue, Ontario in green and New Zealand in red. Because of different scales of DNSPs, Australian data points are mostly found on the right-hand side and New Zealand's on the left-hand side. Ontario has data points across the full scale, but more among the smaller utilities on the left. Some observations from these plots:

- *Figure 5.5, Standard SFA*: There is no clear funnel-shaped pattern (one common indicator of heteroscedasticity) but the vertical spread of residuals does vary depending on the fitted value range and jurisdiction. A notable feature is vertical stripes, especially among the Australian observations but also for NZ and Ontario. These clusters suggest that some DNSPs have fitted values that are relatively stable over time but exhibit noticeable residual variability, potentially indicating that the models are not fully capturing some specific characteristics of certain DNSPs, or that the predicted values for those DNSPs do not change over time, even though the actual data does. Another feature is the apparently high number of outliers among the Australian and New Zealand DNSPs.

  Figure 5.6, BC95-JTT-HN: Vertical striping patterns are no longer present. Instead, we now observe curved, rightward-bending patterns for each group, particularly for NZ and AUS. This curvature could be related to the way this model handles inefficiency, which is allowed to vary over time by jurisdiction. While this helps the model capture long-run differences between countries, likely reducing the strong vertical patterns seen in the standard models, it might also introduce some distortion if the trend does not perfectly match the actual shape of inefficiency across the full range of observations. The spread of residuals around zero is also narrower compared to the standard model, suggesting more stable predictions. Overall, the BC95-JTT-HN model appears to improve the residual distribution by reducing noise and group-specific clustering, but the curved trends suggest further refinements may still be needed.

Figures 5.7 to 5.10, Kumb90 model variants: These models show clear improvements when compared to both the standard model (Figure 5.5) and the BC95-JTT-HN model (Figure 5.6). The residuals are more symmetrically distributed around zero, especially in the Kumb90-JTT-HN model and its GTC variant. The Kumb90 models show wider dispersion than the BC95 models but represent an improvement over the standard models.

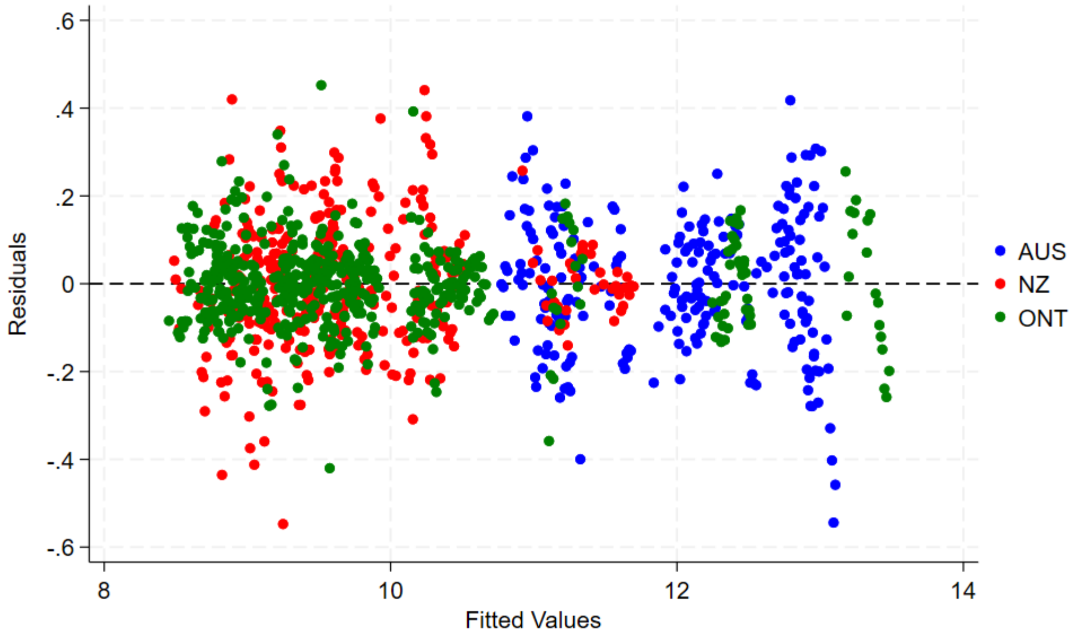
The residual plots indicate that the time-varying models reduced the number of outliers and heteroskedasticity among Australian DNSPs compared to the standard models. Although the BC95 model shows a rightward-bending patterns, it has very few outliers and its residuals are tightly centred on zero.The Kumb90 models, especially Kumb90-JTT-HN and Kumb90-JTT-HN-GTC, display a cleaner residual structure with no visible patterning. These models may offer greater flexibility in capturing inefficiency dynamics across groups, without introducing systematic bias or heteroscedastic patterns.

Figures 5.11 to 5.15 plot the residuals versus fitted values for the five LSE models, including the standard ABR24 LSE model, and the four time-varying inefficiency specifications.

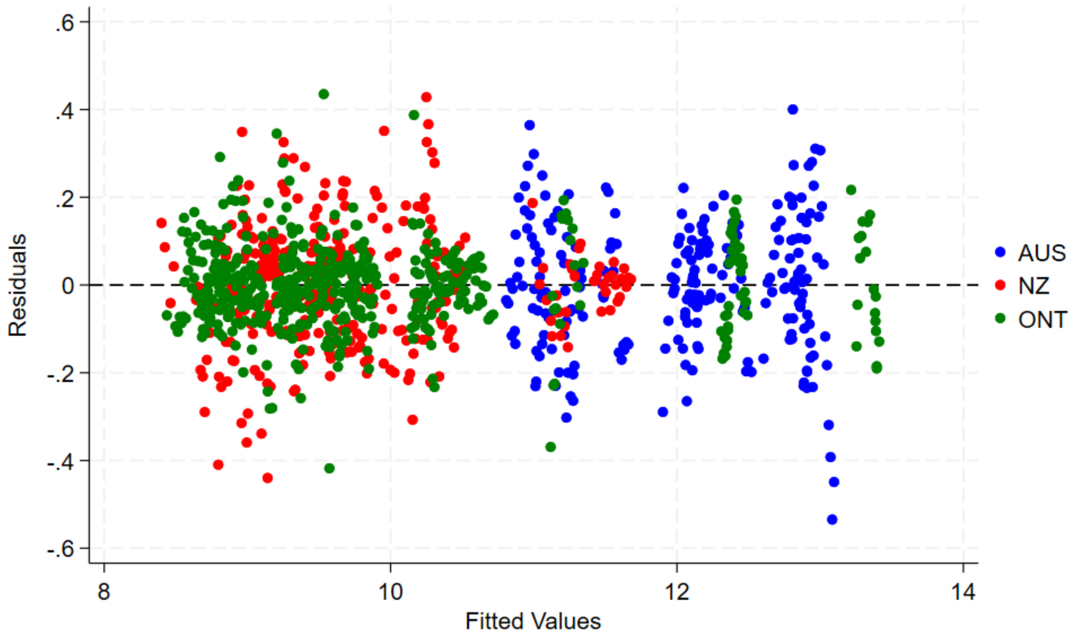
- *Figure 5.11, Standard LSE model*: Vertical striping is present, and outliers are evident. The residuals are mostly concentrated between –0.5 and 1 and appear to be more dispersed than the standard SFA model.
- *Figures 5.12 and 5.15, time-varying LSE models*: These specifications display similar residual dispersion patterns and represent an improvement over the standard models. The number of outliers in the Australian sample is reduced, and the Australian residuals are more tightly centred around zero.

In summary, the time-varying models show improvements in addressing heteroskedasticity in the residuals when compared to the standard models. This suggests that incorporating a time-varying inefficiency component helps mitigate issues related to omitted variables present in the standard specifications.

Figure 5.5 Residuals versus Fitted values: Standard ABR24 SFA Models

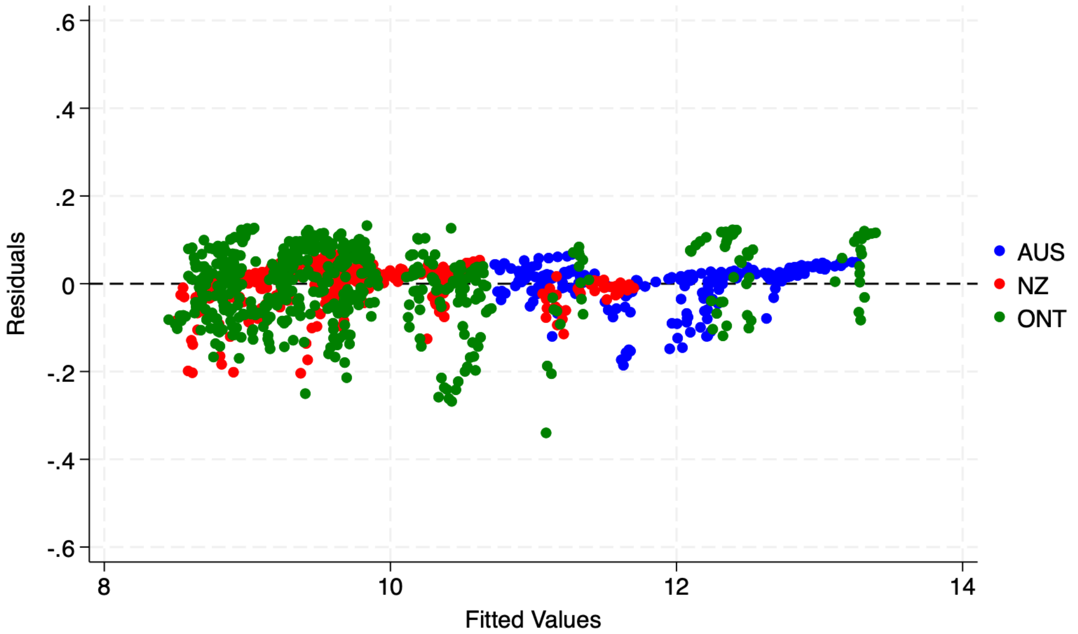


SFACD Long Period

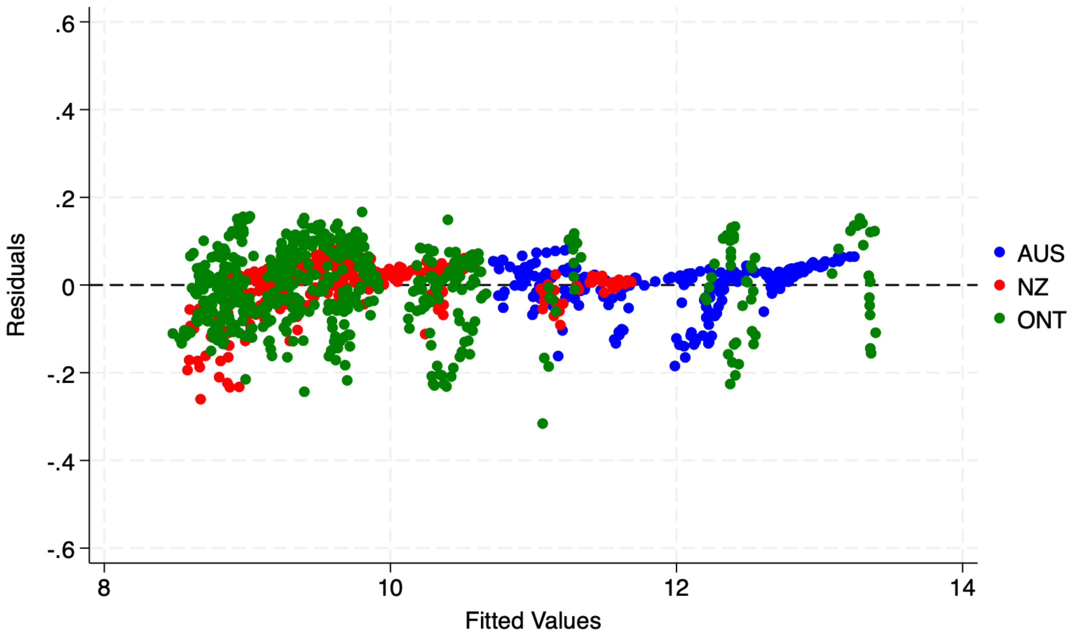


SFATLG Long Period

Figure 5.6 Residuals versus Fitted values: BC95-JTT-HN Models

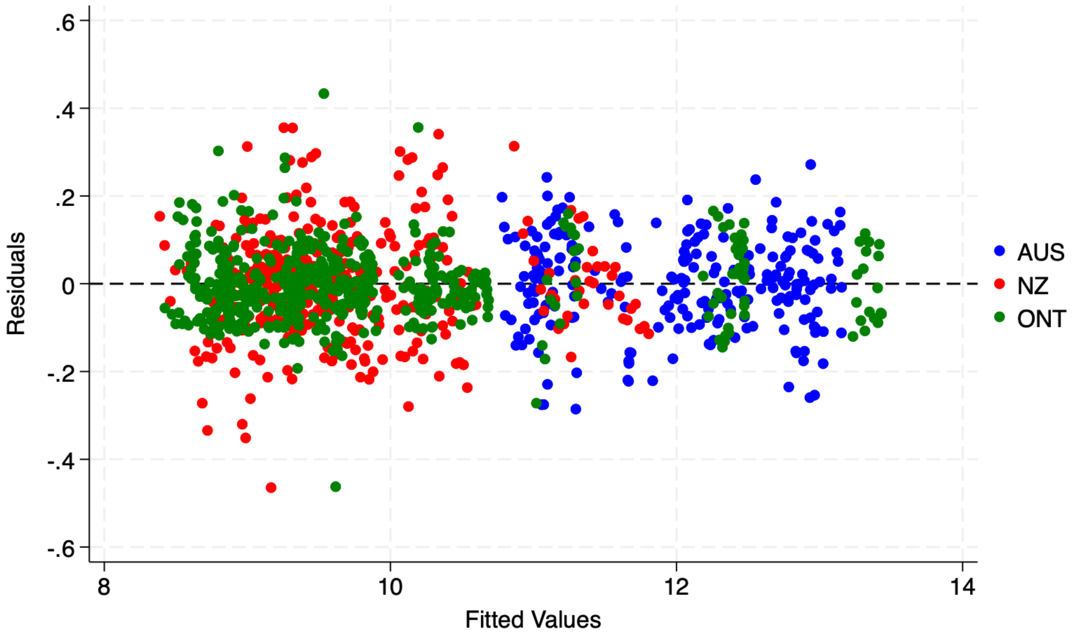


SFACD Long Period

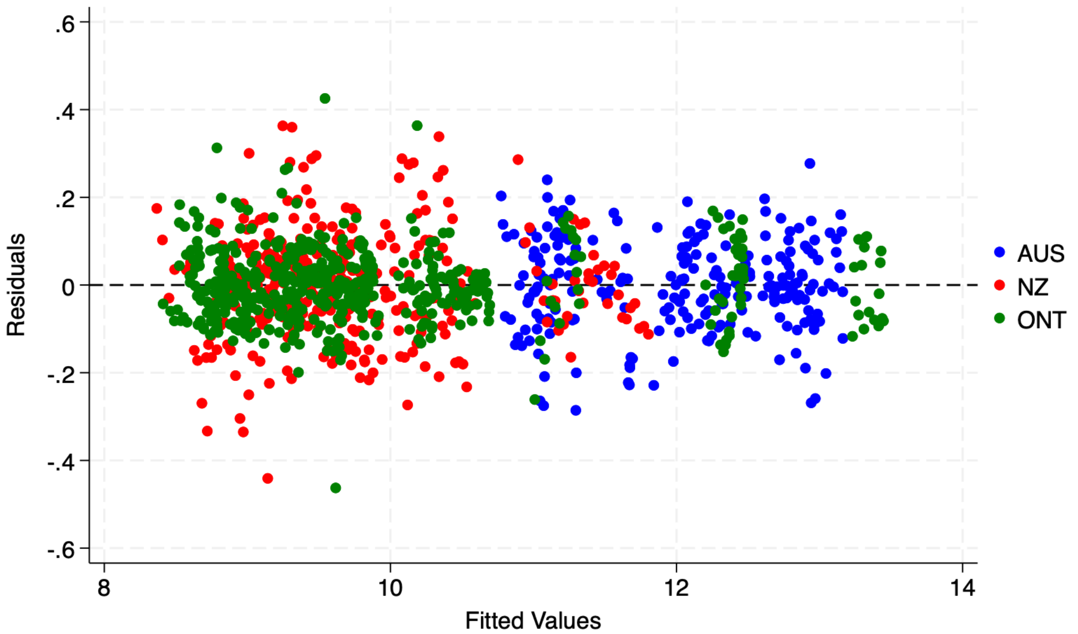


SFATLG Long Period

Figure 5.7 Residuals versus Fitted values: Kumb90-JTT-HN Models

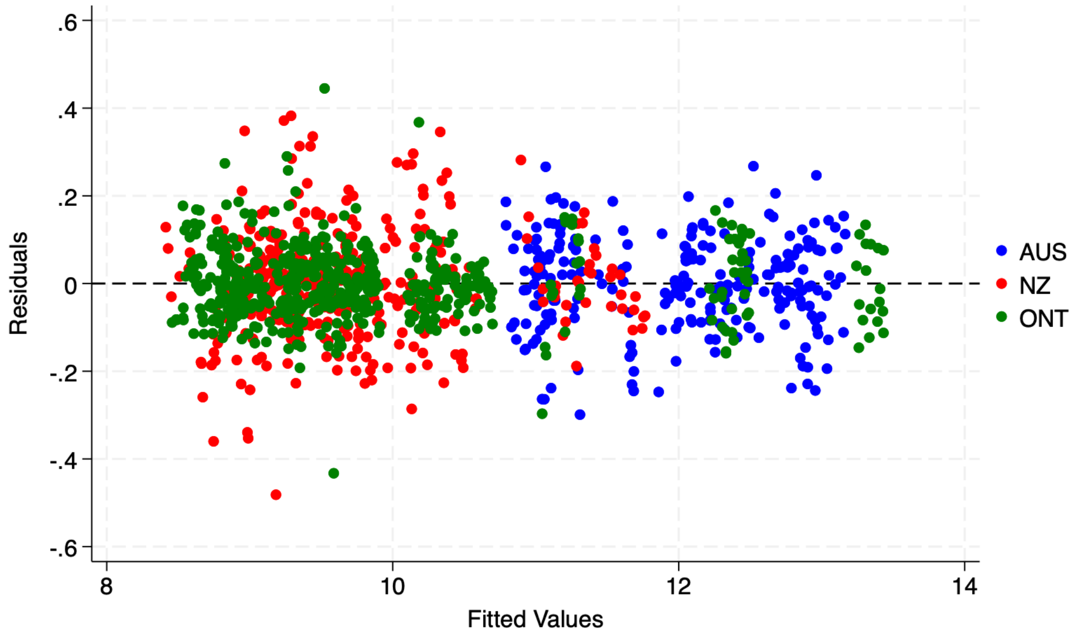


SFACD Long Period

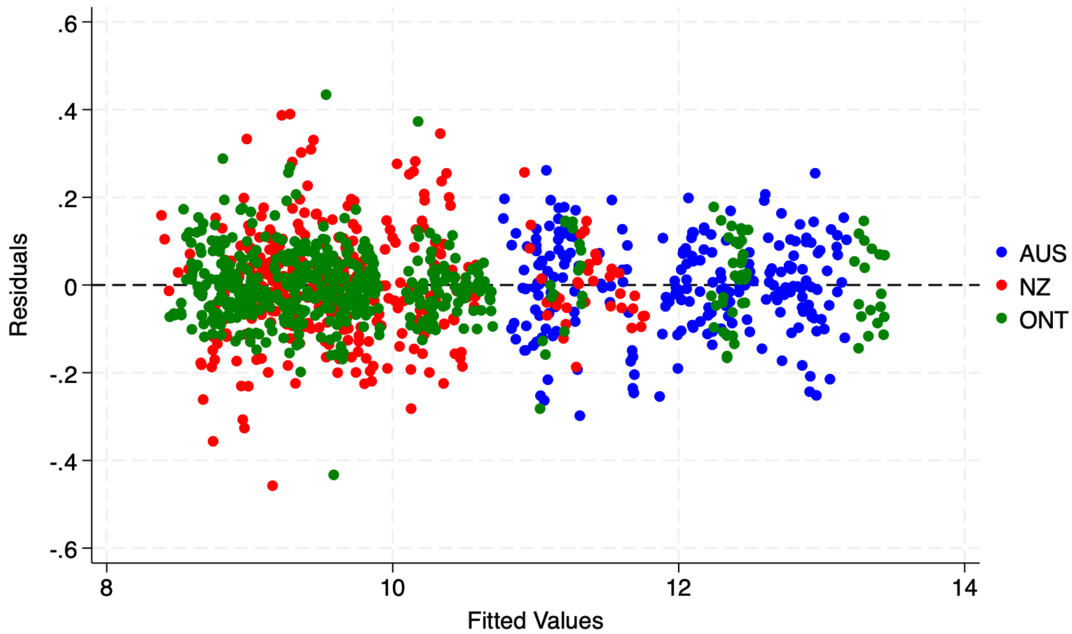


SFATLG Long Period

Figure 5.8 Residuals versus Fitted values: Kumb90-JTT-HN-GTC Models

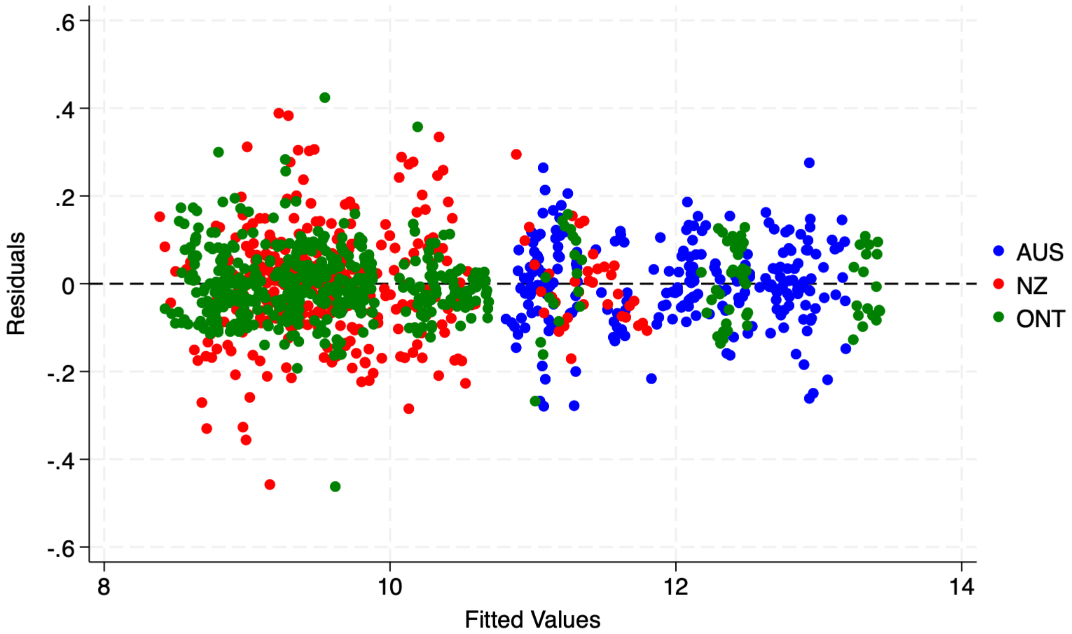


SFACD Long Period

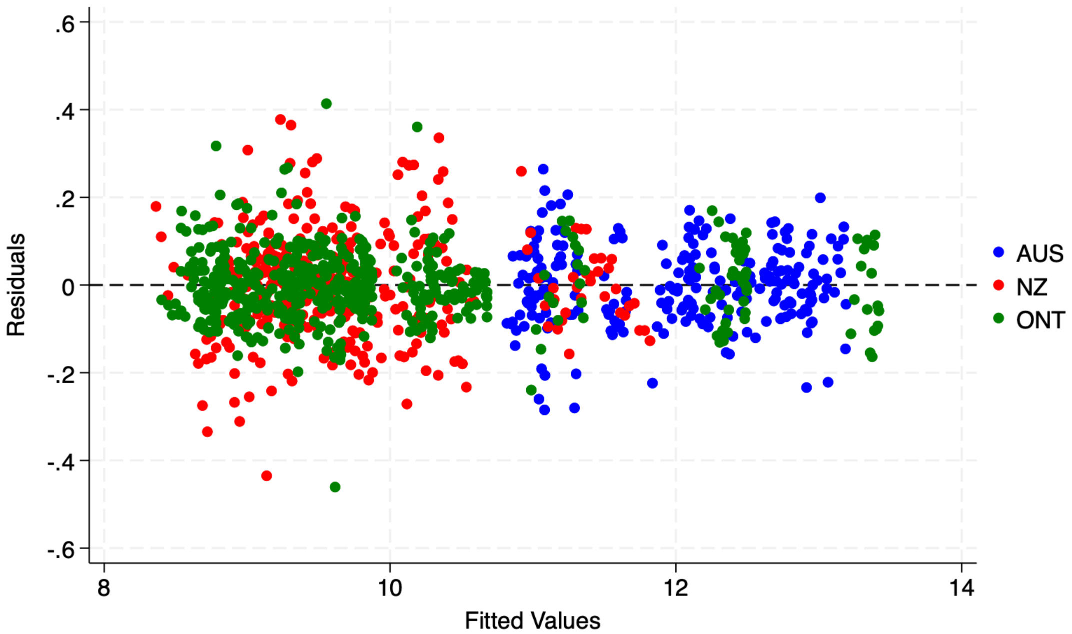


SFATLG Long Period

Figure 5.9 Residuals versus Fitted values: Kumb90-AJTT-HN Models

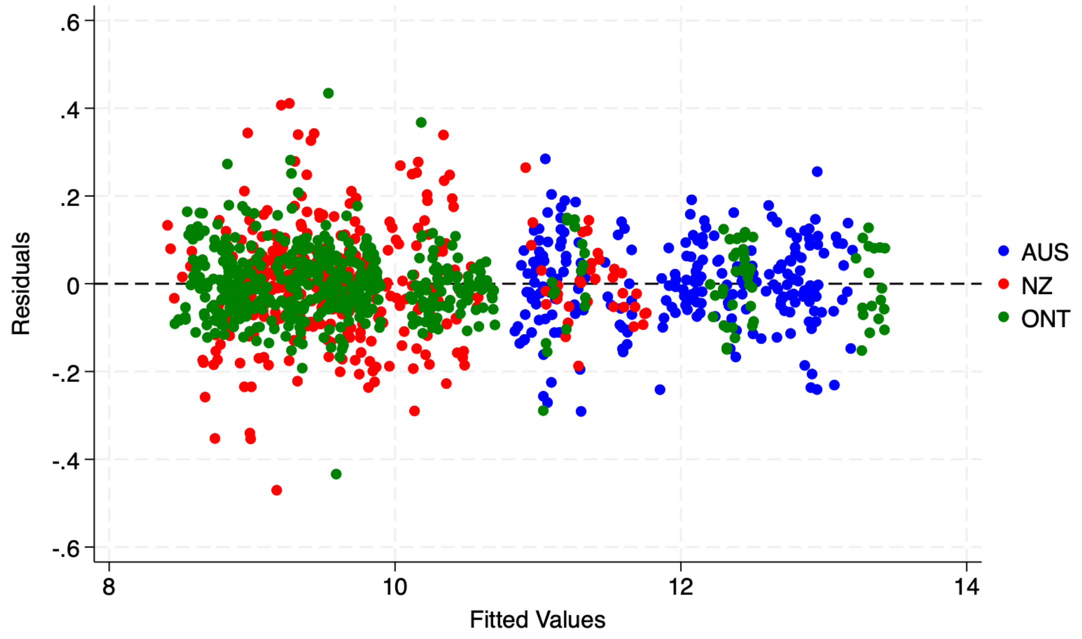


SFACD Long Period

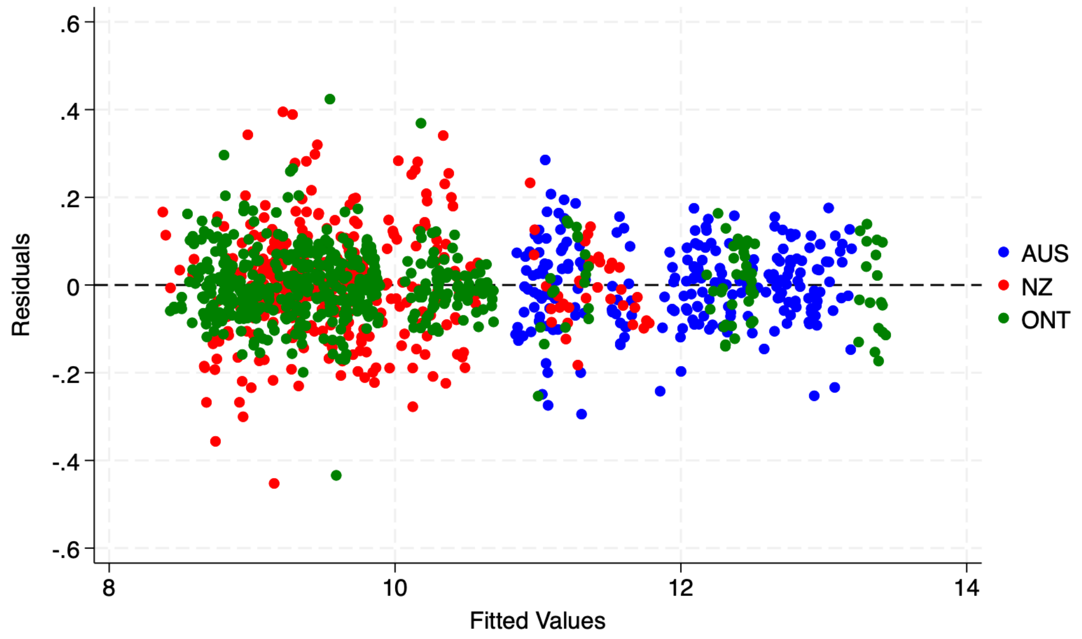


SFATLG Long Period

Figure 5.10 Residuals versus Fitted values: Kumb90-AJTT-HN-GTC Models

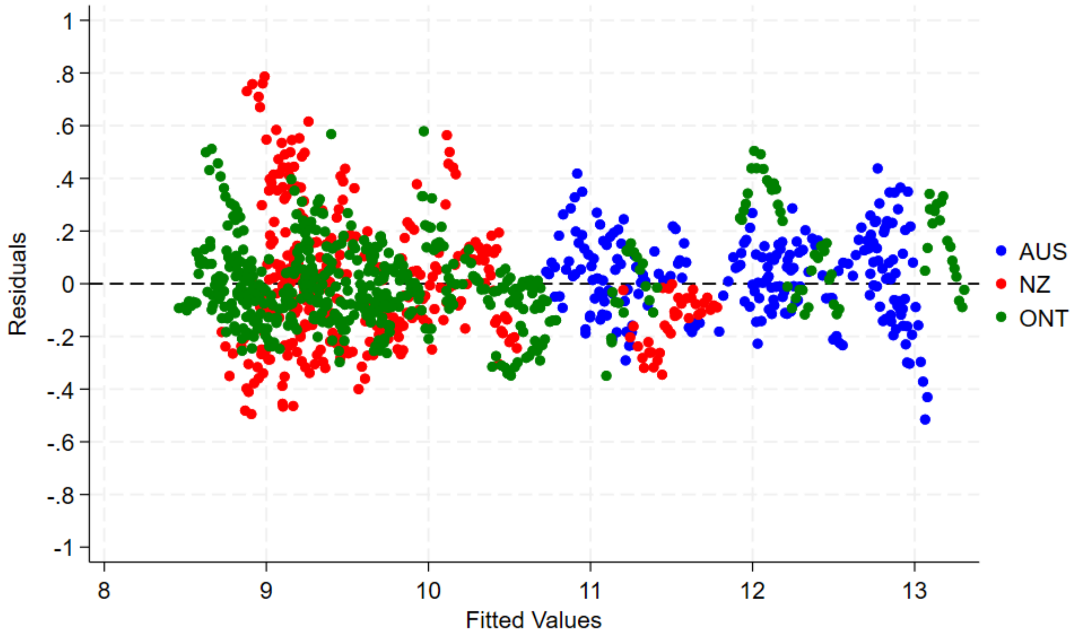


SFACD Long Period

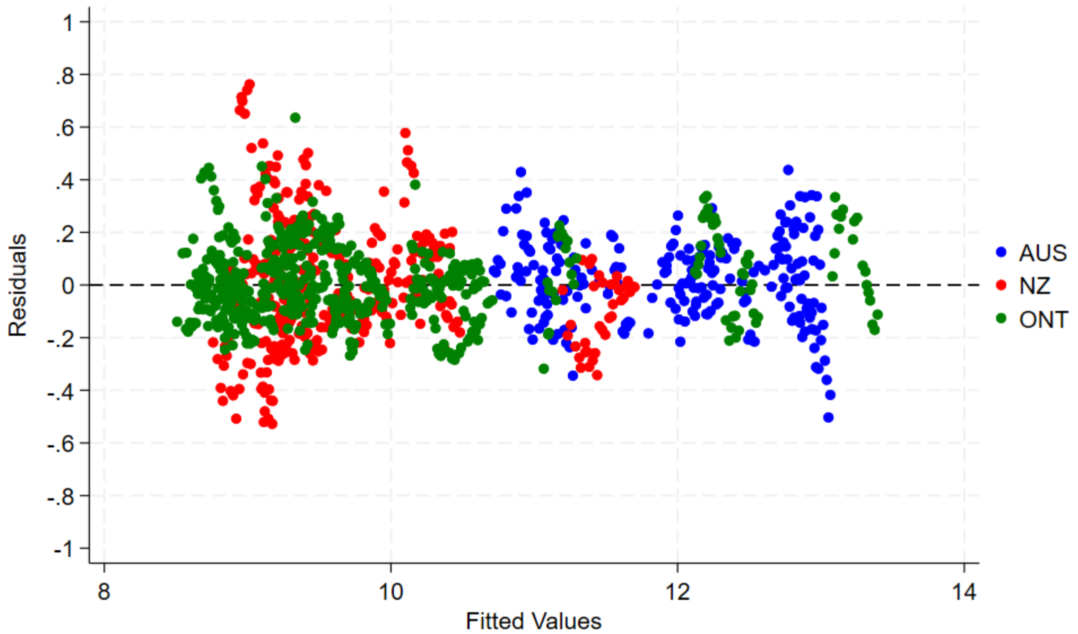


SFATLG Long Period

Figure 5.11 Residuals versus Fitted values: Standard ABR24 LSE Models

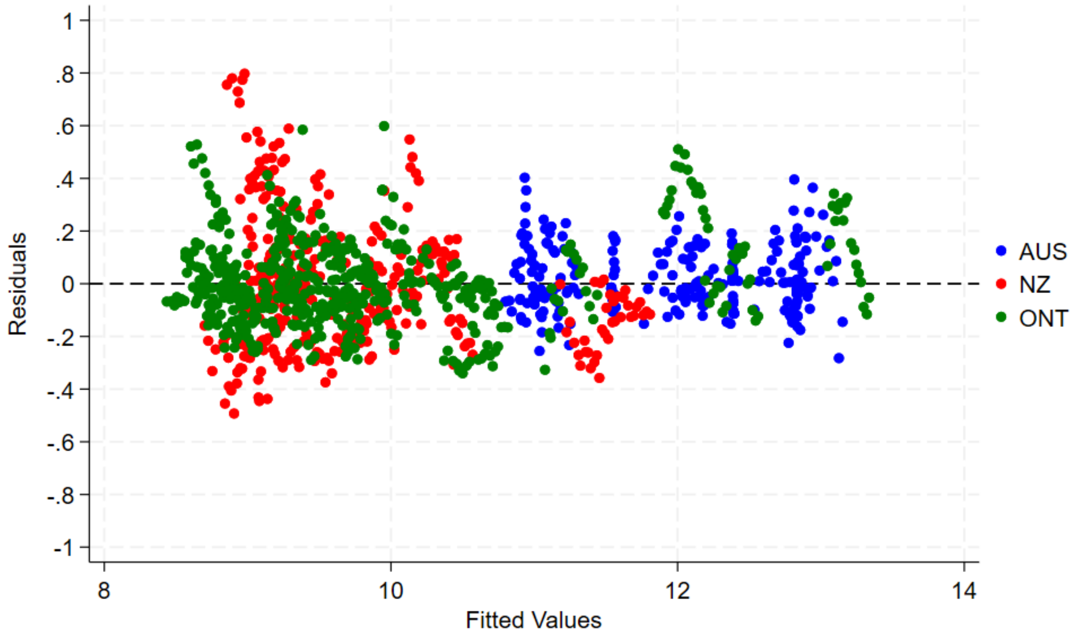


LSECD Long Period

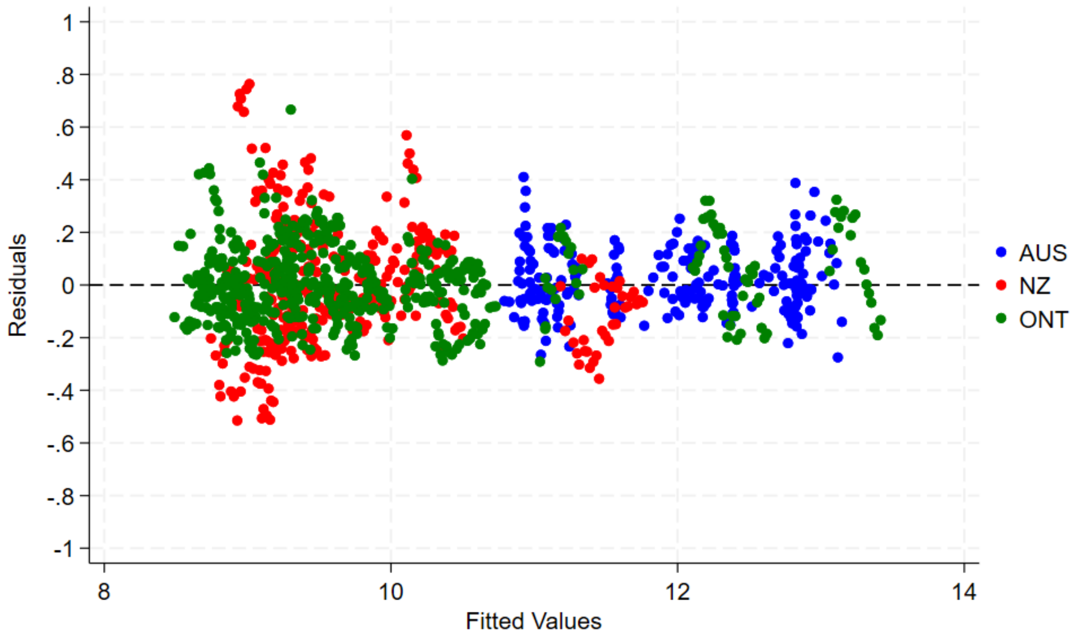


LSETLG Long Period

Figure 5.12 Residuals versus Fitted values: LSE-ADTT Models

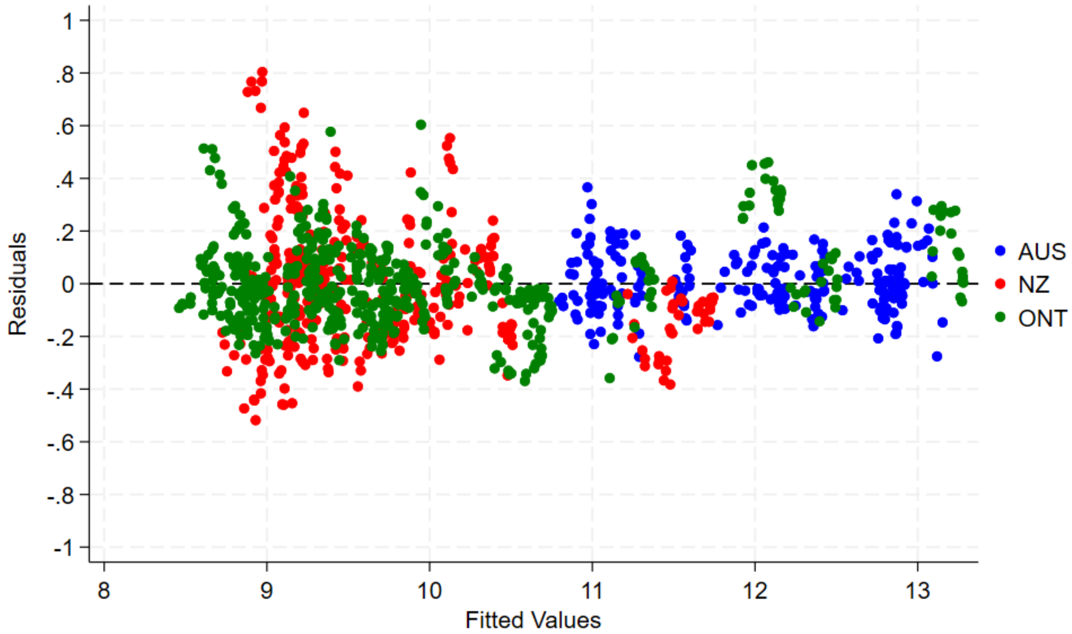


LSECD Long Period

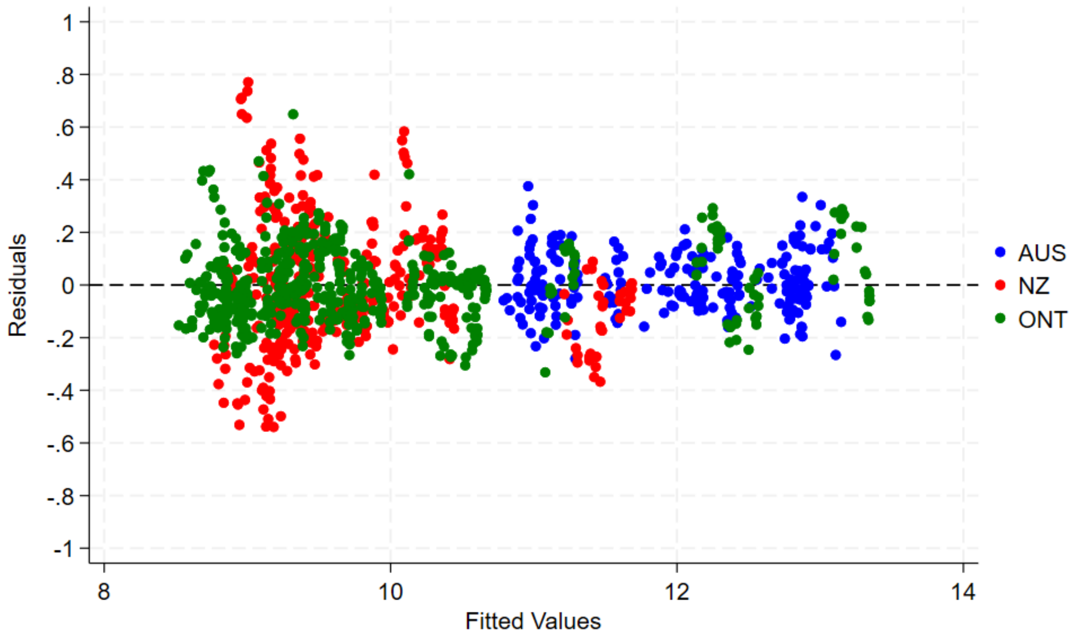


LSETLG Long Period

Figure 5.13 Residuals versus Fitted values: LSE-ADTT-GTC Models

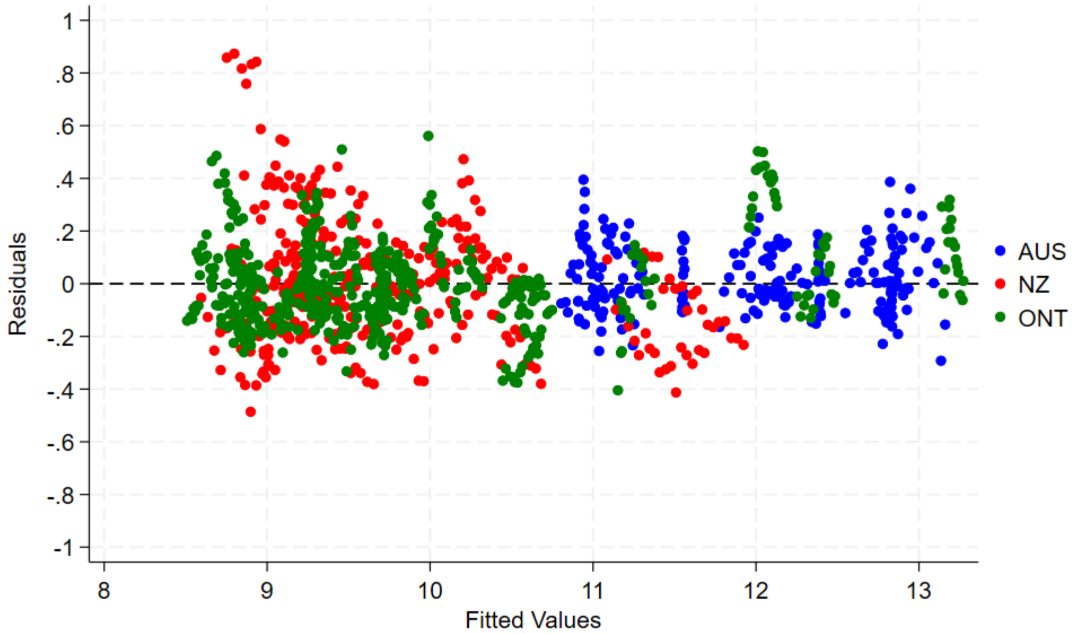


LSECD Long Period

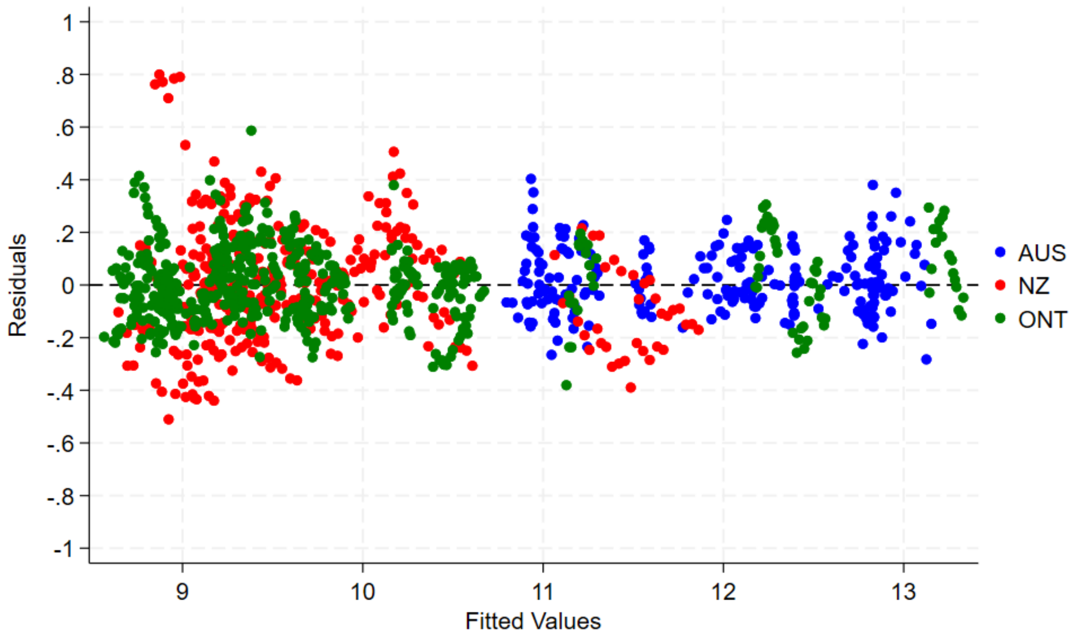


LSETLG Long Period

Figure 5.14 Residuals versus Fitted values: LSE-AJTT Models

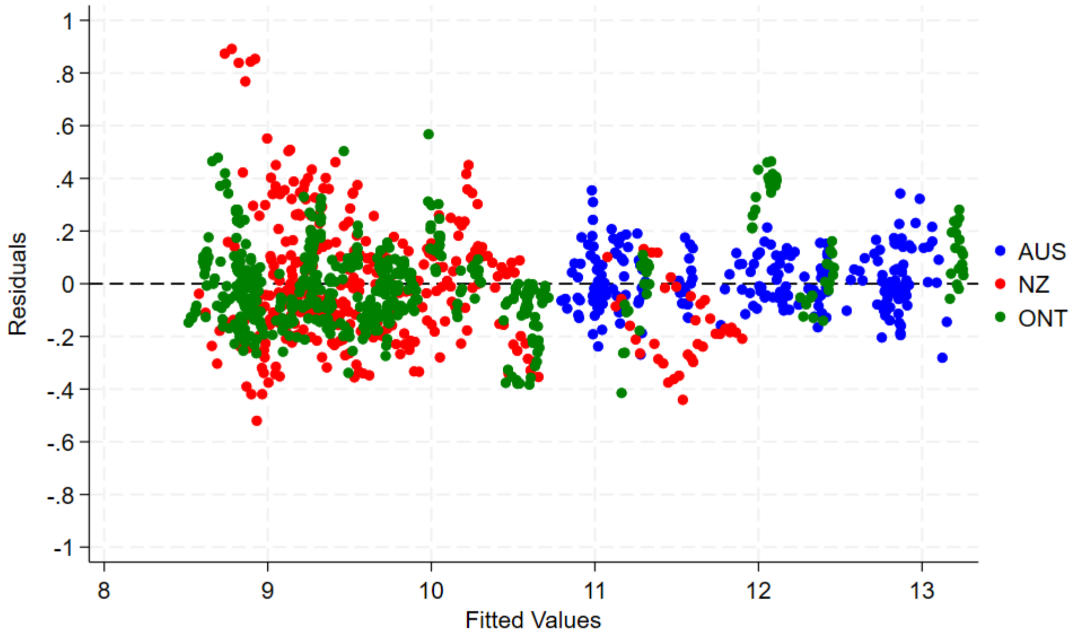


LSECD Long Period

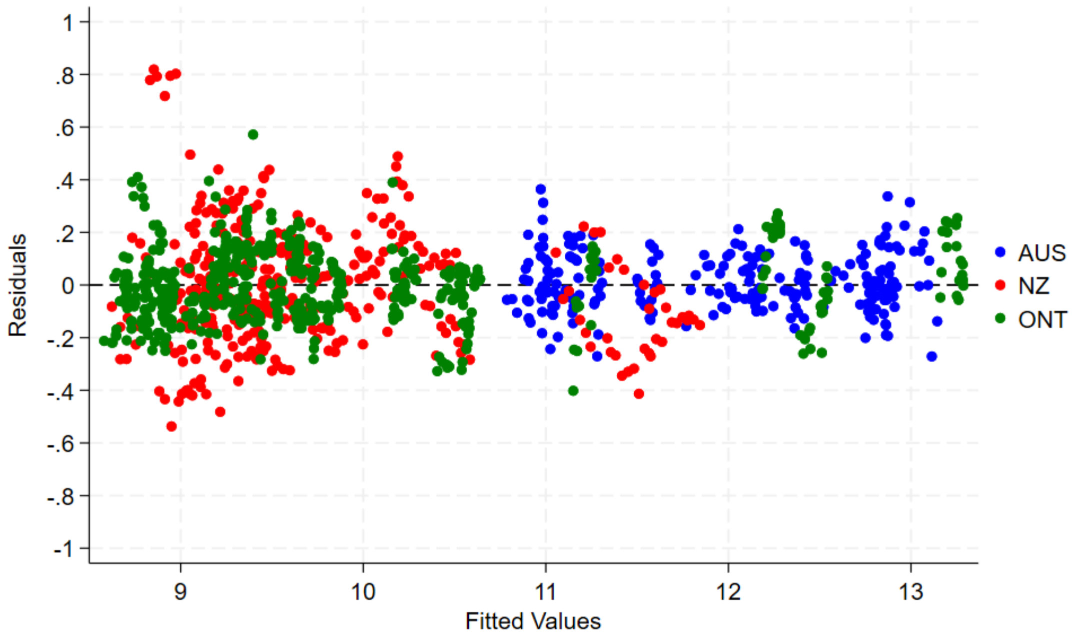


LSETLG Long Period

Figure 5.15 Residuals versus Fitted values: LSE-AJTT-GTC Models



LSECD Long Period



LSETLG Long Period

5.1.7 Consistent Efficiency Scores Performance

To assess the plausibility of the time pattern of efficiency scores estimated with time-varying inefficiency models it is useful to examine their correlation with Opex multilateral partial factor productivity (MPFP) index values, as reported in the AER's Benchmarking Reports (ABRs). Opex MPFP is a non-parametric index method that measures changes in opex partial productivity over time and the relative levels of opex productivity of the firms in the sample. Since the efficiency scores of econometric frontier cost functions and the Opex MPFP index are both designed to measure opex partial productivity levels and trends, we would generally expect a strong positive correlation between the two.

In this analysis, we use the same opex MPFP indexes as those published in ABR24. However, instead of calculating the indexes relative to the most efficient value across all years and DNSPs, as done in the ABRs, we recalculate them relative to the most efficient DNSP in each year. This approach aligns the MPFP measure with the way efficiency scores are defined in the econometric models. We refer to these recalculated indexes as the OPFP-based efficiency measure.

Table 5.8 presents correlation coefficients between Australian efficiency scores and the OPFP-based efficiency for each of the long-period time-varying models examined here, as well as the standard ABR24 long-period models. It also shows the average of the correlation coefficients for the CD and TLG versions of each model. Table 5.9 and 5.10 report the correlation coefficients between the efficiency scores of each Australian DNSP and the OPFP-based efficiency measure for the long-period time-varying models. It also presents the average correlation coefficients for the Cobb–Douglas and Translog versions of each model.

Table 5.8 Correlation between Australian Efficiency Scores and OPFP-based efficiency (Long Period)

	CD	*TLG*	*Avg*
SFA models			
Standard ABR24 Models	0.645	0.691	0.668
BC95- JTT-HN	0.791	0.833	0.812
Kumb90 JTT-HN	0.818	0.818	0.818
Kumb90 JTT-HN-GTC	0.822	0.820	0.821
Kumb90 AJTT-HN	0.774	0.819	0.797
Kumb90 AJTT-HN-GTC	0.764	0.826	0.795
LSE models			
Standard ABR24 Models	0.695	0.750	0.722
LSE-ADTT	0.809	0.863	0.836
LSE-ADTT-GTC	0.806	0.872	0.839
LSE-AJTT	0.786	0.857	0.822
LSE-AJTT-GTC	0.776	0.848	0.812

The Table 5.8 indicates that, relative to the standard models, all time-varying models show stronger correlations with the OPFP-based efficiency scores:

- The efficiency scores of the standard SFA models have reasonably strong correlations (between 0.6 to 0.79) with the MPFP index, particularly in the SFATLG specification. The average of the standard SFA model correlation coefficients is 0.67. This is despite the fact that OPFP-based efficiency varies over time, whereas the standard SFA models have time-invariant efficiency scores for each DNSP. It reflects the strong cross-sectional correlation between the DNSPs' productivity levels estimated by the two methods.
- The BC95-JTT-HN model shows a substantial improvement over the standard SFA model, with an average correlation coefficient of 0.81. This improvement means that the time-patterns of efficiency found by the BC95-JTT-HN models are well correlated with those of the OPFP-based efficiency
- Among the Kumb90 models, the Kumb90-JTT-HN-GTC variant achieves the highest average correlation with OPFP-based efficiency (0.821), followed by the Kumb90-JTT-HN models (0.818). The Kumb90-AJTT-HN and the Kumb90-AJTT-HN-GTC also present very strong correlation (equal or above 0.8).
- The efficiency scores of the standard LSE models show reasonably strong correlations with OPFP-based efficiency, with an average value of 0.72. This is higher than the SFA equivalent and again reflects the strong cross-sectional correlation with the OPFP-based efficiency. All the time-varying LSE alternatives achieve slightly higher correlations, and they are all similar, being between 0.81 and 0.84, representing a large improvement over the standard LSE model.

Tables 5.9 and 5.10 show the correlation coefficients between the OPFP-based efficiency scores and the time-varying model results. These correlations are generally strong for most DNSPs, although a few consistently display weaker alignment.

- In the BC95-JTT-HN model, AND and PCR stand out with weak correlations.
- The Kumb90 specifications are highly consistent with one another, with two DNSPs, TND and SAP, showing weak or negative correlations. SAP records a negative correlation in all Kumb90 models except the Kumb90-AJTT-HN.
- The time-varying LSE models also produce consistent outcomes across specifications, with three DNSPs (CIT, SAP and AND) showing weak correlations. Among these, AND records negative correlations in all four LSE models, while CIT turns negative in the LSE-AJTT and its GTC variant.

Table 5.9 Correlation between Australian DNSPs Efficiency Scores and OPFP-based efficiency SFA Models (Long Period)

	BC95-JTT-HN			*Kumb90-JTT-HN*			*Kumb90-JTT-HN-GTC*		
	CD	*TLG*	*AVG*	*CD*	*TLG*	*AVG*	*CD*	*TLG*	*AVG*
EVO	0.909	0.918	0.914	0.834	0.832	0.833	0.837	0.836	0.836
AGD	0.944	0.948	0.946	0.942	0.941	0.941	0.941	0.941	0.941
CIT	0.509	0.505	0.507	0.751	0.749	0.750	0.749	0.747	0.748
END	0.869	0.883	0.876	0.853	0.851	0.852	0.857	0.855	0.856
ENX	0.480	0.478	0.479	0.758	0.756	0.757	0.768	0.765	0.767
ERG	0.763	0.772	0.767	0.700	0.703	0.701	0.707	0.711	0.709
ESS	0.861	0.876	0.868	0.753	0.758	0.755	0.764	0.768	0.766
JEN	0.693	0.694	0.694	0.793	0.787	0.790	0.783	0.779	0.781
PCR	0.175	0.147	0.161	0.679	0.678	0.679	0.686	0.685	0.686
SAP	0.569	0.582	0.576	-0.257	-0.253	-0.255	-0.259	-0.256	-0.257
AND	0.381	0.371	0.376	0.431	0.429	0.430	0.427	0.426	0.427
TND	0.655	0.631	0.643	0.164	0.160	0.162	0.174	0.170	0.172
UED	0.743	0.779	0.761	0.858	0.858	0.858	0.862	0.862	0.862

Table 5.9 (cont.)

	Kumb90-AJTT-HN			*Kumb90- AJTT-HN-GTC*		
	CD	*TLG*	*AVG*	*CD*	*TLG*	*AVG*
EVO	0.841	0.84	0.84	0.843	0.845	0.844
AGD	0.933	0.935	0.934	0.931	0.932	0.931
CIT	0.751	0.759	0.755	0.747	0.757	0.752
END	0.849	0.841	0.845	0.853	0.849	0.851
ENX	0.749	0.733	0.741	0.759	0.752	0.755
ERG	0.710	0.689	0.700	0.722	0.703	0.713
ESS	0.745	0.907	0.826	0.756	0.910	0.833
JEN	0.809	0.817	0.813	0.801	0.806	0.803
PCR	0.688	0.703	0.695	0.691	0.704	0.697
SAP	0.359	0.363	0.361	0.361	-0.425	-0.032
AND	0.450	0.447	0.449	0.447	0.444	0.445
TND	0.206	0.207	0.206	0.224	0.236	0.230
UED	0.861	0.868	0.864	0.857	0.863	0.860

Table 5.10 Correlation between Australian DNSPs Efficiency Scores and OPFP-based efficiency LSE Models (Long Period)

	LSE-ADTT			LSE-ADTT-GTC		
	CD	*TLG*	*AVG*	*CD*	*TLG*	*AVG*
EVO	0.716	0.746	0.731	0.715	0.743	0.729
AGD	0.863	0.871	0.867	0.864	0.870	0.867
CIT	0.037	0.100	0.068	0.053	0.246	0.149
END	0.861	0.871	0.866	0.862	0.870	0.866
ENX	0.735	0.782	0.759	0.752	0.807	0.779
ERG	0.878	0.876	0.877	0.876	0.880	0.878
ESS	0.824	0.834	0.829	0.828	0.847	0.837
JEN	0.461	0.490	0.476	0.452	0.490	0.471
PCR	0.774	0.808	0.791	0.770	0.807	0.789
SAP	0.299	0.288	0.293	0.301	0.279	0.290
AND	-0.312	-0.295	-0.304	-0.318	-0.230	-0.274
TND	0.528	0.533	0.531	0.509	0.570	0.539
UED	0.818	0.832	0.825	0.818	0.834	0.826

Table 5.10 (cont.)

	LSE-AJTT			*LSE-AJTT-GTC*		
	CD	*TLG*	*AVG*	*CD*	*TLG*	*AVG*
EVO	0.716	0.748	0.732	0.714	0.744	0.729
AGD	0.865	0.873	0.869	0.864	0.873	0.869
CIT	-0.138	0.091	-0.023	-0.232	0.07	-0.081
END	0.862	0.874	0.868	0.862	0.874	0.868
ENX	0.71	0.792	0.751	0.692	0.791	0.741
ERG	0.87	0.872	0.871	0.867	0.872	0.869
ESS	0.816	0.829	0.823	0.817	0.828	0.822
JEN	0.461	0.504	0.482	0.448	0.508	0.478
PCR	0.747	0.813	0.78	0.697	0.813	0.755
SAP	0.305	0.292	0.298	0.305	0.292	0.298
AND	-0.378	-0.325	-0.351	-0.406	-0.322	-0.364
TND	0.454	0.497	0.476	0.414	0.492	0.453
UED	0.818	0.835	0.827	0.821	0.833	0.827

5.1.8 Concluding Comments

Some of the SFA models faced computational or optimisation challenges, especially in the short period and for more complex models. The long-period models considered here successfully produced standard errors despite indications of convergence to a local rather than global maximum. This highlights the need to focus on the long sample period for time-varying inefficiency models, which require large datasets. The Kumb90-JTT-HN was the most stable among them. The LSE models do not require an MLE search algorithm and hence do not have convergence issues.

Among the SFA models, the BC95-JTT-HN and Kumb90-JTT-HN models show the strongest alignment with economic theory, based on the signs and statistical significance (at the 5 per cent level) of output coefficients. The Kumb90-JTT-HN-GTC model also performs well, with all output coefficients significant at the 10 per cent level. However, the Kumb90 models sometimes have a positive sign on the undergrounding variable, particularly in the short-period sample; and the Kumb90-JTT-HN-GTC has a positive undergrounding even in the long period Turning to the LSE time-varying models; all specifications perform equally well regarding the signs and statistical significance of output coefficients.

All tested SFA Translog models improve upon the standard SFA Translog model by reducing the frequency of monotonicity violations, particularly in the long period. Looking at the long period, the Kumb90-JTT-HN and Kumb90-JTT-HN-GTC models have the lowest rate of violations (zero for all jurisdictions), followed by BC95-JTT-HN which has a reasonable low rate of MVs. On the other hand, the Kumb90-AJTT-HN and Kumb90-AJTT-HN-GTC models, although improving to some degree over the standard model, have excessive levels of MVs. All of the four LSE time-varying inefficiency Translog models improve on the standard LSE Translog model in terms of the rate of MVs. These four models perform similarly in terms of the frequency of MVs in the whole sample, but differ in terms of the Australian sub-sample, where the LSE-ADTT-GTC performs the best, followed by the LSE-AJTT and LSE-AJTT-GTC models (which perform equally), then LSE-ADTT.

Model fit is generally comparable across the time-varying inefficiency SFA specifications, as indicated by the Adjusted-Pseudo R^2. Among the Kumb90 models, the Kumb90-JTT-HN specification stands out as the most favourable trade-off between model fit and parsimony. Turning to the LSE models, the LSE-AJTT and LSE-AJTT-GTC models show an improved fit compared to the LSE-ADTT and LSE-ADTT-GTC models and to the standard LSE models.

In terms of residual behaviour, as shown in scatter diagrams, all Kumb90-based models show better homoskedasticity performance than the BC95-JTT-HN model. The time-varying LSE models all show reasonably similar homoskedasticity performance, although the LSE-AJTT-GTC version has perhaps the narrowest scatter.

The correlations of estimated efficiency scores with the Opex MPFP index indicate that the BC95-JTT-HN model has the highest correlation. The Kumb90-JTT-HN model provides some improvement over the standard SFA model. The Kumb90-AJTT-HN and Kumb90-AJTT-HN-GTC do not improve correlation compared to the standard SFA model. The LSE models all produce similar results, a slight improvement over the standard LSE model.

In light of these findings, the models that perform best on-balance are:

(a) BC95-JTT-HN,

(b) Kumb90-JTT-HN.

(c) LSE-AJTT and LSE-AJTT-GTC.

Overall, these models produce results that are consistent with expected output elasticities and the correct signs of coefficients, particularly in the longer-period sample. They also exhibit few or no monotonicity violations, a high level of goodness of fit, satisfactory performance in statistical tests, and stable efficiency scores. The computational stability of these models is assessed in the next section.

These models have different attributes in the way they model changes in efficiency over time. The models in (b) and (c) impose strong restrictions on the shapes of the trends. Furthermore, the model in (b) only accounts for jurisdiction-specific trends in efficiency, whereas the models in (c) allow for DNSP-specific trends. The GTC version of the model in (c) appear to have some advantages in reducing MVs and multicollinearity.

5.2 Sensitivity analysis

In this section, we assess the robustness of the most promising SFA and LSE time-varying models by re-estimating them with alternative datasets to check whether performance remains consistent across different time spans. While the full evaluation was based on the 2006–2023 dataset, here we progressively restrict the period to 2006–2019, 2006–2020, 2006–2021, and 2006–2022, and extend it to 2006–2024. The aim is to determine whether the models remain robust when fewer or more years are included.

Table 5.11 presents some key indicators of model performance for the SFA time-varying models estimated using the long sample period. Note that while we report the pseudo-adjusted R^2 and BIC values, they need to be treated with caution. As previously noted, BIC should only be used for comparing models in the same family (ie, same data, same dependent variable values and same stochastic assumptions). Pseudo-adjusted R^2 can be applied across different model families, provided that the dependent variable values and data are the same. Hence these measures may be unreliable for comparing models using different datasets.

The convergence of these models was not uniformly successful. In some cases, there was the message “cannot compute an improvement -- flat region encountered”. As previously stated, we do not consider that this necessarily means the model solution is unsatisfactory. However, in several other cases the standard errors could not be estimated for the coefficients. This may occur if the MLE routine cannot calculate numerical derivatives at some point. These instances of non-convergence occurred only in Translog models in the following instances:

- BC95-JTT-HN in the 2006-2019 data sample;
- Kumb90-JTT-HN in the 2006-2020 sample;
- Kumb90-JTT-HN-GTC in the 2006-2019 and 2006-2024 sample.

None of the models showed convergence issues when estimated using the 2006–2023 long sample (see Table 5.2).

With respect to coefficient signs and significance, issues were observed only in the Kumb90-AJTT-HN models, specifically concerning the significance of RMD in the Translog specification for the 2006–2019, 2006–2021 and 2006–2022 samples. No other models showed such issues (except in cases where standard errors could not be produced). For the sign of the share of undergrounding variable, the only problems occurred in the Kumb90-JTT-HN-GTC model for the 2006–2020 and 2006–2021 periods. This same model also showed sign issues for the undergrounding variable in the 2006–2023 sample.

On monotonicity violations, no Cobb-Douglas models showed any violations, consistent with the coefficient sign results. In the Translog models, the BC95-JTT-HN specification shows a slight increase in violations as more years are included in the dataset. For the Australian sample, the MV rate was 0 per cent in the 2006–2021 sample, rising to 8.1 per cent in the 2006–2022 sample, 9.8 per cent in 2006–2023 but declined to 8.1 per cent in 2006–2024. The frequency of MVs in the total sample, was 0 per cent in 2006–2019 but increased to between 10 and 15 per cent in the longer sample periods. Nonetheless, these rates remain low and indicate generally robust performance across periods.

For the Kumb90-JTT-HN model, the MV rates tend to decline as more years are included. In the 2006–2019 sample, they were relatively high at 55.5 per cent for the Australian sample and 27.6 per cent for the total sample. In 2006–2021, they fell to 34.1 and 11.0 per cent respectively, and in 2006–2022, to 5.9 and 1.3 per cent. In 2006–2023 and 2006–2024, the MV rate was zero per cent for both samples. The same pattern is observed in the Kumb90-JTT-HN-GTC specification, with MV rates of 56.9 and 30.8 per cent in 2006–2020, falling to 43.8 and 14.1 per cent in 2006–2021, 10.0 and 3.1 per cent in 2006–2022, and zero per cent for both Australian and total sample in 2006–2023.

For the Kumb90-AJTT-HN model, the MV rate is very low in the 2006–2020 sample (zero per cent for Australian DNSPs and 7.7 per cent for the total sample). By contrast, it exceeds 50 per cent for Australian DNSPs in the 2006–2019, 2006–2021, 2006–2022 and 2006–2023 samples. In the 2006–2024 sample, the rate declines to 40.1 per cent; however, this still reflects a relatively high frequency of violations.

For the LSE models, presented in Table 5.12, all specifications consistently show the correct signs for the output and undergrounding share coefficients, as well as statistical significance for the output coefficients. Monotonicity violations appear relatively stable across the sample periods. For the LSE-AJTT model, the MV rate for Australian DNSPs is 7.7 per cent in both the 2006–2019 and 2006–2020 periods, decreasing to 5.8 per cent in 2006–2021, then rising to 10 per cent in 2006–2022 and 11.1 per cent in 2006–2023, but dropping back to 7.7 per cent in 2006-2024. The MV rate for the total sample remains between 18 and 20 per cent across all periods. For the LSE-AJTT-GTC model, the MV rate for the Australian sample is 7.7 per cent in the 2006–2019, 2006–2020, and 2006–2021 periods, increasing to 9.5 per cent in 2006–2022 but dropping to zero in 2006–2023 and increasing to7.7 per cent in 2006–2022. For the total sample, the MV rate remains around 20 per cent in all periods.

Overall, the results suggest that both SFA and LSE time-varying models are generally robust to changes in the sample period. The Kumb90-JTT-HN specifications show the strongest improvements in monotonicity compliance as more years are included, whereas the Kumb90-AJTT-HN specification exhibits excessive violations across most of the samples tested. While the Kumb90-JTT-HN-GTC performs well in terms of monotonicity, it appears to be more sensitive to convergence issues. The BC95 models display modest increases in violations. LSE models maintain consistent coefficient behaviour across all periods, with only moderate variation in monotonicity rates.

Table 5.11 Selected Indicators: SFA models – Long period

		2006-2019		*2006-2020*		*2006-2021*		*2006-2022*		*2006-2024*	
		CD	*TLG*	*CD*	*TLG*	*CD*	*TLG*	*CD*	*TLG*	*CD*	*TLG*
BC95	Convergence*	(b)	x(c)	(b)	✓	✓	✓	✓	(b)	✓	✓
JTT-HN	Output coefs - signs	✓	NA	✓	✓	✓	✓	✓	✓	✓	✓
	- significance**	✓	NA	✓	✓	✓	✓	✓	✓	✓	✓
	Undergrounding sign	✓	NA	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.	0.0%	NA	0.0%	0.0%	0.0%	0.0%	0.0%	8.1%	0.0%	8.1%
	MVs % - Total	0.0%	NA	0.0%	13.0%	0.0%	12.7%	0.0%	14.3%	0.0%	14.4%
	Pseudo-R^2	0.997	NA	0.997	0.977	0.997	0.997	0.997	0.997	0.998	0.997
	BIC	–328.6	NA	–357.4	–416.1	–377.1	–440.8	–398.6	–465.3	–496.0	–545.7
Kumb90	Convergence*	✓	✓	✓	x(c)	✓	✓	(b)	✓	✓	✓
JTT-HN	Output coefs - signs	✓	✓	✓	NA	✓	✓	✓	✓	✓	✓
	- significance**	✓	✓	✓	NA	✓	✓	✓	✓	✓	✓
	Undergrounding sign	✓	✓	✓	NA	✓	✓	✓	✓	✓	✓
	MVs % - Aust.	0.0%	55.5%	0.0%	NA	0.0%	34.1%	0.0%	5.9%	0.0%	0.0%
	MVs % - Total	0.0%	27.6%	0.0%	NA	0.0%	11.0%	0.0%	1.3%	0.0%	0.0%
	Pseudo-R^2	0.994	0.995	0.994	NA	0.994	0.994	0.994	0.994	0.994	0.994
	BIC	–1140.0	–1126.0	–1214.9	NA	–1313.2	–1289.0	–1411.6	–1383.1	–1578.6	–1558.3
Kumb90	Convergence*	✓	x(c)	✓	✓	✓	✓	✓	✓	✓	x(c)
JTT-HN-GTC	Output coefs - signs	✓	NA	✓	✓	✓	✓	✓	✓	✓	NA
	- significance**	✓	NA	✓	✓	✓	✓	✓	✓	✓	NA
	Undergrounding sign	✓	NA	✓	x	✓	x	✓	✓	✓	NA
	MVs % - Aust.	0.0%	NA	0.0%	56.9%	0.0%	43.8%	0.0%	10.0%	0.0%	NA
	MVs % - Total	0.0%	NA	0.0%	30.8%	0.0%	14.1%	0.0%	3.1%	0.0%	NA
	Pseudo-R^2	0.994	NA	0.994	0.994	0.994	0.994	0.994	0.994	0.994	NA
	BIC	–1102.0	NA	–1173.3	–1157.5	–1266.2	–1244.4	–1361.9	–1335.6	–1518.3	NA

Table 5.11 (cont.)

		2006-2019		*2006-2020*		*2006-2021*		*2006-2022*		*2006-2024*	
		CD	*TLG*	*CD*	*TLG*	*CD*	*TLG*	*CD*	*TLG*	*CD*	*TLG*
Kumb90	Convergence*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
AJTT-HN	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance**	✓	X	✓	✓	✓	X	✓	X	✓	✓
	Undergrounding sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.	0.0%	56.6%	0.0%	0.0%	0.0%	73.1%	0.0%	63.3%	0.0%	40.1%
	MVs % - Total	0.0%	49.6%	0.0%	7.7%	0.0%	50.0%	0.0%	37.7%	0.0%	20.7%
	Pseudo-R^2	0.995	0.995	0.994	0.994	0.994	0.995	0.994	0.995	0.994	0.994
	BIC	–1102.9	–1081.4	–1163.6	–1143.3	–1163.7	–1261.0	–1371.3	–1359.1	–1556.4	–1546.0

Table 5.12 Selected Indicators: LSE models – Long period

		2006-2019		2006-2020		2006-2021		2006-2022		2006-2024	
		CD	*TLG*	*CD*	*TLG*	*CD*	*TLG*	*CD*	*TLG*	*CD*	*TLG*
LSE-AJTT	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Undergrounding sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.	0.0%	7.7%	0.0%	7.7%	0.0%	5.8%	0.0%	10.0%	0.0%	7.7%
	MVs % - Total	0.0%	18.7%	0.0%	18.5%	0.0%	18.4%	0.0%	19.3%	0.0%	19.1%
	Pseudo-R^2	0.982	0.984	0.981	0.984	0.981	0.984	0.981	0.984	0.981	0.984
LSE-AJTT-GTC	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Undergrounding sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.	0.0%	7.7%	0.0%	7.7%	0.0%	7.7%	0.0%	9.5%	0.0%	7.7%
	MVs % - Total	0.0%	19.8%	0.0%	19.8%	0.0%	19.9%	0.0%	20.1%	0.0%	21.4%
	Pseudo-R^2	0.981	0.984	0.981	0.984	0.981	0.984	0.981	0.984	0.981	0.984

Notes: * (a) "convergence not achieved"; (b) "cannot compute an improvement -- flat region encountered"; (c) Failed to estimate some or all parameter standard errors.
** ✓ indicates significance at 0.05 level; # indicates significance at 0.10 level.

6 Applications of Time-Varying Inefficiency Models

This chapter explains how the results of the time-varying inefficiency models could be used to produce estimates of the trends in Opex PFP which can be decomposed into the separate effects of changes in cost efficiency, technical change, and changes in scale efficiency. This decomposition is particularly useful to the regulator, for whom changes in cost efficiency as well as the levels of cost efficiency are key metrics of the effectiveness of regulation.

This section also outlines our initial thinking on how the base year efficiency of DNSPs could be calculated when the econometric model has time-varying inefficiency. To date, the benchmarking roll forward model (BRFM) has been used to estimate the base year efficient opex from the period-average efficiency. Here the implications of time-varying inefficiency are considered for this process.

6.1 Reliability of Time Varying Efficiency Scores

Before turning to the applications of time-varying inefficiency models, the first step is to examine the reliability of the efficiency scores produced by the time-varying inefficiency models. This includes the estimated time profile of efficiency score estimates for each DNSP, and especially those estimated for the last year of the sample period. This section first looks at the confidence intervals for time-varying efficiency score estimates over the sample period. Then it compares the time profiles for efficiency scores obtained using different models and methods.

6.1.1 Confidence intervals

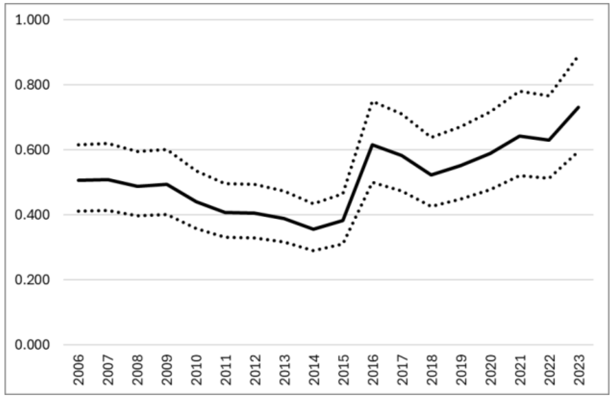
In some applications of regression models, predicted values may be less precise at the extremes of the data because the standard error of a prediction increases as the observation moves further from the mean of the explanatory variables. With efficiency scores varying over time, we consider whether the confidence intervals of efficiency scores for the first and last years of the sample are wider than in middle years. To assess this, we examine the confidence intervals of the proposed models. Two representative models are used for this purpose, the BC95-JTT-HN model,[50] and the LSE-AJTT-GTC model, and we focus only on the long sample TLG models.

Figures 6.1 and 6.2 present the efficiency scores and associated 95% confidence intervals for the 13 DNSPs and for the average, under the BC95-JTT-HN and LSE-AJTT-GTC long-period TLG models, respectively. In both cases, the solid line shows the estimated efficiency scores, while the dotted lines represent the upper and lower bounds of the confidence intervals.

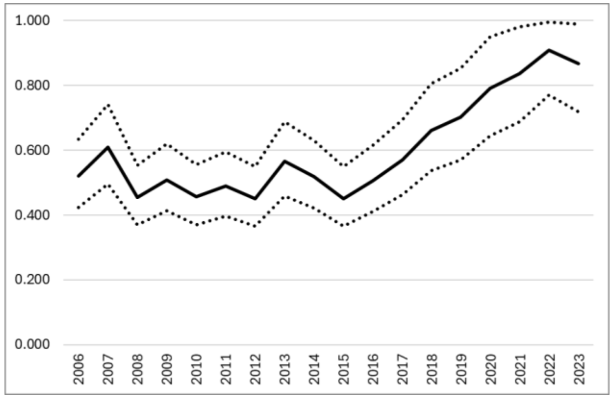
[50] Confidence intervals for the conditional expectation $E[\exp(-u_i) \mid \varepsilon_i]$ are not available *sfpanel, kumb90* models.

Figure 6.1 Efficiencies Score Confidence Interval: BC95-JTT-HN, TLG, Long Period

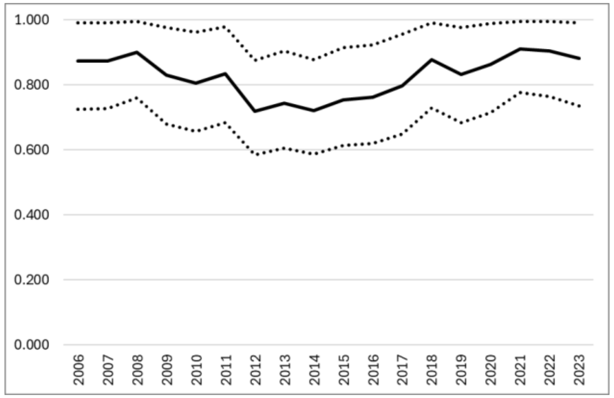
EVO



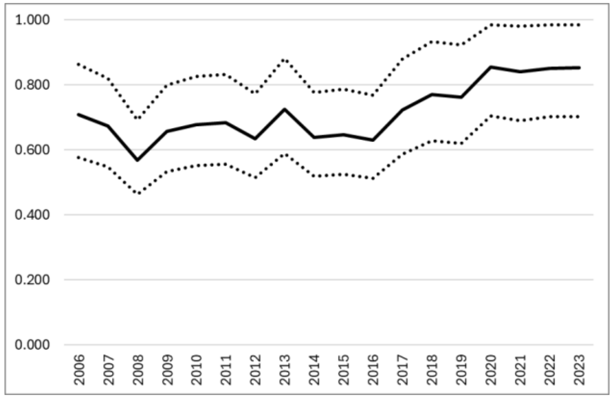
AGD



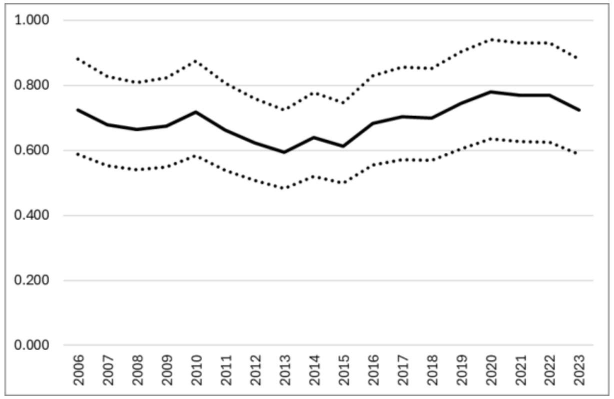
CIT



END



ENX



ERG

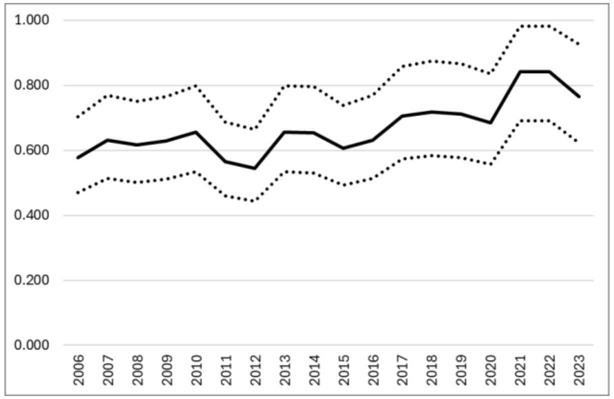
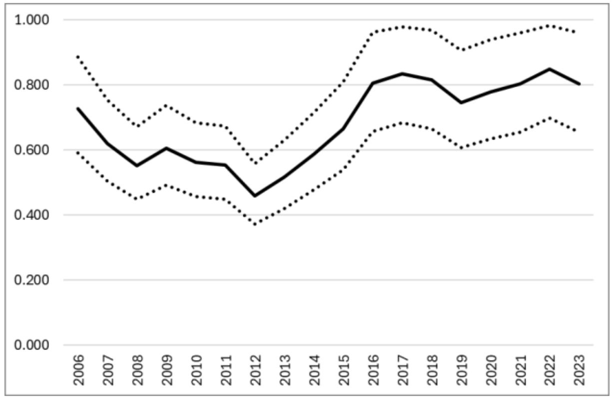
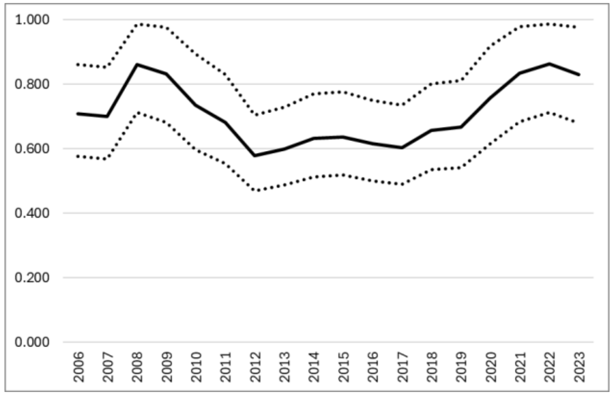


Figure 6.1 (cont.)

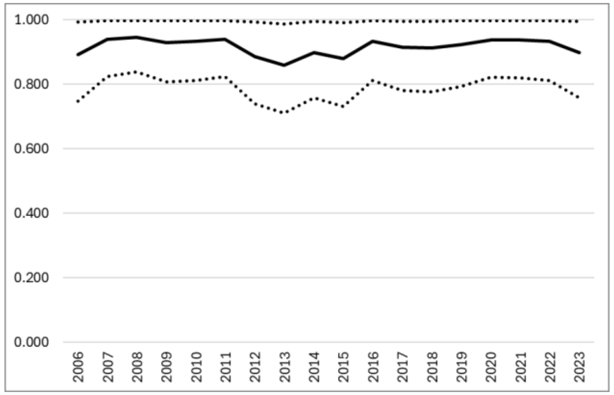
ESS



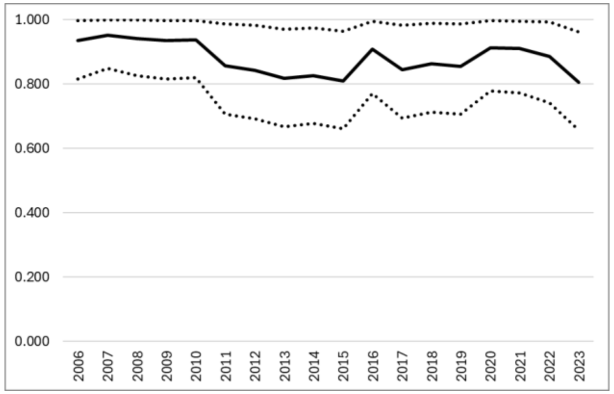
JEN



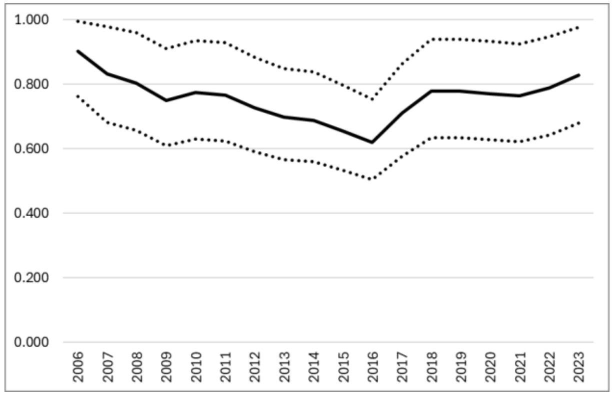
PCR



SAP



AND



TND

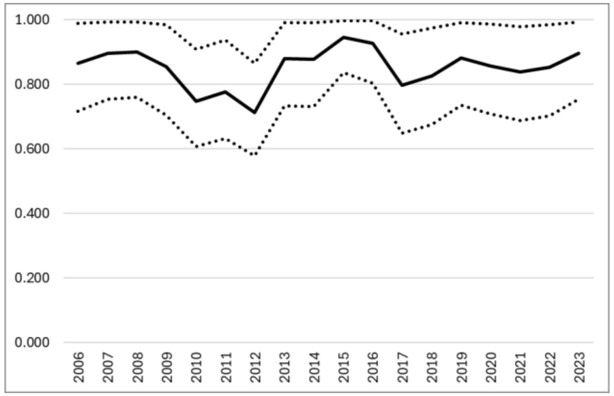
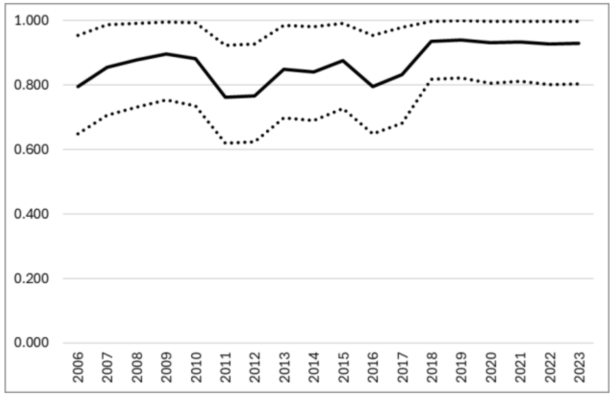


Figure 6.1 (cont.)

UED



Australia - Average

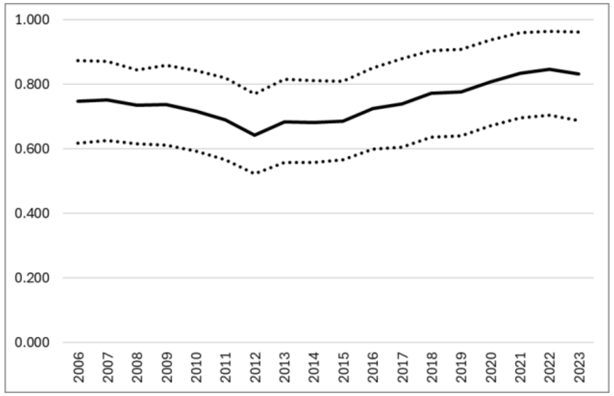
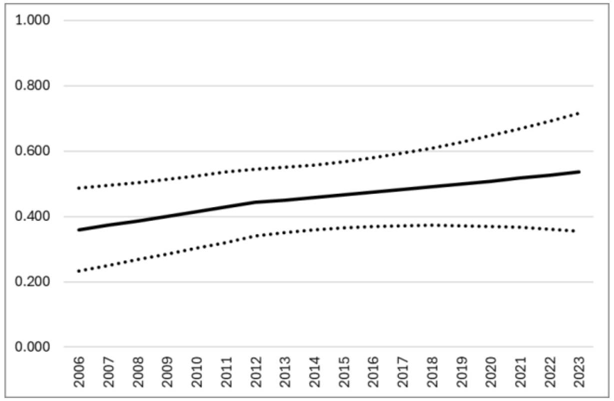
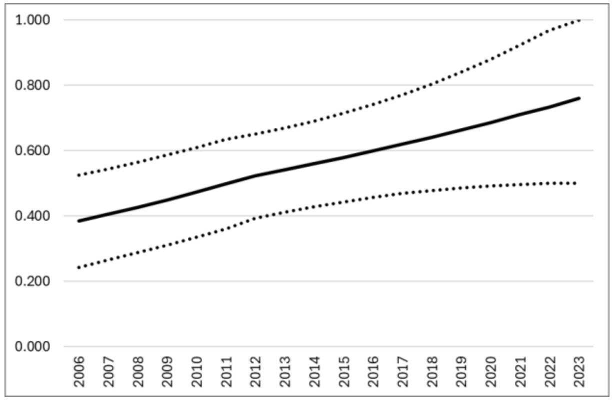


Figure 6.2 Efficiencies Score Confidence Interval: LSE-AJTT-GTC, TLG, Long Period

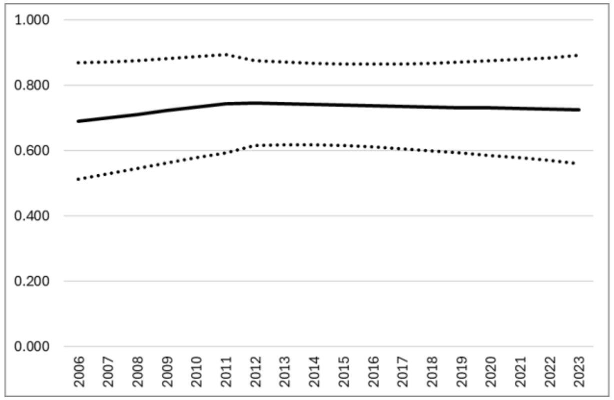
EVO



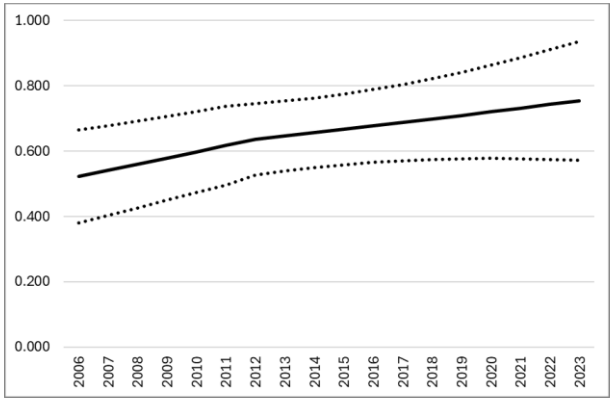
AGD



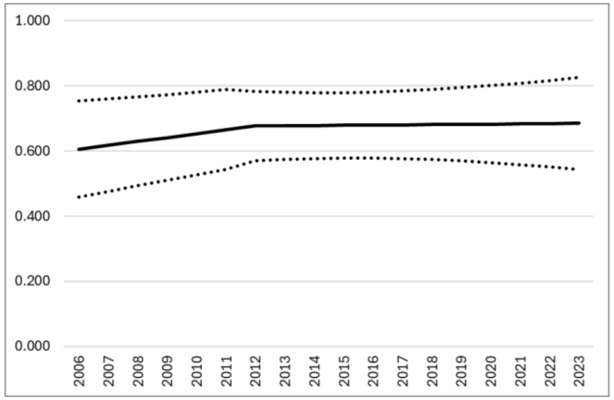
CIT



END



ENX



ERG

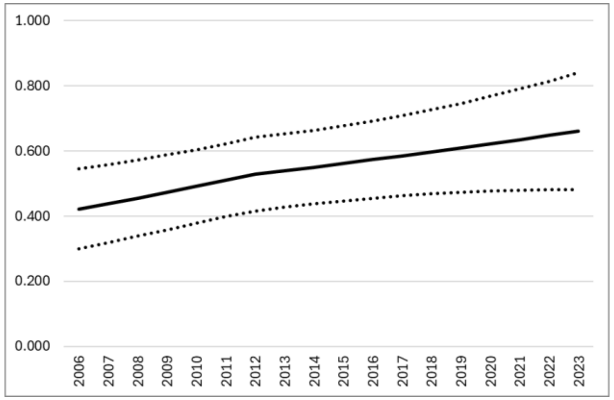
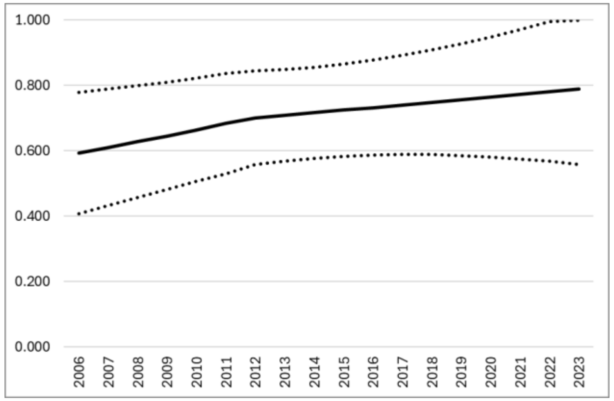
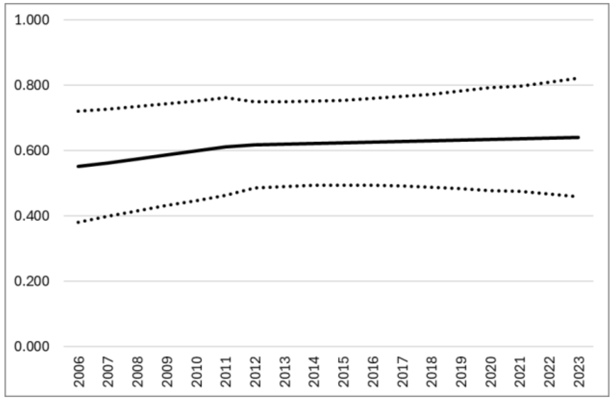


Figure 6.2 (cont.)

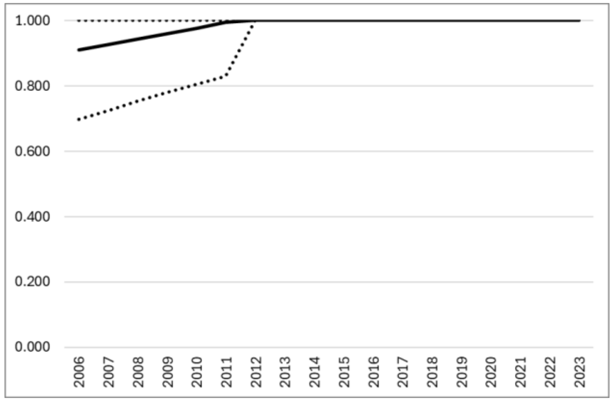
ESS



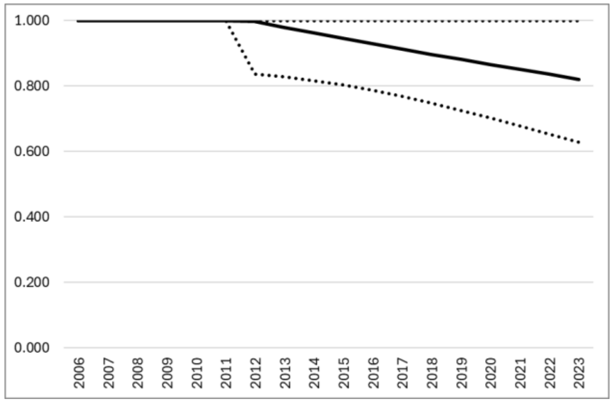
JEN



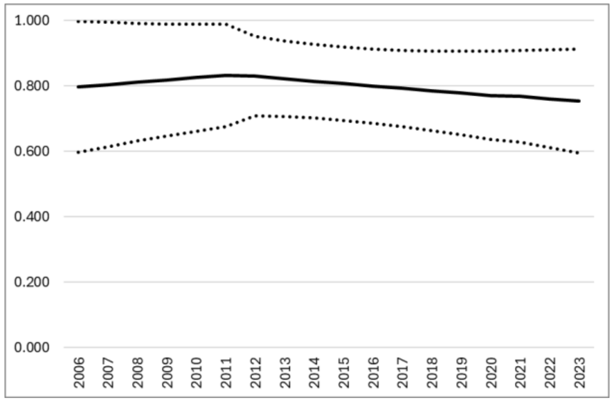
PCR



SAP



AND



TND

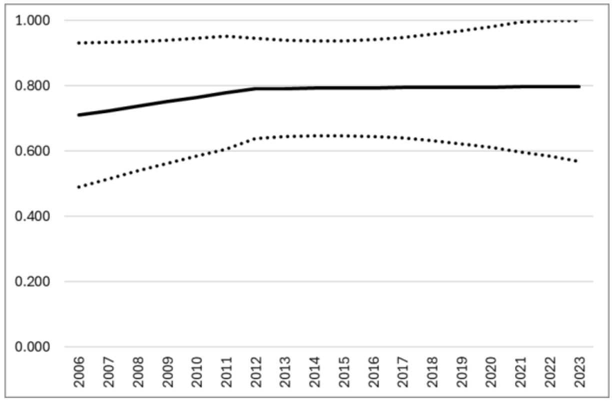
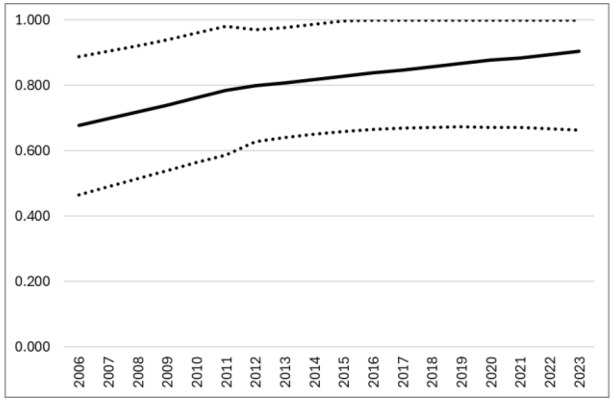
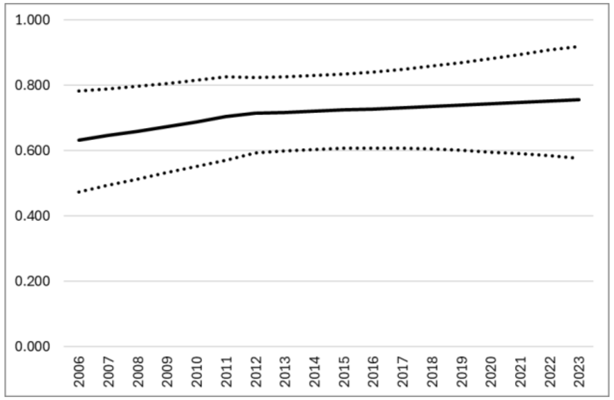


Figure 6.2 (cont.)

UED	Australia - Average
	

The results show that:

- In the BC95-JTT-HN model (Figure 6.1), the width of the confidence intervals varies over time but does not consistently widen at the sample endpoints. The average gap between the efficiency score and the lower bound is –17.3% in 2006, –17.5% in 2015, and –17.2% in 2023. The corresponding average gaps to the upper bound were 16.7%, 18.1%, and 15.8%, respectively. These results indicate that there is no consistent evidence of reduced precision at the sample endpoints.
- In the LSE-AJTT-GTC model (Figure 6.2), the behaviour of the confidence intervals differs. The average gap between the efficiency score and the lower bound was –25.0% in 2006, –16.2% in 2015, and –23.8% in 2023. The corresponding average gaps to the upper bound were 23.6%, 15.0%, and 21.5%, respectively. These results indicate that the intervals widened substantially towards the end, suggesting lower precision in the later years.

Overall, the analysis indicates that for the selected DNSPs, the SFA model (BC95-JTT-HN) shows no evidence of reduced precision in efficiency score estimates at the sample endpoints, with confidence interval widths remaining broadly consistent over time. The LSE model (LSE-AJTT-GTC) exhibits some widening of intervals towards the end of the period, suggesting a decline in precision in later years. However, given that efficiency scores are averaged across models in the final assessment, any endpoint imprecision is unlikely to materially affect the reliability of the resulting efficiency estimates. As such, while the patterns observed warrant awareness, they do not present a significant concern for the application of the last year of the time-varying efficiency score for measuring the efficient opex.

6.1.2 Comparison of Efficiency Time Profiles

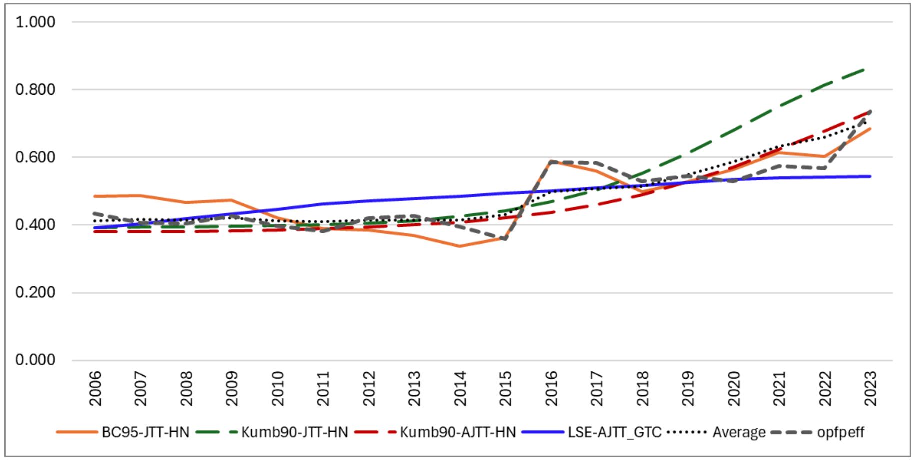
A key question is whether the time-varying models produce efficiency score patterns that adequately reflect the true efficiency trends of individual DNSPs. To assess this, we compare the estimated efficiency scores from the BC95-JTT-HN, Kumb90-JTT-HN, Kumb90-AJTT-HN and LSE-AJTT-GTC models over the long period, for each DNSP and for the industry average. We selected the most promising models, excluding LSE-AJTT as it produced results similar to LSE-AJTT-GTC, which we retained to highlight the effect of the GTC variable. The Kumb90-AJTT-HN model was also included to show the impact of the Australian time-trend variables on the DNSPs' efficiency profiles.

These results are also compared with the OPFP efficiency measure (outlined in section 5.1.7) and with the average of the four econometric models. Figures 6.3 and 6.4 present the estimated efficiency scores from the four models, their average, and the OPFP series.[51]

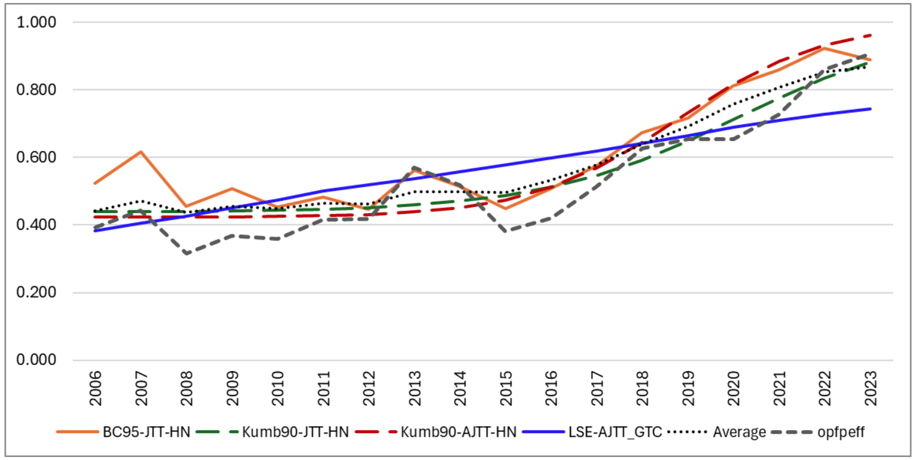
[51] Note that the average is shown for comparison only and should not be taken to suggest that all of these four models should be used for the regulatory application of benchmarking models.

Figure 6.3 Efficiency Score Comparisons: CD, Long Period

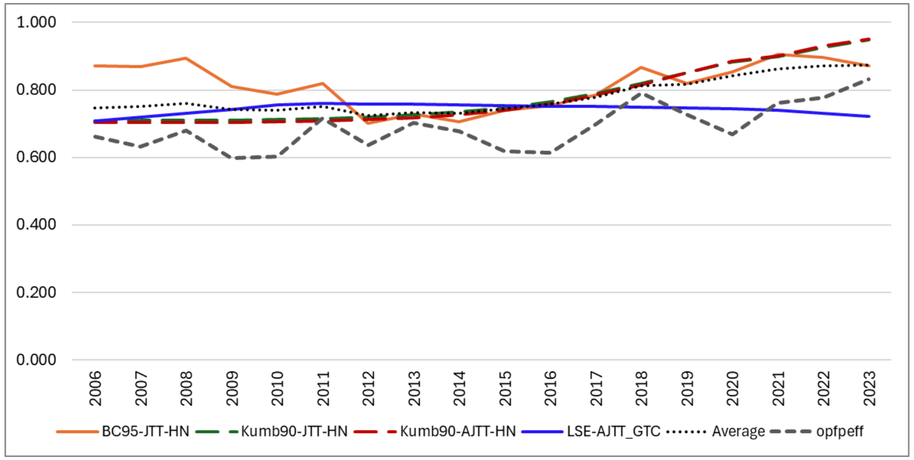
EVO



AGD



CIT



END

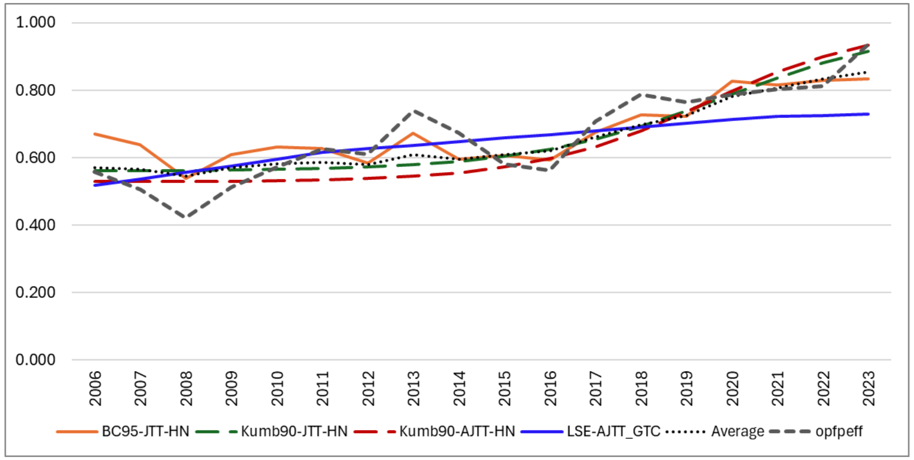
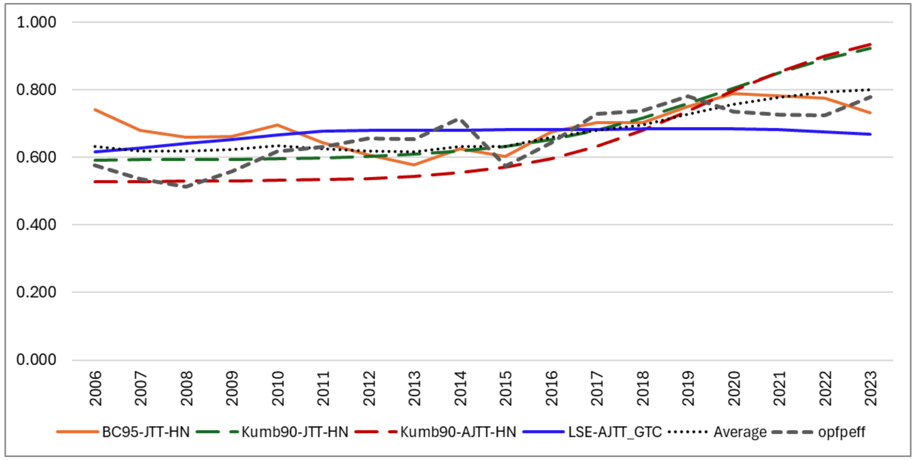
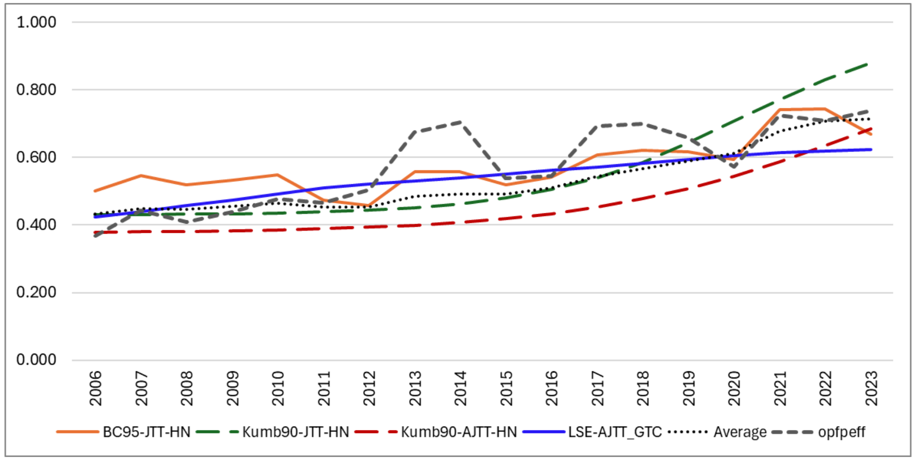


Figure 6.3 (cont.)

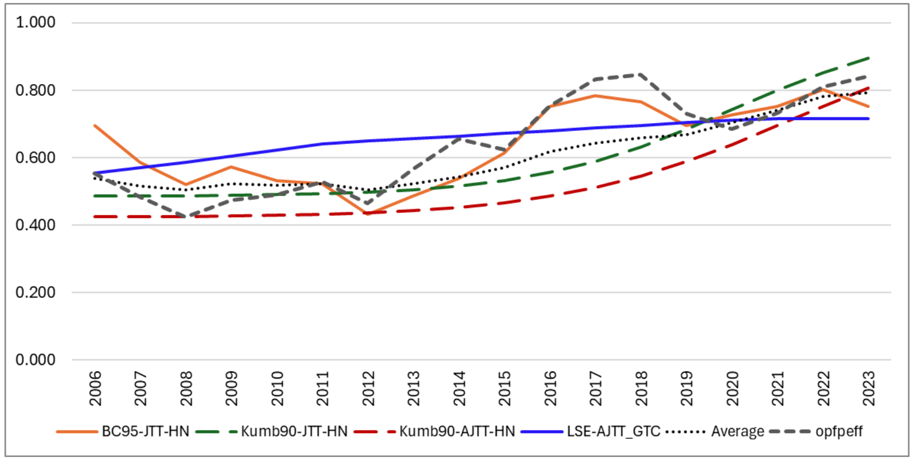
ENX



ERG



ESS



JEN

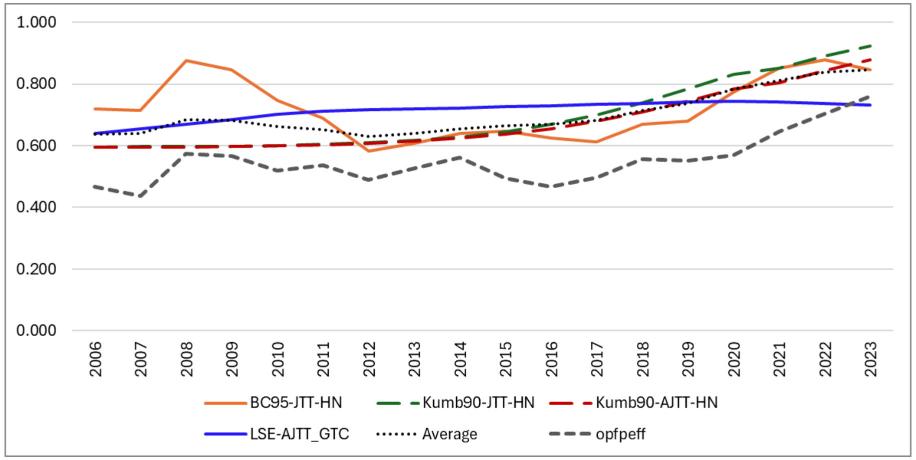
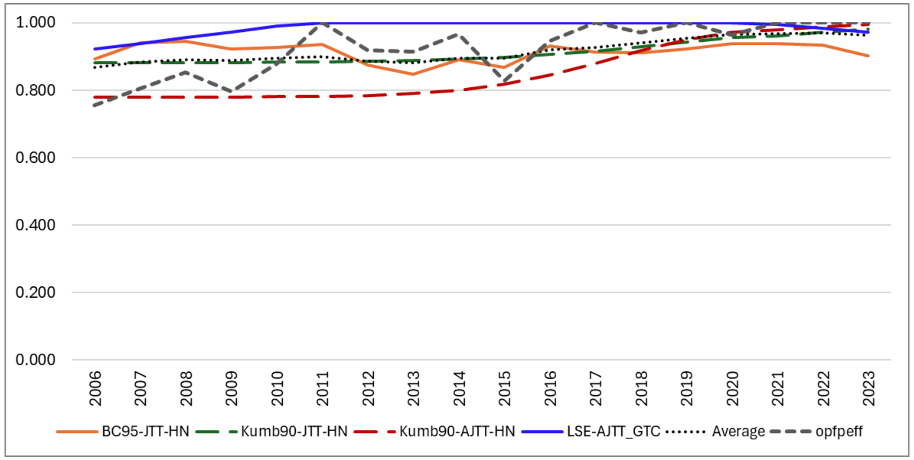
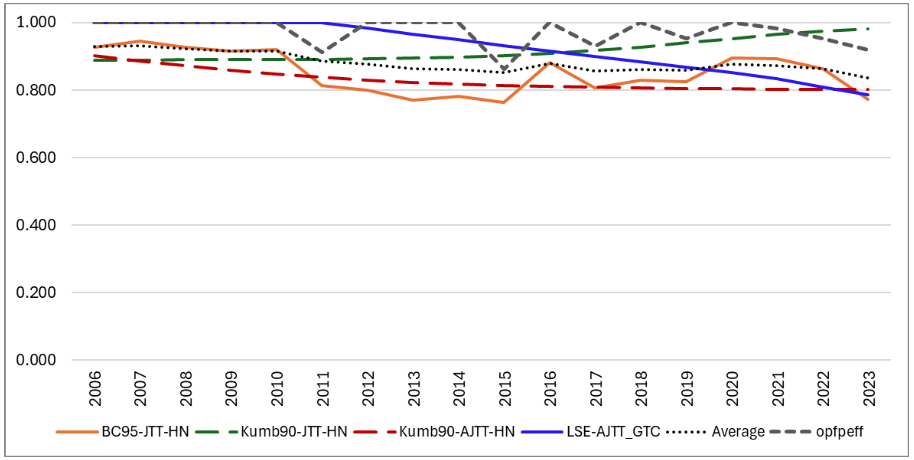


Figure 6.3 (cont.)

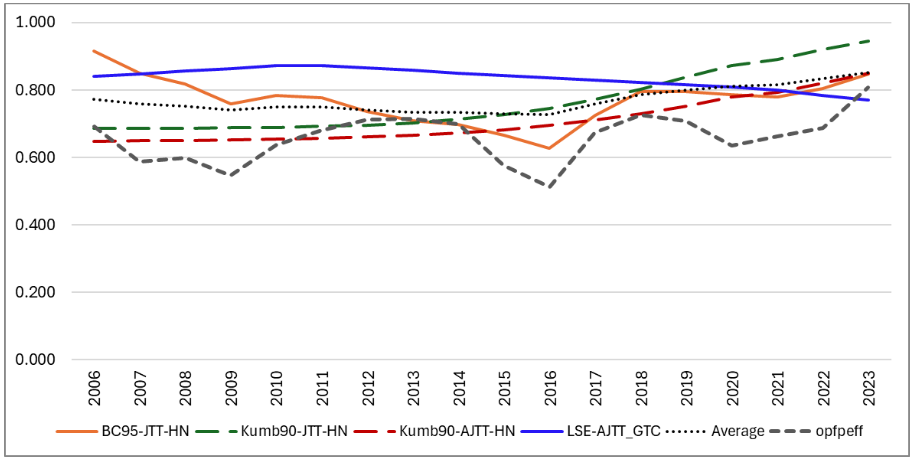
PCR



SAP



AND



TND

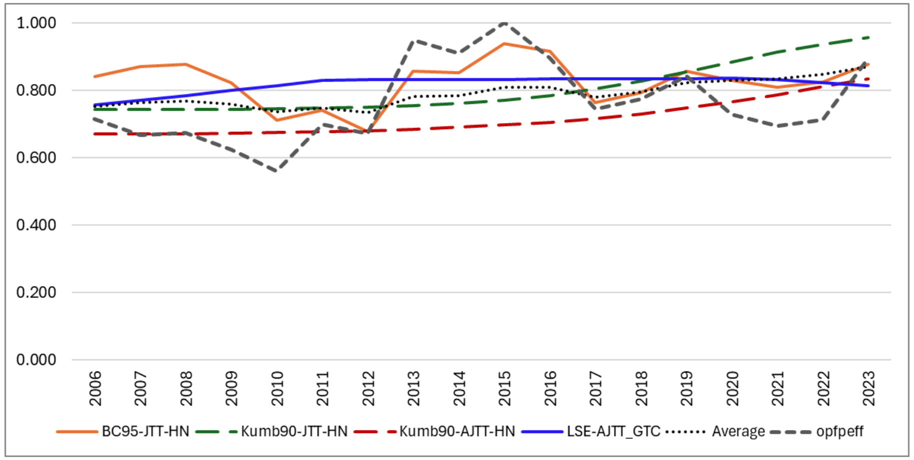
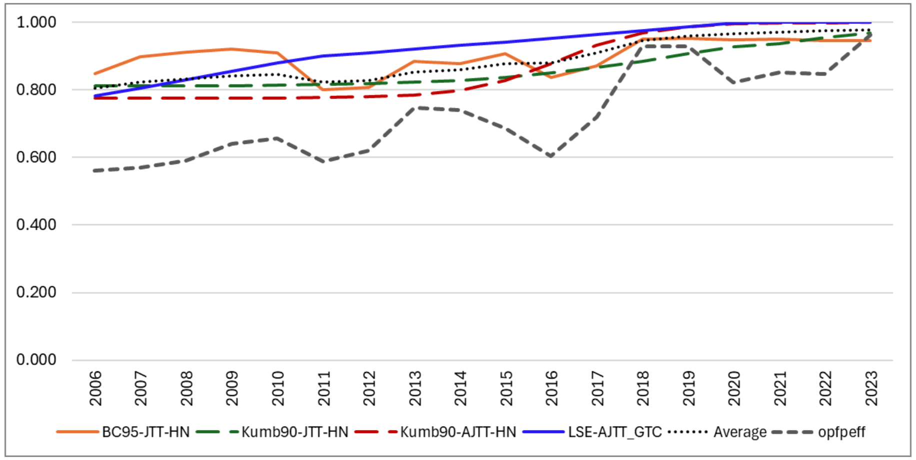


Figure 6.3 (cont.)

UED



Australia - AVG

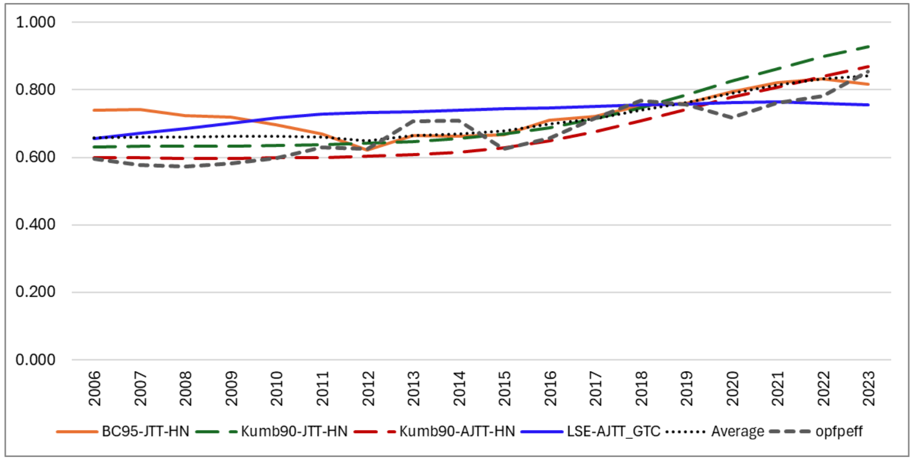
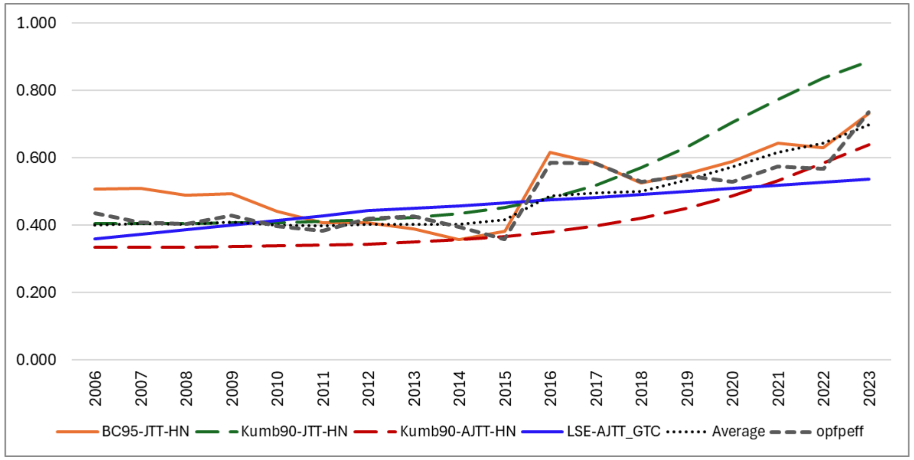
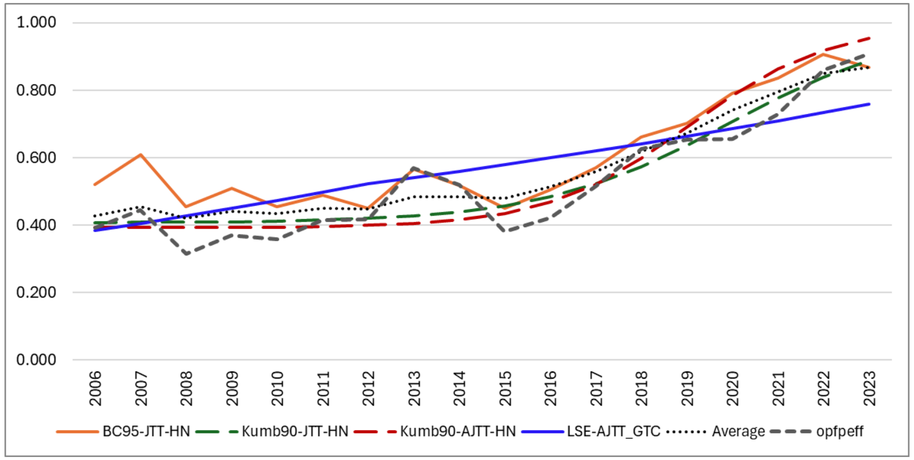


Figure 6.4 Efficiencies Score Comparison: TLG, Long Period

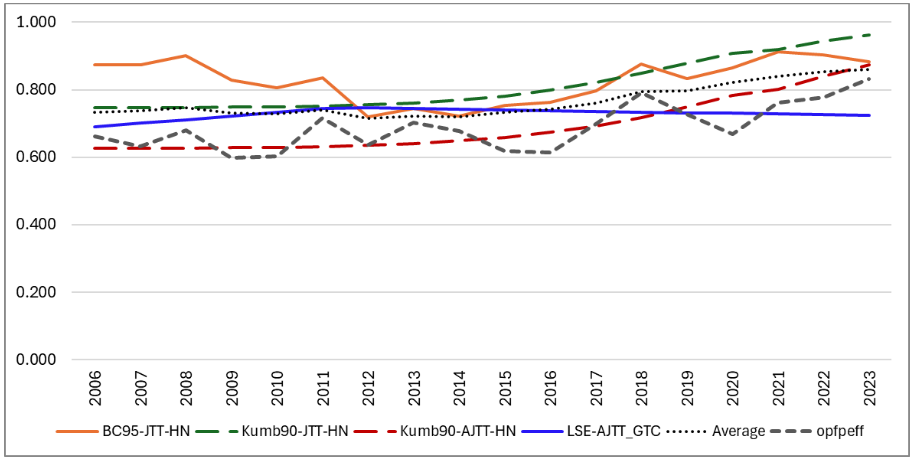
EVO



AGD



CIT



END

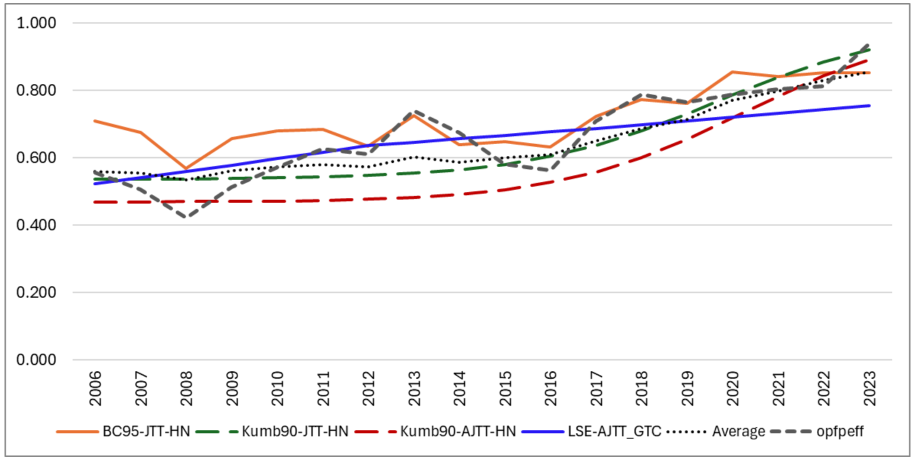
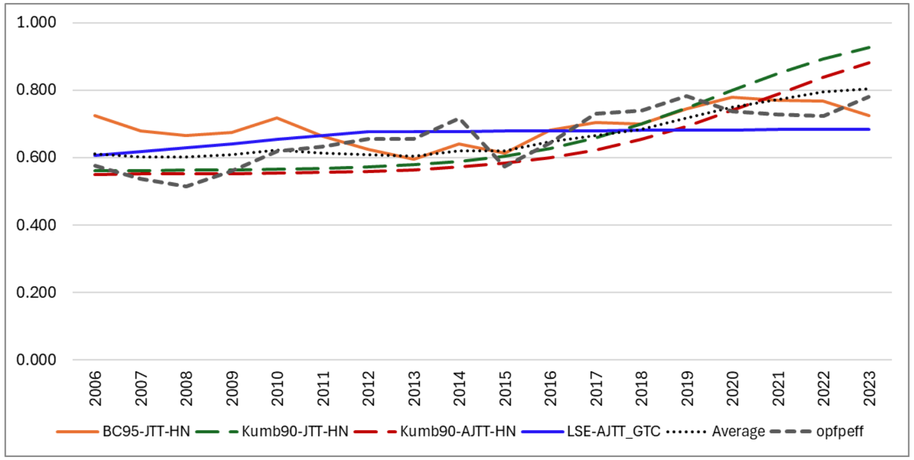
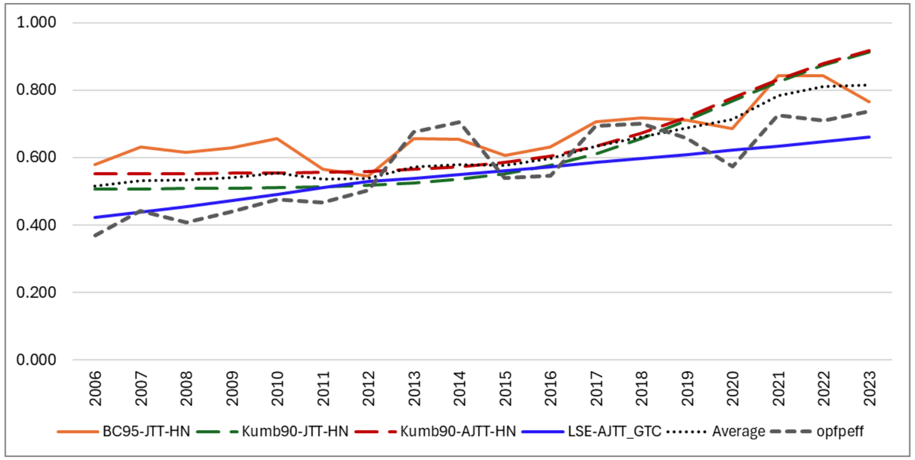


Figure 6.4 (cont.)

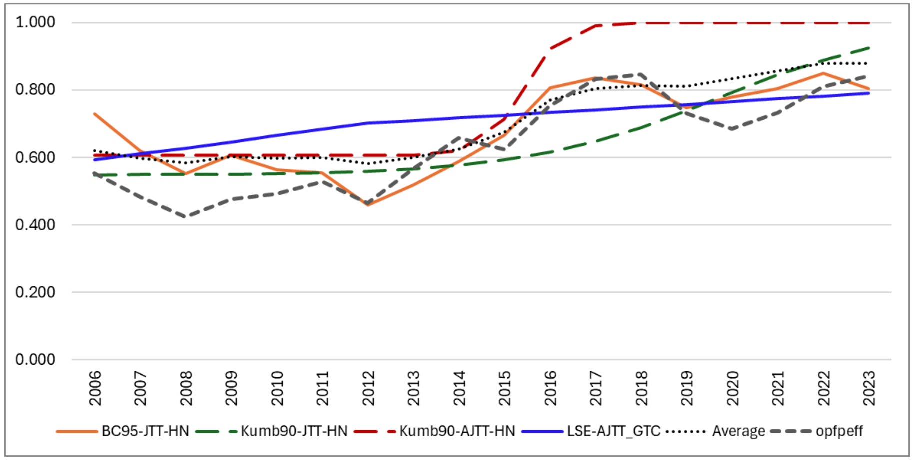
ENX



ERG



ESS



JEN

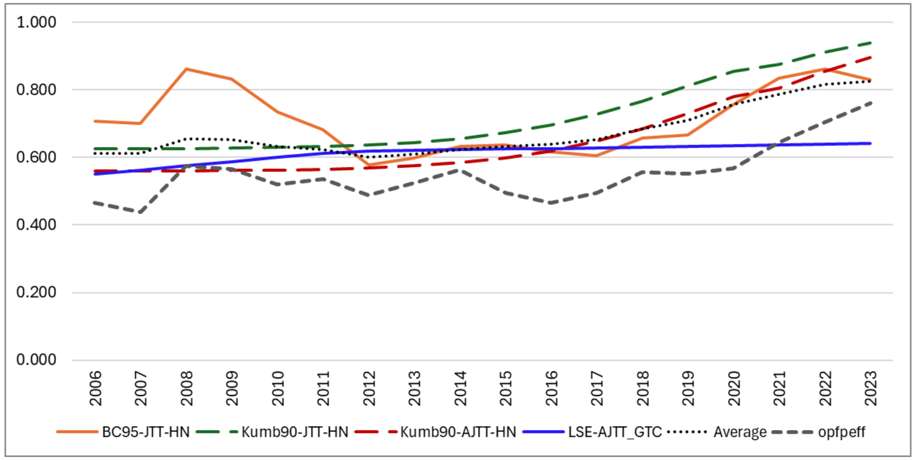
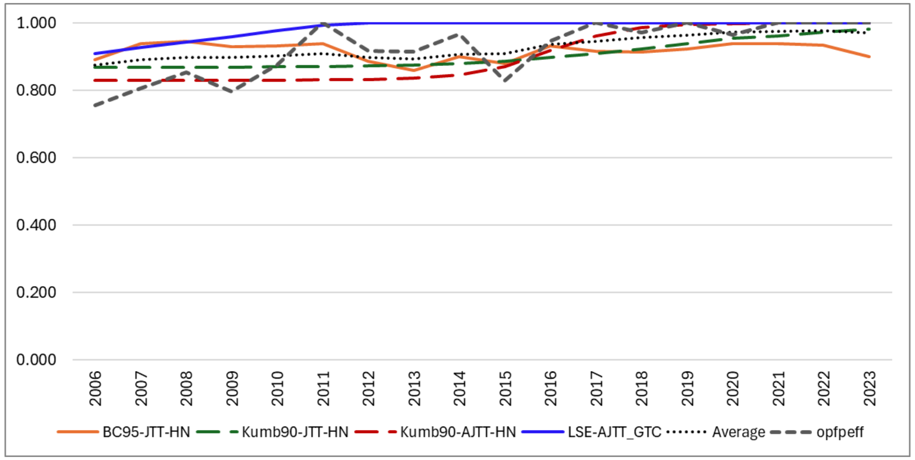
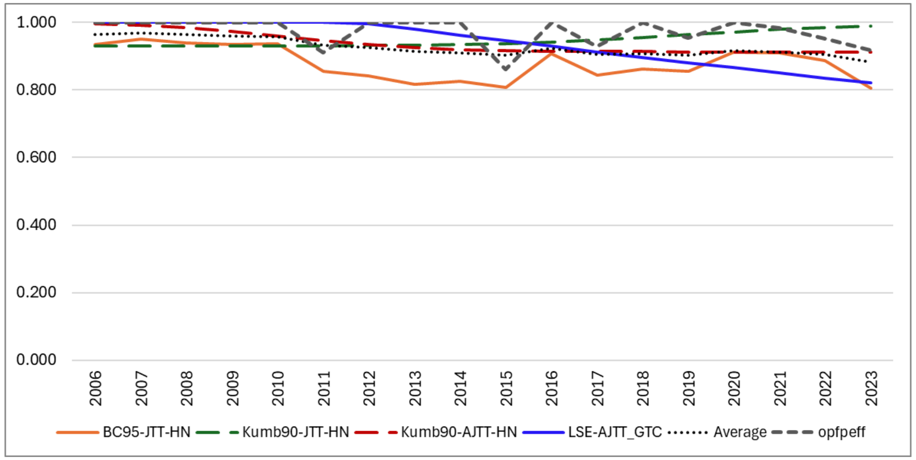


Figure 6.4 (cont.)

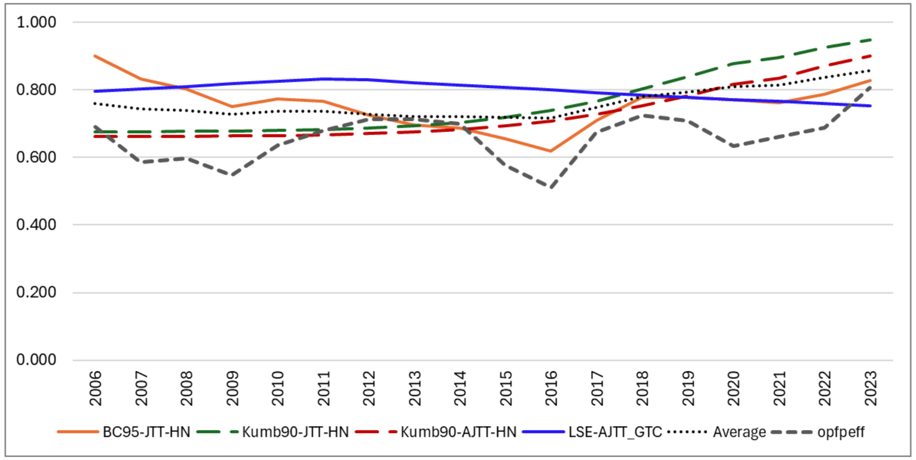
PCR



SAP



AND



TND

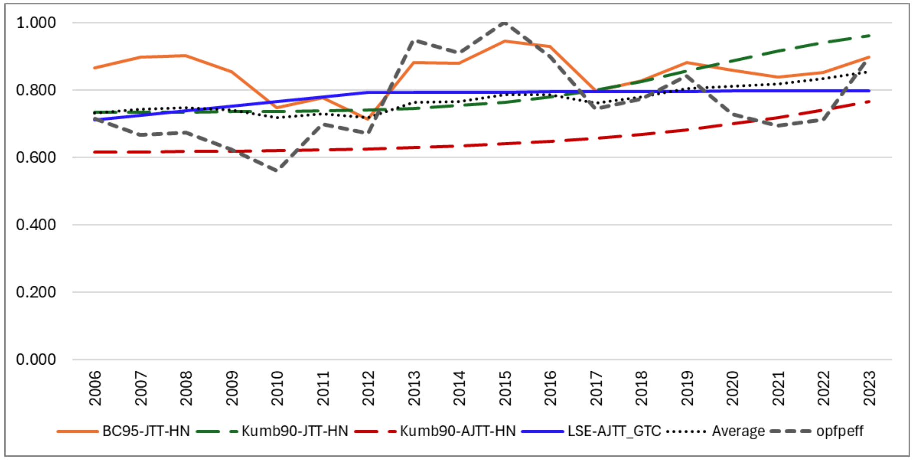
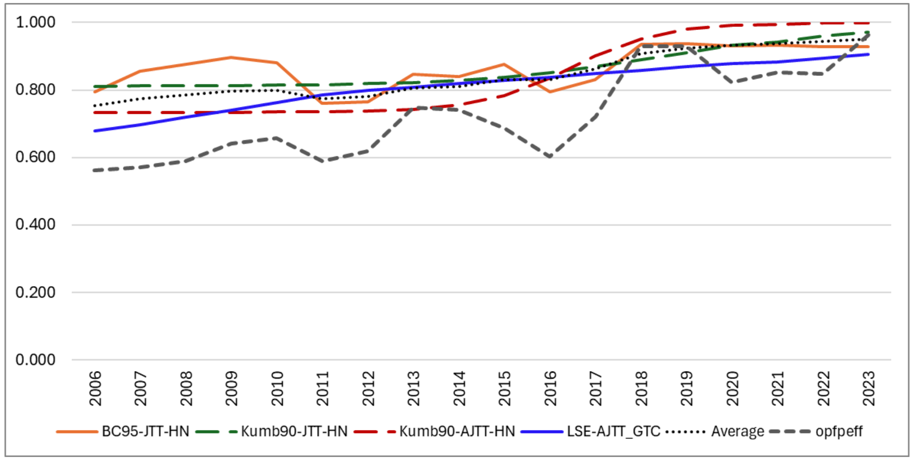
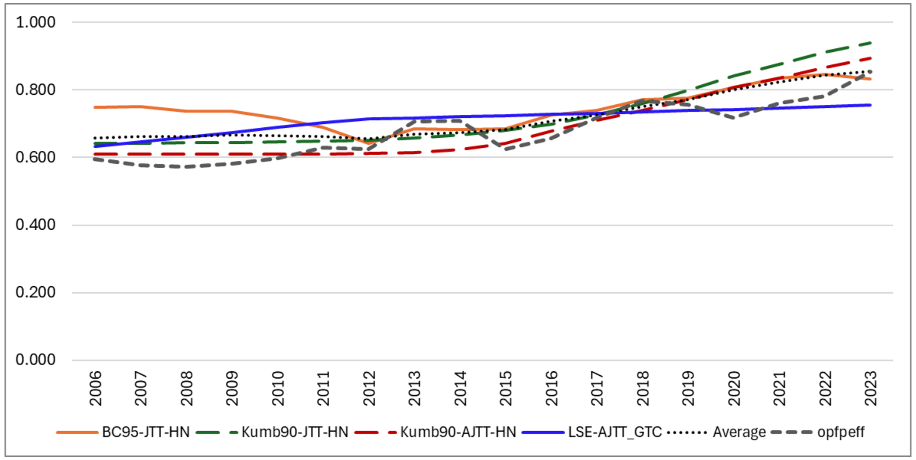


Figure 6.4 (cont.)

UED



Australia - AVG



The following observations can be made:

- The BC95-JTT-HN model produces patterns that are closely aligned with the OPFP efficiency series, suggesting that both methods capture comparable shifts in efficiency over time. However, the levels are not always consistent. On average, for both Cobb-Douglas and Translog models, the BC95-JTT-HN model yields higher efficiency scores than the OPFP.
- The Kumb90-AJTT-HN specification allows for a wide variety of profile shapes. In most cases, the resulting profiles are similar to those from the more restricted Kumb90-JTT specification. The main exceptions are ESS (only for Translog), PCR, SAP and UED, where the Kumb90-AJTT-HN produces efficiency score profiles that diverge considerably from the Kumb90-JTT-HN.
- The LSE models consistently generate concave efficiency profiles, although for some DNSPs the shape is close to linear.
- When considering average efficiency scores, overall, the Kumb90-AJTT-HN model produces the lowest scores, while the BC95-JTT-HN model produces the highest. In the Cobb–Douglas models, the Kumb90-AJTT-HN model produces the lowest scores for 10 DNSPs and for the industry as a whole. In contrast, the BC95-JTT-HN model produces the highest average scores for seven DNSPs. In the Translog models, the Kumb90-AJTT-HN model produces the lowest scores for seven DNSPs and for the industry as a whole. In contrast, the BC95-JTT-HN model produces the highest average scores for seven DNSPs and for the industry overall.
- When considering efficiency scores in 2023, the LSE-AJTT-GTC model produces the lowest scores overall, while the Kumb90-JTT-HN model produces the highest. In the Cobb–Douglas models, the LSE-AJTT-GTC model produces the lowest scores for 10 DNSPs and for the industry as a whole. In contrast, the Kumb90-JTT-HN model produces the highest scores for eight DNSPs and for the industry overall. In the Translog models, the LSE-AJTT-GTC model produces the lowest scores for 10 DNSPs and for the industry as a whole. In contrast, the Kumb90-JTT-HN model produces the highest scores for eight DNSPs and for the industry overall.

Table 6.1 reports the average annual rate of change in efficiency scores from 2006 to 2023 for each DNSP. The values are the average of the Cobb-Douglas and Translog models. Generally, across the four econometric specifications, the BC95-JTT-HN model produces the lowest average rate of change, while the Kumb90 models produce the highest. The Kumb90-HN-AJTT model generally yields an average rate of change closest to that of the OPFP efficiency measure. The LSE-AJTT-GTC model's rates of change tend to be closest to the average of the four econometric models.

Table 6.1 Average Annual Growth Rate of Efficiency Scores, 2006–2023

	BC95-JTT-HN	*Kumb90 JTT-HN*	*Kumb90 AJTT-HN*	*LSE AJTT-GTC*	*OPFPeff*	*AVG (Econometric Models)*
EVO	2.1%	4.7%	3.9%	2.2%	3.1%	3.3%
AGD	3.1%	4.4%	5.0%	4.0%	4.9%	4.1%
CIT	0.1%	1.6%	1.9%	0.2%	1.4%	0.9%
END	1.2%	3.1%	3.6%	2.1%	3.1%	2.5%
ENX	-0.1%	2.8%	2.6%	0.6%	1.8%	1.5%
ERG	1.7%	3.9%	3.3%	2.4%	4.1%	2.8%
ESS	0.6%	3.4%	3.4%	1.6%	2.5%	2.2%
JEN	0.9%	2.5%	2.6%	0.9%	2.9%	1.8%
PCR	0.1%	0.7%	1.3%	0.5%	1.6%	0.6%
SAP	-1.0%	0.5%	-0.6%	-1.3%	-0.5%	-0.6%
AND	-0.5%	2.0%	1.7%	-0.4%	0.9%	0.7%
TND	0.2%	1.6%	1.3%	0.6%	1.3%	0.9%
UED	0.8%	1.1%	1.7%	1.6%	3.2%	1.3%
AVG	0.6%	2.3%	2.3%	0.9%	2.1%	1.5%

Overall, the estimated efficiency time-paths differ substantially across the four econometric time-varying models. Each approach imposes restrictions on the shape of the time profile of efficiencies. The BC95-JTT-HN and Kumb90-AJTT-HN are the most flexible, LSE-AJTT-GTC model is somewhat more restrictive, and the Kumb90-JTT-HN is the most restrictive, since all DNSPs' inefficiency trends are the same.

One approach to improve robustness is to base the average not on a single SFA and LSE model, but on two different SFA models and one LSE model; if the models are theoretically sound and perform well, averaging their estimates is likely to yield more reliable result.[52]

6.2 Decomposing productivity changes

An estimated cost function with time-varying inefficiency can be used to decompose productivity changes into the separate effects of cost efficiency changes (CEC), technical change (TC), scale efficiency changes (SEC), and the effect of changes in OEFs (OEFC). The decomposition of a cost function is presented in Coelli et al (2003: 41–42). Here the application is an opex cost function rather than a total cost function, and the decomposition is the rate of change in Opex PFP rather than the rate of change in TFP.

Consider the Translog cost frontier:

[52] Several alternative methods for combining model results exist beyond simple averaging. These include techniques that explicitly account for model uncertainty, such as Bayesian model averaging and the Minimum Quadratic Loss approach, as discussed in Economic Insights (2017b). Another option is inverse-variance weighting, which assigns weights to each parameter being averaged which are proportional to the inverse of each parameter's variance (eg, the parameters being averaged may be the efficiency score of one DNSP estimated using different models). Further analysis is required to determine the method most appropriate for this application.

$$c_{it} = \beta_0 + \sum_{m=1}^{M} \beta_m y_{mit} + 0.5 \sum_{m=1}^{M} \sum_{l=1}^{M} \beta_{ml} y_{mit} y_{lit} + \varphi z_{it} + \lambda t + u_{it}(t) + v_{it} \quad (6.2.1)$$

where v_{it} is an idiosyncratic normally distributed and independent random disturbance term, and u_{it} is a measure of inefficiency, which may be the fixed and time-varying DNSP-specific effects in the LSE models, or a one-sided stochastic inefficiency term which is itself determined by *z*-variables, as in the SFA models. In the Cobb-Douglas specification, all β_{ml} parameter are restricted to equal zero.

After estimating the opex cost function, the Opex PFP change between period 0 and period 1 for firm *n* can be calculated as:

$$\ln\left(\frac{Opex\ PFP_{i1}}{Opex\ PFP_{i0}}\right) = \sum_{m=1}^{M} \left(\frac{[(1-e_{i1})/e_{i1}]\varepsilon_{mi1} + [(1-e_{i0})/e_{i0}]\varepsilon_{mi0}}{2}\right)(y_{i1} - y_{i0}) \quad (6.2.2)$$

$$-0.5\left[\frac{\partial c_{i0}}{\partial t} + \frac{\partial c_{i1}}{\partial t}\right] - \varphi(z_{i1} - z_{i0}) + \ln\left(\frac{\theta_{i1}}{\theta_{i0}}\right)$$

where:

- c_{it} is the log of real opex, the dependent variable of the cost function, for firm *i* in period *t*;
- y_{mit} is the log of output *m* for firm *i* in period *t*;
- *t* is the time variable(s) used to measure technical change, often a simple time trend;
- $\varepsilon_{mit} = \partial c_{it}/\partial y_{mit}$, the scale elasticity of opex with respect to output *m* for firm *i* in period *t*; and $e_{it} = \sum_{i=1}^{M} \varepsilon_{mit}$.
- z_{it} is the log OEF variable for firm *i* in period *t*;
- θ_{it} is the opex cost efficiency score for firm *i* in period *t*; and: $\ln(\theta_{i1}/\theta_{i0}) = -(u_{i1} - u_{i0})$, where u_{it} is a measure of inefficiency, since $\theta_{it} = \exp(-u_{it})$.

The four terms on the right-hand-side of (6.2) are measures of: (i) scale efficiency change (SEC); (ii) technical change (TC); (iii) the effect of change in the OEF (OEFC); and (iv) opex cost efficiency change (CEC).

In the SFA models that include *t* as a regressor to measure technical change, $\partial c_{it}/\partial t = \lambda$. In the SFA models that use the GTC formulation of technical change, in place of λt, the formulation is: $\sum_p \lambda_p I_p$, were the subscript *p* refers to periods, each comprising three years, I_p is a dummy variable for that period, and λ_p is a parameter to be estimated. For the first three years, there is no dummy variable. The λ_p coefficient represents the cumulative effect of technical change from the first year of the sample to the first year of period p. In the first year of a period to which an I_p dummy variable applies: $\partial c_{it}/\partial t = \lambda_p - \lambda_{p-1}$; and in all other years $\partial c_{it}/\partial t = 0$.[53]

The cost efficiency scores are calculated as shown in equation (4.5.4) for the LSE time-varying models and for the SFA models they are calculated using the Battese-Coelli (1988) method. That is: $\theta_{it} = E[\exp(u_{it})|(v_{it} + u_{it})]$. These statistics are returned by the Stata programs used here for estimating SFA models.

The elasticities of cost with respect to an individual output m are given by:

$$\varepsilon_{mit} = \beta_m + 0.5 \sum_{l=1}^{M} \beta_{ml} y_{lit} \tag{6.2.3}$$

In the Cobb-Douglas formulation, since all $\beta_{ml} = 0$: $\varepsilon_{mit} = \beta_m$.

The four decomposition components and the change in Opex PFP were calculated for each of the selected time-varying models.[54] Figures 6.5 to 6.9 show the average contribution of each effect (SEC, TC, OEFC and CEC) to the Opex PFP rate of change, together with the average Opex PFP rate of change over the full sample period (2006–2023), for each of the four preferred time-varying models. The values represent the average of the CD and TLG specifications. Key observations include:

- The average SEC over 2006–2023 is only a small component of productivity change for all models. In the Kumb-AJTT-HN model it is mostly negative, suggesting that DNSPs have, on average, become less scale efficient over time, operating further away from the point of minimum efficient scale. By contrast, the other models show mostly small positive values, suggesting limited improvements in scale efficiency.
- The average TC is larger in magnitude for the Kumb models. The Kumb90-JTT-HN model consistently records the largest effect, followed by the Kumb90-AJTT-HN model. The BC95-JTT-HN model ranks third, although its magnitude is considerably lower than that of the Kumb models. The LSE-AJTT-GTC model consistently shows the lowest magnitude of TC.
- The average effect of change in the undergrounding variable (OEFC) is very small for the Kumb models, indicating that these specifications attribute limited influence of

[53] In the LSE models, an additional term is included in technical change. The term $min(.)$ in equation (4.5.4), which is subtracted off the firm-specific effects to derive inefficiency measures, needs to be added back to the technical change term.

[54] As noted, we selected the most promising models, excluding LSE-AJTT, which produced results similar to LSE-AJTT-GTC, and adding Kumb90-AJTT-HN to illustrate the impact of the Australian time-trend variables. Noting, however, that in the ABR24 data sample, the Kumb90-AJTT-HN has an excessive frequency of MVs for Australian DNSPs. Section 5.2 shows there is some improvement in its MV performance in the ABR25 data sample (2006-2024). Hence, it is a potentially feasible model in future, larger, data samples.

underground on shifts in the cost frontier. In contrast, the LSE-AJTT-GTC and, more markedly, the BC95-JTT-HN model record substantially higher values, with the latter consistently assigning the greatest magnitude to this effect.

- The average CEC varies considerably across models and DNSPs, reflecting both differences in how each specification estimates cost efficiency and the heterogeneous performance of the individual DNSPs. Overall, the Kumb models record the largest average improvements, followed by the LSE-AJTT-GTC model, while the BC95-JTT-HN model consistently reports the lowest values.
- The average Opex PFP change also varies considerably across models and DNSPs. Broadly speaking, however, all models are consistent in indicating positive change for most DNSPs, with only a few recording small or moderate negative values.

Overall, the Kumb90 models yield results that diverge noticeably from the other specifications. They generally report negative values for scale efficiency change (SEC), larger magnitudes of technical change (TC), smaller effects from the undergrounding variable, and higher average values of cost efficiency change (CEC).

Figure 6.5 Scale Efficiency Change (SEC) - Average 2006-2023

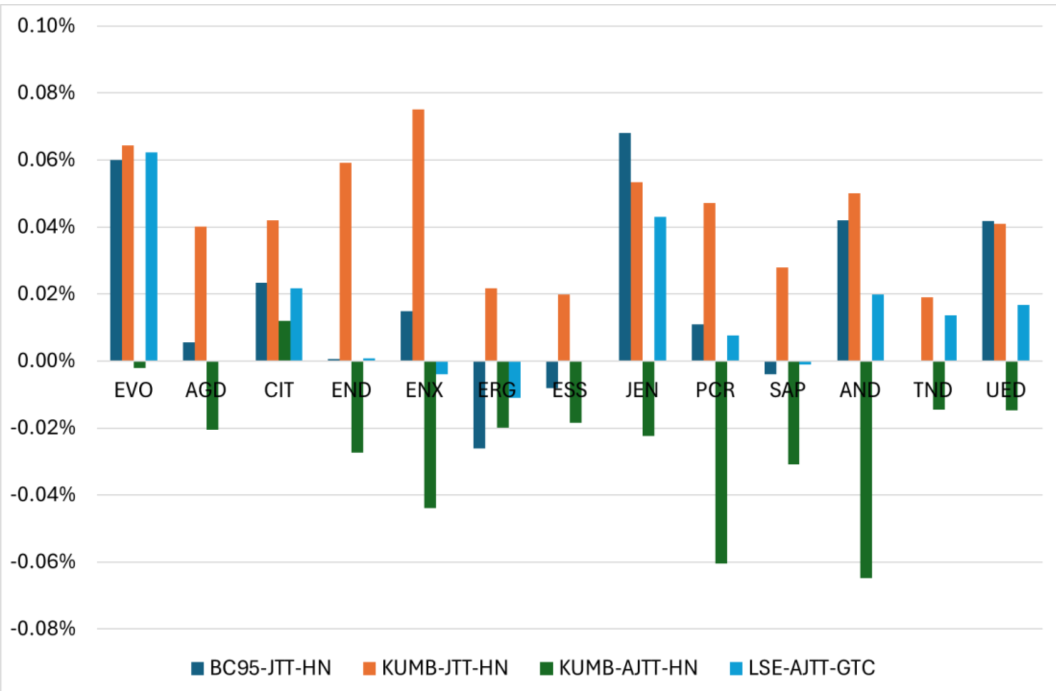


Figure 6.6 Technical Change (TC) - Average 2006-2023

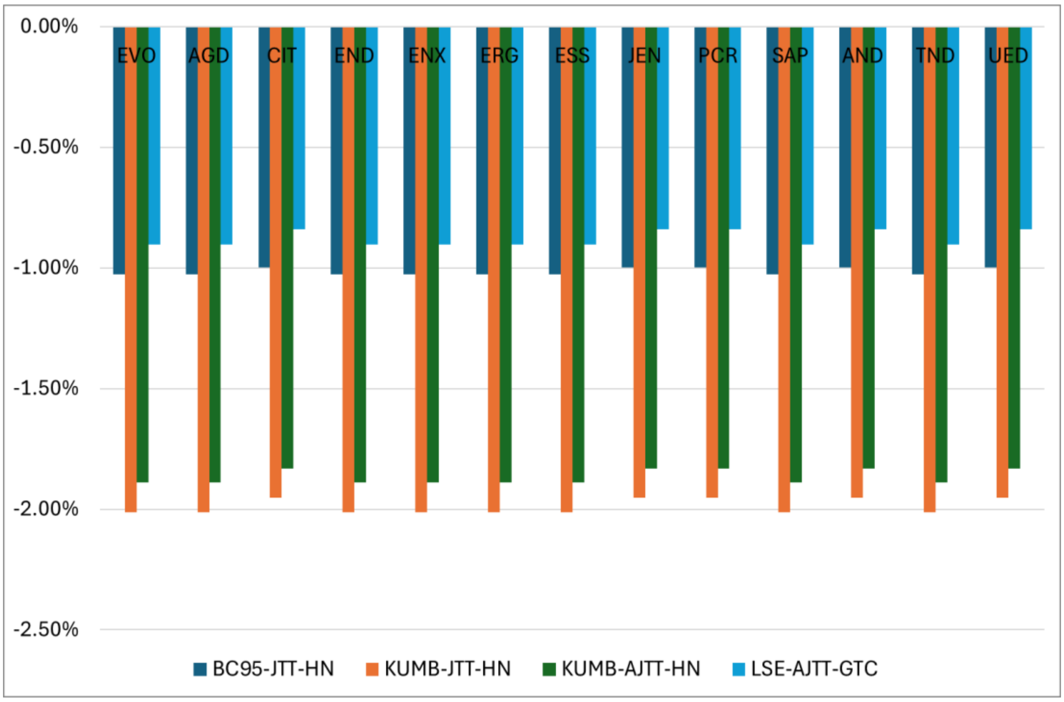


Figure 6.7 Effect of Change in the UG variable (OEFC) - Average 2006-2023

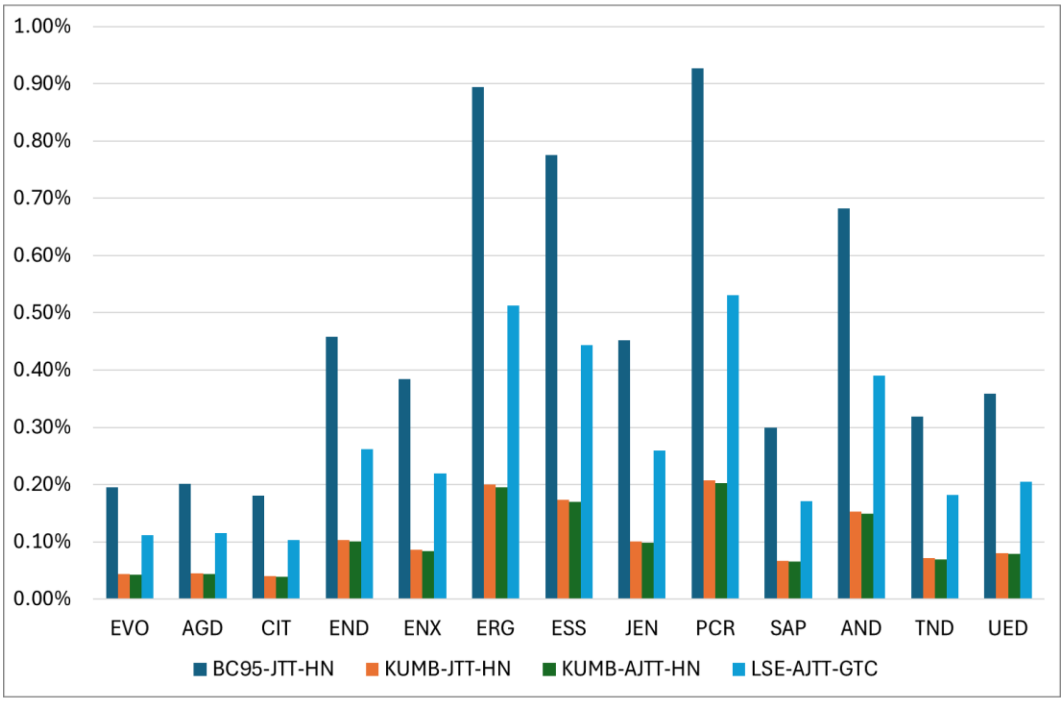


Figure 6.8 Cost Efficiency Change (CEC) - Average 2006-2023

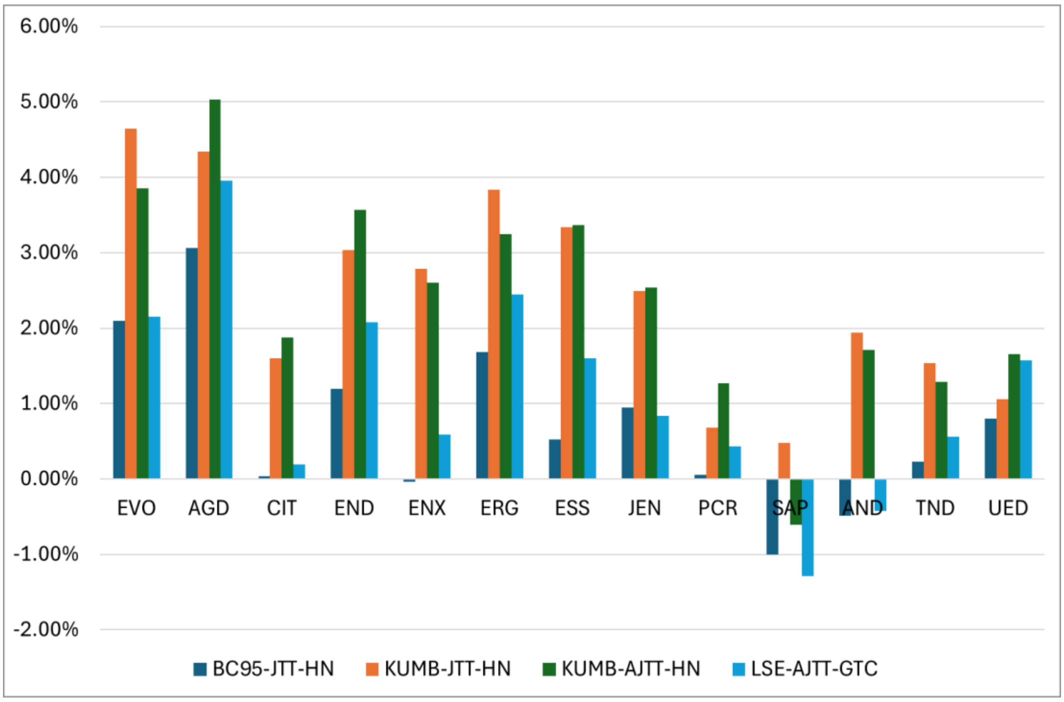
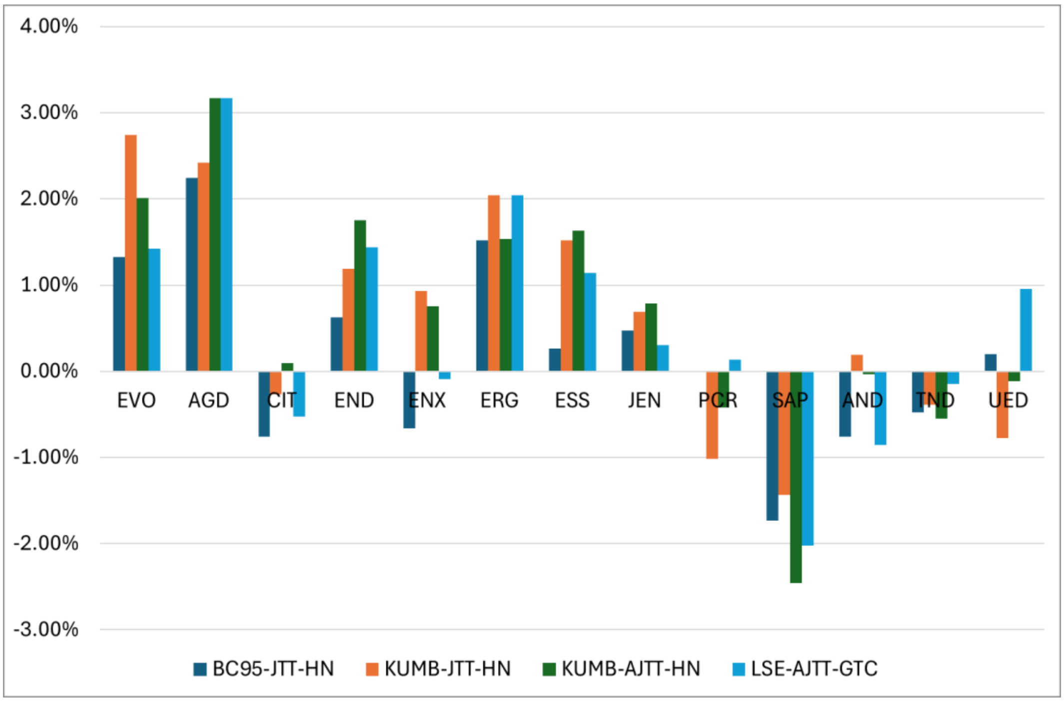


Figure 6.9 Opex PFP change - Average 2006-2023



6.3 The benchmarking roll-forward model (BRFM)

This section outlines how base-year efficiency for DNSPs can be calculated when the econometric model includes time-varying inefficiency. Essential Energy, Endeavour and Ausgrid all expressed the need for clarity in how a model with jurisdictional time trends or time-varying inefficiency would be used in the benchmarking roll-forward model (BRFM). Endeavour suggested that the BRFM be reviewed in this consultation process.

6.3.1 The standard BRFM

When applying the benchmarking results to determine whether an efficiency adjustment is needed to a DNSP's base-year opex, the AER uses a procedure referred to as the benchmarking roll-forward model (BRFM). In the standard opex cost function, efficient opex is calculated by multiplying observed opex by the efficiency score.[55] Since the efficiency score does not vary over time it is applied to the sample average real opex (deflated by the opex price index in the base year period) to obtain a sample average efficient level of opex. This average efficient opex is taken to be representative of the mid-point of the estimation sample.[56] Then, this efficient opex is rolled forward to the base year using the parameters of the econometric model to account for the drivers of efficient real opex over the relevant period. Because real opex refers to nominal opex deflated by the opex input price index, it is not necessary to make an adjustment for the differential growth rates between opex input prices and the CPI.

The period-average efficient opex can be calculated as:[57]

$$C_{PA}^{e} = C_{PA} \times \min\left\{\frac{\theta_{PA}}{ACP}, 1\right\} \tag{6.3.1}$$

where C_{PA} is the real period-average opex; C_{PA}^{e} is the efficient period-average opex; θ_{PA} is the period-average estimated efficiency score; and ACP is the 'adjusted comparison point' defined as:

$$ACP = \frac{0.75}{1 + OEF\%} \tag{6.3.2}$$

Here OEF% is an ex-post adjustment for the combined effect of OEFs on the relevant DNSP's efficient opex, and 0.75 is a discretionary adjustment by the regulator to account for factors such as modelling imprecision and the combined effect of any unaccounted for OEFs.

---

[55] Inefficiency is assumed to be constant over time, and although a time trend variable is included, it is used as a control variable for the effects of various factors, such as TC and OEFC and not only CEC. Therefore, it is not interpreted as a measure of CEC.

[56] It is not, in general, equal to actual opex in the mid-sample year. The roll-forward is from the period-average to the base year.

[57] Another way to calculate the period-average efficient opex is to take the average of the estimated efficient opex in each year. That is, $C_{PA}^{e} = \min\left[\left(\frac{1}{T}\sum_{t=1}^{T} \theta_t \, C_t\right)/ACP \, , C_{PA}\right]$, where θ_t is the efficiency score and C_t is the real opex in year t, and T is the number of years in the sample. This approach yields different values to (6.3.1), although the differences, in most cases, do not appear to be material. It may warrant further examination.

The roll-forward to the base-year is performed as follows:

$$C^e_{BY} = C^e_{PA} \times \dot{C} \qquad \text{where: } \dot{C} = \dot{Y} - \dot{P} \tag{6.3.3}$$

where subscript BY represents the base year; the dot over a variable represents the cumulative change in that variable from the period-average to the base year; $\dot{C}$ is the predicted change in opex (at constant base year prices); Y is the total output index; P denotes the opex partial factor productivity of the relevant DNSP.

The rate of change of the output index (denoted $\dot{Y}$) is calculated as follows:

$$\dot{Y} = \frac{1}{e_Y} \sum_{m=1}^{M} \varepsilon_m . \dot{y}_m \tag{6.3.4}$$

where y_m is output m, $\dot{y}_m$ is its growth rate, ε_m is the elasticity of real opex with respect to output m; and $e_Y \equiv \sum_{m=1}^{M} \varepsilon_m$. The rate of change in opex partial factor productivity (denoted $\dot{P}$) *for an efficient firm* is calculated as follows:

$$\dot{P} = (1 - e_Y)\dot{Y} - \varphi\dot{z} - \lambda(t - s) \tag{6.3.5}$$

where $\dot{z}$ is the rate of change in the OEF, φ is the elasticity of real variable cost with respect to the OEF, and λ is the parameter on the time-trend variable used to measure technical change. The rate of change in the efficiency score is not relevant here because (6.3.5) measures the cumulative rate of change in efficient opex.

The specifications proposed in this report allow for time-varying efficiency scores. As a result, if the base year falls within the benchmarking period it is theoretically possible to calculate base-year efficient opex directly by multiplying the efficiency score for the base year by the real opex in that same year, rather than first calculating period-average efficient opex and projecting it forward. However, in practice, base years under resets are often outside (after) the benchmarking period, and efficiency estimated for the last year of the data sample cannot be assumed to match the base year. Hence, the BRFM needs to be retained at least to bridge the period between the last year of the sample and the base-year used in a reset.[58]

To incorporate the time varying efficiency models, the BRFM must account for changes in opex partial productivity *for an efficient firm* which, as shown in equation (6.3.5), includes the separate effects of changes in scale efficiency, the effect of OEF changes, and technical change).

[58] In this section, we use the BRFM to bridge the period between the sample-average year and the base year (assumed to be 2023). The alternative approach, bridging from the final year of the sample to the base year, may also be considered in future work.

6.3.2 Equivalent Formulation of the BRFM

There is an alternative sequence to the BRFM calculation, which involves carrying out the ACP adjustment at the end of the process. Hence, equation (6.3.1) would be $C_{PA}^{ue} = C_{PA} \times \theta_{PA}$ (where the superscript *ue* refers to the 'unadjusted-efficient' opex). Then after deriving the unadjusted efficient base year opex using equations (6.3.3) and (6.3.5), the adjusted efficient base-year opex would be obtained by: $C_{BY}^{e} = \min(C_{BY}^{ue}/ACP\,, C_{BY})$. Here C_{BY}^{ue} is the unadjusted base-year efficient opex, and C_{BY} is the actual base year opex. The usefulness of this formulation is that one can calculate the implied base-year efficiency: $\theta_{BY} = C_{BY}^{ue}/C_{BY}$. This can be compared to the efficiency score estimate for the base-year yielded by time-varying inefficiency models (for this purpose assuming the base year is the final year of the sample period).

Table 6.2 presents the BRFM-implied efficiency scores for 2023 for each of the selected time-varying models.[59] Table 6.3 presents the estimated efficiency scores in 2023 produced by the time-varying models. The correlation coefficients between the 2023 efficiency scores in Tables 6.2 and 6.3 are presented in the first row of Table 6.4. In the second row of Table 6.4, the correlation coefficients between the change in efficiency from the mid-period to 2023 are shown. Table 6.4 also presents, in the third row, the correlation coefficients between the BRFM-implied efficiency scores presented in Table 6.2 and the with Opex MPFP-implied efficiency scores in 2023.

Table 6.2 BRFM-implied 2023 efficiency scores

	BC95-JTT-HN		*KUMB90-JTT-HN*		*KUMB90-AJTT-HN*		*LSE-AJTT-GTC*	
	CD	*TLG*	*CD*	*TLG*	*CD*	*TLG*	*CD*	*TLG*
EVO	0.733	0.777	0.854	0.871	0.761	0.645	0.614	0.589
AGD	1.000	0.995	1.000	0.975	1.000	0.985	0.848	0.832
CIT	0.901	0.919	0.953	0.985	0.934	0.829	0.743	0.716
END	0.869	0.923	0.965	0.944	0.928	0.844	0.764	0.765
ENX	0.749	0.756	0.837	0.810	0.794	0.769	0.658	0.644
ERG	0.693	0.810	0.747	0.840	0.616	0.874	0.597	0.602
ESS	0.788	0.847	0.831	0.912	0.718	1.000	0.732	0.791
JEN	0.886	0.875	0.962	1.000	0.929	0.904	0.803	0.663
PCR	0.885	0.891	1.000	1.000	0.950	1.000	0.882	0.877
SAP	0.756	0.785	0.910	0.939	0.815	0.922	0.749	0.756
AND	0.876	0.865	0.986	0.987	0.905	0.941	0.868	0.821
TND	0.904	0.938	0.986	0.980	0.870	0.801	0.821	0.783
UED	1.000	1.000	1.000	1.000	1.000	1.000	0.970	0.823
AVG	0.849	0.876	0.925	0.942	0.863	0.886	0.773	0.743

[59] As noted, we selected the most promising models, excluding LSE-AJTT, which produced results similar to LSE-AJTT-GTC, and adding Kumb90-AJTT-HN to illustrate the impact of the Australian time-trend variables.

Table 6.3 Time-varying 2023 efficiency scores

	BC95-JTT-HN		*KUMB90-JTT-HN*		*KUMB90-AJTT-HN*		*LSE-AJTT-GTC*	
	CD	*TLG*	*CD*	*TLG*	*CD*	*TLG**	*CD*	*TLG*
EVO	0.684	0.732	0.867	0.887	0.734	0.639	0.545	0.535
AGD	0.888	0.868*	0.882	0.888	0.961	0.953	0.744	0.760
CIT	0.873	0.882	0.949	0.962	0.952	0.874	0.721	0.725
END	0.833	0.852	0.916	0.921	0.934	0.891	0.729	0.754
ENX	0.733	0.723	0.923	0.926	0.873	0.880	0.668	0.684
ERG	0.670	0.765	0.880	0.914	0.686	0.918	0.623	0.661
ESS	0.751	0.804	0.896	0.923	0.806	1.000	0.716	0.790*
JEN	0.846	0.830	0.924	0.940	0.879	0.895	0.733	0.641
PCR	0.901	0.899	0.981	0.981	0.994	1.000	0.973	1.000
SAP	0.771	0.805	0.982	0.990	0.802	0.913	0.785	0.821
AND	0.846	0.829	0.944	0.949	0.850	0.902	0.769	0.754
TND	0.877	0.897	0.956	0.960	0.834	0.765	0.814	0.798
UED	0.947	0.928	0.969	0.973	1.000	0.999	1.000	0.905
AVG	0.817	0.832	0.928	0.936	0.870	0.895	0.755	0.756

* The DNSP or model is affected by excessive MVs.

Tables 6.2 and 6.3 show that, on average, the BRFM-implied 2023 efficiency scores are higher than the time-varying 2023 scores for the BC95-JTT-HN model. For the Kumb90-AJTT-HN, the BRFM-implied scores are slightly lower. For the LSE-AJTT-GTC model the BRFM-implied scores are higher under the CD specification but lower under the TLG specification. For the Kumb90-JTT-HN results are, on average, very similar across both approaches.

Table 6.4 shows that the correlation between the BRFM-implied and model efficiency scores in 2023 is very strong for the BC95-JTT-HN, Kumb90-AJTT-HN and LSE-AJTT-GTC specifications, while the Kumb90-JTT-HN model shows a moderate correlation. In contrast, the correlations for the change in efficiency from the mid-period to 2023 are very strong across all models. When comparing the BRFM-implied efficiency scores with Opex MPFP in 2023, the correlation is moderate for the SFA models and strong for the LSE model.

Table 6.4 Correlation Coefficients between Efficiency Scores

	BC95-JTT-HN		*KUMB90-JTT-HN*		*KUMB90-AJTT-HN*		*LSE-AJTT-GTC*	
	CD	*TLG*	*CD*	*TLG*	*CD*	*TLG*	*CD*	*TLG*
2023 BRFM	0.948	0.896	0.563	0.511	0.903	0.919	0.888	0.866
Eff. Change	0.995	0.990	0.879	0.926	0.919	0.950	0.897	0.873
2023 OPFP	0.597	0.608	0.662	0.585	0.616	0.561	0.722	0.846

These results provide additional confidence in the reliability of the efficiency estimates from the selected models.

6.4 Concluding Comments

Unlike the standard current models, the time-varying models presented in this report allow for the decomposition of productivity changes into changes in efficiency, technical change, and other factors. They can also enable the efficient base-year opex to be calculated directly if the base year is the same as the last year of the sample period. This can remove the need to rely on the BRFM framework, or reduce its role to bridging between the last year of the sample and the base-year.

A comparison of the estimated efficiency trends produced by several feasible time-varying inefficiency models indicates that a degree of uncertainty exists in the efficiency scores estimated by these models at any point in time. This is due to differences in the time-patterns of efficiency estimated using different models. We suggest that extending the current approach to include two different SFA models alongside an LSE model (rather than only one SFA and one LSE model, as currently applied) may improve the reliability of the estimated efficiency scores.

We also show that the BRFM provides a means to mitigate such uncertainties and strengthen the robustness of the results. The BRFM can still be applied to the period-average efficiency scores produced by time-varying inefficiency models, and an estimate of the efficiency score at the end of the sample period can be derived from it. This provides additional estimates of the period-end efficiency score, which are arguably not constrained by restrictive assumptions relating to the efficiency time profile. Hence, the period-end efficiency score of each DNSP can be obtained by averaging over more estimates.

Overall, while these findings highlight the potential value of incorporating time-varying inefficiency models into the opex assessment framework, they remain preliminary. We reiterate that these are initial views and further analysis is required.

Appendix A: Development of the Current Opex Modelling Approach

This section revisits the reasoning behind the current specification of the opex benchmarking models, original consultant Economic Insights' reasons for choosing a data sample with two overseas jurisdictions, and the use of jurisdictional-specific effects and their interpretation.

A.1 Background

In 2011 and 2012, the Australian Energy Market Commission (AEMC) completed reviews into the regulatory use of total factor productivity (TFP) measurement and, more generally, energy network regulation (AEMC, 2011; 2012). Among the outcomes of these reviews were:

- obligations on Australian electricity DNSPs to provide detailed data to the AER to support benchmarking, and
- obligations on the AER to produce an annual electricity network benchmarking reports and to have regard to the benchmarking outcomes when making regulatory decisions on DNSPs' allowable costs.

As part of developing its *Expenditure forecast assessment guidelines for electricity distribution and transmission*, the AER consulted on its benchmarking method in late 2012 and 2013. This included workshops with stakeholders. On behalf of the AER, Economic Insights prepared consultation papers on the specification of outputs, inputs, and operating environment factors (OEFs) for calculating total factor productivity (Economic Insights, 2013a, 2013b, 2013c).

The output and input specifications developed through this consultation were informed by previous benchmarking studies conducted by Economic Insights and Pacific Economics Group (PEG) in Australia, New Zealand and Ontario. The AER's benchmarking was first applied in 2014 for the NSW and ACT utilities.

A.2 Outputs Specification

Economic Insights (2013a) initially proposed to define outputs as including:

- *customer numbers* to represent the "relatively fixed services the DNSP supplies. These are activities the DNSP has to undertake regardless of the level of energy delivered and include connection related infrastructure (eg having more residential customers may require more local distribution transformers and low voltage mains), customer calls, etc" (Economic Insights, 2013a: iii).
- *system capacity* which is "approximated by the product of circuit line length and the total capacity of distribution level transformers" (Economic Insights, 2013a: iii). However, Economic Insights (2014: 10–11) stated that a disadvantage of the combined system capacity measure was that its multiplicative nature potentially advantaged large

DNSPs. Another criticism is that this measure places insufficient emphasis on demand-side outcomes. It decided to replace system capacity with two outputs:

  - circuit length, which indicates the distances over which power is carried, and
  - peak supply capacity measured by ratcheted maximum demand (RMD). This measure aims to give DNSPs credit for the capacity required to meet historical peak demand, rather than for any excess capacity installed beyond actual user needs.

- *energy deliveries/throughput*, although less important as a cost driver "reflects the main output customers are charged for and maintains consistency with earlier economic benchmarking studies, nearly all of which have included throughput as an output" (Economic Insights, 2013a: iv).
- *duration of customer interruptions* (ie, total customer minutes off-supply), as a negative output to measure reliability performance.

This output specification can be compared to the PEG (2008) study, prepared for the Ontario Energy Board (OEB), which used as outputs: the number of end-user customers, the total delivery volume, and the total circuit km of distribution line. The Economics Insights (2009) study for New Zealand's Commerce Commission used as outputs: customer connections, energy throughput, and two measures of system capacity ("system line capacity" and "overall system capacity"). PEG (2013) used as outputs: customer numbers, total kWh deliveries, "system capacity peak demand" (ie, ratcheted maximum demand) and average line length. The output specification adopted by Economic Insights in 2014 for Australian DNSP benchmarking is similar to the PEG 2013 specification, except it adds the reliability output.

A.3 Input specification

Economic Insights (2013a) initially proposed to define inputs as:

- Network services opex deflated by and opex price index. Opex is adjusted to remove accounting items not reflecting current year input use.
- Capital inputs were initially of three kinds: (i) overhead lines, measured in MVA–kilometres (sum of kilometres of line by voltage class multiplied by the weighted average MVA rating for each class) (ii) underground cables, measured in MVA-km, and (iii) transformers and other capital, measured by MVA of transformers. However, capital input specification was revised following the 2013 consultation process, as discussed below.

This input specification is the same as that used by Economic Insights (2009). PEG (2013) used two inputs: deflated opex and capital inputs. The latter was measured by firstly estimating a base-year capital stock by deflating the value of assets and then using the perpetual inventory method for subsequent years. This involved adding the gross additions of fixed assets deflated by asset prices and allowing for an economic depreciation rate. Economic Insights preferred to measure capital inputs using physical capital measures rather than deflated accounting-based measures of the written-down value of capital assets. Physical measures were considered as more robust data because accounting-based measures can be inconsistent for a range of reasons. Physical measures also better reflect the time-profile of asset services capability than do accounting depreciation profiles.

Economic Insights (2014) changed the specification of capital in two important respects:

- Overhead lines were separated into overhead subtransmission lines (lines of 33 kV and higher) and distribution lines (lines of less than 33 kV). Underground cables were similarly separated into subtransmission and distribution cables. This disaggregation improves accuracy of input measurement because it "allows for the fact that subtransmission lines and cables account for a high proportion of most DNSPs' total overhead MVAkms and total underground MVAkms, respectively, but a much smaller proportion of overhead lines and underground cables asset values and annual user costs" (Economic Insights, 2014: 13). Since input weights are based on cost shares, separating subtransmission from distribution ensures that the former has smaller, more cost-reflective influence on the input index, compared to the dominant influence it will have when they are aggregated.
- The Transformers and Other capital input was redefined to exclude the *first stage zone substation transformer capacity* of those systems that have two stage transformation from the higher voltages (mainly in NSW and Qld). "The MVA quantity of Transformers and other capital inputs is now the sum of single stage transformation at the zone substation level, the second stage of two stage transformation at the zone substation level, and distribution transformer capacity. … The purpose of this modification is to allow for the more complex system structures and different transmission/distribution boundaries some states have inherited relative to others" (Economic Insights, 2014: 13).

A.4 Prices and values

Constructing output and input indexes also requires information on the prices or values of the individual inputs and outputs in order to calculate the index weights.

The opex price index, used to deflate the value of opex, is a weighted average of the ABS Wage Price Index (WPI) and five Producer Price Indexes covering business, computing, secretarial, legal and accounting, and public relations services. This index and the weights used followed earlier research by PEG (2006), which aligned nine major components of opex with related price indexes, with weights based on the cost shares. Economic Insights considered two alternative ABS labour price indexes, Average weekly ordinary time earnings

(AWOTE) and the WPI, ultimately adopting the latter which is unaffected by compositional changes in the industry labour force. This implies that the implicit labour quantity is 'quality adjusted'. The opex price indexes used for New Zealand and Ontario DNSPs are constructed from analogous sub-indexes as developed by Economic Insights and PEG.

The annual user cost of capital inputs is taken to be the return on capital and return of capital for each of the capital input components "calculated in a way which approximates the corresponding building blocks calculations". The input price for each of the three capital components is then derived by dividing their annual user cost by their respective physical quantity proxy. By using an exogenous user cost of capital covering the return on and return of capital it ensures consistency with other building blocks calculations" (2013a: v).

A.5 Operating environment factors

Economic Insights (2013b, 2013c) identified various likely OEFs:

(a) Network density and network dispersion measures;

(b) The extent of undergrounding. This is included as a conditioning variable in the econometric opex model;

(c) Climatic factors

(d) Terrain indicators

(e) Vegetation encroachment or management measures;

(f) Distribution network complexity. This is accounted for by excluding the first stage of two stage transformation at the zone substation level.

The current AER benchmarking framework incorporates customer density effects by including line length and customer numbers explicitly as outputs. "This means that DNSPs who have low customer density, for instance, receive credit for their longer line lengths whereas this would not be the case if output was measured by only one output such as throughput" (Economic Insights, 2013c: 18–19). While some studies have used DNSPs' service areas to measure density, a better measure of the areas actually supplied is the areas within a certain distance of distribution lines. Hence, circuit length is the preferred measure of the spatial dimension of the network.

Differences in energy density are sometimes captured by including customer mix as an OEF, or by including a high-level disaggregation of customers or energy throughput outputs. These approaches do not capture all energy density differences; e.g., the effects of climatic or affluence differences. The AER output specification, including both customer numbers and energy delivered, also partly captures energy density differences.

In the AER framework, some climatic effects are accounted for by *ex post* adjustments of efficiency scores. These include exposure to termite infestation (which mostly affects Ergon and Essential) and exposure to cyclones (affecting only Ergon). These were quantified by Sapere and Merz (2018).

Economic Insights (2013c: 20) suggested that with reliability included as an output it is "important to either include climatic effects as an operating environment factor or to exclude severe weather related impacts from reliability measures and associated restoration costs from the input side.".

Economic Insights (2013c: 20) observed that the "terrain a DNSP has to operate in can significantly affect its costs. Hilly areas are typically more expensive to service than flat areas and forested areas will also incur higher vegetation management costs. … However, all DNSPs face a range of terrain conditions over their service areas and it is often difficult to average these in an easily quantifiable way. There is also a dearth of indicators for terrain conditions that lend themselves to use in economic benchmarking". Economic Insights (2013b) suggested three metrics for further consideration:

- *Bushfire risk*: Number of days over 50 per cent of service area subject to severe or higher bushfire danger rating (NSW rating standard);
- *Rural proportion*: Percentage of line length classified as rural;
- *Vegetation encroachment*: Percentage of route line length requiring active vegetation management.

The AER currently makes an *ex post* adjustment to efficiency scores for bushfire risk, specifically focussed on the higher compliance requirements of Victorian DNSPs, and for division of responsibility for vegetation management. Sapere-Merz identified vegetation management as a material OEF, but they were unable to quantify a vegetation OEF at that time. They also noted that network topology was likely to be a material OEF but was not assessed in their study.

A.6 Benchmarking methodologies

In assessing alternative benchmarking methods, Economic Insights decided not to use data envelopment analysis (DEA). Although DEA has the advantage of not requiring "the specification of a functional form for the frontier or a distributional form for the inefficiency effects … it has the disadvantage that it is deterministic in nature and hence the efficiency scores obtained can be quite sensitive to the effects of random factors and data errors" (Economic Insights, 2014: 7).

Productivity Index Number (PIN) methods are also deterministic and hence affected by data noise. However, this is mitigated to some extent by using firm-level averages over the sample period. A key advantage of PINs is they can be calculated using very small data sets, are transparent, reproducible and robust. Two PIN methods have been used:

- For making comparisons of TFP and partial factor productivity (PFP) between DNSPs, the Multilateral Törnqvist index method of Caves, Christensen and Diewert (1982) was used.
- For examining time series patterns of PFP and PFP for individual DNSPs or the industry, the Törnqvist index was initially used (Economic Insights, 2017a). However, in 2020 Economic Insights switched to using the Multilateral Törnqvist index method for time series comparisons because of the large variability in the reliability output. "The standard time-series indexes are less able to accurately capture the impact of these large percentage changes because they do not satisfy the transitivity property. This can lead to the standard time-series indexes being subject to some degree of 'drifting' higher or lower after a spike in the [reliability] variable" (Economic Insights, 2020: 6).

Two econometric methods for estimating opex cost functions were adopted:

- "least squares econometrics (LSE)" which is a version of corrected ordinary least squares where the inefficiency effects are estimated via firm-specific fixed effects. Specifically, the difference between the firm-specific fixed effect and the minimum fixed effect in the sample.[60]
- stochastic frontier analysis (SFA), in which the inefficiency effects are estimated via a one-sided random effect.

Both methods have the advantage of allow for random noise in the data. They also allow for variable returns to scale and OEFs can be included. Each has its own disadvantages:

- The LSE model does not directly estimate the frontier function. An average cost function is estimated and it is assumed that the frontier function "is a parallel shift (in logarithms) of the average function" (Economic Insights, 2014: 8).
- SFA requires a particular distributional form to be assumed for the inefficiency effects, and compared to LSE it "tends to be more data hungry and hence more unstable when applied to small data sets" (Economic Insights, 2014: 8).

While these models performed well up to and including the 2021 data, the inclusion of data from 2022 onwards has introduced some issues, most notably, an increase in the frequency of monotonicity violations and convergence problems in the SFATLG models.

[60] Both "LSE" and "SFA" are SFA approaches using different estimators: a variant of corrected ordinary least squares (COLS) and SFA using maximum likelihood estimation (MLE). The key distinction is that LSE does not require parametric assumptions for the error terms, while SFA, which uses an MLE estimator, does (e.g., truncated normal for inefficiency and normal for noise).

A.7 Data Sample

Economic Insights (2014: 30) observed that, at the time of the first Economic Benchmarking study, the Australian dataset—13 DNSPs over eight years, totalling 104 observations—did not provide enough variation to estimate even a simple opex cost-function model. To obtain robust and reliable econometric results, additional cross-sectional observations were required.

New Zealand was the most obvious choice, with a database structured similarly to the AER's benchmarking RIN and comparable variable coverage. The Canadian province of Ontario also has a long history of DNSP productivity analysis. Despite climate differences, including Ontario DNSPs adds valuable additional observations to the dataset. US data were considered but ultimately not used because the available datasets are patchy, use different definitions from those in Australia, New Zealand and Ontario, and often cover vertically integrated utilities, raising significant cost-allocation issues.

All financial series from New Zealand and Canada were converted to Australian dollars using OECD GDP purchasing power parities (PPPs). Given potential differences in opex definitions across countries, country dummy variables for New Zealand and Ontario were included. Including country dummies allows the cost function to control for cross-country differences such as any conversion discrepancies not fully addressed by the PPP adjustment, opex coverage, systematic operating environment effects.

Initially, the econometric models were estimated using the period from 2006 to the latest year of data available. In 2019, Economic Insights began to estimate these models using two sample periods: the full period from 2006 to the latest year, and the shorter period from 2012 to the latest year. This reflected the expectation that changes in the regulatory framework for assessing opex, and particularly the introduction of benchmarking, may significantly affect the efficiency scores of Australian DNSPs. Therefore, averaging the efficiency scores obtained using the sample of more recent data and those obtained using the longer sample effectively increases the weight given to more recent data when assessing DNSPs' cost efficiency.

In 2023, the definition of opex for benchmarking purposes was revised following the AER's final decision on differences in capitalisation practices. It was determined that these differences should be addressed by allocating 100% of corporate overheads expenditure to the Opex series used for benchmarking (AER, 2023).

Appendix B: Correlations of Output Elasticities

This appendix analyses the correlation between total output cost elasticity and DNSP size, measured by customer numbers and circuit length, across both the standard and time-varying models, focusing only on the Australia sample. When comparing DNSPs against their peers (that is, in a between-firm context), a positive correlation implies that larger DNSPs tend to have higher output cost elasticities, while a negative correlation indicates that larger DNSPs tend to have lower elasticities. A positive correlation is expected. Smaller DNSPs are more likely to face economies of scale (indicated by a total output cost elasticity substantially less than 1). Larger DNSPs are assumed to have already captured most available scale economies and thus operate closer to constant returns to scale, with output cost elasticities approaching 1. Some of the largest DNSPs may face decreasing returns to scale (indicated by a total output cost elasticity greater than 1).

Table B.1 presents the correlations between size and total elasticity for the standard and time varying models in this report. Figures B.1 to B.6 show these correlations for the SFATLG models, and Figures B.7 to B.11 show the corresponding results for the LSETLG models.

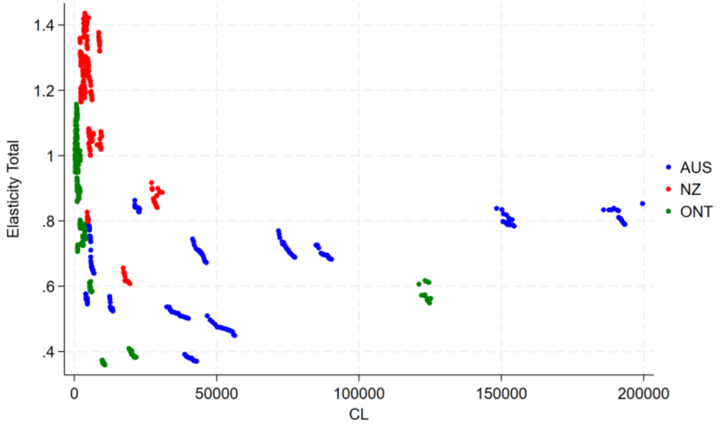
Table B.1 Correlation between DNSP Size and Total Elasticity

	Long Period		*Short Period*	
	Circuit Length	*Customer No.*	*Circuit Length*	*Customer No.*
Standard SFATLG	0.419	-0.643	0.256	-0.770
BC95- JTT-HN	0.745	0.280	0.537	0.341
Kumb90 JTT-HN	0.373	-0.684	0.033	-0.684
Kumb90 JTT-HN-GTC	0.557	-0.499	-0.296	-0.963
Kumb90 AJTT-HN	0.812	0.376	-0.078	0.589
Kumb90 AJTT-HN-GTC	0.620	0.632	-0.713	-0.766
Standard LSETLG	0.844	0.303	0.703	0.071
LSE-ADTT	0.816	0.628	0.701	0.697
LSE-ADTT-GTC	0.812	0.612	0.710	0.673
LSE-AJTT	0.752	0.701	0.661	0.579
LSE-AJTT-GTC	0.708	0.655	0.763	0.594

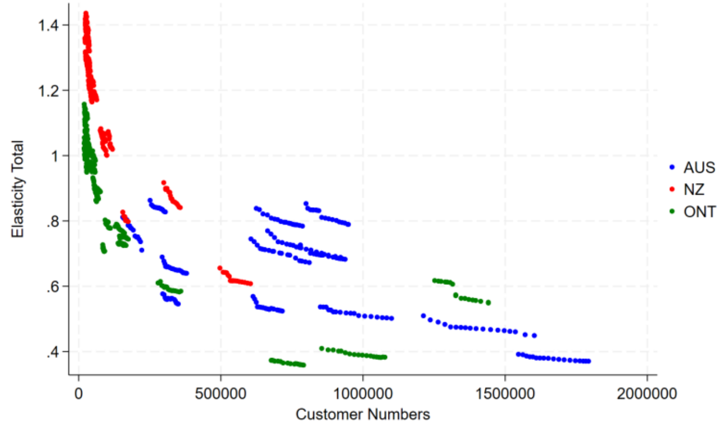
For the SFA models in the long sample period, the only specification to consistently produce positive correlations between total output elasticity and size measures for the Australian sample are BC95-JTT-HN model, Kumb90-AJTT-HN and Kumb90-AJTT-HN-GTC. The standard model, Kumb90-JTT-HN, and Kumb90-JTT-HN-GTC all produced positive correlations only with circuit length in the long period.

Considering the LSE models in the long sample period, the correlation between total output elasticity and DNSP size is consistently positive, and for scale measured by customer numbers, is consistently stronger in the time-varying inefficiency LSE models than in the standard LSE specification. This indicates a clearer positive relationship between firm size and cost efficiency.

Figure B.1 Elasticity Total x Size: Standard ABR24 SFA Models

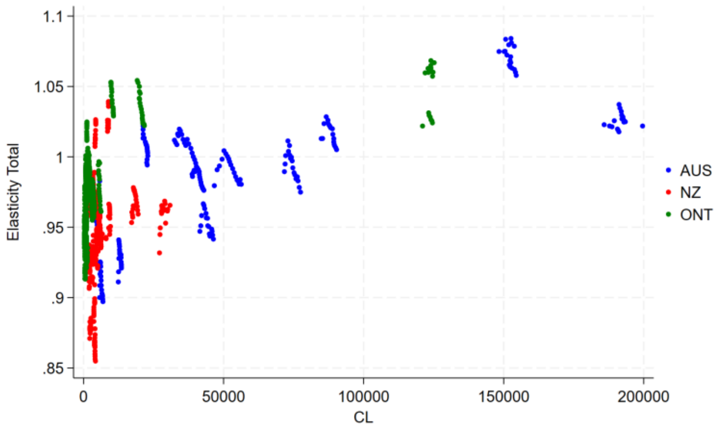


Circuit Length-SFATLG Long Period

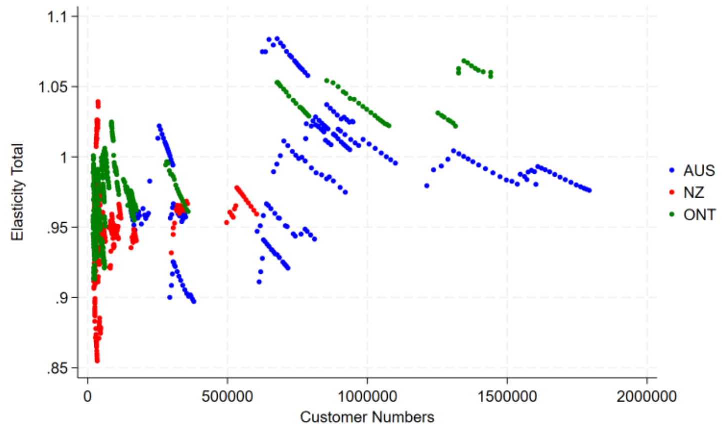


Customer Number-SFATLG Long Period

Figure B.2 Elasticity Total x Size: BC95-JTT-HN Models

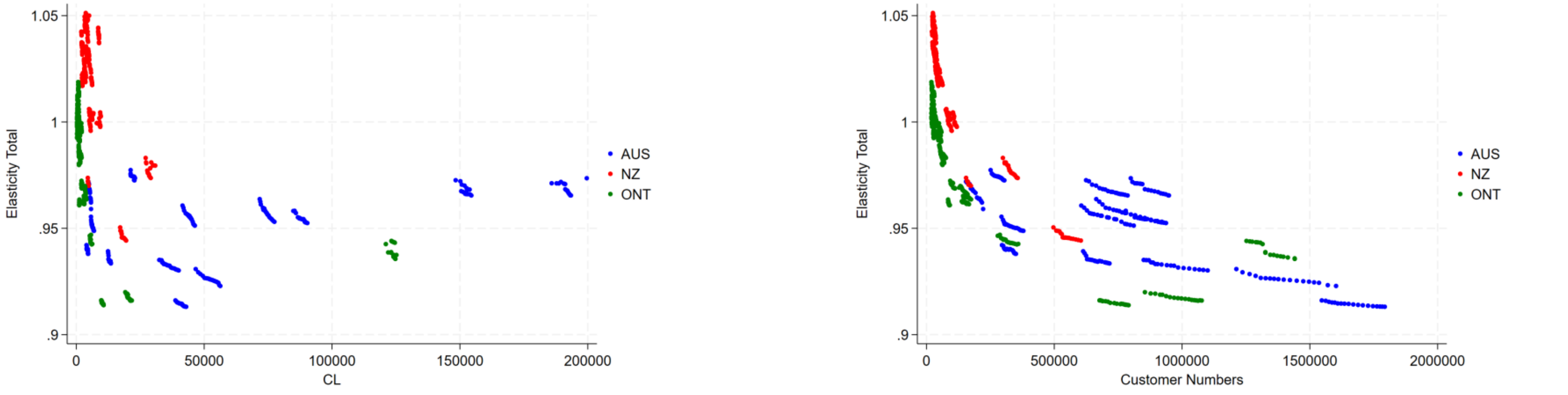


Circuit Length-SFATLG Long Period



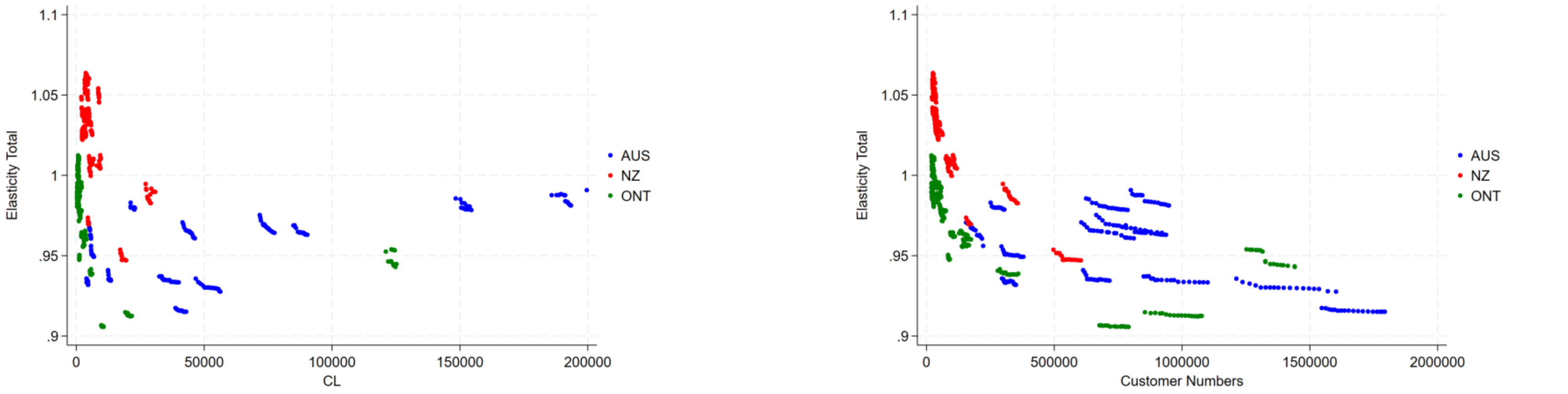
Customer Number-SFATLG Long Period

Figure B.3 Elasticity Total x Size: Kumb90-JTT-HN Models



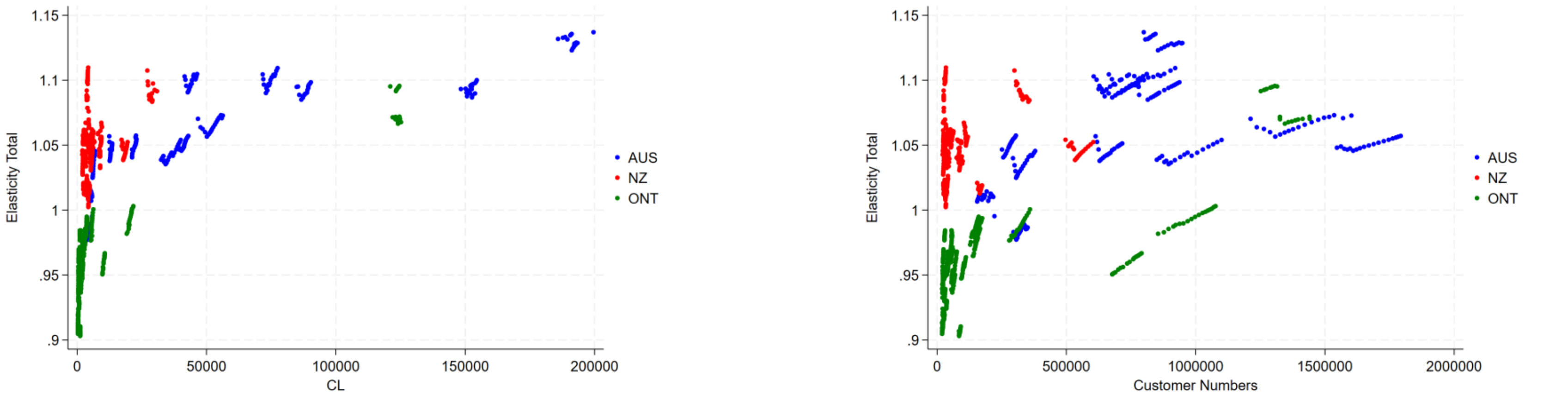
Circuit Length-SFATLG Long Period	*Customer Number-SFATLG Long Period*

Figure B.4 Elasticity Total x Size: Kumb90-JTT-HN-GTC Models



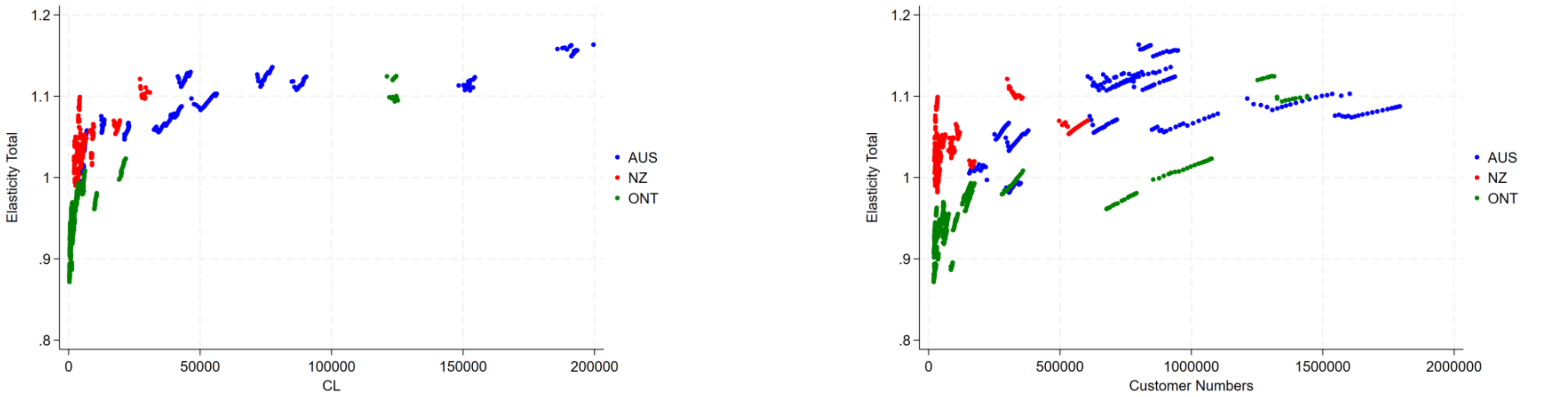
Circuit Length-SFATLG Long Period	*Customer Number-SFATLG Long Period*

Figure B.5 Elasticity Total x Size: Kumb90-AJTT-HN Models



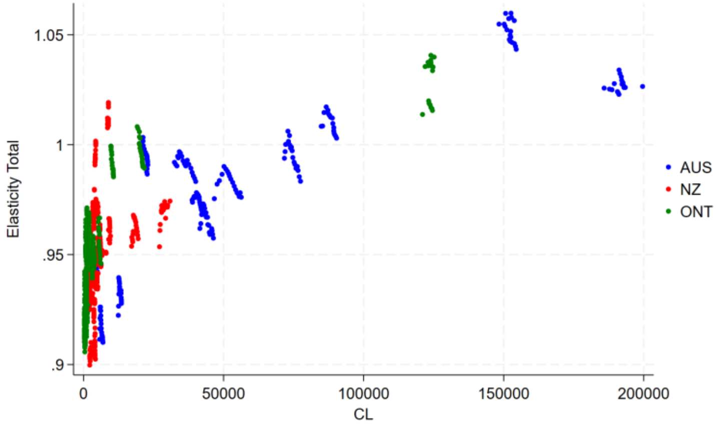
Circuit Length-SFATLG Long Period | *Customer Number-SFATLG Long Period*

Figure B.6 Elasticity Total x Size: Kumb90-AJTT-HN-GTC Models

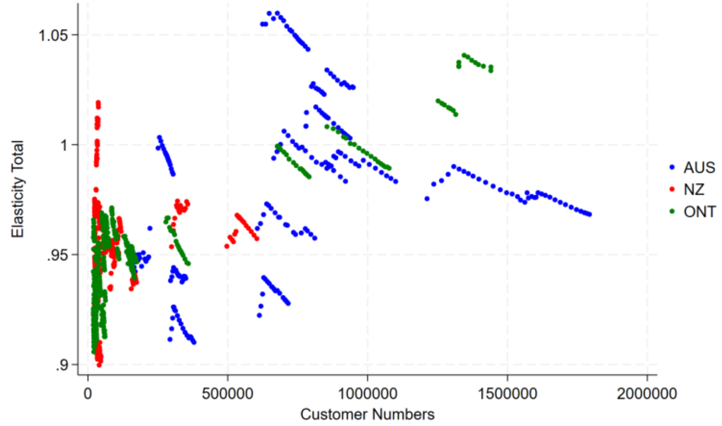


Circuit Length-SFATLG Long Period | *Customer Number-SFATLG Long Period*

Figure B.7 Elasticity Total x Size: Standard ABR24 LSE Models

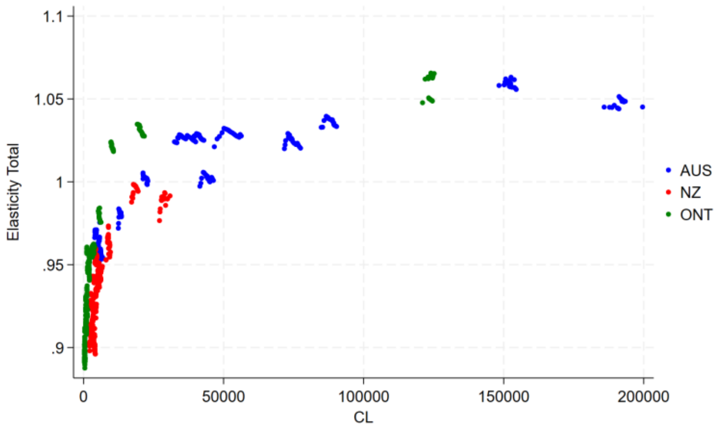


Circuit Length-LSETLG Long Period

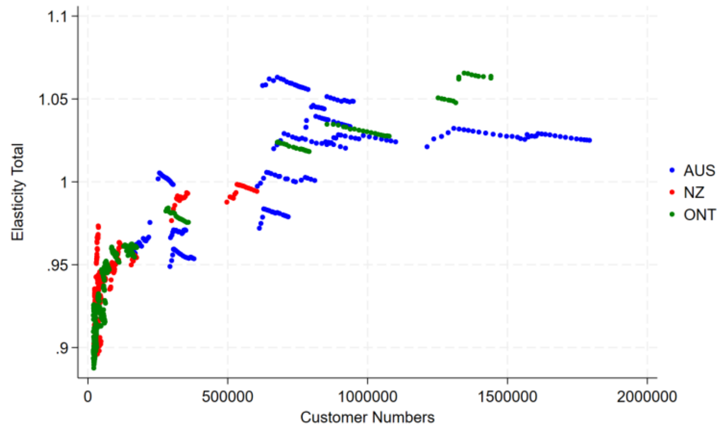


Customer Number-LSETLG Long Period

Figure B.8 Elasticity Total x Size: LSE-ADTT Models

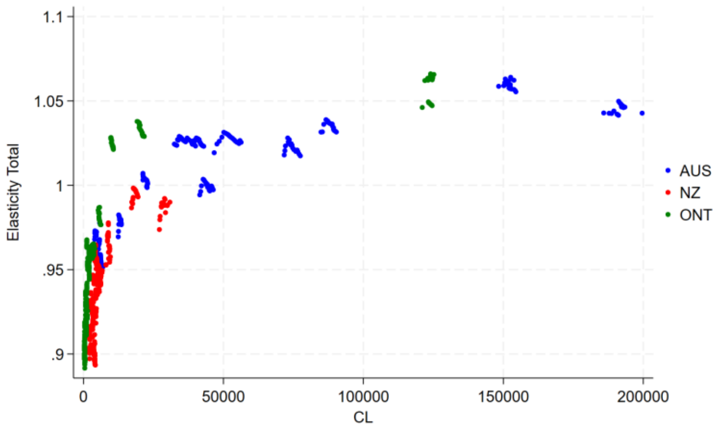


Circuit Length-LSETLG Long Period

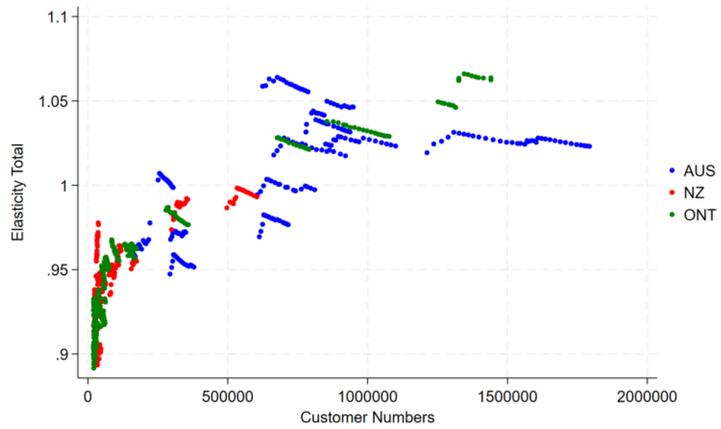


Customer Number-LSETLG Long Period

Figure B.9 Elasticity Total x Size: LSE-ADTT-GTC Models

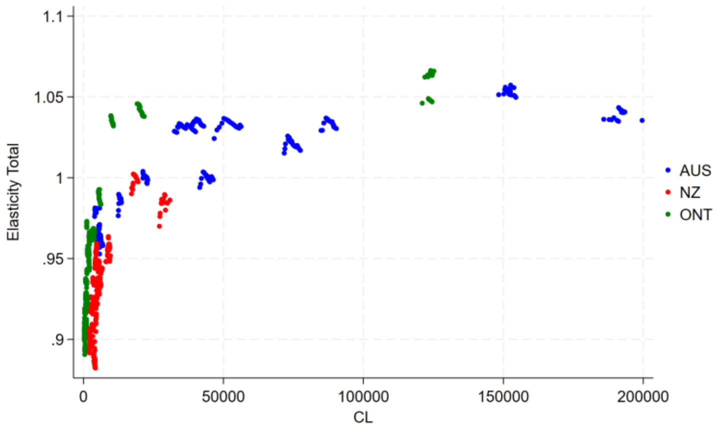


Circuit Length-LSETLG Long Period

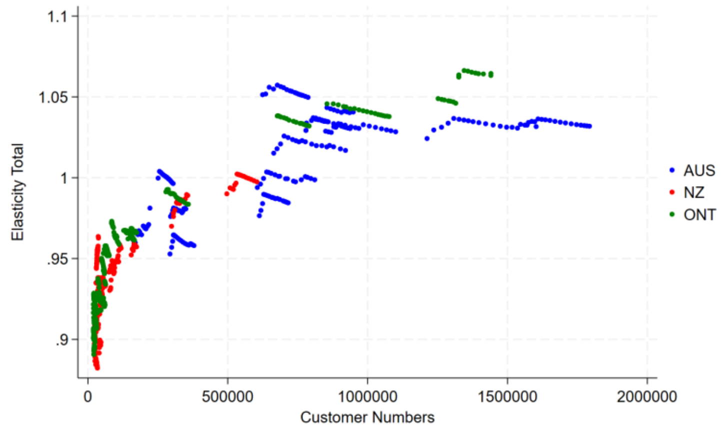


Customer Number-LSETLG Long Period

Figure B.10 Elasticity Total x Size: LSE-AJTT Models

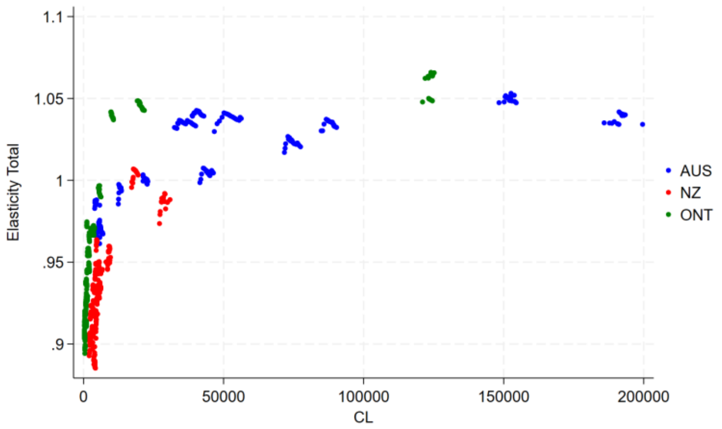


Circuit Length-LSETLG Long Period

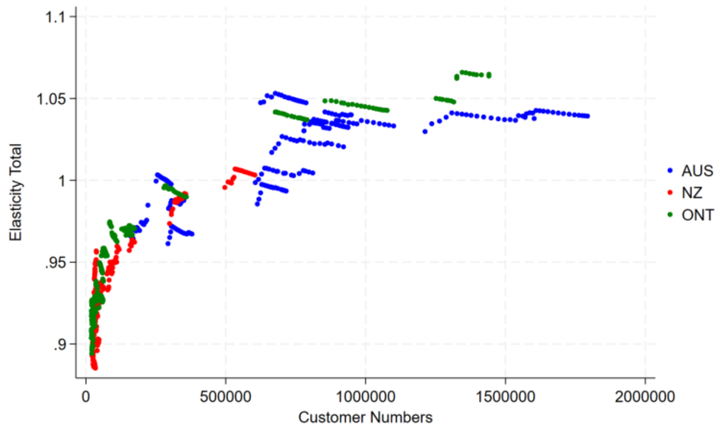


Customer Number-LSETLG Long Period

Figure B.11 Elasticity Total x Size: LSE-AJTT-GTC Models



Circuit Length-LSETLG Long Period



Customer Number-LSETLG Long Period

Appendix C: Correlations between Efficiency Scores

Table C.1 presents the correlation between the average efficiency scores for individual DNSPs from each time-varying model and the corresponding standard model. While no particular result is expected, since the aim of time-varying models is not necessarily to replicate the efficiency scores of the standard specification, this comparison provides insight into how the introduction of time-varying inefficiency affects the previously estimated scores.

Table C.1 Correlation of Average Efficiency Scores with respective Standard Models

	Long Period		Short Period	
	SFACD	*SFATLG*	*SFACD*	*SFATLG*
BC95- JTT-HN	0.975	0.705	0.937	0.202
Kumb90 JTT-HN	0.962	0.745	0.991	0.335
Kumb90 JTT-HN-GTC	0.961	0.725	0.986	0.304
Kumb90 AJTT-HN	0.967	0.674	0.909	0.154
Kumb90 AJTT-HN-GTC	0.955	0.621	0.935	-0.317
Kumb90-AJTTnz-HN	0.937	0.627	0.958	0.377
Kumb90-AJTTnz-HN-GTC	0.903	0.481	0.961	0.259

Interestingly, the correlations between the time-varying and standard model efficiency scores are very high for the Cobb-Douglas specifications (all above 0.9). In contrast, the correlations are notably lower for the Translog specifications, particularly in the short period. Note that, the standard long-period SFATLG specification is affected by monotonicity violations in 79.5 per cent of the Australian sample and is therefore excluded from the efficiency score average in the benchmarking report. Therefore, weaker correlation of efficiency scores with that model cannot be viewed as a shortcoming.

Similarly, the short period standard SFATLG model is also deficient, since it failed to converge, and is also excluded from the efficiency score average. Hence, divergence in efficiency scores between the time-varying and standard models is greatest in the specifications known to be problematic. This supports the view that the time-varying models may be correcting deficiencies in the efficiency scores of those standard models.

Table C.2 presents the correlations between the average efficiency scores from each time-varying inefficiency LSE model and the standard model.

Table C.2 Correlation of Average Efficiency Scores with respective Standard Models

	Long Period		Short Period	
	LSECD	*LSETLG*	*LSECD*	*LSETLG*
LSE-ADTT	1.000	0.989	1.000	0.938
LSE-ADTT-GTC	0.999	0.986	0.999	0.939
LSE-AJTT	0.998	0.980	0.999	0.953
LSE-AJTT-GTC	0.997	0.966	0.999	0.901

The correlations between the average efficiency scores from the time-varying LSE alternatives and those from the standard LSE models are very high, exceeding 0.9 in all cases, indicating a strong degree of consistency. In some instances, such as the Cobb-Douglas LSE-ADTT models, the correlation reaches 1.0, indicating no difference in the efficiency scores. Once again, the short-period LSE-TLG specification shows the lowest correlations (although still above 0.9). The standard short-period LSE-TLG model is affected by 48.7 per cent MVs in the Australian sample.

References

Aigner DJ, Lovell CAK and Schmidt P (1977) Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics* 6: 21–37.

Alvarez A, Amsler C, Orea L, et al. (2006) Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. *Journal of Productivity Analysis* 25: 201–212.

Australian Energy Market Commission (AEMC) (2011) *Review into the use of total factor productivity for the determination of prices and revenues: Final report*. 30 June.

Australian Energy Market Commission (AEMC) (2012) *Economic Regulation of Network Service Providers, and Price and Revenue Regulation of Gas Services: Rule Determination*. 29 November.

Australian Energy Regulator (AER) (2023) *How the AER will assess the impact of capitalisation differences on our benchmarking: Final Guidance note*. May.

Baltagi BH and Griffin JM (1988) A General Index of Technical Change. *Journal of Political Economy* 96(1).

Battese G and Coelli T (1988) Prediction of firm-level technical efficiencies with a generalized frontier function and panel data. *Journal of Econometrics* 38: 387–399.

Battese GE and Coelli TJ (1992) Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. In: *International Applications of Productivity and Efficiency Analysis*. Springer, pp. 149–165.

Battese GE and Coelli TJ (1995) A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics* 20(2): 325–332.

Belotti F and Ilardi G (2017) *Consistent inference in fixed-effects stochastic frontier models*. 1147, Working Paper, October.

Belotti F, Daidone S, Ilardi G, et al. (2012) Stochastic frontier analysis using Stata. *Stata Journal*: 1–39.

Caudill SB, Ford JM and Gropper DM (1995) Frontier estimation and firm-specific inefficiency measures. *Journal of Business & Economic Statistics*. Epub ahead of print 1995.

Caves DW, Christensen LR and Diewert WE (1982) Multilateral comparisons of output, input, and productivity using superlative index numbers. *Economic Journal* 92: 73–86.

Chen Y-Y, Schmidt P and Wang H-J (2014) Consistent estimation of the fixed effects stochastic frontier model. *Journal of Econometrics* 181(2): 65–76.

Coelli T, Estache A, Perelman S, et al. (2003) *A Primer on Efficiency Measurement for Utilities and Transport Regulators*. WBI Development Studies.

Colombi R, Kumbhakar SC, Martini G, et al. (2014) Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. *Journal of Productivity Analysis* 42(2): 123–136.

Cornwell C, Schmidt P and Sickles RC (1990) Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics* 46(1–2): 185–200.

Cuesta RA (2000) A Production Model with Firm-Specific Temporal Variation in Technical Inefficiency: With Application to Spanish Dairy Farms. *Journal of Productivity Analysis* 13: 139–158.

Economic Insights (2009) *Electricity Distribution Industry Productivity Analysis: 1996–2008*. Prepared for Commerce Commission by Denis Lawrence, Erwin Diewert, John Fallon and John Kain, September.

Economic Insights (2013a) *Economic Benchmarking of Electricity Network Service Providers*. Report prepared for Australian Energy Regulator by Denis Lawrence and John Kain, June.

Economic Insights (2013b) *Measurement of Outputs and Operating Environment Factors for Economic Benchmarking of Electricity Distribution Network Service Providers*. 16 April. Briefing Notes prepared by Denis Lawrence and John Kain for Australian Energy Regulator.

Economic Insights (2013c) *Outputs and Operating Environment Factors to be Used in the Economic Benchmarking of Electricity Distribution Network Service Providers*. Briefing Notes prepared by Denis Lawrence and John Kain for Australian Energy Regulator, 20 February.

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

Economic Insights (2015) *Response to Consultants' Reports on Economic Benchmarking of Electricity DNSPs*. Report prepared by Denis Lawrence, Tim Coelli and John Kain for the Australian Energy Regulator, 22 April.

Economic Insights (2017a) *Economic Benchmarking Results for the Australian Energy Regulator's 2017 DNSP Benchmarking Report*. Prepared for the Australian Energy Regulator by Denis Lawrence, Tim Coelli and John Kain, October.

Economic Insights (2017b) *Topics in efficiency benchmarking of energy networks: Choosing the model and explaining the results*. Report prepared by Denis Lawrence, John Fallon, Michael Cunningham, Valentin Zelenyuk, Joseph Hirschberg for The Netherlands Authority for Consumers and Markets, December. Available at: https://www.acm.nl/sites/default/files/documents/2020-08/economic-insights-topics-in-efficiency-benchmarking-of-energy-networks-choosing-the-model-and-explaining-the-results.pdf.

Economic Insights (2019) *Economic Benchmarking Results for the Australian Energy Regulator's 2019 DNSP Annual Benchmarking Report*. Report prepared by Denis Lawrence, Tim Coelli and John Kain for Australian Energy Regulator, 16 October.

Economic Insights (2020) *Economic Benchmarking Results for the Australian Energy Regulator's 2020 DNSP Annual Benchmarking Report*. Prepared for Australian Energy Regulator by Denis Lawrence, Tim Coelli and John Kain, October.

Frontier Economics (2025) *Memo: Response to AER's Phase 1 consultation on potential mis-specification issues – electricity distribution opex cost functions*. 27 February.

Greene W (2005a) Fixed and Random Effects in Stochastic Frontier Models. *Journal of Productivity Analysis* 23(1): 7–32.

Greene W (2005b) Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics* 126(2): 269–303.

Jorgenson DW (1967) The Theory of Investment Behavior. In: Ferber R (ed.) *Determinants of Investment Behavior*. Nat. Bureau of Econ. Research.

Kumbhakar S, Ghosh S and McGuckin T (1991) A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms. *Journal of Business & Economic Statistics* 9(3): 279–286.

Kumbhakar SC (1990) Production frontiers, panel data, and time-varying technical inefficiency. *Journal of Econometrics* 46(1–2): 201–211.

Kumbhakar SC, Lien G and Hardaker JB (2014) Technical efficiency in competing panel data models: a study of Norwegian grain farming. *Journal of Productivity Analysis* 41(2): 321–337.

Kumbhakar SC, Wang H-J and Horncastle AP (2015) *A Practitioner's Guide to Stochastic Frontier Analysis Using Stata*. Cambridge University Press.

Kumbhakar SC, Parmeter CF and Zelenyuk V (2020a) Stochastic Frontier Analysis: Foundations and Advances I. In: Ray SC, Chambers R, and Kumbhakar S (eds) *Handbook of Production Economics*. Singapore: Springer, pp. 1–40.

Kumbhakar SC, Parmeter CF and Zelenyuk V (2020b) Stochastic Frontier Analysis: Foundations and Advances II. In: Ray SC, Chambers R, and Kumbhakar S (eds) *Handbook of Production Economics*. Singapore: Springer, pp. 1–38.

Meeusen W and van den Broeck J (1977) Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review* 18(2): 435–444.

Nguyen BH, Sickles RC and Zelenyuk V (2022) Efficiency Analysis with Stochastic Frontier Models Using Popular Statistical Softwares. In: Chotikapanich D, Rambaldi AN, and Rohde N (eds) *Advances in Economic Measurement*. Singapore: Springer Nature Singapore, pp. 129–171. Available at: https://link.springer.com/10.1007/978-981-19-2023-3_3 (accessed 24 April 2023).

Pacific Economics Group (PEG) (2006) *TFP Research for Victoria's Power Distribution Industry: 2005 Update*. Report prepared for the Essential Services Commission, November.

Pacific Economics Group (PEG) (2008) *Benchmarking the Costs of Ontario Power Distributors*. 20 March.

Pacific Economics Group (PEG) (2013) *Productivity and Benchmarking Research in Support of Incentive Rate Setting in Ontario : Final Report to the Ontario Energy Board*. 5 November.

Pitt M and Lee L (1981) The measurement and sources of technical inefficiency in the Indonesian weaving industry. *Journal of Development Economics* 9: 43–64.

Quantonomics (2023) *Opex Cost Function Development – Options for Incorporating Time-Varying Measures of DNSP Inefficiency (unpublished)*. Memorandum prepared for the AER by Michael Cunningham, Joe Hirschberg, Alice Giovani, 1 October.

Quantonomics (2024a) *Economic Benchmarking Results for the Australian Energy Regulator's 2024 DNSP Annual Benchmarking Report*. Report prepared for Australian Energy Regulator by Michael Cunningham, Joseph Hirschberg and Alice Giovani, 30 September.

Quantonomics (2024b) *Electricity Distribution Opex Cost Function: Potential Misspecification Issues*. Memorandum prepared forthe AER by Michael Cunningham, Alice Giovani and Joe Hirschberg, 21 November.

Quantonomics (2024c) *Memorandum: Electricity Distribution Opex Cost Function: Potential Misspecification Issues*. Memo prepared for AER by Michael Cunningham, Alice Giovani and Joe Hirschberg, 21 November. Available at: https://www.aer.gov.au/system/files/2024-11/Quantonomics%20-%20Memorandum%20-%20Opex%20Cost%20Function%20Development%20-%20November%202024_0.pdf.

Ritter C and Simar L (1994) *Pitfalls of normal-gamma stochastic frontier models*. STAT Discussion Papers 9417.

Ryan DL and Wales TJ (2000) Imposing local concavity in the translog and generalized Leontief cost functions. *Economics Letters* 67(3): 253–260.

Sapere research group and Merz Consulting (2018) *Independent review of Operating Environment Factors used to adjust efficient operating expenditure for economic benchmarking*. Report for the Australian Energy Regulatory by Simon Orme, Dr. James Swansson, Geoff Glazier, Ben Kearney, Dr. Howard Zhang, August.

Schmidt P and Sickles RC (1984) Production Frontiers and Panel Data. *Journal of Business & Economic Statistics* 2(4): 356–374.

Sickles RC (2005) Panel estimators and the identification of firm-specific efficiency levels in parametric, semiparametric and nonparametric settings. *Journal of Econometrics* 126(2): 305–334.

Sickles RC and Zelenyuk V (2019) *Measurement of Productivity and Efficiency: Theory and Practice*. 1st edn Cambridge University Press. Available at: https://www.cambridge.org/core/product/identifier/9781139565981/type/book (accessed 12 October 2022).

Sickles RC, Song W and Zelenyuk V (2020) Econometric analysis of productivity: theory and implementation in R. In: *Handbook of Statistics*. Elsevier, pp. 267–297.

Wang H and Schmidt P (2002) One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *Journal of Productivity Analysis* 18(2): 289–296.

Wang H-J (2002) Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model. *Journal of Productivity Analysis* 18: 241–253.