

Electricity Distribution Benchmarking Opex Model Development

Report prepared for Australian Energy Regulator

22 September 2026

Michael Cunningham, Alice Giovani, Harsh Nathani, Valentin Zelenyuk, Joe Hirschberg

Executive Summary

This report builds on the earlier consultation report, *Electricity Distribution Benchmarking Opex Model Development* (the Phase 2 report—Quantonomics 2025b). The consultation focuses on the performance of econometric opex cost function models, particularly Translog specifications, and the high frequency of monotonicity violations observed in recent annual benchmarking reports.

The Phase 2 report examined alternative model specifications and estimation techniques to improve the performance of opex cost function models, as well as approaches to incorporating time-varying inefficiency and decomposing time trends into technical change and changes in distribution network service provider (DNSP) opex cost efficiency. This report, referred to as Phase 3, responds to stakeholder comments on the Phase 2 report, investigates further model specifications and refinements, and evaluates their application within the Benchmarking Roll Forward Model (BRFM).

Phase 3 develops more stable models, including refinements to starting values for the stochastic frontier analysis (SFA) models that partly address computational issues faced in the Phase 2 report, and tests alternative specifications that allow greater flexibility in efficiency profiles over time. More parsimonious models, such as Parsimonious Translog (TLGP) are also assessed. Rather than using a fixed model specification, TLGP models use a fixed procedure to sequentially eliminate statistically insignificant second-order output terms in the TLG models. The results generally favour TLGP over the full Translog (TLG), as TLGP retains statistically relevant second-order terms while reducing complexity, multicollinearity and computational burden, with no substantial loss of model fit.

Most tested models successfully address the monotonicity violation issue, allowing a broader set of models to proceed to final evaluation against the criteria established in Phase 2. Overall, the assessment favours the BC95 and LSE models over the Kumb90 models. The preferred models, in order, are:

- BC95-JTT-HN, BC95-JTT-HN-GTC and BC95-JTT-HN-QTT;
- LSE-ADTT-GTC-QTT, LSE-ADTT-GTC and LSE-AJTT-QTT.

Each of these model types performs adequately in both Cobb-Douglas (CD) and TLG functional forms, which are both currently used in the AER's BRFM. They also perform adequately in the TLGP functional form, which is preferred, for the reasons given.

The report also considers the application of the models within the BRFM, including calculating base-year efficient opex under two alternative approaches: ie, the Period-Average approach, which averages efficiency scores over the sample period, and the Period-End approach, which uses the final-year efficiency score and rolls it forward to the base year. The results show that the differences between the two approaches vary considerably across model

families and specifications. The period-end method on average produces higher adjustments than the period-average method, with the largest difference observed for the LSE models, a more moderate difference for the BC95 models, and broadly similar average adjustments for the Kumb90 models. The results also show that, where no efficiency adjustment is required under one approach, no adjustment is generally required under the other, indicating robustness in the treatment of the most efficient DNSPs.

Contents

Executive Summary	i
1 Introduction	1
2 Summary of submissions.....	3
2.1 Main Points in Submissions	3
2.2 Frontier Economics report.....	6
3 Conceptual Issues	15
3.1 Reasonableness of Efficiency Scores and Trends	15
3.2 Output Elasticities of Opex MPFP	19
3.3 Methods of increasing the flexibility of efficiency time trends.....	20
3.4 Cobb-Douglas, Hypothesis Tests and Parsimony	24
3.5 Stability with sample changes	28
3.6 Short sample period & existing specifications	29
4 Preliminary evaluation of estimated models.....	31
4.1 Convergence and Computational Issues.....	32
4.2 Consistency with Economic Theory or Industry Knowledge.....	35
4.3 Goodness-of-fit & Performance on Specification Tests.....	44
4.4 Efficiency Score Profiles.....	55
4.5 Parsimony	124
4.6 Sensitivity analysis for alternative sample periods.....	126
4.7 Summary Assessment	130
5 Decomposition of Productivity Changes.....	133
5.1 Outline of the Method.....	133
5.2 Results.....	134
6 Benchmarking Roll-forward Model (BRFM) Issues.....	150
6.1 Options and Submissions	150
6.2 Discussion of Selected Issues	151
6.3 Efficiency Incentives	154
6.4 Analysis of BRFM Options	155

7	Network Utilisation Metrics	172
7.1	Network Utilisation Measures	172
7.2	Incorporating Hosting Capacity in Benchmarking	174
7.3	Other Justice and Equity Centre Submissions.....	175
Appendix A	Review of Frontier Economics' calculation of BC95 efficiency scores	177
A.1	Background	177
A.2	Frontier Economics' observations and alternative method	177
A.3	Testing Frontier Economics' Method.....	178
A.4	Conclusions	180
Appendix B	Kumb90 Efficiency Time Trend Specification.....	181
B.1	Centred and final-year time-trend specifications	181
B.2	Kumb90 Efficiency Specification	181
B.3	Implications for the choice of normalisation.....	182
Appendix C	Benchmarking Roll-forward Method	184
C.1	BRFM method using Period-Average Efficiency	184
C.2	BRFM method using Period-End Efficiency.....	185
Appendix D	Sensitivity Analysis for Alternative Sample Periods Results	187
References		197

1 Introduction

This report is part of the Australian Energy Regulator's (AER) consultation on further developing and improving the opex cost benchmarking model. In late 2025, the AER published the consultation document *Electricity Distribution Benchmarking Opex Model Development* (Quantonomics 2025b). In March 2026, the AER received six submissions from DNSPs: Ausgrid, AusNet, Ergon/Energex (jointly), Evoenergy, Jemena and South Australia Power Networks (SAPN). A submission with a community/consumer focus was also received from the Justice and Equity Centre. Several DNSP submissions (including Evoenergy, Ergon/Energex, and Ausgrid) engaged Frontier Economics for advice, endorsing its findings and recommendations. Frontier Economics' report engaged fully in the detail of the econometric modelling.

The purposes of this technical report are, firstly, to respond to the submissions on the AER's 'Phase 2' consultation on opex cost model development. Secondly, by so doing, further improve the time-varying inefficiency models considered for inclusion in the benchmarking framework, and make recommendations to the AER.

This technical report summarises and responds to the submissions received, focusing on the conceptual and methodological arguments raised by stakeholders. It assesses the merits of the approaches proposed by stakeholders and explains how the issues raised affect the evaluation of alternative efficiency models. The report also presents empirical results for the models tested using the proposed approaches. These results are used to compare the performance of the various alternative models, considering both their statistical performance and practical suitability for regulatory benchmarking. In addition, we discuss the applicability of the tested models within the benchmarking roll-forward model, and whether the standard time-invariant models should be replaced or supplemented by time-varying efficiency models.

This report is organised as follows.

- Chapter 2 summarises the submissions received, including the Frontier Economics report.
- Chapter 3 discusses important conceptual issues raised in submissions. Many of these relate to the criteria for evaluating the alternative models. Some relate to whether the short sample period should be retained and whether the standard models should be retained alongside the time-varying inefficiency models.
- Chapter 4 presents the evaluation of the alternative econometric opex models with time-varying inefficiency tested in this study. Details of the econometric models are presented in *Attachment A: Time-Varying Inefficiency Models Results*. Chapter 4 presents conclusions on the preferred models based on the comparative evaluation.

- Chapter 5 presents the results of decomposing changes in opex partial factor productivity (opex PFP) into the separate effects of changes in technology, scale, efficiency and the effects of changes in OEFs.
- Chapter 6 discusses issues relating to how efficiency score results from time-varying inefficiency models can be used within the benchmarking roll-forward model (BRFM). It also presents some results on testing alternative time-varying inefficiency models using alternative roll-forward approaches.
- Chapter 7 discusses the issues raised by the Justice and Equity Centre relating to network utilisation metrics.

2 Summary of submissions

In this section, we summarise the main points raised by stakeholders and identify the matters that are treated as accepted positions, as well as those requiring further assessment. Quantonomics thanks all stakeholders for their submissions. The submissions have been valuable in clarifying the key areas of agreement and disagreement, and in refining the issues and models considered in this report.

2.1 Main Points in Submissions

Differing views were expressed on the merits of the preferred models presented in the Quantonomics (2025b) consultation document (“Phase 2”). Frontier Economics was critical of some detailed aspects of the models put forward by Quantonomics and proposed changes or extensions to those models. Frontier Economics identified issues that it considered needed to be addressed before the models presented by Quantonomics could be evaluated properly. These issues include the approach used to select starting values, the treatment of time-trend variables, and the calculation of efficiency scores. Further, Frontier Economics argued that its extended and more flexible models should be preferred to the Quantonomics models. With the incorporation of its suggested refinements, Frontier Economics submitted that time-varying efficiency models should replace the existing time-invariant efficiency models. Frontier Economics’ report is discussed further in section 2.2.

Jemena and AusNet had different perspectives in terms of incorporating time-varying efficiency models, including:

- supporting some of the time-varying efficiency models proposed by Quantonomics, but rejecting others;
- recommending that the time-varying models should supplement rather than replace the existing models. Jemena argued that these models are more parameter-intensive and may face convergence or stability issues over time. Further, a larger model set also reduces the risk of benchmarking results being unduly influenced by the inclusion or exclusion of individual models in any given year. AusNet similarly recommended that the AER should gradually transition away from time-invariant models.

SA Power Networks, while not forming a view on the substance of Frontier Economics’ concerns, believed they should be investigated.

The submissions appear to agree with some of the positions taken in Phase 2, although the following points are mainly noted in the Ausgrid submission:

- (a) Time-varying efficiency models should be used in opex benchmarking.
- (b) The multi-criteria approach for evaluating model performance is supported as “appropriate and consistent with good practice for selecting between econometric

models” (Ausgrid 2026, 9). Frontier Economics proposed some adjustments to the criteria, which are discussed in section 2.2.2.

- (c) The Colombi *et al* four-components model and the FECSS model are not suitable for the AER’s regulatory benchmarking purposes.
- (d) The inclusion of additional outputs, input prices, or heteroscedastic error structures should not be further pursued at this stage.
- (e) Use of the half-normal assumption in SFA models, instead of truncated-normal, is a pragmatic trade-off.

In this report, these are taken to be accepted positions at present. Some other points of agreement among submitters are examined further in this report:

- Jemena also argued that the short sample period should no longer be used for estimating opex efficiency as the use of time-varying efficiency models make those models redundant. Frontier Economics’ preferred models all use the long sample period. The issue of whether the short sample period should be retained is discussed in section 3.6.
- Most submitters agreed that the current BRFM approach, which relies on sample period-average efficiency scores, should be retained. There was general agreement that alternatives based on estimated efficiency at the end of the sample period need further examination. AusNet was cautious about the continued use of the current approach and suggested that the current consultation is “an opportunity for the AER to reconsider the role of the BRFM framework if time-varying inefficiency models are incorporated into the benchmarking approach.” Frontier Economics raised specific issues for consideration if the BRFM was applied using estimated efficiency at the end of the sample, including the interaction of the BRFM with the Efficiency Benefit Sharing Scheme (EBSS). Issues relating to the use of time-varying efficiency models with the BRFM are examined in chapter 6.

Some of the issues on which submissions differed include:

- As noted earlier, on the merits of particular models proposed by Quantonomics, and on the merits of the existing time-invariant efficiency models, Jemena expressed different views from those of Frontier Economics. Jemena supported some of the Quantonomics Phase 2 models but not others, whereas Frontier Economics rejected the LSE models and made revisions to the efficiency time trend assumed in the SFA models. Section 2.2 presents Frontier Economics’ main positions. The empirical issues are addressed in *Attachment A: Time-Varying Inefficiency Models Results*.
- Related to this issue, Frontier Economics argues that all existing time-invariant efficiency models should be abandoned, whereas Jemena supports retaining the

existing long-sample period models. Jemena’s position is that the time-varying inefficiency models should supplement rather than replace the existing models. This issue is discussed in section 3.6.

- Ausgrid sees the GTC specification of technical change, which allows step changes in technology every three years, as having merit, and Frontier Economics uses the GTC specification of technology change in its models. However, Jemena sees problems with the GTC approach because it could introduce volatility in the calculated average rate of technical change between the sample mean and the end period, and this volatility could flow through to the BRFM.

The Justice and Equity Centre submission pointed out that this consultation process could be more accessible to consumer groups and advocates. Whilst this is a technical report, we have endeavoured to make it accessible. The Justice and Equity Centre also proposed network utilisation metrics for evaluating DNSP performance, and recommended that the AER develop a framework for identifying efficient curtailment thresholds for customer energy resources (CER) at each stage of the energy transition. These issues are examined in section 7.

Other submitters, such as Ausgrid, also raised the concern that new outputs, such as provision of access to CER and grid stability, are omitted outputs. Frontier Economics also considers CER to be an omitted variable. However, it states:

... outputs such as the provision of CER services and supporting grid stability are multi-faceted, meaning that there is unlikely to be one single measure that captures the effect of the output completely. ... [and] even if measurable outputs could be defined (e.g., through identification of suitable proxy variables), it may be challenging to obtain consistent measures of those outputs across all DNSPs. ... By way of example, based on preliminary investigations, we found that there are no consistent data across the three jurisdictions on the volume of solar exports or on investment in network capacity to support solar exports (potential proxy variables for the output of facilitating two-way power flows) (Frontier Economics 2026, 34–35).

In 2023, the AER consulted on approaches to measuring hosting capacity services. With respect to benchmarking, the AER decided, at that time, not to “develop an interim export services operating environment factor (OEF) as there is insufficient data available to do so” (2023, 6). However, it committed to “initiate a review of how the benchmarking models can be updated for export services by 2027, or earlier if sufficient data becomes available” (2023, 6). Given this context, the issue of measuring CER services is out of scope for this process, since it will be addressed by the AER in another process in 2027.

AusNet emphasised that the AER should improve its methods for adjusting benchmarking results for the effects of operating environment factors (OEFs). We note that improving the quantification of existing and new OEFs is part of the AER’s benchmarking development

program (AER 2025, 24), but it is outside of the scope of the present project of improving the performance of the econometric opex cost function models.

2.2 Frontier Economics report

This section summarises the main positions taken in the Frontier Economics report. Our responses are mostly in the remaining sections of the report.

2.2.1 Implementation errors and suggested improvements

Frontier Economics made certain criticisms of Quantonomics' implementation of its models and also proposed some suggested improvements. Addressing these points, we have re-estimated the models presented in the previous report, which are presented in *Attachment A: Time-Varying Inefficiency Models Results*.

Time trend specification for efficiency

A key criticism of implementation was that Quantonomics adopted different types of time trends in the various BC95 models tested. In BC95-JTT-HN, the efficiency time trend was expressed as calendar years (ie, 2006, 2007, etc.), whereas in BC95-JTT-HN-GTC the efficiency time trend term was centred. In the latter case, the time trend was zero at the midpoint of the sample. Frontier Economics noted that the first approach, used in the BC95-JTT-HN model, implies that the expected DNSP inefficiency is effectively time-invariant over the benchmarking period.

We agree that the time trend should equal zero at some point in the sample, with other observations representing the number of years (positive or negative) from that base period. In the base period, the inefficiency distribution is then half-normal under all of the SFA specifications tested.¹

Frontier Economics proposes to define the time trend as the observation year minus the final year. This will be zero in the final year of the benchmarking period. Frontier Economics applied this only to some models, whereas for models with a time trend term that is zero at the midpoint of the sample, this time trend specification was retained. There appears to be an inconsistency in the way this approach was applied. Frontier Economics calculated the time trend as:

- $t + 1 - (\text{final year})$ for Ontario DNSPs,
- $t - (\text{final year})$ for Australian and New Zealand DNSPs

where t denotes the year variable after adjusting for differences in calendar and financial years across Victorian DNSPs, New Zealand and Ontario and (final year) is 2023 for the models

¹ Note that, for the Kumb90 models, the inefficiency term follows a half-normal distribution in all years, rather than only in the base year.

tested in the Phase 2 report.² The correct approach should be $(t - (\text{final year}))$ for all observations. This is because the Ontario data are already adjusted by adding half a year to reflect their reporting on a calendar year basis. By adding one full year to the Ontario time trend, Frontier Economics' approach effectively shifts the Ontario observations one full year ahead, rather than preserving the intended half-year lag.

We accept Frontier Economics' criticism, although we have tested the use of different base periods and found that for some models, the results are not invariant to the base period chosen.³ We have re-run all BC95 models using a time trend that is equal to zero in the final year of the benchmarking period (ie, FY2023), while applying the correction to the Ontario time trend. For the Kumb90 models, we find that having the base year in the middle of the sample period yields more stable models (see Appendix B). For greater consistency, we have also updated the included GTC dummy variables (six variables that denote successive blocks of 3-year periods) so that the last GTC dummy (for 2021 to 2023) is now excluded rather than the first GTC dummy variable (for 2006 to 2008). This means the implicit coefficient for the final period is equal to zero under the revised time trend specification. The results for the re-estimated models are presented in *Attachment A: Time-Varying Inefficiency Models Results*. The estimated models are assessed in Section 4 of this report.

Specification test

Frontier Economics noted that Quantonomics incorrectly applied the *linktest* specification test. "When applying the linktest, the user should specify the form of the model being tested, so the structure of that model can be adopted in the linktest regression. When no such specification is provided by the user after running *sfp* estimation, the linktest command by default uses the BC92 time-varying decay model specification to run the test" (Frontier Economics 2026, 46). We accept this criticism and have adopted this correction in the models presented in this report.

Convergence & Starting Values

While Quantonomics had found that some models did not converge, Frontier Economics claims that the models will converge when more appropriate starting values are used. Parameter starting values are used to initiate numerical optimisation of Maximum Likelihood estimation. Frontier Economics offers a multi-step procedure for obtaining starting values, which it claims yields solutions for models that did not converge using our approach.

²t is constructed from the reported year variable after adjusting for differences in reporting periods across jurisdictions, so that all observations are expressed on a comparable financial-year basis. The adjustment adds 0.5 to the reported year for Victorian DNSPs before 2021 and for all Ontario observations, where data were reported on a calendar-year basis, and subtracts 0.25 from the reported year for all New Zealand observations, where data are reported on a year-to-March basis.

³ This appears to be because for the SFA models, the inefficiency time trends enter the model nonlinearly.

Their proposed method is:

- (i) Estimate a simplified Cobb-Douglas version of the model using OLS.⁴
- (ii) Use the estimated parameters (including variance of the disturbances) as starting values in estimating a time-invariant Cobb-Douglas stochastic frontier analysis (SFA) model with a half-normal inefficiency distribution. The OLS residual variance is used as the starting value for the total disturbance variance, while the variance-share parameter is initialised at 0.1, together providing initial values for the disturbance and inefficiency variances.
- (iii) Use the estimated parameters (including the variances of both the disturbance and the inefficiencies) as starting values, and set the starting values to zero for all parameters determining the inefficiency time pattern, and estimate a time-varying Cobb-Douglas SFA model, with the starting values for the newly introduced Translog terms set to zero.
- (iv) Use the estimated parameters (including the variances of both the disturbance and the inefficiencies and the inefficiency time-trend parameters) as starting values to estimate a time-varying Translog SFA model.

Frontier Economics' method differs from Quantonomics' approach in two respects: first, Quantonomics did not use the estimated variances of the disturbance and inefficiency terms from step (ii) as starting values for step (iii). Second, Quantonomics applied a similar multi-step procedure to the TLG model, rather than using the time-invariant CD model as the source of starting values.

We consider that Frontier Economics' approach has merit because it is simple to implement and can improve the convergence performance of some models. The SFA models presented in *Attachment A: Time-Varying Inefficiency Models Results* therefore use this approach. In addition, we have performed a sensitivity analysis of starting values, and incorporated further settings for the maximum likelihood (ML) optimisation process.⁵

Calculation of efficiency scores

In models where the post-estimation *bc* command fails to produce efficiency scores for all observations, Frontier Economics is able to estimate the missing efficiency scores using an alternative calculation method. Our assessment of this method is presented in Appendix A below. We conclude that Frontier Economics' alternative method for calculating efficiency

⁴The use ordinary least squares (OLS) estimates as parameter starting values is well-established in the econometrics literature generally, and in SFA literature in particular, because the OLS estimates are consistent estimators under a wide range of modelling conditions and can provide suitable starting values for more sophisticated models that require numerical optimization.

⁵ All final SFA models were estimated using the following *additional* ML optimisation options: `repeat(20) rescale ltol(0) tol(1e-7)`.

scores can be accepted as a valid approach to estimating BC95 efficiency scores, as it reproduces Stata's standard results to numerical precision for all non-missing scores.

Note, however, that the failure of the Stata command to calculate efficiency scores appears to reflect difficulties in reliably estimating some DNSP-specific efficiency time trend coefficients. The missing efficiency scores should therefore be interpreted as a symptom of the underlying model estimation problem, rather than as the problem of computing efficiency scores *per se*. Frontier Economics' method addresses the issue of missing efficiency scores, but it does not resolve the underlying model estimation instability. This issue remains evident in some specifications, as indicated by very large coefficient estimates or missing standard errors for some coefficients.

Impacts of applying Frontier Economics' recommendations

We re-estimated the models presented in the Phase 2 report to incorporate Frontier Economics' recommendations discussed above. The results are presented in Attachment A. The main impacts of applying these changes, compared with the approach used in the Phase 2 report, are summarised below.

- *BC95 Models:*
 - The revised starting values approach considerably improves the computational performance of the BC95 models. In the Phase 2 report, considering the longer estimation period from 2006 to 2023, only one BC95 model, BC95-JTT-HN, could be properly estimated. By contrast, under the revised approach presented in *Attachment A: Time-Varying Inefficiency Models Results*, four alternative BC95 model specifications produced estimation results that were at least partially satisfactory. The AJTT variants of the BC95 models continue to present some computational issues.
 - For the BC95-JTT-HN model, which was the only BC95 model that produced valid results in the Phase 2 report, the MVs remain broadly consistent with the previous results. The residual plots and efficiency trends are also generally aligned, and the ranking of efficiency scores remains unchanged. The average efficiency score across Australian DNSPs is slightly lower under the revised approach, by 2.1 per cent.
- *Kumb90 models:*
 - The Phase 2 report already considered four Kumb90 specifications. These models also run successfully under the revised starting values and ML optimisation approach, although some indications of computational issues, such as unusual coefficient or standard error estimates, remain (see section 4.1 and *Attachment A: Time-Varying Inefficiency Models Results*).

- *LSE models:*
 - Frontier Economics' proposed change to basing the time trend at zero in 2023, compared to being centred at the period-average, was also applied to LSE models. When applied to these models, the output coefficients, year coefficient, efficiency scores and monotonicity violations remain unchanged with the revised efficiency time-trend definition. The coefficients on the DNSP-specific and jurisdictional time trend variables also remain unchanged. Only the constant term and the coefficients on the DNSP-specific and jurisdictional dummy variables are changed.

2.2.2 Comments on Evaluation Criteria

Whilst the use of multiple criteria for model evaluation was supported, Frontier Economics has specific views on the criteria, which are summarised as follows.

1. *Sign and significance of output coefficients:* Frontier Economics agrees that it is essential that a model has positive output elasticities that are all significant at the 5 per cent level.
2. *Frequency of monotonicity violations:* The AER's criteria to date are that if monotonicity violations (MVs) for a given Australian DNSP exceed 50 per cent of its observations, the model is considered invalid for that DNSP. And if a model is invalid for 50 per cent or more of Australian DNSPs (i.e., 7 out of 13 DNSPs), it is considered invalid overall. In the Phase 2 report, we discarded some models on this basis and generally preferred models with as few MVs as possible, although no formal threshold was defined. Frontier Economics proposes alternative criteria for the purposes of model evaluation. If MVs occur in fewer than 10 per cent of Australian observations, the model is acceptable. If MVs are between 10 per cent and 25 per cent, there is a "moderate issue". If MVs occur in greater than 25 per cent of observations, the model should be rejected. In some cases, this criterion will be less restrictive and in some cases more restrictive than the criterion we adopted. Whilst the Frontier Economics criterion is of interest, we have decided to retain our approach of rejecting models only where MVs occur in more than 50 per cent of Australian observations. However, we flag a model as having moderate monotonicity issues where either: (i) one or more Australian DNSPs have MVs in 50 per cent or more of their observations; and/or (ii) the model has MVs in more than 25 per cent and up to 50 per cent of Australian observations.
3. *Reasonableness of output weights:* Frontier Economics agrees that the sum of the estimated elasticities should be close to, but usually slightly lower than, 1. However, they disagree with other aspects of how this criterion is applied by Quantonomics. Little reliance is placed on checking the reasonableness of individual output elasticities

by Frontier Economics, because: (a) the high degree of collinearity between some output variables may mean that individual output elasticities are estimated with relatively high uncertainty; and (b) there is a lack of accurate priors about what they should be.

It is argued that output weight priors cannot be reliably obtained from the output index weights: “We have explained recently our concerns that the output weights used to construct the MPFP and MTFP indices are unreliable because they have been estimated using a flawed approach” (Frontier Economics 2026, 46). Instead, they look for consistency with the output elasticities estimated by the best-performing time-varying models. Our response to Frontier Economics’ views on reasonableness of output weights as a criterion is in section 3.2.

4. *Coefficient of underground share*: Frontier Economics proposes a slightly weaker criterion than Quantonomics used: “A coefficient that is not negative, but statistically insignificant, is acceptable as it can result from imprecise estimates rather than a more severe issue (i.e., mis-specification). A positive and statistically significant coefficient would, however, raise concerns” (Frontier Economics 2026, 56). In our view, the coefficient on underground share should also be assessed by reference to its economic significance, in addition to negativity of sign and/or statistical significance. For the purpose of estimating efficiency, the key issue is whether the estimated coefficient is consistent with economic knowledge, which indicates that the coefficient should be negative,⁶ not only whether it is statistically different from zero. On this basis, the underground share variable should have a negative effect on opex, and as in the Phase 2 report, a negative but statistically insignificant coefficient on underground share is also considered acceptable (Quantonomics 2025b, 22).
5. *Convergence*: Frontier Economics uses the same criterion as Quantonomics, but as discussed above, they propose a different approach to setting parameter starting values. This reduces the instances of non-convergence of numerical optimisation routines for likelihood maximisation, and hence fewer SFA models estimated under MLE are to be rejected on this criterion. In our view, this criterion should not be limited to cases where models fail to converge but encompass whether computational issues affect the reliability of the estimated parameters and standard errors, even where convergence is technically achieved.
6. *Goodness-of-fit*: Frontier Economics, consistently with Quantonomics, uses the Pseudo-adjusted R-squared statistic and the Bayesian Information Criterion (BIC), where

⁶ Underground cables have higher capital cost, but generally have lower operating and maintenance costs, than overhead lines. Underground cables require less maintenance associated with pole inspection and replacement, vegetation management and storm damage, for example, but when maintenance is required, it will generally be more costly for underground cables.

available. Frontier Economics notes (as Quantonomics also noted in the Phase 2 report) that “the BIC of the BC95 models should not be compared to the BIC of the Kumb90 models or the standard SFA model as the log likelihood values are calculated in a different manner” (Frontier Economics 2026, 60).

7. *Performance on specification tests*: As discussed above, Frontier Economics points out that Quantonomics had implemented the *linktest* incorrectly. Frontier Economics also relies on statistical tests of the joint significance of additional parameters to conclude that the Translog specification consistently outperforms the Cobb-Douglas specification. We discuss the topic of statistical tests of joint significance in section 3.4.
8. *Residual plots*: Frontier Economics, consistently with Quantonomics, uses residual plots to show whether the residuals have a reasonably constant variance with the mean consistently at zero, and there is no pattern in the residuals. “This indicates that the model is not systematically misestimating true values in a predictable manner, indicating that the model is not adequately explaining the full structure of the data” (Frontier Economics, 2026: 62).
9. *Stability to sample changes*: Whereas Quantonomics tests the stability of model results over different sub-samples, Frontier Economics tests stability by adding new ‘simulated’ data (with an assumed level of efficiency) to the sample and testing how well the model performs with those new data. We discuss Frontier Economics’ views on the use of sub-samples and its approach to out-of-sample testing in section 3.5.
10. *Consistency of efficiency scores & reasonableness of efficiency time trend*: Frontier Economics rejects comparing efficiency scores to those of previous benchmarking studies, because it regards those studies as mis-specified. They do consider consistency checks against the best-performing models. Frontier Economics agrees “that the overall trend in Opex MPFP is reasonable, and estimates should be consistent with that trend” (Frontier Economics 2026, 46). In addition to this test, they examine whether “the pattern of efficiency scores over time appears *plausible*” (46, emphasis added). This criterion is used to reject Kumb90 models, because they all converged to a common efficiency level of 100 per cent when they had a positive trend. It is also used to reject the LSE models because the implied linear trend in inefficiencies is regarded as too restrictive and implausible. Our response to Frontier Economics’ views on these issues is in section 3.1.
11. *Parsimony*: Frontier Economics (2026, 47) says it places “only minor emphasis” on this criterion. Our response to this position is in section 3.4.

2.2.3 More Flexible Modelling Approaches

This section summarises some of the specific modelling choices made by Frontier Economics in its empirical analysis. As mentioned, Frontier Economics is critical of the restrictiveness of

the inefficiency time profiles allowed for in most of the models estimated by Quantonomics. One of Frontier Economics' proposed approaches to increasing the flexibility of efficiency time trends was to adapt an approach taken by Quantonomics to technical change. In addition to linear time trends to reflect the rate of technical change in the cost frontier, Quantonomics also considered a generalised technical change (GTC) index among the determinants of the cost frontier. This divides the sample into six separate periods, represented by five dummy variables, with the remaining (omitted) period used as the base period. Frontier Economics uses an equivalent specification for the efficiency time trend, which it calls the generalised efficiency change (GEC). "This allows the rate of efficiency change to become faster or slower, or even change direction, over time" (Frontier Economics 2026, 15).

In section 4, we test models that include the GEC specification in the efficiency change determinants in SFA models (ie, BC95 and Kumb90). In section 3.3, we discuss issues of identification that may arise with this method in some models, and also note some modelling variations available using the GEC, other than those presented by Frontier Economics. Section 3.3 also discusses some other methods of making efficiency trends more flexible.

Frontier Economics regards the Kumb90 models estimated by Quantonomics as having a restrictive time profile of efficiency adjustment over time, or, in the case of the model with separate time trends for each Australian DNSP, producing implausible results. Frontier Economics applies the Kumb90 models with the GEC specification of efficiency change combined with the GTC specification of technical change.

Frontier Economics also regards the LSE time-varying models as imposing an unduly restrictive time pattern of efficiency changes, namely that they occur at a constant rate. The evidence from Opex PFP analysis, they argue, suggests that this assumption is unrealistic. To increase its flexibility, Frontier Economics tests a "spline specification in the DNSP-specific efficiency trend: a continuous piecewise-linear trend with a kink in 2015". The choice of the break point at 2015 is based on: (a) "a clear kink in residual patterns around 2015"; (b) it being about the time when benchmarking started; and (c) a change in the direction of the Opex MPFP index is also seen at around this time (Frontier Economics 2026, 15). Further, Frontier Economics tested a GTC time-trend variant of this model with the New Zealand time trend added, rather than the Ontario time trend. However, Frontier Economics ultimately recommended abandoning all LSE models, implying that it did not consider the spline specification to provide satisfactory results.

After evaluating the models it tested, Frontier Economics found the following three models to outperform the others:

- *BC95-JTT-HN-GTC*. This is a model Quantonomics rejected because it failed to produce efficiency scores for all the Australian DNSPs. However, with Frontier Economics' alternative coding for calculating efficiency scores, this problem did not arise. We have replicated this model using Frontier Economics' proposed approaches,

and the results are presented in Section 2.2 of *Attachment A: Time-Varying Inefficiency Models Results*.

- *BC95-AGEC-HN-GTC*. This BC95 model combines the GTC specification for technical change with the GEC specification for inefficiency change for Australian DNSPs. We have replicated this model, and the results are presented in Section 3.2 of *Attachment A: Time-Varying Inefficiency Models Results* under the name BC95-AGECJTT-HN-GTC).
- *Kumb-AGECJUR-HN-GTC*. This Kumb90 model combines the GTC specification for technical change with the GEC specification for inefficiency change for Australian DNSPs, while the jurisdictions have a linear time trend. We have replicated this model, and the results are presented in Section 3.4 of *Attachment A: Time-Varying Inefficiency Models Results* (under the name Kumb90-AGECJTT-HN-GTC).

In this report, we evaluate these models against the others previously tested. We also explore some other approaches to introducing greater flexibility into efficiency time trends.

3 Conceptual Issues

Before evaluating models (in section 4), this section discusses several issues arising from the submissions that are important to how the models should be implemented or evaluated.

3.1 Reasonableness of Efficiency Scores and Trends

This section 3.1 outlines the views in submissions on the application of the benchmarking roll-forward model (BRFM), and also discusses Frontier Economics' arguments about some time-varying frontier methods having implausible or excessively restrictive time patterns of efficiency. We argue that these two issues are closely linked and that there appears to be an inconsistency in Frontier Economics' arguments on these two subjects.

3.1.1 The benchmarking roll-forward model (BRFM)

The current opex econometric models produce estimates only of a single efficiency score for each DNSP, which can be interpreted as its average efficiency over the chosen sample period. To derive the estimated efficient opex at the end of the sample period, or for a year after the end of the sample period (ie, the base year of a price reset), the AER uses the BRFM.

Section 6.3 of *Electricity Distribution Benchmarking Opex Model Development* (Quantonomics 2025b) discusses some of the issues and options that time-varying inefficiency models raise for the application of the benchmarking roll-forward model. If opex econometric models incorporate time-varying efficiency, the available options include:

- (a) Averaging the estimated efficiency scores over the sample period and applying the BRFM as before.
- (b) Using the efficiency estimated at the end of the sample period (if the base year is the same as the last year of the sample period) or, if the base year is after the last year of the sample, using the BRFM to bridge the shorter gap between the last year of the sample period and the base year. The latter would most often be the case.

In addition, both options might be combined. We stated:

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 (Quantonomics 2025b, 143).

In the DNSP submissions on the use of BRFM for calculating base-year efficient opex using time-varying inefficiency models, most DNSPs would prefer to retain the current BRFM applied from the sample midpoint to the base year. Continuing to apply the current BRFM would require averaging the time-varying efficiency scores. Evoenergy and Ergon/Energex

observe that the current method is well-established. A differing view is AusNet, with concerns that:

... retaining the current BRFM framework and applying it using sample-average efficiency scores estimated from time-varying models may introduce systematic bias. For DNSPs exhibiting improving efficiency over time, averaging efficiency scores across the sample period may mask recent performance improvements which are most relevant for regulatory decision making. Conversely, DNSPs whose efficiency has deteriorated over time may be inadvertently advantaged, as sample-average efficiency scores may be overstated by earlier, less relevant performance levels.

Stakeholders generally suggest that the alternative of rolling forward the estimated efficient opex in the last year of the sample to the base year would need to be thoroughly analysed and consulted on (SAPN, Ausgrid).

Frontier Economics (2026) also raised several issues not previously considered. Although we considered the potential for inaccuracy if the assessment of efficient opex is based on the efficiency score in the last year of the sample, Frontier Economics raises some specific issues to be considered in this regard. They also raise some questions relating to the practicalities of making adjustments for operating efficiency factors (OEFs), and questions relating to the effects on businesses' incentives to make cost efficiencies. Concerning the incentive effects, especially, Frontier Economics (2026, 74) suggests there are "a number of complex issues that the AER would need to take into account, and address, before changing its current approach of rolling forward from the middle of the benchmarking period." Frontier Economics also recommends that the AER should retain its current BRFM approach; and before making any changes to its BRFM approach, the AER should clarify these issues and consult with stakeholders on them.

If accepted, the submissions (including the Frontier Economics report) that the current BRFM should be retained imply that only the sample average efficiency score for each DNSP would be used in deriving the estimated efficient base-year opex.

3.1.2 Criterion of reasonableness of efficiency time trends

Frontier Economics accepts the reasonableness of the estimated efficiency time trend as one criterion for evaluating the comparative performance of models. They test this in two ways. Firstly, "the overall trend in Opex MPFP is reasonable, and estimates should be consistent with that trend" (Frontier Economics 2026, 46). Second, the pattern of efficiency scores over time should be 'plausible'.

On the first of these two criteria, whilst we accept that Opex PFP movements are a relevant comparator for the estimated efficiency time trend (and present these comparisons in section 4.4 below), they should be interpreted with some caution. Opex PFP is not a direct measure of efficiency change. It may also reflect technical change, scale effects, and differences between

the output specifications used in the PFP analysis and the output specification used in the econometric benchmarking models. In particular, the PFP analysis assumes constant returns to scale, whereas the benchmarking models allow for variable returns to scale.⁷ It does not account for any impact of the share of underground cables on opex, and includes additional output drivers, such as energy throughput and reliability, which may affect the observed Opex PFP trend independently of efficiency change.

The second criterion is used by Frontier Economics to reject several models. The Kumb90 models with jurisdictional efficiency time trends (JTT), presented by Quantonomics, are rejected by Frontier Economics because:

- *The time profile of efficiency scores is overly restrictive:* If the efficiency scores of DNSPs tend to change monotonically over the sample period, then they will converge. If they tend to increase monotonically, they will converge to 100 per cent; whereas if they tend to decrease monotonically, they will converge to some level less than 100 per cent. (Frontier Economics 2026, 27).
- *Comparative efficiency levels are fixed:* The efficiency score rankings of DNSPs never change.

The Kumb90 models with the AJTT efficiency trends (with a separate trend for each Australian DNSP and common time trends for NZ and Ontario DNSPs, respectively) do not have the restriction that efficiency rankings are fixed. However, they are also rejected by Frontier Economics based on what they refer to as the plausibility of efficiency trends:

... the restrictiveness of the specification is apparent ... with implausible efficiency trends for Essential Energy over the long benchmarking period, among other DNSPs. While the AJTT version is less restrictive than the JTT version, the formulation of the Kumb90 efficiency adjustment (as applied by Quantonomics) appears too restrictive to yield intuitive or plausible trends in efficiency over time. (Frontier Economics 2026, 40).

The LSE models are all rejected by Frontier Economics under this criterion because they “assume a constant rate of efficiency change over the whole benchmarking period” despite “allowing for differences in efficiency time trends across Australian DNSPs.” The cited concern is that this specification “allows no possibility for DNSPs to catch up more or less quickly over time, which is a strong assumption and does not align with the highly nonlinear trajectories of the Opex MPFP indices” (Frontier Economics 2026, 30).

These statements are inconsistent with the way we interpreted the derived efficiency scores in the LSE models. In these models, efficiency scores are calculated in proportion to the DNSP that has the highest efficiency. In time-varying inefficiency applications, a firm that is the most efficient in one year, may not be the most efficient in another year, and hence the efficiency

⁷ The econometric models consistently estimate returns to scale to be close to constant.

scores can be calculated against a different comparator in different years. Figures 4.5.2, 4.5.4, 4.5.6 and 4.5.8 of the ‘Phase 2’ report showed that, with the linear time trend effects in those LSE models,⁸ not all the efficiency scores evolve with a constant linear slope over the benchmarking period. We noted that the LSE models with DNSP-specific effects that are linearly related to time consistently generated concave efficiency profiles (Quantonomics 2025b, 132). Hence, based on our calculation of efficiency scores in the LSE models presented in the Phase 2 report, the efficiency scores do not necessarily change at a constant rate over, or by constant increments, the sample period.

We agree that the LSE time-varying efficiency specifications are comparatively restrictive and cannot capture the type of year-to-year variation in relative efficiency seen in the Opex PFP results. In this report, we examine several methods to introduce more flexibility into the LSE efficiency time-profiles. Frontier Economics tested a continuous piecewise-linear DNSP-specific efficiency trend with one ‘knot’ as an extension of the LSE models. In this report we test this approach and other reasonable extensions to the LSE model, including piecewise-linear DNSP-specific efficiency trends with two knots, and DNSP-specific effects that vary quadratically with time. These extended models are not necessarily more restrictive than some of the SFA models tested. Hence, we do not accept Frontier Economics’ outright rejection of all LSE models due to lack of plausible efficiency time trends.

Lack of a plausible time trend, although given considerable emphasis, is not the only grounds used by Frontier Economics for rejecting the LSE models. Frontier Economics also argues that they have a much poorer fit to the data than either the BC95 or Kumb90 models. We examine the empirical results in Section 4.3.

In summary, in many cases where models presented by Quantonomics are rejected by Frontier Economics, the criterion of ‘plausible’ trends in efficiency scores is used as either the main reason or one of the main reasons given.

The principal purpose of introducing time-varying inefficiency into the opex cost function models is to address problems that have emerged in the standard opex models, especially with some models exhibiting an excessive frequency of monotonicity violations (MVs), and in some cases, failing to converge, or converging to a model with unacceptable parameter values. Remedying these issues must be among the primary criteria for model evaluation.

We accept that a good model should produce efficiency scores that are reasonably correlated with the Opex MPFP index, since it is a relevant comparator measure, subject to the caveats noted earlier. On the other hand, Frontier Economics’ proposed test of ‘plausibility’ of the efficiency time trend appears to be too subjective and somewhat vague, as what is ‘plausible’ for some may be considered as not ‘plausible’ by others. Most importantly, too much weight is given to this criterion, since, as noted above, Frontier Economics recommends retaining the

⁸ See Attachment A to this report for the re-estimated versions of the same models.

existing BRFM. In that case, only the average of all estimated efficiency scores over the sample period is used. The accuracy of estimated efficiency scores in every sample year would be of lesser importance than the accuracy of the estimation of the average efficiency score.

In summary, Frontier Economics has given an unduly heavy weight to its views on the ‘plausibility’ of efficiency time trends, which cannot be easily reconciled with its recommendation that only the average of the efficiency scores be used by the AER. Therefore, its use of this ground to reject many of the models presented by Quantonomics cannot be supported.

We propose to use the closeness of fit of the estimated efficiency scores with movements in the Opex MPFP index as one criterion to be weighed alongside other criteria. It is a useful criterion, but as discussed previously and in section 3.2, it is only an approximate guide to the reliability of the estimated time patterns of efficiency scores. Further, if the current BRFM method is retained, this criterion is less important than it would be if the efficiency estimates at the end of the sample period were relied on. Hence, we give greater emphasis to successfully addressing the problems of convergence and excessive frequency of monotonicity violations whilst producing robust models that consistently meet the essential requirements.

3.2 Output Elasticities of Opex MPFP

As previously mentioned, Frontier Economics places limited weight on assessing the reasonableness of individual output elasticities. They argue that (a) collinearity between output variables can lead to imprecise estimates of individual elasticities, and (b) it is difficult to establish reliable benchmarks to assess the reasonableness of individual opex cost function output elasticities. Values from previous cost functions, it is argued, should not be used as benchmarks, as those models are mis-specified. Nor should the index-number-based TFP output weights be used, since it is argued that those estimates were derived using a flawed approach.

We understand that output weights relevant to total costs do not necessarily reflect the output weights relevant to opex. The same outputs should generally be relevant. However, in an econometric exercise, the determinants of opex and of capital costs may differ in degrees of importance. For this reason, we placed limited importance on comparisons of output elasticities with index weights. They were used primarily as a consistency check to diagnose any potential issues and potentially explain why the benchmarking results differ. There were no firm prior assumptions on the appropriate values for the opex cost function output weights.

Frontier Economics, on behalf of Evoenergy, argued in its response to the 2025 output weights memorandum that the Leontief production functions used to define the MTFP output weights were mis-estimated (Evoenergy 2025). These concerns were raised again in submissions on the Phase 2 report (Frontier Economics 2026, 46). We do not agree with the claim that the method of estimating index output weights is flawed. We responded to this claim in detail in

the memorandum *Non-reliability Output Index Weights ABR25: Supplementary Analysis* (Quantonomics 2025d). The method has been independently reviewed by the Centre for Efficiency and Productivity Analysis (Peyrache 2024), which found the established method to be sound. The output weights were re-estimated in 2025, using both the established method and the alternative methods proposed by Peyrache in the memo *Nonreliability Output Index Weights ABR25* (Quantonomics 2025c). The standard method was developed by Denis Lawrence and Erwin Diewert (2006), the latter being one of the world's leading economists in the fields of index numbers and productivity measurement.

Notably, there is an *apparent* inconsistency between Frontier Economics' rejection of the output index weights as 'flawed', whilst also claiming that the trends in Opex MPFP are a reasonable basis for assessing the reasonableness of Opex efficiency score trends. The Opex MPFP index is a ratio of the multilateral output index to the opex input index, where the multilateral output index is constructed using the same weights that Frontier Economics regards as 'flawed'. However, this position will *not* be an inconsistency if the outputs are so closely correlated that variations in the output weights do not strongly influence Opex MPFP trends. The log of customer numbers *is* highly correlated with the log of RMD, although the log of circuit length is not as highly correlated with either log-customer numbers or log-RMD.⁹ Hence, the output weights are likely to significantly influence the measured Opex MPFP trends.

In summary, we do not agree that the output index weights are flawed. We did not use the output index weights as reliable benchmarks for the reasonableness of the opex cost function output elasticities, only as a relevant comparator. For consistency, and as previously stated, the Opex MPFP index should be viewed as a relevant comparator of the time pattern of estimated efficiency scores, but not the key test of their reliability. We propose to use the output index weights as a relevant cross-check to the estimated parameter values of the opex cost function, without expecting any close matching of values. We also propose to use the Opex MPFP as a relevant comparator to the estimated efficiency scores, without expecting a close fit. These positions are consistent with our previous positions.

3.3 Methods of increasing the flexibility of efficiency time trends

3.3.1 Use of GEC with GTC

Frontier Economics defined the Generalised Efficiency Change (GEC) as the same set of dummy variables used by Quantonomics for the GTC index. These represent a set of groups of sequential three-year periods. Two of Frontier Economics' preferred models used both the GEC and GTC together in the models. These are:

⁹ In the 2006 to 2023 data sample, these correlations are 0.980, 0.851 and 0.771 respectively.

- SFA-BC95-AGEC-HN-GTC, and
- SFA-Kumb-AGECJUR-HN-GTC.

This raises the question as to whether the use of this same set of dummy variables twice in these models may give rise to issues of multicollinearity and/or identification of the effects of interest. This discussion will focus on the BC95 model, but similar issues are likely to apply to the Kumb90 model.

In the Phase 2 report, the basic cost model was expressed as:

$$c_{it} = \alpha_0 + x'_{it}\beta + u_{it} + v_{it} \quad (3.1)$$

where $i = 1, \dots, N$ indexes the DNSPs and $t = 1, \dots, T$ represents the periods; x_{it} is a vector of independent variables which includes the outputs and the OEFs (such as the jurisdictional dummy variables and the share of undergrounding). The other symbols all have the same meanings as in the Phase 2 report. In the Battese and Coelli (1995) variant of SFA, the ‘pre-truncated mean’ of the inefficiency distribution is a function of certain variables, z , which can vary over time or between firms:

$$u_{it} \sim N^+(\mu_{it}, \sigma_u^2) \quad \text{with: } \mu_{it} = \delta_0 + z'_{it}\delta, \quad (3.2)$$

We noted that an equivalent way of expressing the inefficiency term u_{it} , is:

$$u_{it} = \delta_0 + z'_{it}\delta + \varpi_{it} \quad (3.3)$$

where ϖ_{it} is a random variable obtained from a truncation of a normally-distributed random variable with zero mean and constant variance, with the truncation point determined by: $\varpi_{it} \geq -\delta_0 - z'_{it}\delta$ (Battese and Coelli 1995, 327). Using (3.3), equation (3.1) can be written as:

$$c_{it} = \alpha_0 + x'_{it}\beta + \delta_0 + z'_{it}\delta + \varpi_{it} + v_{it} \quad (3.4)$$

Let the vector of explanatory variables at observation (i, j) be: $x_{it} = \{y_{it}, z_{it}\}$, where y_{it} is the set of cost drivers such as log outputs, and z_{it} is the set of dummy variables used to define both GTC and GEC. And let δ_0 be restricted to zero. Then:

$$c_{it} = \alpha_0 + y'_{it}\beta_y + z'_{it}\beta_z + z'_{it}\delta + \varpi_{it} + v_{it}, \quad (3.5)$$

This shows that the z_{it} vector appears twice on the right-hand side. This will give rise to a computational identification problem associated with estimating the parameter vectors β_z and δ , which may be practically impossible to separately identify with accuracy. This is similar to the computational issue associated with adequately identifying both α_0 , the intercept of the frontier model, and δ_0 , the intercept of the function determining μ_{it} , which is the reason why

the parameter δ_0 was restricted to zero. Thus, when the GTC and GEC set of dummy variables are used together, the parameter vectors β_z and δ may be poorly estimated.

Since these parameters are central to the inferences to be drawn from the model, in terms of quantifying efficiency changes and technical change, this is a serious deficiency. Furthermore, the implied multicollinearity may reduce the precision of the estimated coefficients and, in turn, may impact the reliability of efficiency score estimates.

For these reasons, rather than using GTC and GEC in the same models, it is likely to be preferable to use either:

- (a) a GTC specification of technical change and continuous jurisdictional or DNSP-specific time trends for efficiency change, or
- (b) continuous jurisdictional time trends for technical change and a GEC specification for efficiency change.

Modelling results using GEC variables are presented in section 3 of the *Attachment A: Time-Varying Inefficiency Models Results*.

3.3.2 LSE with Spline Functions and Piecewise Regression

Frontier Economics (2026, 44) tests some LSE models with a continuous piecewise time trend term for efficiency change. It is formulated as (mostly using Frontier Economics' notation):

$$\begin{aligned} lvc_{it} = & \beta_0 + \beta_1 ly1_{it} + \beta_2 ly2_{it} + \beta_3 ly3_{it} + \beta_4 lz1_{it} + \beta_5 NZ_i + \beta_6 Ont \\ & + GTC' \beta_7 + \sum_{j=2}^{13} \alpha_i d_i^j + \sum_{j=1}^{13} \delta_i dtm_{it}^j + \sum_{j=1}^{13} \gamma_i dtp_{it}^j + \beta_8 nz_yr_{it} + \varepsilon_{it} \end{aligned} \quad (3.6)$$

where:

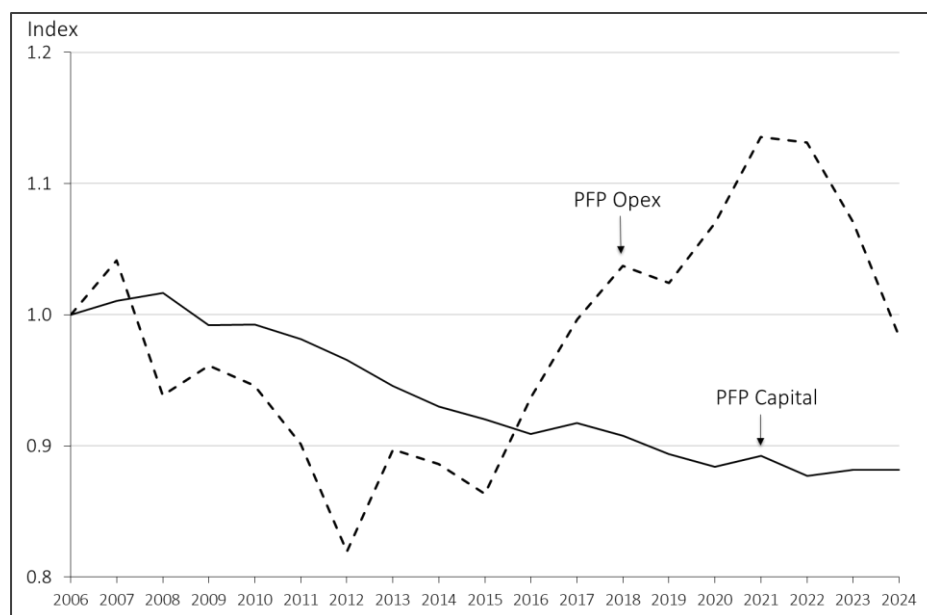
- $d_i^j = 1$ if $i = j$ and 0 otherwise
- $dtm_{it}^j = \min(t - 2015, 0)$ if $i = j$ and 0 otherwise
- $dtp_{it}^j = \max(t - 2015, 0)$ if $i = j$ and 0 otherwise,
- and j indexes the 13 Australian DNSPs, which form a subset of the full sample of DNSPs, indexed by i .

In this piecewise expression for the DNSP-specific trends, 2015 is the “knot” at which the slope of each function changes from δ_i to γ_i for DNSP i .¹⁰ We test this structural break with all the LSE specifications.

Figure 3.1 shows the industry average multilateral Törnqvist PFP indexes from the 2025 annual benchmarking analysis. The chosen knot year of 2015 corresponds with a year that could be viewed as a structural break (although another candidate may be 2012). Figure 3.1 shows that there is a second candidate break point, which is 2022. For this reason, we also test models that have two structural breaks (2015, 2022) with the LSE specifications.

It would be possible to include piecewise-linear efficiency time trends for Australian DNSPs in the BC95 and Kumb90 SFA specifications. However, Frontier Economics did not test this approach and we do not propose to test it either. Results for LSE models using piecewise regression are presented in *Attachment A: Time-Varying Inefficiency Models Results*.

Figure 3.1 DNSP industry partial factor productivity indexes, 2006–2024



3.3.3 Quadratic Time Trends

Another way of increasing the flexibility of the efficiency change, as an alternative to Frontier Economics’ GEC method, is to include quadratic functions of time trends rather than linear time trends. In the Phase 2 report, we did not test quadratic time trends because the more complex models tended to have more convergence problems. However, Frontier Economics’ method of choosing starting values reduces the convergence problems, and more complex

¹⁰ Spline functions are piecewise polynomial functions. Piecewise regression is a method of estimating linear piecewise functions that allow the relationship between variables to change at one or more specified points, known as knots or break points.

models can be tested. Further, in some cases the ‘quadratic trend’ models may be simpler than the GEC models.

Recall the basic cost model in equation (3.1), with variables as defined there. In the case of the BC95 model where $u_{it} \sim N^+(\mu_{it}, \sigma_u^2)$, the function determining μ_{it} can be generalised to:

$$\mu_{it} = \sum_j \delta_j(d_{it}^j t) + \sum_j \gamma_j(d_{it}^j t^2) \quad (3.7)$$

where j is a jurisdiction or an individual DNSP, depending on whether the JTT or AJTT set of time trends is used, d_{it}^j is a dummy variable equal to 1 if DNSP i is a member of j and 0 otherwise; t is the time variable; and δ_j and γ_j are parameters to be estimated.

In the case of the Kumb90 model, where $u_{it} = g(z_{it}) \cdot u_i$, and $u_i \sim N^+(0, \sigma_u^2)$, the function $g(z_{it})$ can be specified as a transformation of a quadratic function of t , similar to Kumbhakar’s original formulation:

$$g(z_{it}) = \left[1 + \exp\left(\delta_0 + \sum_j \delta_j(d_{it}^j t) + \sum_j \gamma_j(d_{it}^j t^2)\right) \right]^{-1} \quad (3.8)$$

The LSE model with quadratic inefficiency trend terms is in the spirit of Cornwell, Schmidt and Sickles (1990). Here, the cost function can be specified as:

$$c_{it} = \alpha'_0 + x'_{it}\beta + \sum_j \theta_j d_{it}^j + \sum_j \delta_j(d_{it}^j t) + \sum_j \gamma_j(d_{it}^j t^2) + v_{it} \quad (3.9)$$

Here, the inefficiency term u_{it} is defined as:

$$u_{it} = \sum_j \theta_j d_{it}^j + \sum_j \delta_j(d_{it}^j t) + \sum_j \gamma_j(d_{it}^j t^2) \\ - \min \left[\sum_j \theta_j d_{it}^j + \sum_j \delta_j(d_{it}^j t) + \sum_j \gamma_j(d_{it}^j t^2) \right]$$

Efficiency estimates are then obtained using $\exp(-u_{it})$. In the interests of finding models that are more flexible in terms of the time profiles of efficiency which they can accommodate, these models can be tested as possible alternative means of achieving such flexibility. Modelling results when using quadratic specifications are presented in section 4 of *Attachment A: Time-Varying Inefficiency Models Results*.

3.4 Cobb-Douglas, Hypothesis Tests and Parsimony

3.4.1 Should the Cobb-Douglas be retained?

This section discusses Frontier Economics’ argument that preference should be given to Translog models over Cobb-Douglas (CD) models. It is argued that CD models should only

be used as a fallback if the Translog models cannot be used due to serious statistical problems, because the CD models “fit the data more poorly than the Translog specifications. Therefore, these models are unnecessary provided the recommended Translog models can be estimated reliably” (Frontier Economics 2026, 5).

The key grounds for this conclusion are hypothesis tests of the joint statistical insignificance of the parameters on the higher-order output terms included in the Translog model, but not the CD models. That is, the test of whether these additional terms are jointly equal to zero versus the alternative hypothesis that at least one is significantly non-zero. Frontier Economics (2026, 64) states that such tests indicate that these higher-order terms are statistically significant “in 17 of the 18 models estimated”.

Frontier Economics has previously presented a similar argument with regard to the standard models. In the memo ‘Evoenergy – Benchmarking Limitations’ (2023a), Quantonomics disagreed with this argument because the CD models performed well against the parallel TLG models on some measures of goodness-of-fit, and hence the TLG model was not a better fit *in all cases*. This question needs to be re-examined in the context of time-varying inefficiency models.

Of key importance in this issue is the criterion of parsimony: If two models have the same explanatory performance, then the simpler of the two models should be preferred. This criterion is also discussed in the next section. Goodness-of-fit statistics often have a penalty for degrees of freedom lost. However, the penalty is not the same between different goodness-of-fit statistics, and need not be the same as that embodied in the hypothesis test of the joint insignificance of parameters. Hence, these statistics do not necessarily yield the same conclusion. Although a group of additional variables may not be jointly statistically insignificant, the degree to which they improve the explanatory power of the model may not exceed the penalty in the goodness-of-fit statistic for degrees of freedom lost. The Bayesian Information Criterion (BIC) is a preferred measure of fit, which appropriately penalises larger model size.¹¹ Importantly, the penalty for loss of parsimony in the BIC is greater than the 0.05 significance test used by Frontier Economics for the joint test of parameters. Although the BIC measure of fit is preferred, it is not available for the ‘panel-corrected standard errors’ estimation method used for the LSE models. Hence, the Pseudo-Adjusted R^2 statistic, which has a weaker penalty for loss of parsimony, is used to compare the fit of different LSE models, and also to

¹¹ Among the goodness-of-fit statistics we use here, the pseudo-Adjusted R^2 statistic has a comparatively small adjustment for loss of parsimony, whereas the BIC has a greater penalty for loss of parsimony. In light of its relatively weak penalty, Greene questions whether the adjusted R^2 “will necessarily lead the analyst to the correct model (assuming it is among the ones considered) as the sample size increases” (Greene 2012, 179). He observes that either the BIC or Akaike’s Information Criterion are to be preferred (Greene 2012, 180) because of their greater penalty. Cameron and Trivedi (2010, 359) state that the “BIC gives a larger model-size penalty and is generally better [than the AIC], especially if smaller models are desired.”

compare fit across model types, whereas BIC is used to compare fit within each of the two SFA model types considered.

Translog models may also be more prone to overfitting than CD models. In addition, insignificant second-order coefficients in the Translog specification, if retained, may affect the robustness of the results. In this context, the CD models provide less flexible but more parsimonious and possibly more robust estimates. In this report, we also consider a third alternative, which is to remove the insignificant second-order coefficients in the Translog models—here called ‘parsimonious Translog’. Because the insignificant coefficients may change from year to year as new data are added, applying this method in each annual update would require a *fixed procedure* for successively eliminating the insignificant second-order terms, rather than a fixed specification for the second-order terms.

The question of whether the CD model should be retained could be viewed as an empirical question based on statistical tests applied to the time-varying inefficiency models. We consider the following tests:

- Hypothesis tests of the joint insignificance of the additional Translog parameters.
- Comparative goodness-of-fit statistics for parallel CD and TLG models, focusing on the BIC measure that appropriately penalises loss of parsimony.
- The degree of multicollinearity, as indicated by variance inflation factors (VIFs). This is relevant because the purpose of the models is to draw inferences from the estimated parameters, and it is well known that high degrees of multicollinearity can make statistical inferences less reliable.

However, we note that empirical testing may not produce consistent results when the opex cost function modelling is updated annually, potentially changing the preferred models and affecting the averaging calculation. Accordingly, in addition to the statistical criteria outlined above, consistency over time, or robustness to sample changes, is also an important consideration when assessing whether the CD models should be retained. In this regard, retaining the CD models while also estimating the TLG models may provide a more stable modelling framework, particularly given that the latter is likely to produce excessive monotonicity violations for some DNSP at some point. We also note that, as shown in section 4.4 and in *Attachment A: Time-Varying Inefficiency Models Results*, the efficiency score rankings and trends broadly aligned across the TLG and CD models, and it might be argued that the differences in estimated values are not highly material. These considerations should be taken into account alongside the empirical test results when evaluating whether the Cobb-Douglas models should be retained.

3.4.2 Parsimony

Criterion of Parsimony

It was noted in section 2.2.2 that Frontier Economics does not give much weight to the criterion of parsimony. We have a different view. As noted by Greene (2012, 180): “All else given, simplicity does have some appeal” (Greene 2012, 180). For this reason, goodness-of-fit statistics such as BIC penalise loss of degrees of freedom, giving some weight to parsimony.

Our criterion of parsimony also includes that a “preferred model should also be straightforward to apply, e.g., with reliable software packages and minimal troubleshooting required” (Quantonomics 2025b, 24). This is a crucial requirement in the AER’s regulatory context. It is the aim of the benchmarking models that they can be reliably applied in successive years without needing to modify the models’ specifications (or in the case of the parsimonious models tested here, that there is a clearly defined procedure that is consistently applied when simplifying the model specifications). This supports confidence in the predictability and reliability of the efficiency assessments for which the models are used. More complicated models, chosen because they have marginally better performance in a given sample (possibly due to in-sample overfitting the data), may not be as robust relative to other specifications when applied over time, when additional data is added to the sample. There is a reasonable expectation that the criterion of parsimony is also consistent with this emphasis on robustness of the models over time.

Methods for Improving Parsimony

To date, it has been considered desirable to maintain consistent model specifications from year-to-year. An alternative approach, based on consistency of method, would allow certain insignificant higher-order terms to be dropped from a preferred TLG model. Retaining these terms may reduce the precision and reliability of the results, as coefficients estimated with low accuracy and high uncertainty can affect the overall performance of the model.

We therefore applied a procedure from Wang and Zelenyuk (2024) whereby a simplified TLG model is developed (hereafter called TLGP to indicate the parsimonious version). This parsimonious version of the TLG model is developed using the following procedure:

- (1) Simplification is limited to the higher-order TLG variables.
- (2) If there are any insignificant higher-order TLG coefficients, then the first step is to drop the variable with the smallest t-statistic in absolute value, and re-estimate the model.
- (3) This procedure is iterated until all of the remaining higher-order TLG terms are statistically significant.

We have also adopted this procedure for the Quadratic Models (QTT). Analogously, simplification is limited to the higher-order inefficient variables. If there are any insignificant

higher order efficient coefficients, then the first step is to drop the variable with the smallest t-statistic in absolute value, and re-estimate the model. This procedure is iterated until all of the remaining higher-order efficient terms are statistically significant.

3.5 Stability with sample changes

In the Phase 2 report, we evaluated the stability of the time-varying inefficiency models by using sub-periods and testing whether the models perform similarly as additional years of data are added to the sample. Frontier Economics rejects this method because:

We already know that the standard models are capable of producing reasonable-looking estimates for subsets of the historical data, even though the underlying model is mis-specified, since the incidence of statistical problems (e.g., monotonicity violations) have become more severe over time. As such, testing the capacity of the models to use subsets of old data (in effect, testing what the model would have outputted had it been used in 2019, for instance) does not address the primary concern. (Frontier Economics 2026, 62).

We disagree with this statement because the method we used tested the stability of the new time-varying models, not only the standard time-invariant models. It directly addresses the primary concern that a suitable model should not perform more poorly against the model evaluation criteria, including with regard to the frequency of MVs and the computational stability of the model as additional years of data are added to the sample.

Frontier Economics argued that a preferred model should be able to fit and explain new data. They proposed an alternative approach to assess stability to sample changes by simulating opex data for future years consistent with an assumed level of efficiency. They constructed artificial observations for additional years using the BC95-JTT model and assuming the efficiency scores of each DNSP remain constant. They then estimate the models with datasets that include this artificial data, and test how closely the estimated efficiency scores of the model correspond to the assumed efficiency scores in the simulated observations.

This method has severe shortcomings as discussed below. Further, whilst there may be merit in using a conventional form of out-of-sample testing, this would not involve testing the out-of-sample fit of efficiency scores because true efficiency scores are not known, as is also their distribution.

Out-of-sample validation is a way of testing the predictive performance of a model and its underlying assumptions. This typically involves excluding some of the sample data when estimating the model, then using the model to make predictions for the data excluded from the sample. The accuracy of these out-of-sample predictions can then be directly computed using goodness-of-fit statistics. Models with better out-of-sample prediction performance are naturally preferred over those with inferior out-of-sample predictions. After validation, the models are re-estimated with the full dataset.

In Economic Insights (2015), a 10-fold cross-validation procedure was used, which is a well-established approach in statistics. The models were estimated on randomly selected subsamples of the dataset. In each case, a randomly selected sub-sample representing 10 per cent of the observations was set aside for model validation, and the remaining observations were used for estimation. This random selection procedure created 10 subsamples, and in turn, each was excluded from the sample for estimation. Statistics were then reported separately for both the estimation and validation samples. It may be argued that setting aside data at the end of the sample period is a more stringent test. This is because when randomly selecting the subsample to be excluded, data on both sides of each observation are used to help determine the forecast. Furthermore, excluding the most recent period(s) has the advantage of likely being the most informative for future periods, so a model better at predicting the most recent period(s) may be better in predicting the future.

Frontier Economics' approach is inconsistent with these practical model validation procedures because it relied on artificial 'data' constructed using one of the models, and assumed constant efficiency scores in the out-of-sample period. In our view, the central purpose of out-of-sample testing is to test the model against *actual* data not used in estimation rather than simulated data, which may be sensitive to the selected assumptions of the simulated data generating process. Testing models against arbitrarily constructed 'data' can be useful for various theoretical explorations and could be implemented as well, as a complementary strategy. Such an approach must ensure that many alternative scenarios are explored to achieve robustness of conclusions, because there is no guarantee that any particular assumptions adopted for generating the data are consistent with the true data generation process.

The method we adopted of testing the models using various data sub-samples is well-suited to the question being addressed, which was whether the model performance against the criteria set out in the report is satisfactory as additional years of data are added to the sample. If not, there is a higher likelihood that there may be problems in estimating the model as additional data are added to the sample in future annual benchmarking reports. We continue to use the same method in this report.

3.6 Short sample period & existing specifications

This section addresses:

- whether the short sample period should be used with the time-varying inefficiency models
- whether the standard time-invariant inefficiency models should continue to be used, and
- whether the short sample period should continue to be used with the standard models.

3.6.1 Short sample period with time-varying inefficiency models

As stated in section 2.1, some submissions suggested that the short sample period (2012–2024) should not be used for time-varying models. This position is consistent with Quantonomics’ observation that:

... 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) datasets so that short-period analyses are less suitable in this context. (Quantonomics 2025b, iv)

For this reason, we recommend that the time-varying inefficiency models be estimated with the long sample period only.

3.6.2 Retention or Discontinuance of Standard Models

Frontier Economics argued that the existing time-invariant inefficiency models should be discontinued. This contrasted with Jemena’s view that the time-varying inefficiency models should, at least initially, supplement rather than replace the existing models (with the potential for eventual replacement pending ongoing performance).

We agree with Jemena’s view that the standard models should be continued alongside the time-varying inefficiency models until the latter are well established and have the confidence of the AER and stakeholders. This will enable the AER and stakeholders to compare results with previous benchmarking reports using the standard models, and since the AER has a statutory obligation to produce benchmarking results each November, it will reduce the risk that the AER would otherwise bear if the new models incur any issues with an expanded data sample that require troubleshooting. Once the time-varying inefficiency models are considered to be performing consistently well over some agreed transition period, the standard models can be discontinued.

3.6.3 Retention of short-period for standard models

We recommend that the short-period Translog models be discontinued. These are the models that have consistently produced the highest MVs, and the SFA versions of these models have recently encountered convergence issues. Both the LSE and SFA short-period Translog models should be discontinued. The short-period LSE and SFA Cobb-Douglas models could be retained for the time being as they appear to perform well in estimating period-average efficiencies for the more recent short period.

4 Preliminary evaluation of estimated models

In this study, the models presented in the Phase 2 report are re-estimated using Frontier Economics' suggested approaches, including modified time trends where needed, the corrected *linktest*, and Frontier Economics' starting values approach. Frontier Economics' proposed models using general efficiency change (GEC) are also replicated, namely:

- SFA-BC95-AGEC-HN-GTC, renamed in this report as BC95-AGECJTT-HN-GTC; and
- SFA-Kumb-AGECJUR-HN-GTC, renamed in this report as Kumb90-AGECJTT-HN-GTC.

Further, in responding to the criticism that some of the Phase 2 models had restrictive patterns of efficiency change, we have estimated several additional specifications for the inefficiency term, including a quadratic time trend (QTT) for inefficiency (rather than linear), and for the LSE models, spline or piecewise regression (PW) specifications, reproducing and extending Frontier Economics' use of this approach.

We revised the time-trend specifications by using final-year normalisation for the BC95 and LSE models, while retaining the centred trend for Kumb90.¹² For all GTC models, rather than omitting the first GTC dummy variable, we now omit the final GTC dummy variable.

All models were estimated for 2006 to 2023 using the Cobb-Douglas and Translog functional forms. As a third approach, we also estimated parsimonious TLG models. These are derived using a standardised method of simplification:

- If any of the higher-order log output terms of the TLG model are insignificant, then the term with the smallest t-statistic (in absolute value) is dropped, and the model re-estimated.
- This procedure is repeated with the resulting model, and this iterative procedure continues until all of the remaining higher-order log output terms are statistically significant.

The quadratic time trend models were estimated using the same iterative approach, retaining only statistically significant quadratic inefficiency trend terms.

In total, 33 different specifications were tested. Detailed model results are provided in *Attachment A, Time-Varying Inefficiency Models Results*.

Evaluation of the models is based on multi-criteria analysis against the evaluation criteria set out in section 2.3 of the Phase 2 report as modified in section 3 above. This section summarises the performance of each model against these criteria:

¹² See a discussion on Kumb90 time trend specifications in Appendix B.

- Convergence and computational issues in section 4.1
- Consistency with economic theory and industry knowledge in section 4.2
- Goodness-of-fit and performance on specification tests in section 4.3
- Consistency of efficiency score time-profiles in section 4.4
- Parsimony in section 4.5.
- Stability to sample changes is addressed in section 4.6.

Section 4.7 presents an overall evaluation based on the results in the preceding sections. The evaluation has been carried out using both rankings and scaled metrics, and different weighting schemes have been tested to combine evaluations across criteria. These approaches result in similar rankings of models. Section 4.7 also presents our overall findings on preferred models.

4.1 Convergence and Computational Issues

Frontier Economics' proposed revised starting-value approach materially improved computational performance. We also carried out robustness testing, assessed alternative starting values and included additional maximum likelihood optimisation settings, which provided some further improvements.¹³

Table 4.1.1 presents the performance of the tested models. The LSE models are not included in the table because none exhibited the issues assessed. A ✓ indicates that the model did not exhibit any issues in relation to the criterion assessed in the relevant column. A ! indicates that an issue was identified, but it raises only moderate concerns and is not, by itself, sufficient to rule out the model. A X indicates that a material issue was identified and the model should therefore be ruled out. The issues identified and assessed are:

- *Convergence*: An exclamation mark indicates that the model reported the message “cannot compute an improvement, flat region encountered” after the iteration process. This warning is not treated as a serious problem where the model still produces complete and plausible estimates, as it may simply indicate that the likelihood function is relatively flat near the reported solution. Nevertheless, it is flagged because it may suggest some uncertainty regarding the precision or stability of the optimisation result.
- *Missing estimates*: A cross indicates that the model failed to estimate one or more coefficients, standard errors or z-statistics, resulting in missing values. This is treated as a serious problem because it prevents a complete assessment of the model

¹³ We also assessed the time-invariant Cobb-Douglas models used to generate starting values for the time-varying specifications. These models converged properly in all cases and showed no evidence of computational problems.

parameters and associated statistical inference and may indicate estimation failure or inadequate parameter identification.

- *Unusually High Coefficient Value:* Assesses whether the model produces unusually large coefficient estimates. Where identified with an exclamation mark, the issue relates to DNSP-specific time-trend (from AJTT models) coefficients exceeding 1.¹⁴ This is not treated as a serious problem by itself, but is flagged because unusually large estimates may indicate weak identification.
- *Large Standard Errors:* Assesses whether the model produces unusually large standard errors for any coefficient other than the constant term. An issue is identified with an exclamation mark where a standard error exceeds 1. This is not treated as a serious problem by itself, but is flagged because large standard errors indicate imprecisely estimated coefficients and may reflect weak identification or multicollinearity.

In Phase 2, for the SFA models, only one BC95 model and four Kumb90 specifications were successfully estimated for 2006 to 2023. Under the revised approach, including the new specifications tested:

- Five BC95 and four Kumb90 specifications fully satisfied the computational performance and convergence criteria.
- Six Kumb90 specifications produced partially satisfactory results.
- Three BC95 specifications failed the assessment. In addition to producing unusually large inefficiency time-trend coefficients, the specifications failed to estimate standard errors and z-statistics for some of these coefficients. The BC95-AJTT-HN specification also generated a flat-region warning following optimisation. For this reason, these three models are excluded from the assessments that follow.

Overall, the revised approach substantially increased the number of estimable models, but the specifications affected by potential identification issues or abnormal standard errors should not be preferred where more stable alternatives are available.

¹⁴ In these LSE models, where the dependent variable is in log form, the coefficient on the time trend is related to the negative of the annual *rate of change* of the efficiency score. In light of the typically estimated range of efficiency scores, and since they are upper-bounded at 1, such exceptionally large increases (in the order of 100 per cent in a single year) are not considered economically plausible.

Table 4.1.1: Convergence and Computational Issues

<i>Model</i>	<i>Functional Form</i>	<i>Convergence</i>	<i>Missing Estimates</i>	<i>High Coefficient</i>	<i>Large SE</i>
<u><i>Revised Phase 2 Models</i></u>					
BC95-JTT-HN	CD	✓	✓	✓	✓
	TLG	✓	✓	✓	✓
	TLGP	✓	✓	✓	✓
BC95-JTT-HN-GTC	CD	✓	✓	✓	✓
	TLG	✓	✓	✓	✓
	TLGP	✓	✓	✓	✓
BC95-AJTT-HN	CD	!	X	!	✓
	TLG	!	X	!	✓
	TLGP#				
BC95-AJTT-HN-GTC	CD	✓	X	!	✓
	TLG	✓	X	!	✓
	TLGP#				
Kumb90-JTT-HN	CD	✓	✓	✓	✓
	TLG	✓	✓	✓	✓
	TLGP	✓	✓	✓	✓
Kumb90-JTT-HN-GTC	CD	✓	✓	✓	✓
	TLG	✓	✓	✓	✓
	TLGP	✓	✓	✓	✓
Kumb90-AJTT-HN	CD	✓	✓	!	✓
	TLG	✓	✓	!	!
	TLGP	✓	✓	!	!
Kumb90-AJTT-HN-GTC	CD	✓	✓	!	✓
	TLG	✓	✓	!	!
	TLGP	✓	✓	!	!
<u><i>Generalised Efficiency Change</i></u>					
BC95-AGECJTT-HN	CD	✓	✓	✓	✓
	TLG	✓	✓	✓	✓
	TLGP	✓	✓	✓	✓
BC95-AGECJTT-HN-GTC	CD	✓	✓	✓	✓
	TLG	✓	✓	✓	✓
	TLGP	✓	✓	✓	✓
Kumb90-AGECJTT-HN	CD	✓	✓	✓	!
	TLG	✓	✓	✓	!
	TLGP	✓	✓	✓	!
Kumb90-AGECJTT-HN-GTC	CD	✓	✓	✓	!
	TLG	✓	✓	✓	!
	TLGP	✓	✓	✓	!

Table 4.1.1: (cont.)

<i>Model</i>	<i>Functional Form</i>	<i>Convergence</i>	<i>Missing Estimates</i>	<i>High Coefficient</i>	<i>Large SE</i>
<i>Quadratic Time Trend Models</i>					
BC95-JTT-HN-QTT	CD	✓	✓	✓	✓
	TLG	✓	✓	✓	✓
	TLGP	✓	✓	✓	✓
BC95-JTT-HN-GTC-QTT	CD	✓	✗	!	✓
	TLG	✓	✗	!	✓
	TLGP	✓	✗	✓	✓
Kumb90-JTT-HN-QTT	CD	✓	✓	✓	✓
	TLG	✓	✓	✓	✓
	TLGP	✓	✓	✓	✓
Kumb90-JTT-HN-GTC-QTT	CD	✓	✓	✓	!
	TLG	✓	✓	✓	!
	TLGP	✓	✓	✓	!
Kumb90-AJTT-HN-QTT	CD	✓	✓	✓	✓
	TLG	✓	✓	✓	✓
	TLGP	✓	✓	✓	✓
Kumb90-AJTT-HN-GTC-QTT	CD	✓	✓	!	!
	TLG	✓	✓	!	!
	TLGP*				

Notes: *The TLGP model is not included because all second-order terms in the TLG specification are statistically significant.[#] The TLGP specification was not pursued because both the corresponding CD and full TLG specifications failed.

4.2 Consistency with Economic Theory or Industry Knowledge

Table 4.2.1 reports the main output coefficients and their sum, which indicates returns to scale. These represent output elasticities, which for the CD models are constant. For the TLG and TLGP models, the output elasticities vary across observations, and the values presented in Table 4.2.1 are those evaluated at the sample mean output values. This means that the results presented in the Table are local results at the sample mean, rather than sample-wide scale properties.

In Table 4.2.1, the models rejected in Section 4.1 are not included. The main findings are:

- All coefficients are statistically significant at the 5 per cent level.
- Customer numbers is the dominant cost driver in 87 of 89 models, with coefficients ranging from 0.380 to 0.646. The exceptions are the Translog BC95-JTT-HN and BC95-AGECJTT-HN models, where RMD is dominant and Customers ranks second.
- The RMD coefficient ranges from 0.112 to 0.407. It is the second-largest driver in 69 models and the least in 18 models. The Circuit length coefficient ranges from 0.124 to 0.283. It ranks second in 19 models and least in 70 models.

- The sum of the coefficients, representing returns to scale, ranges from 0.954 to 1.009. It is slightly above 1.0 in three Translog models: Kumb90-AJTT-HN, Kumb90-AJTT-HN-GTC, and Kumb90-AJTT-HN-GTC-QTT.

Table 4.2.1: Output coefficients and returns to scale

<i>Model</i>	<i>Functional Form</i>	<i>Cust Num</i>	<i>CircLen</i>	<i>RMD</i>	<i>RTS</i>
<u><i>Revised Phase 2 Models</i></u>					
BC95-JTT-HN	CD	0.516	0.161	0.297	0.973
	TLG	0.400	0.148	0.407	0.956
	TLGP	0.408	0.154	0.393	0.955
BC95-JTT-HN-GTC	CD	0.533	0.138	0.309	0.980
	TLG	0.433	0.132	0.395	0.960
	TLGP	0.434	0.124	0.403	0.961
Kumb90-JTT-HN	CD	0.470	0.213	0.291	0.974
	TLG	0.489	0.223	0.275	0.987
	TLGP	0.487	0.222	0.277	0.986
Kumb90-JTT-HN-GTC	CD	0.463	0.224	0.286	0.973
	TLG	0.505	0.222	0.260	0.987
	TLGP	0.528	0.207	0.252	0.986
Kumb90-AJTT-HN [#]	CD	0.586	0.153	0.253	0.992
	TLG	0.619	0.221	0.168	1.008
	TLGP	0.564	0.203	0.225	0.992
Kumb90-AJTT-HN-GTC [#]	CD	0.588	0.165	0.239	0.992
	TLG	0.640	0.221	0.148	1.009
	TLGP	0.564	0.198	0.229	0.991
LSE-ADTT	CD	0.544	0.227	0.194	0.964
	TLG	0.380	0.230	0.345	0.954
	TLGP	0.394	0.233	0.328	0.955
LSE-ADTT-GTC	CD	0.575	0.223	0.170	0.967
	TLG	0.432	0.225	0.300	0.957
	TLGP	0.437	0.224	0.296	0.957
LSE-AJTT	CD	0.616	0.217	0.134	0.967
	TLG	0.470	0.216	0.269	0.955
	TLGP	0.479	0.216	0.261	0.956
LSE-AJTT-GTC	CD	0.646	0.211	0.112	0.969
	TLG	0.527	0.208	0.222	0.957
	TLGP	0.524	0.208	0.225	0.957
LSE-AJT _{Tx} -GTC ⁽¹⁾	CD	0.636	0.209	0.124	0.968
	TLG	0.502	0.206	0.249	0.957
	TLGP	0.505	0.207	0.255	0.958
<u><i>Generalised Efficiency Change</i></u>					
BC95-AGECJTT-HN	CD	0.518	0.168	0.286	0.972
	TLG	0.393	0.161	0.400	0.954
	TLGP	0.398	0.165	0.391	0.954

Table 4.2.1: (cont.)

<i>Model</i>	<i>Functional Form</i>	<i>Cust Num</i>	<i>CircLen</i>	<i>RMD</i>	<i>RTS</i>
BC95-AGECJTT-HN-GTC	CD	0.530	0.144	0.304	0.978
	TLG	0.421	0.142	0.395	0.958
	TLGP	0.429	0.141	0.389	0.959
Kumb90-AGECJTT-HN #	CD	0.537	0.183	0.271	0.992
	TLG	0.504	0.214	0.280	0.998
	TLGP	0.546	0.185	0.267	0.998
Kumb90-AGECJTT-HN-GTC #	CD	0.573	0.186	0.232	0.991
	TLG	0.560	0.206	0.229	0.995
	TLGP	0.612	0.175	0.209	0.996
<i>Quadratic Time Trend Models</i>		-			
BC95-JTT-HN-QTT	CD	0.533	0.146	0.300	0.979
	TLG	0.438	0.134	0.390	0.962
	TLGP	0.437	0.132	0.393	0.962
Kumb90-JTT-HN-QTT	CD	0.481	0.210	0.286	0.976
	TLG	0.508	0.219	0.263	0.991
	TLGP	0.529	0.196	0.267	0.992
Kumb90-JTT-HN-GTC-QTT#	CD	0.453	0.216	0.300	0.969
	TLG	0.518	0.231	0.238	0.988
	TLGP	0.528	0.229	0.229	0.987
Kumb90-AJTT-HN-QTT	CD	0.597	0.149	0.248	0.994
	TLG	0.611	0.187	0.196	0.993
	TLGP	0.545	0.207	0.237	0.989
Kumb90-AJTT-HN-GTC-QTT#	CD	0.380	0.227	0.370	0.977
	TLG	0.549	0.283	0.170	1.002
LSE-ADTT-QTT	CD	0.549	0.225	0.190	0.965
	TLG	0.391	0.227	0.337	0.955
	TLGP	0.410	0.227	0.319	0.956
LSE-ADTT-GTC-QTT	CD	0.577	0.222	0.168	0.967
	TLG	0.439	0.223	0.295	0.957
	TLGP	0.442	0.224	0.291	0.957
LSE-AJTT-QTT	CD	0.623	0.215	0.129	0.967
	TLG	0.481	0.214	0.261	0.956
	TLGP	0.487	0.214	0.255	0.956
LSE-AJTT-GTC-QTT	CD	0.641	0.208	0.120	0.969
	TLG	0.534	0.207	0.217	0.958
	TLGP	0.530	0.207	0.221	0.958
LSE-AJT Tx-QTT ⁽²⁾	CD	0.644	0.205	0.120	0.969
	TLG	0.480	0.209	0.267	0.956
	TLGP	0.487	0.209	0.260	0.956
LSE-AJT Tx-GTC-QTT ⁽³⁾	CD	0.632	0.205	0.131	0.969
	TLG	0.510	0.205	0.243	0.958
	TLGP	0.512	0.207	0.240	0.958

Table 4.2.1: (cont.)

<i>Model</i>	<i>Functional Form</i>	<i>Cust Num</i>	<i>CircLen</i>	<i>RMD</i>	<i>RTS</i>
<i>Piecewise Regression Models</i>					
LSE-ADTT-PW-2015	CD	0.574	0.212	0.181	0.967
	TLG	0.422	0.212	0.323	0.958
	TLGP	0.437	0.214	0.308	0.959
LSE-ADTT-PW-2015-2022	CD	0.576	0.211	0.180	0.968
	TLG	0.421	0.211	0.325	0.958
	TLGP	0.437	0.213	0.309	0.959
LSE-AJTT-PW-2015	CD	0.638	0.205	0.127	0.969
	TLG	0.508	0.201	0.249	0.959
	TLGP	0.512	0.203	0.245	0.959
LSE-AJTT-PW-2015-2022	CD	0.639	0.204	0.127	0.970
	TLG	0.508	0.200	0.251	0.959
	TLGP	0.511	0.202	0.246	0.959

Notes: # As noted in Section 4.1, the model may be subject to identification issues. ⁽¹⁾ The LSE-AJTTx-GTC model differs from the LSE-AJTT-GTC model by including an additional Ontario-specific trend. ⁽²⁾ The LSE-AJTTx-QTT differs from the LSE-AJTT-QTT model by excluding the generic time trend ('yr') and including an Ontario-specific trend (with efficiency time trends entering the model quadratically). ⁽³⁾ The LSE-AJTTx-GTC-QTT model differs from the LSE-AJTT-GTC-QTT model by including an additional Ontario-specific trend (with efficiency time trends entering the model quadratically).

Table 4.2.2 reports the estimated coefficient and statistical significance of the underground-share variable (UGShare) across the models tested. A ✓ indicates that the UGShare coefficient is negative and an X indicates that it is positive, regardless of statistical significance. Main comments are:

- UGShare is positive in six of the 89 functional forms in Table 4.2.2, all from the Kumb90 family and using a TLG or TLGP specification. In all cases, the coefficient is not statistically significant at the 5 per cent level. In three instances, the coefficient is positive in the TLG model but becomes negative under the parsimonious Translog specification, although it remains statistically insignificant. These six functional forms, where the UGShare is positive, were excluded from further assessment.
- UGShare is negative but not statistically significant at the 5 per cent level in 19 models, all from the Kumb90 family. It is negative and statistically significant in the 64 remaining models.
- The Kumb90-JTT-HN-GTC-QTT and Kumb90-AJTT-HN-GTC-QTT specifications produced satisfactory results only under the CD functional form.

Table 4.2.2: Coefficient and significance of underground share

<i>Model</i>	<i>Functional. Form</i>	<i>UGShare</i>	<i>z-stat</i>	<i>Assessment</i>
<u><i>Revised Phase 2 Models</i></u>				
BC95-JTT-HN	CD	-0.128	-7.37	✓
	TLG	-0.122	-5.94	✓
	TLGP	-0.112	-6.59	✓
BC95-JTT-HN-GTC	CD	-0.152	-8.00	✓
	TLG	-0.138	-5.53	✓
	TLGP	-0.152	-7.61	✓
Kumb90-JTT-HN	CD	-0.063	-2.21	✓
	TLG	-0.017	-0.53	✓
	TLGP	-0.024	-0.76	✓
Kumb90-JTT-HN-GTC	CD	-0.037	-1.27	✓
	TLG	0.009	0.29	✗
	TLGP	-0.000	-0.01	✓
Kumb90-AJTT-HN [#]	CD	-0.065	-2.12	✓
	TLG	-0.011	-0.34	✓
	TLGP	-0.032	-1.13	✓
Kumb90-AJTT-HN-GTC [#]	CD	-0.030	-0.92	✓
	TLG	0.015	0.47	✗
	TLGP	-0.009	-0.31	✓
LSE-ADTT	CD	-0.091	-3.56	✓
	TLG	-0.102	-3.62	✓
	TLGP	-0.090	-3.45	✓
LSE-ADTT-GTC	CD	-0.092	-4.14	✓
	TLG	-0.093	-3.70	✓
	TLGP	-0.091	-3.89	✓
LSE-AJTT	CD	-0.100	-4.39	✓
	TLG	-0.103	-4.10	✓
	TLGP	-0.100	-4.24	✓
LSE-AJTT-GTC	CD	-0.104	-5.00	✓
	TLG	-0.101	-4.33	✓
	TLGP	-0.102	-4.71	✓
LSE-AJT Tx-GTC	CD	-0.109	-5.17	✓
	TLG	-0.111	-4.70	✓
	TLGP	-0.108	-4.92	✓
<u><i>Generalised Efficiency Change</i></u>				
BC95-AGECJTT-HN	CD	-0.122	-6.90	✓
	TLG	-0.110	-5.36	✓
	TLGP	-0.104	-6.16	✓

Table 4.2.2: (cont.)

<i>Model</i>	<i>Functional. Form</i>	<i>UGShare</i>	<i>z-stat</i>	<i>Assessment</i>
BC95-AGECJTT-HN-GTC	CD	-0.146	-7.69	✓
	TLG	-0.127	-5.55	✓
	TLGP	-0.133	-7.26	✓
Kumb90-AGECJTT-HN [#]	CD	-0.056	-1.81	✓
	TLG	-0.013	-0.37	✓
	TLGP	-0.032	-0.94	✓
Kumb90-AGECJTT-HN-GTC [#]	CD	-0.041	-1.31	✓
	TLG	0.008	0.22	✗
	TLGP	-0.011	-0.34	✓
<i>Quadratic Time Trend Models</i>				
BC95-JTT-HN-QTT	CD	-0.145	-8.88	✓
	TLG	-0.145	-7.63	✓
	TLGP	-0.147	-8.69	✓
Kumb90-JTT-HN-QTT	CD	-0.064	-2.26	✓
	TLG	-0.015	-0.47	✓
	TLGP	-0.034	-1.13	✓
Kumb90-JTT-HN-GTC-QTT [#]	CD	-0.043	-1.51	✓
	TLG	0.016	0.46	✗
	TLGP	0.014	0.42	✗
Kumb90-AJTT-HN-QTT	CD	-0.064	-2.11	✓
	TLG	-0.017	-0.50	✓
	TLGP	-0.023	-0.68	✓
Kumb90-AJTT-HN-GTC-QTT [#]	CD	-0.033	-1.14	✓
	TLG	0.046	1.44	✗
	TLGP	-0.092	-3.65	✓
LSE-ADTT-QTT	CD	-0.103	-3.71	✓
	TLG	-0.094	-3.59	✓
	TLGP	-0.093	-4.16	✓
LSE-ADTT-GTC-QTT	CD	-0.093	-3.74	✓
	TLG	-0.091	-3.92	✓
	TLGP	-0.113	-5.49	✓
LSE-AJTT-QTT	CD	-0.110	-4.51	✓
	TLG	-0.107	-4.68	✓
	TLGP	-0.113	-5.58	✓
LSE-AJTT-GTC-QTT	CD	-0.112	-4.74	✓
	TLG	-0.108	-4.95	✓
	TLGP	-0.113	-5.49	✓
LSE-AJT Tx-QTT	CD	-0.110	-4.51	✓
	TLG	-0.107	-4.68	✓
	TLGP	-0.113	-5.58	✓
LSE-AJT Tx-GTC-QTT	CD	-0.112	-4.74	✓
	TLG	-0.108	-4.95	✓
	TLGP	-0.113	-5.49	✓

Table 4.2.2: (cont.)

<i>Model</i>	<i>Functional Form</i>	<i>UGShare</i>	<i>z-stat</i>	<i>Assessment</i>
<i>Piecewise Regression Models</i>				
LSE-ADTT-PW-2015	CD	-0.108	-4.85	✓
	TLG	-0.119	-4.85	✓
	TLGP	-0.109	-4.77	✓
LSE-ADTT-PW-2015-2022	CD	-0.109	-5.00	✓
	TLG	-0.120	-4.99	✓
	TLGP	-0.111	-4.91	✓
LSE-AJTT-PW-2015	CD	-0.114	-5.69	✓
	TLG	-0.118	-5.35	✓
	TLGP	-0.114	-5.55	✓
LSE-AJTT-PW-2015-2022	CD	-0.115	-5.84	✓
	TLG	-0.119	-5.50	✓
	TLGP	-0.115	-5.70	✓

Notes: # As noted in Section 4.1, the model may be subject to identification issues.

Table 4.2.3 reports the frequency of MVs across the TLG and TLGP models that were not ruled out in Tables 4.1.1, 4.2.1 and 4.2.2. CD models are not presented because their output elasticities are fixed and, as shown in Table 4.2.1, none produce negative output elasticities and hence have no MVs.

Table 4.2.3: Frequency of Monotonicity Violations (MV's)

<i>Model</i>	<i>Functional Form</i>	<i>Australia Sample (%)</i>	<i>Total Sample (%)</i>	<i>No. Aust. DNSPs MV*</i>	<i>Assessment</i>
<i>Revised Phase 2 Models</i>					
BC95-JTT-HN	TLG	8.5	14.9	1	!
	TLGP	0.0	13.1	0	✓
BC95-JTT-HN-GTC	TLG	0.4	13.4	0	✓
	TLGP	0.0	13.2	0	✓
Kumb90-JTT-HN	TLG	0.0	0.0	0	✓
	TLGP	0.0	0.0	0	✓
Kumb90-JTT-HN-GTC	TLG	0.0	0.0	0	✓
	TLGP	0.0	0.0	0	✓
Kumb90-AJTT-HN#	TLG	55.1	27.0	7	X
	TLGP	0.0	5.8	0	✓
Kumb90-AJTT-HN-GTC#	TLG	0.0	7.9	0	✓
	TLGP	0.0	7.9	0	✓
LSE-ADTT	TLG	11.1	18.7	1	!
	TLGP	3.4	14.4	0	✓
LSE-ADTT-GTC	TLG	0.0	16.7	0	✓
	TLGP	0.0	15.9	0	✓
LSE-AJTT	TLG	8.1	19.9	1	!
	TLGP	7.7	19.7	1	!
LSE-AJTT-GTC	TLG	8.1	22.9	1	!
	TLGP	8.5	23.1	1	!
LSE-AJTtx-GTC	TLG	8.5	21.4	1	!
	TLGP	8.5	21.8	1	!

Table 4.2.3: (cont.)

<i>Model</i>	<i>Functional Form</i>	<i>Australia Sample (%)</i>	<i>Total Sample (%)</i>	<i>No. Aust. DNSPs MV*</i>	<i>Assessment</i>
<i>Generalised Efficiency Change</i>					
BC95-AGECJTT-HN	TLG	5.1	14.4	1	!
	TLGP	0.0	13.3	0	✓
BC95-AGECJTT-HN-GTC	TLG	1.3	13.6	0	✓
	TLGP	15.4	6.0	1	!
Kumb90-AGECJTT-HN [#]	TLG	0.0	0.0	0	✓
	TLGP	0.0	0.0	0	✓
Kumb90-AGECJTT-HN-GTC [#]	TLGP	0.0	0.0	0	✓
<i>Quadratic Time Trend Models</i>					
BC95-JTT-HN-QTT	TLG	3.4	14.1	0	✓
	TLGP	3.8	14.2	1	!
Kumb90-JTT-HN-QTT	TLG	0.0	0.0	0	✓
	TLGP	0.0	0.0	0	✓
Kumb90-AJTT-HN-QTT [#]	TLG	56.0	22.5	7	X
	TLGP	17.5	6.0	2	!
LSE-ADTT-QTT	TLG	5.6	18.2	1	!
	TLGP	3.4	17.0	0	✓
LSE-ADTT-GTC-QTT	TLG	0.4	15.9	0	✓
	TLGP	0.4	15.2	0	✓
LSE-AJTT-QTT	TLG	8.5	20.2	1	!
	TLGP	7.7	20.1	1	!
LSE-AJTT-GTC-QTT	TLG	8.5	23.3	1	!
	TLGP	10.3	24.0	1	!
LSE-AJTtx-QTT	TLG	7.7	19.9	1	!
	TLGP	7.7	20.1	1	!
LSE-AJTtx-GTC-QTT	TLG	8.5	21.8	1	!
	TLGP	15.4	23.5	2	!
<i>Piecewise Regression Models</i>					
LSE-ADTT-PW-2015	TLG	7.7	21.0	1	!
	TLGP	7.3	18.6	1	!
LSE-ADTT-PW-2015-2022	TLG	7.7	20.9	1	!
	TLGP	6.8	18.4	1	!
LSE-AJTT-PW-2015	TLG	9.0	21.8	1	!
	TLGP	15.4	23.0	2	!
LSE-AJTT-PW-2015-2022	TLG	9.0	21.7	1	!
	TLGP	15.4	23.0	2	!

Notes: * Number of Australian DNSPs where MVs occur in more than 50 per cent of the observations;

As noted in Section 4.1, the model may be subject to identification issues.

As set out in Section 2.2.2, we reject models only where MVs occur in more than 50 per cent of the observations for at least 7 of the 13 Australian DNSPs. However, we flag a model as having moderate monotonicity issues where either: (i) one or more Australian DNSPs have

MVs in 50 per cent or more of their observations; and/or (ii) the model has MVs in more than 25 per cent of Australian observations and is not rejected on the first criterion. A flagged model should not be preferred where alternatives are available.

Table 4.2.3 shows that:

- Of the 53 reported specifications, 22 are accepted, 29 are flagged as having moderate monotonicity issues, and two are rejected. The two rejected specifications are the TLG versions of Kumb90-AJTT-HN and Kumb90-AJTT-HN-QTT.
- The parsimonious Translog specification tends to improve, or at least mostly not harm, monotonicity performance:
 - Across the 25 paired TLG and TLGP models, TLGP improves the overall assessment for six models compared to TLG, leaves the assessment unchanged for 17, and worsens the assessment for two: BC95-AGECJTT-HN-GTC and BC95-JTT-HN-QTT. Examples are the improvements obtained in the worst-performing models: (i) for Kumb90-AJTT-HN, the Australian MV rate falls from 55.1 per cent to zero, and the number of DNSPs with excessive MVs falls from seven to zero; (ii) for Kumb90-AJTT-HN-QTT, the Australian MV rate falls from 56.0 per cent to 17.5 per cent, and the number of DNSPs with excessive MVs falls from seven to two.
 - TLGP reduces the proportion of Australian observations with MVs (compared to TLG) in 11 of the 25 paired models (unchanged in seven, higher in seven).
- GTC improves monotonicity for several BC95 and LSE-ADTT specifications.¹⁵
 - For example, the Australian MV rate falls from 8.5 to 0.4 per cent between TLG BC95-JTT-HN and BC95-JTT-HN-GTC, and from 11.1 per cent to zero between TLG LSE-ADTT and LSE-ADTT-GTC. The corresponding TLGP models in the BC95 case both have zero Australian MVs, whereas in the LSE-ADTT case, it improves from 3.4 per cent to zero per cent.
 - For the Kumb90-AJTT specification, GTC TLGP models retain zero Australian MVs.
- The main monotonicity concerns are concentrated in the Kumb90-AJTT TLG specifications, including their quadratic time trend versions. In contrast, the Kumb90-JTT specifications generally perform well on this criterion.

¹⁵ There are some instances in which the inclusion of GTC variables increases the frequency of MVs, particularly in certain Kumb90 quadratic specifications. However, these specifications were ruled out in the preceding assessment sections.

- None of the 30 LSE specifications shown in Table 4.2.3 is rejected. However, 24 are flagged as having moderate monotonicity issues. Only six are accepted without concerns. These flags generally arise because one or two DNSPs exceed the DNSP-level threshold, rather than because MVs are widespread across the Australian sample.

4.3 Goodness-of-fit & Performance on Specification Tests

Table 4.3.1 presents the results of goodness-of-fit tests across models that were not ruled out in previous sections.¹⁶

Table 4.3.1: Goodness-of-fit Results

<i>Model</i>	<i>Functional Form</i>	<i>BIC</i>	<i>RMSE</i>	<i>Pseudo-adjusted R²</i>
<i>Revised Phase 2 Models</i>				
BC95-JTT-HN	CD	-346.5	0.051	0.999
	TLG	-419.0	0.054	0.998
	TLGP	-432.2	0.053	0.998
BC95-JTT-HN-GTC	CD	-352.5	0.056	0.998
	TLG	-427.2	0.052	0.998
	TLGP	-440.4	0.057	0.998
Kumb90-JTT-HN	CD	-1503.3	0.101	0.994
	TLG	-1477.2	0.100	0.994
	TLGP	-1482.3	0.101	0.994
Kumb90-JTT-HN-GTC	CD	-1441.8	0.103	0.994
	TLGP	-1436.8	0.102	0.994
Kumb90-AJTT-HN [#]	CD	-1470.6	0.098	0.994
	TLGP	-1470.3	0.098	0.994
Kumb90-AJTT-HN-GTC [#]	CD	-1410.8	0.100	0.994
	TLGP	-1409.4	0.100	0.994
LSE-ADTT	CD		0.186	0.980
	TLG		0.173	0.982
	TLGP		0.173	0.982
LSE-ADTT-GTC	CD		0.185	0.980
	TLG		0.172	0.983
	TLGP		0.172	0.983
LSE-AJTT	CD		0.181	0.981
	TLG		0.167	0.984
	TLGP		0.167	0.984
LSE-AJTT-GTC	CD		0.178	0.981
	TLG		0.165	0.984
	TLGP		0.165	0.984
LSE-AJT Tx-GTC	CD		0.178	0.981
	TLG		0.164	0.984
	TLGP		0.164	0.984

¹⁶ The residual plots for each tested model are presented in Attachment A.

Table 4.3.1: (cont.)

<i>Model</i>	<i>Functional Form</i>	<i>BIC</i>	<i>RMSE</i>	<i>Pseudo-adjusted R²</i>
<u><i>Generalised Efficiency Change</i></u>				
BC95-AGECJTT-HN	CD	-335.9	0.050	0.999
	TLG	-412.1	0.052	0.998
	TLGP	-425.8	0.051	0.998
BC95-AGECJTT-HN-GTC	CD	-333.2	0.055	0.998
	TLG	-409.6	0.049	0.999
	TLGP	-404.3	0.054	0.998
Kumb90-AGECJTT-HN [#]	CD	-1433.3	0.103	0.994
	TLG	-1403.0	0.102	0.994
	TLGP	-1426.0	0.103	0.994
Kumb90-AGECJTT-HN-GTC [#]	CD	-1420.1	0.103	0.994
	TLGP	-1415.5	0.102	0.994
<u><i>Quadratic Time Trend Models</i></u>				
BC95-JTT-HN-QTT	CD	-368.1	0.054	0.998
	TLG	-444.5	0.063	0.998
	TLGP	-458.4	0.063	0.998
Kumb90-JTT-HN-QTT	CD	-1492.4	0.101	0.994
	TLG	-1468.7	0.100	0.994
	TLGP	-1492.7	0.100	0.994
Kumb90-JTT-HN-GTC-QTT [#]	CD	-1453.0	0.102	0.994
Kumb90-AJTT-HN-QTT	CD	-1465.2	0.097	0.995
	TLGP	-1439.0	0.096	0.995
Kumb90-AJTT-HN-GTC-QTT	CD	-1411.9	0.098	0.994
LSE-ADTT-QTT	CD		0.186	0.980
	TLG		0.172	0.983
	TLGP		0.173	0.983
LSE-ADTT-GTC-QTT	CD		0.185	0.980
	TLG		0.172	0.983
	TLGP		0.172	0.983
LSE-AJTT-QTT	CD		0.180	0.982
	TLG		0.167	0.984
	TLGP		0.167	0.984
LSE-AJTT-GTC-QTT	CD		0.177	0.982
	TLG		0.165	0.984
	TLGP		0.165	0.984
LSE-AJT Tx-QTT	CD		0.176	0.982
	TLG		0.164	0.984
	TLGP		0.164	0.984
LSE-AJT Tx-GTC-QTT	CD		0.176	0.982
	TLG		0.164	0.984
	TLGP		0.164	0.984

Table 4.3.1: (cont.)

<i>Model</i>	<i>Functional Form</i>	<i>BIC</i>	<i>RMSE</i>	<i>Pseudo-adjusted R²</i>
<i>Piecewise Regression Models</i>				
LSE-ADTT-PW-2015	CD		0.098	0.981
	TLG		0.096	0.983
	TLGP		0.096	0.983
LSE-ADTT-PW-2015-2022	CD		0.099	0.980
	TLG		0.097	0.983
	TLGP		0.096	0.983
LSE-AJTT-PW-2015	CD		0.098	0.982
	TLG		0.095	0.984
	TLGP		0.095	0.984
LSE-AJTT-PW-2015-2022	CD		0.099	0.982
	TLG		0.096	0.984
	TLGP		0.096	0.984

Notes: (1) Here, RMSE refers to the sample root mean squared residual, calculated as the square root of the sample mean of the squared differences between observed and fitted log costs. For the LSE models, fitted log costs are obtained from the model's linear prediction. The same calculation is used for the SFA models, where the residual is the composite residual relative to the estimated cost frontier and therefore reflects both statistical noise and inefficiency, rather than the noise component alone.

As noted in Section 4.1, the model may be subject to identification issues.

The main comments are:

- Unlike the other metrics shown in Table 4.3.1, RMSE does not include any penalty for additional variables. However, it helps to shed light on the other goodness-of-fit results. RMSE values range from 0.049 to 0.186. The BC95 models generally record the lowest RMSE values, with RMSE values between 0.049 and 0.063. The piecewise LSE models also perform relatively well, with RMSE values between 0.095 and 0.099. This is broadly comparable to or slightly lower than the Kumb90 models, which range from 0.096 to 0.103. The standard and quadratic LSE models record the highest RMSE values, ranging from 0.164 to 0.186. Within the LSE family, the TLG and TLGP specifications consistently record lower RMSE values than the corresponding CD models.
- Pseudo-adjusted R^2 values are high across all model families, ranging from 0.980 to 0.999. The BC95 models generally provide the strongest in-sample fit, with values between 0.998 and 0.999, followed by the Kumb90 models, with values between 0.994 and 0.995, and then the LSE models with values between 0.980 and 0.984. The piecewise LSE models have similar pseudo-adjusted R^2 as other LSE models (ranging from 0.980 to 0.984), even though their RMSE values were lower, which may be because they have many more parameters. Within the LSE family, the TLG and TLGP specifications generally improve fit relative to the corresponding CD models.

- BIC results should be compared only within each SFA model family:¹⁷

BC95 family:

- BIC values range from -458.4 to -333.2 . The TLG and TLGP consistently outperform the corresponding CD specifications (ie, have lower BIC). The TLGP form generally improves on the TLG, except for BC95-AGECJTT-HN-GTC, for which the TLG specification has the lower BIC. Adding a quadratic time trend substantially improves BIC for the JTT models, while the effect of adding GTC depends on the underlying specification.
- BC95-JTT-HN-QTT records the strongest BIC performance, with the TLGP specification ranked first at -458.4 , followed by the TLG specification at -444.5 . The other JTT models provide intermediate results, while the weakest BIC performance is associated with the CD forms of the AGE CJTT specifications.

Kumb90 family:

- BIC values range from -1503.3 to -1403.0 . CD specifications generally have lower BIC values than TLG models, meaning that in the Kumb90 group of models, the CD models often have a better fit than TLG. The TLGP models typically have a lower BIC than the TLG, although they still do not usually outperform CD. The exception is Kumb90-JTT-HN-QTT, for which the TLGP has a marginally lower BIC than CD, at -1492.7 compared with -1492.4 . Overall, the simpler JTT time trend specification performs best with the Kumb90 models, while the addition of AJTT and GTC terms generally worsens BIC.
- Kumb90-JTT-HN provides the strongest overall BIC performance, with its CD specification recording the lowest BIC value, at -1503.3 . The weakest results are concentrated in the AGE CJTT and AJTT-GTC specifications, particularly the Kumb90-AGE CJTT-HN TLG model and the AJTT-HN-GTC models.

Table 4.3.2 shows likelihood ratio (LR) tests of restricted models versus less restricted models. For example, TLGP is a restricted version of TLG. The LR test determines whether the restrictions in the TLGP model are statistically valid, under the maintained assumptions about the model and data, by testing the null hypothesis that the parameter vector satisfies the constraints. If the associated p-value is less than 0.05, the constraints result in a statistically significant deterioration in model fit. LR tests are only available for the SFA models, since the LSE models are not estimated using the maximum likelihood method. Hence, Table 4.3.2 does not include LSE models. Further, models Kumb90-JTT-HN-GTC-QTT and Kumb90-AJTT-HN-GTC-QTT are not included in Table 4.3.2 because both the TLG and TLGP versions of these models failed in the preceding tests.

¹⁷ BIC is not available for the panel-corrected standard error estimators used for the LSE models.

Table 4.3.2: Likelihood Ratio Tests (Prob > chi2)

<i>Model</i>	<i>TLGP v. TLG</i>	<i>CD v. TLGP</i>	<i>CD v. TLG</i>	<i>Indicated function</i>
<u><i>Revised Phase 2 Models</i></u>				
BC95-JTT-HN	0.6591	0.0000	0.0000	TLGP
BC95-JTT-HN-GTC	0.6785	0.0000	0.0000	TLGP
Kumb90-JTT-HN	0.1696	0.0158	0.0145	TLGP
Kumb90-JTT-HN-GTC	n.a.	0.0000	n.a.	TLGP
Kumb90-AJTT-HN [#]	n.a.	0.0001	n.a.	TLGP
Kumb90-AJTT-HN-GTC [#]	n.a.	0.0002	n.a.	TLGP
<u><i>Generalised Efficiency Change</i></u>				
BC95-AGECJTT-HN	0.8610	0.0000	0.0000	TLGP
BC95-AGECJTT-HN-GTC	0.0001	0.0000	0.0000	TLG
Kumb90-AGECJTT-HN [#]	0.2827	0.0355	0.0684	TLGP
Kumb90-AGECJTT-HN-GTC [#]	n.a.	0.0092	n.a.	TLGP
<u><i>Quadratic Time Trend Models</i></u>				
BC95-JTT-HN-QTT	0.9582	0.0000	0.0000	TLGP
Kumb90-JTT-HN-QTT	0.4010	0.0008	0.0054	TLGP
Kumb90-AJTT-HN-QTT	0.2766	0.0027	0.0037	TLGP

Note: n.a. (not applicable) means the TLG version of this model has been ruled out in preceding tests.

[#] As noted in Section 4.1, the model may be subject to identification issues.

Among the 13 SFA specifications in Table 4.3.2, TLGP ranks first in 12. TLG ranks first only for BC95-AGECJTT-HN-GTC. Overall, the results consistently favour TLGP over CD, and in the great majority of cases, TLGP is preferred over TLG. This indicates that removing statistically insignificant second-order terms generally preserves model fit.

Table 4.3.3 shows Wald tests of restricted models versus less restricted models, where the restricted models are nested within the less restricted models through restrictions on parameters. This is an alternative to the LR test, and is applied to the LSE models (since the LR tests are not available for the LSE models). The test statistics have a similar interpretation as those in Table 4.3.2. If the associated p-value is less than 0.05, the restrictions result in a statistically significant deterioration in model fit.

In all of the 15 LSE specifications in Table 4.3.3, TLGP ranks above the TLG and CD models. TLG ranks second in all models. These results consistently favour TLGP over the TLG and CD models.

Table 4.3.3: Wald Tests (Prob > chi2)

<i>Model</i>	<i>TLGP v. TLG</i>	<i>CD v. TLGP</i>	<i>CD v. TLG</i>	<i>Indicated function</i>
<u><i>Revised Phase 2 Models</i></u>				
LSE-ADTT	0.4944	0.0000	0.0000	TLGP
LSE-ADTT-GTC	0.9797	0.0000	0.0000	TLGP
LSE-AJTT	0.9206	0.0000	0.0000	TLGP
LSE-AJTT-GTC	0.9764	0.0000	0.0000	TLGP
LSE-AJT Tx-GTC	0.9324	0.0000	0.0000	TLGP
<u><i>Quadratic Time Trend Models</i></u>				
LSE-ADTT-QTT	0.5911	0.0000	0.0000	TLGP
LSE-ADTT-GTC-QTT	0.9806	0.0000	0.0000	TLGP
LSE-AJTT-QTT	0.9279	0.0000	0.0000	TLGP
LSE-AJTT-GTC-QTT	0.9706	0.0000	0.0000	TLGP
LSE-AJT Tx-QTT	0.9227	0.0000	0.0000	TLGP
LSE-AJT Tx-GTC-QTT	0.9117	0.0000	0.0000	TLGP
<u><i>Piecewise Regression Models</i></u>				
LSE-ADTT-PW-2015	0.5383	0.0000	0.0000	TLGP
LSE-ADTT-PW-2015-2022	0.5196	0.0000	0.0000	TLGP
LSE-AJTT-PW-2015	0.8831	0.0000	0.0000	TLGP
LSE-AJTT-PW-2015-2022	0.8750	0.0000	0.0000	TLGP

Table 4.3.4 presents the results of statistical tests. Models ruled out in the preceding assessment stages were excluded. The main comments are:

- *Linktest results:* The *linktest* is a specification test implemented in Stata. All BC95 CD models fail the *linktest*, while the corresponding TLG and TLGP models pass. This suggests that the CD may be too restrictive and that including second-order terms addresses the detected misspecification. All reported Kumb90 models pass the *linktest* across all reported functional forms. The *linktest* is unavailable for the LSE models.
- *Joint significance of the Translog parameters:* The second-order parameters are jointly significant in all reported TLG models except Kumb90-AGECJT-HN. They are jointly significant in every TLGP model, confirming that the retained terms collectively add explanatory power and supporting the simplification process.
- *Multicollinearity:* The average variance inflation factor (VIF) and the Condition Number show the degree of multicollinearity.
 - Multicollinearity is severe in the full TLG models, with average VIFs ranging from 249.7 to 714.1 and condition numbers from 367.0 to 2,547.3. TLGP substantially reduces average VIFs in most cases, to between 19.0 and 260.0, although some condition numbers remain high. Adding GTC generally reduces the full TLG average VIF. For example, it falls from 714.1 to 547.5 for BC95-

- JTT-HN and from 405.4 to 370.6 for LSE-AJTT. Its effect on condition numbers and TLGP specifications is less consistent.
- Average VIFs are considerably lower in most CD models, generally ranging from 9.4 to 26.3. The exception is LSE-ADTT-GTC-QTT, with an average VIF of 318.1.
 - The ADTT and AJTT specifications generally produce similar average VIFs, although AJTT tends to have higher condition numbers where GTC is not included. For example, the TLGP condition number increases from 1,738.5 for LSE-ADTT to 2,246.8 for LSE-AJTT. Once GTC is included, the differences between ADTT and AJTT condition numbers generally become much smaller and the differences between the two specifications narrow.
 - The piecewise LSE models generally record lower average VIFs than the corresponding standard and quadratic LSE specifications, particularly for the CD and TLGP functional forms. However, condition numbers remain high for the ADTT and AJTT versions.
 - *Residual normality:* The interquartile range (IQR) test shows the percentage of severe outliers. Severe outliers comprise about 0.0002 per cent of a normally distributed population. Therefore, the presence of severe outliers indicates non-normality. Table 4.3.4 also presents the Shapiro-Wilk test of normality, which is a conventional test, but may not be well suited to a panel data context.¹⁸ For this reason, we regard the IQR test as the more reliable of the two tests shown.
 - The null hypothesis of normally distributed residuals is rejected for every reported model and functional form. The frequency of severe outliers is generally higher for the BC95 models than for the Kumb90 models, with the LSE models mostly in between. The TLG and TLGP functional forms generally produce the same or lower proportion of severe outliers than the corresponding CD model, although there are some exceptions.
 - The non-normality of residuals is not unusual in economic panel data where residual distributions may contain heavy tails or outliers. The result indicates that statistical inference should be interpreted with caution.

Overall, the statistical tests generally favour the TLGP specification. It retains jointly significant second-order parameters and materially reduces average VIFs relative to the full Translog form.

¹⁸ This is because it requires each data point to be entirely independent of the others, whereas in panel data, residuals may be more closely related to others within the same group, than to those in other groups.

Table 4.3.4: Statistical Tests

Model	Functional Form	linktest	Multicollinearity		Normality of Residuals		TLG Joint Parameters Significance
			Avg. VIF	Cond. No.	IQR (% severe outliers)	Shapiro–Wilk W test*	
Revised Phase 2 Models							
BC95-JTT-HN	CD	X	25.0	1084.2	1.37%	X	
	TLG	✓	714.1	1641.0	1.09%	X	✓
	TLGP	✓	260.0	1436.4	1.09%	X	✓
BC95-JTT-HN-GTC	CD	X	16.6	22.3	1.64%	X	
	TLG	✓	547.5	367.0	1.09%	X	✓
	TLGP	✓	190.0	183.3	0.82%	X	✓
Kumb90-JTT-HN	CD	✓	25.0	1084.2	0.27%	X	
	TLG	✓	714.1	1641.0	0.27%	X	✓
	TLGP	✓	257.7	1547.0	0.27%	X	✓
Kumb90-JTT-HN-GTC	CD	✓	16.6	22.3	0.27%	X	
	TLGP	✓	24.0	36.9	0.27%	X	✓
Kumb90-AJTT-HN [#]	CD	✓	25.0	1084.2	0.27%	X	
	TLGP	✓	32.8	1331.4	0.27%	X	✓
Kumb90-AJTT-HN-GTC [#]	CD	✓	16.6	22.3	0.46%	X	
	TLGP	✓	24.0	36.9	0.27%	X	✓
LSE-ADTT	CD		14.8	1303.3	0.27%	X	
	TLG		414.3	1957.6	0.18%	X	✓
	TLGP		30.5	1738.5	0.09%	X	✓
LSE-ADTT-GTC	CD		13.4	44.0	0.46%	X	
	TLG		377.9	506.6	0.36%	X	✓
	TLGP		33.0	79.9	0.36%	X	✓

Table 4.3.4: (cont.)

<i>Model</i>	<i>Functional Form</i>	<i>linktest</i>	<i>Multicollinearity</i>		<i>Normality of Residuals</i>		<i>TLG Joint Parameters Significance</i>
			<i>Avg. VIF</i>	<i>Cond. No.</i>	<i>IQR</i> (% severe outliers)	<i>Shapiro–Wilk</i> <i>W test*</i>	
LSE-AJTT	CD		14.7	1704.3	0.55%	X	
	TLG		405.4	2541.1	0.46%	X	✓
	TLGP		30.1	2246.8	0.46%	X	✓
LSE-AJTT-GTC	CD		13.4	44.7	0.55%	X	
	TLG		370.6	508.8	0.55%	X	✓
	TLGP		27.6	66.6	0.55%	X	✓
LSE-AJTT-GTCx	CD		18.6	45.9	0.55%	X	
	TLG		367.1	512.9	0.55%	X	✓
	TLGP		32.0	67.3	0.55%	X	✓
<i>Generalised Efficiency Change</i>							
BC95-AGECJTT-HN	CD	X	25.0	1084.2	1.46%	X	
	TLG	✓	714.1	1641.0	1.18%	X	✓
	TLGP	✓	260.0	1436.4	1.18%	X	✓
BC95-AGECJTT-HN-GTC	CD	X	16.6	22.3	1.55%	X	
	TLG	✓	547.5	367.0	1.00%	X	✓
	TLGP	✓	48.2	71.8	0.82%	X	✓
Kumb90-AGECJTT-HN [#]	CD	✓	25.0	1084.2	0.18%	X	
	TLG	✓	714.1	1641.0	0.18%	X	X
	TLGP	✓	23.0	1224.8	0.18%	X	✓
Kumb90-AGECJTT-HN-GTC [#]	CD	✓	16.6	22.3	0.36%	X	
	TLGP	✓	25.5	8139.6	0.46%	X	✓

Table 4.3.4: (cont.)

Model	Functional Form	linktest	Multicollinearity		Normality of Residuals		TLG Joint Parameters Significance
			Avg. VIF	Cond. No.	IQR (% severe outliers)	Shapiro–Wilk W test*	
<u>Quadratic Time Trend Models</u>							
BC95-JTT-HN-QTT	CD	X	25.0	1084.2	1.00%	X	
	TLG	✓	714.1	1641.0	0.36%	X	✓
	TLGP	✓	259.4	1449.1	0.36%	X	✓
Kumb90-JTT-HN-QTT	CD	✓	25.0	1084.2	0.27%	X	
	TLG	✓	714.1	1641.0	0.18%	X	✓
	TLGP	✓	23.0	1224.8	0.18%	X	✓
Kumb90-JTT-HN-GTC-QTT [#]	CD	✓	16.6	22.3	0.27%	X	
Kumb90-AJTT-HN-QTT	CD	✓	25.0	1084.2	0.27%	X	
	TLG	✓	714.1	1641.0	0.27%	X	✓
	TLGP	✓	46.2	1440.4	0.27%	X	✓
Kumb90-AJTT-HN-GTC-QTT	CD	✓	16.6	22.3	0.27%	X	
LSE-ADTT-QTT	CD		16.9	1312.3	0.36%	X	
	TLG		406.4	1960.3	0.27%	X	✓
	TLGP		37.5	1739.8	0.27%	X	✓
LSE-ADTT-GTC-QTT	CD		318.1	524.5	0.46%	X	
	TLG		371.4	507.5	0.36%	X	✓
	TLGP		36.2	248.3	0.36%	X	✓
LSE-AJTT-QTT	CD		18.6	1720.9	0.55%	X	
	TLG		397.9	2544.5	0.55%	X	✓
	TLGP		31.6	2251.1	0.55%	X	✓
LSE-AJTT-GTC-QTT	CD		17.5	46.6	0.55%	X	
	TLG		364.4	509.8	0.55%	X	✓
	TLGP		31.1	259.1	0.55%	X	✓

Table 4.3.4: (cont.)

<i>Model</i>	<i>Functional Form</i>	<i>linktest</i>	<i>Multicollinearity</i>		<i>Normality of Residuals</i>		<i>TLG Joint Parameters Significance</i>
			<i>Avg. VIF</i>	<i>Cond. No.</i>	<i>IQR</i> (% severe outliers)	<i>Shapiro–Wilk W test*</i>	
LSE-AJTTx-QTT	CD		26.3	44.5	0.55%	X	
	TLG		386.2	500.9	0.55%	X	✓
	TLGP		37.3	82.5	0.55%	X	✓
LSE-AJTTx-GTC-QTT	CD		22.4	47.7	0.55%	X	
	TLG		361.1	513.8	0.55%	X	✓
	TLGP		33.2	67.4	0.55%	X	✓
<i>Piecewise Regression Models</i>							
LSE-ADTT-PW-2015	CD		11.3	1303.1	0.55%	X	
	TLG		315.9	1956.7	0.64%	X	✓
	TLGP		27.6	1735.8	0.64%	X	✓
LSE-ADTT-PW-2015-2022	CD		9.4	1314.2	0.55%	X	
	TLG		252.6	1962.7	0.64%	X	✓
	TLGP		22.5	1742.6	0.64%	X	✓
LSE-AJTT-PW-2015	CD		11.3	1702.7	0.55%	X	
	TLG		311	2539.7	0.55%	X	✓
	TLGP		23.2	2246.2	0.55%	X	✓
LSE-AJTT-PW-2015-2022	CD		9.5	1712.9	0.55%	X	
	TLG		249.7	2547.3	0.55%	X	✓
	TLGP		19	2254.7	0.55%	X	✓

Note: * The Shapiro-Wilk test may not be reliable in the present context.

4.4 Efficiency Score Profiles

Attachment A presents the efficiency scores and rankings for each model. Overall, the estimated rankings and average efficiency levels are broadly consistent with the time-invariant models.¹⁹ This section presents the efficiency time-profile for each DNSP by comparing the model-based results with the OPFP series in a series of charts. Models ruled out in the preceding assessment stages were excluded.

The OPFP series is presented only as a useful comparator. However, it is based on a different method and a different output specification, assumes constant returns to scale and does not account for the effects of undergrounding. Therefore, it is not expected that the econometric models should estimate the same pattern of opex efficiency.

4.4.1 EVO

For EVO, Figures 4.4.1.1 to 4.4.1.9 show that:

- EVO's OPFP declined from 2006 to a low in 2014 and 2015, followed by a sharp increase in 2016, and then declined again until 2018. OPFP subsequently improved in most years to 2023. The efficiency score trends for the BC95 specifications closely reproduce the OPFP pattern, although their scores are consistently higher and show more growth.
- Most Kumb90 specifications exhibit a smooth U-shaped pattern, with efficiency declining slightly or remaining broadly stable until approximately 2010–2012 and increasing thereafter, particularly after 2016. The AGEJTT specifications display a more stepwise trajectory, declining until around 2012–2014 and then increasing through a series of sharp rises and plateaus.
- The LSE results show two broad patterns. The standard and GTC specifications follow a relatively smooth upward trend, with some flattening in the latest years. In contrast, the quadratic and piecewise specifications display a U-shaped pattern, declining until approximately 2015 and increasing strongly thereafter. This latter pattern more closely reflects the broad decline-and-recovery trajectory observed in the other model families.

¹⁹ Further analysis of the efficiency profiles, including correlations between efficiency scores and OPFP, expected unconditional efficiency shapes, and DNSP implied inefficiency distributions, is presented in Attachment A.

Figure 4.4.1.1: EVO BC95 vs OPFP Trends – Cobb-Douglas Models

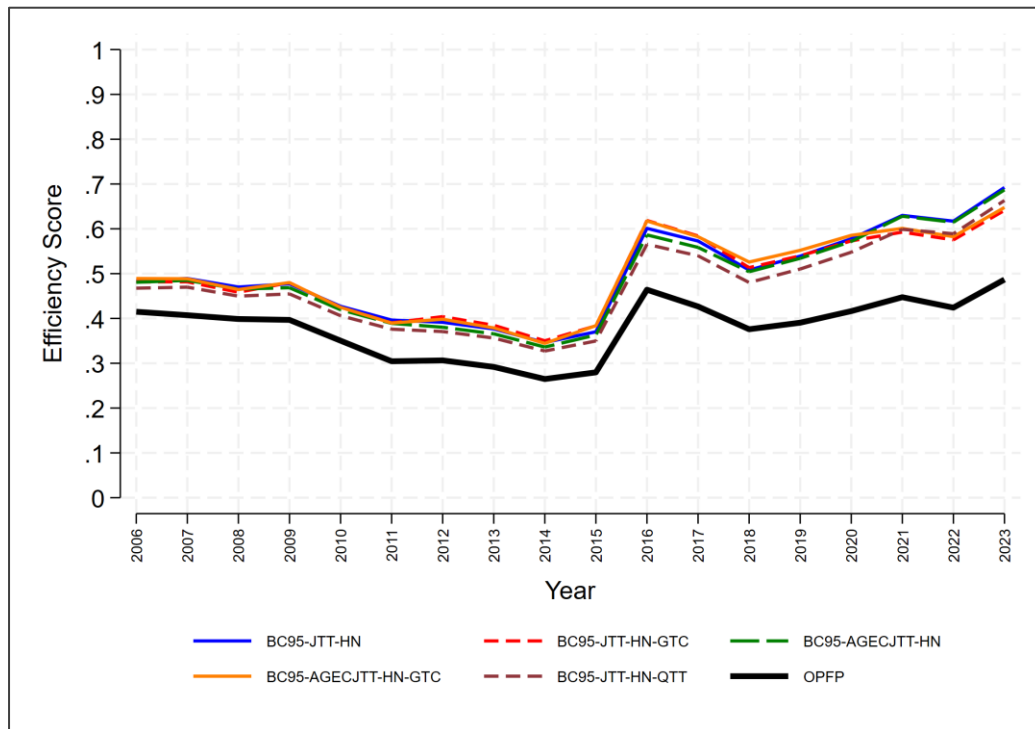


Figure 4.4.1.2: EVO BC95 vs OPFP Trend – Translog Models

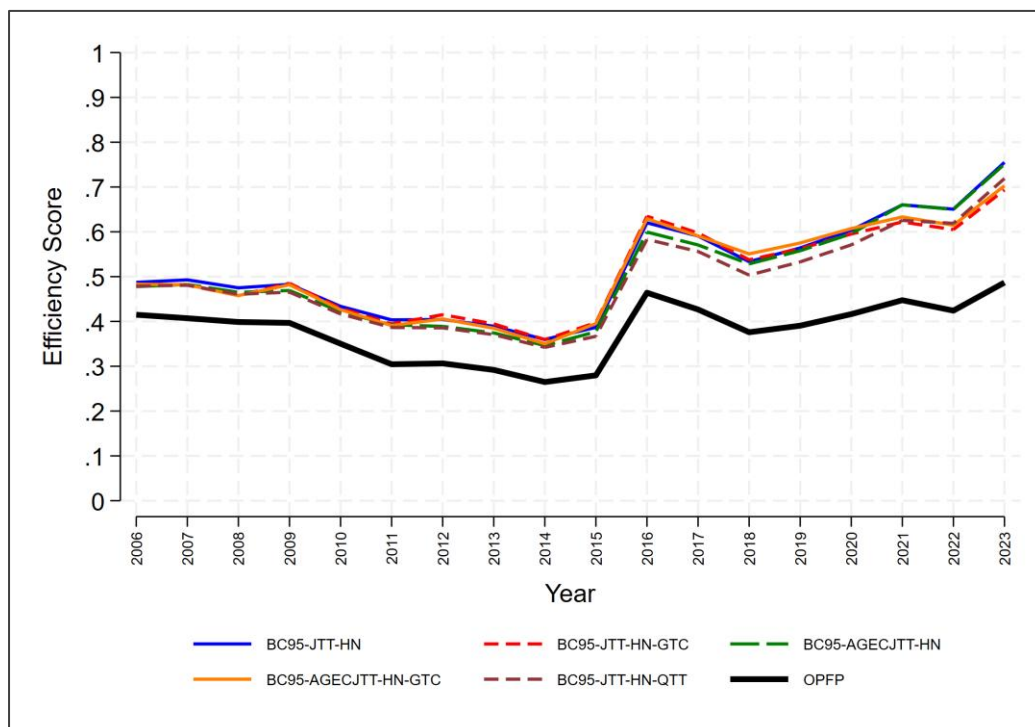


Figure 4.4.1.3: EVO BC95 vs OPFP Trend – Parsimonious Translog Models

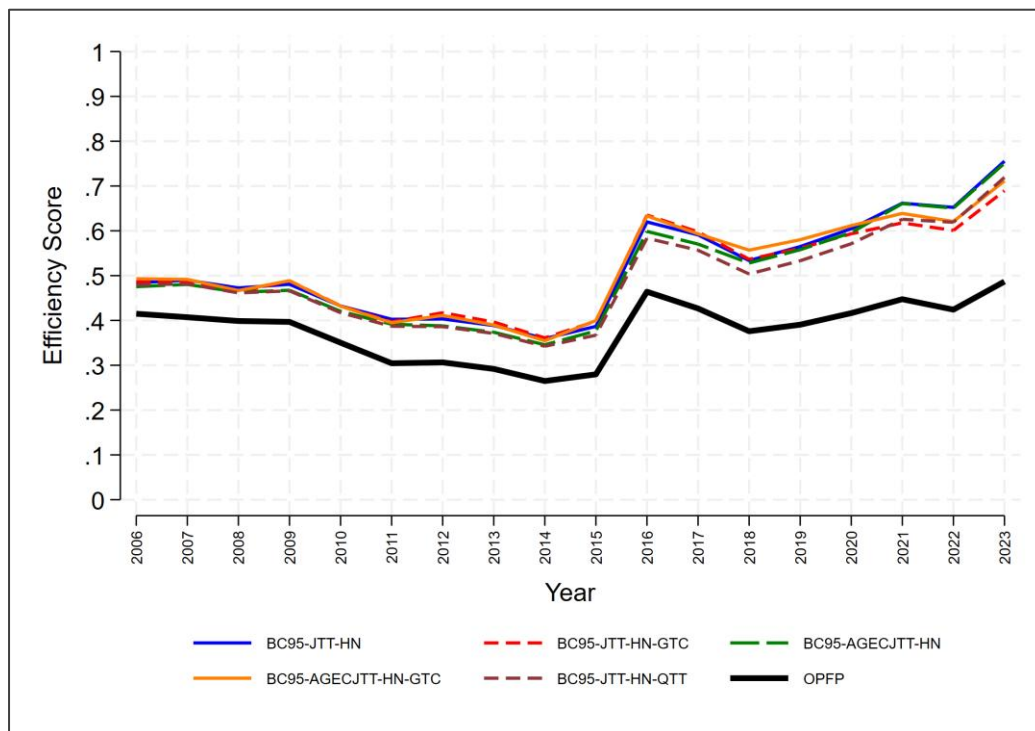


Figure 4.4.1.4: EVO Kumb90 vs OPFP Trends – Cobb-Douglas Models

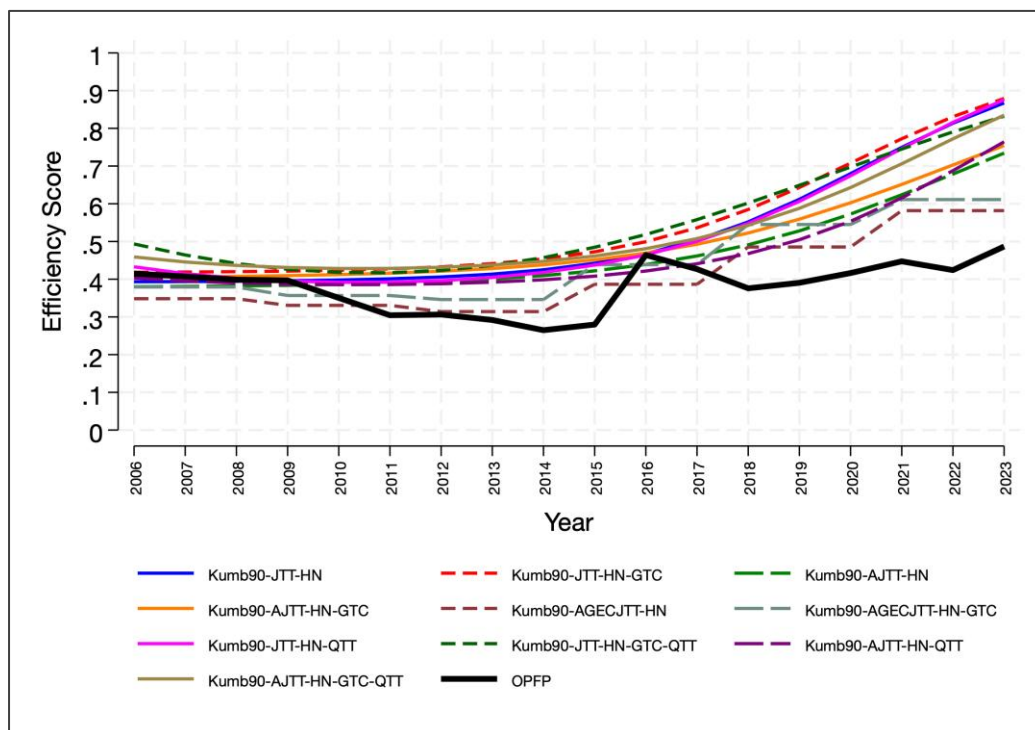


Figure 4.4.1.5: EVO Kumb90 vs OPFP Trends – Translog Models

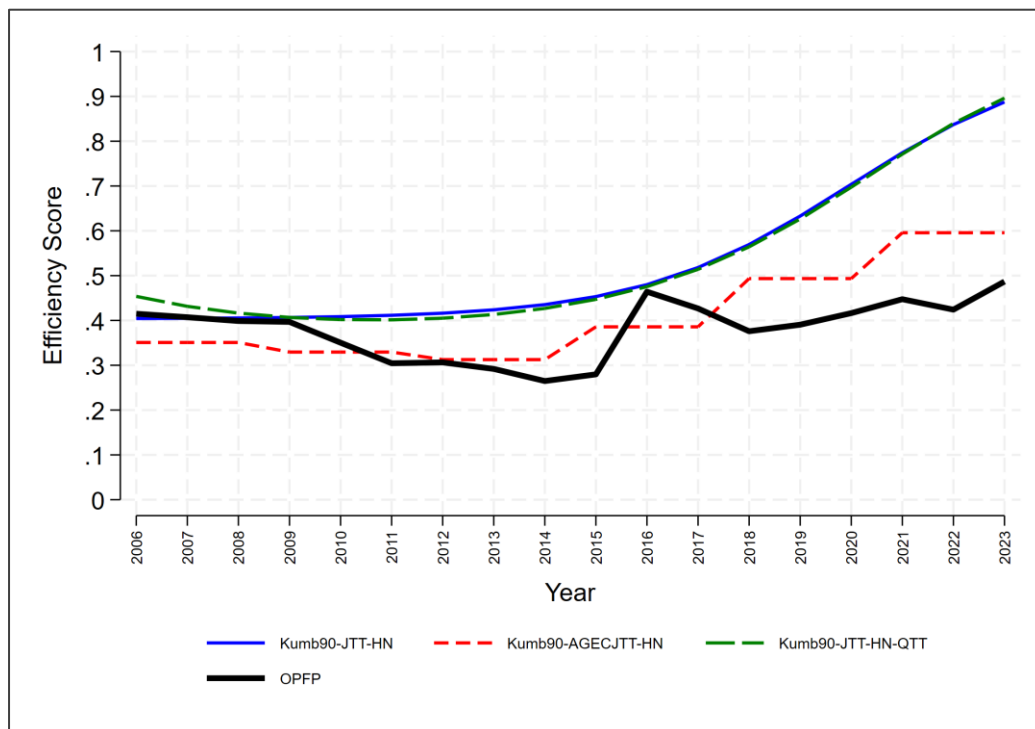


Figure 4.4.1.6: EVO Kumb90 vs OPFP Trends – Parsimonious Translog Models

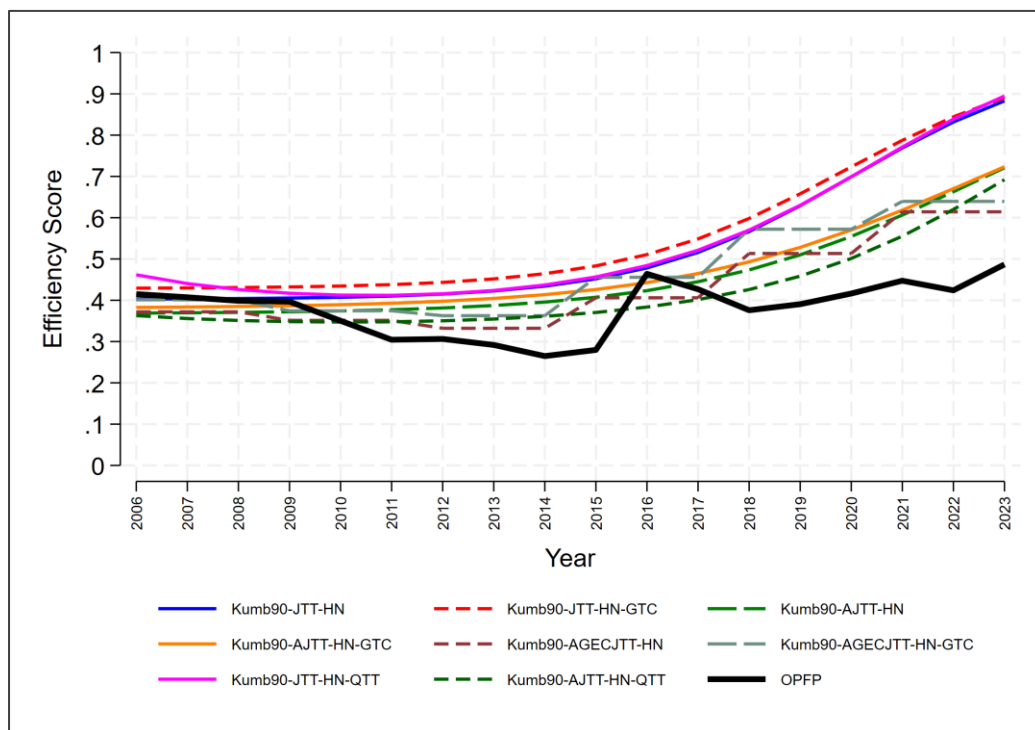


Figure 4.4.1.7: EVO LSE vs OPFP Trends – Cobb-Douglas Models

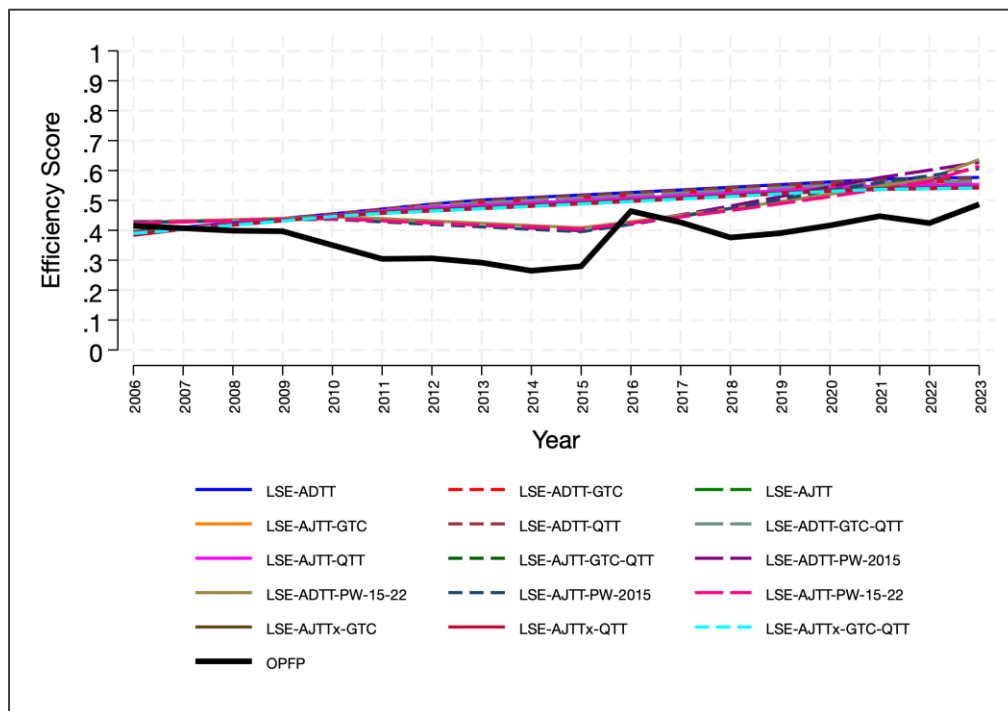


Figure 4.4.1.8: EVO LSE vs OPFP Trends – Translog Models

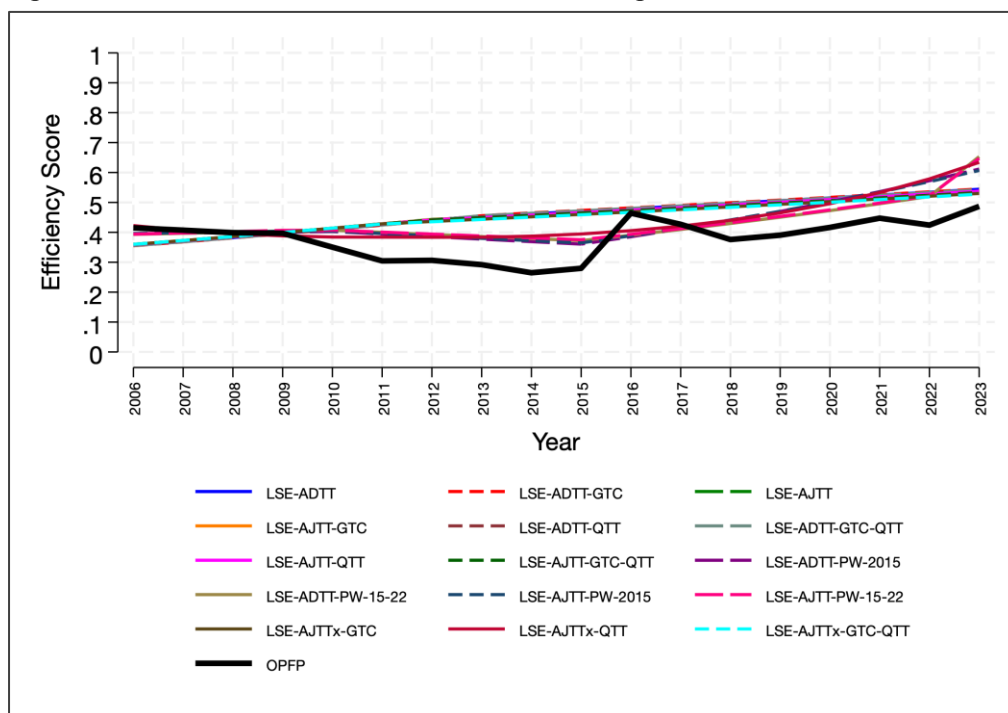
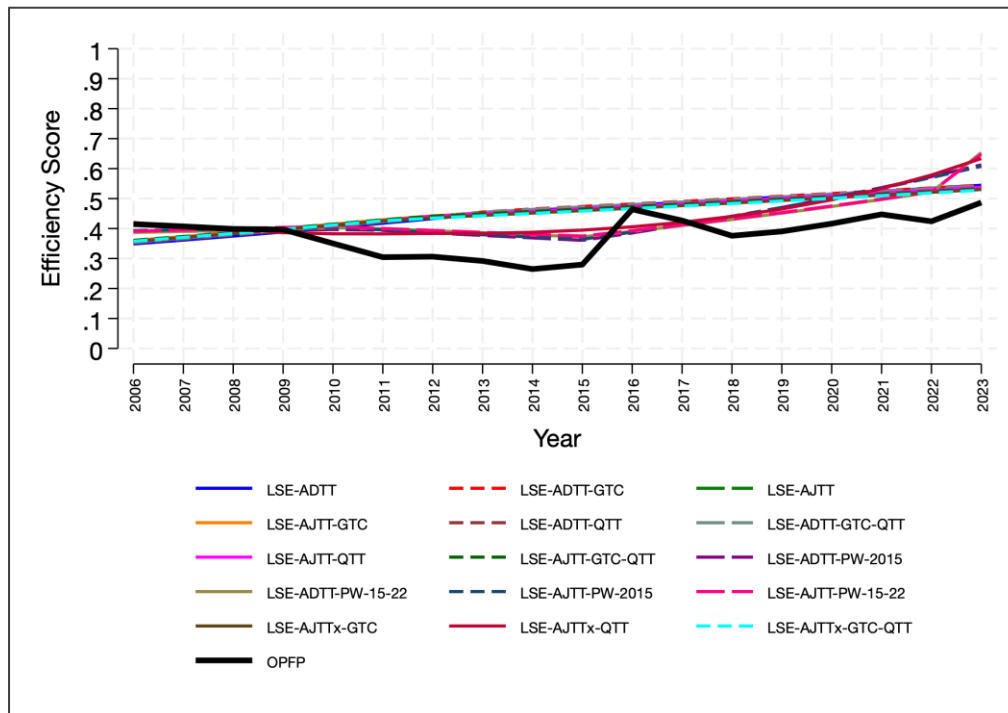


Figure 4.4.1.9: EVO LSE vs OPFP Trends – Parsimonious Translog Models



4.4.2 AGD

For AGD, Figures 4.4.2.1 to 4.4.2.9 show that:

- OPFP fluctuated without a clear trend from 2006 to 2015, with peaks in 2007 and 2013 and a low in 2015. It then increased strongly and consistently to 2022 before declining in 2023. The efficiency score trends for the BC95 specifications closely replicate these movements but consistently produce higher scores and higher growth.
- Most Kumb90 specifications decline slightly or remain stable until around 2010 to 2012, then increase, particularly after 2015. The AJTT specifications initially show accelerating growth until around 2018 to 2019, before efficiency to rise at a progressively slower rate towards one. The JTT specifications increase more moderately, while the AGE CJTT specifications show a stepwise pattern, declining slightly until around 2012 to 2014 and then increasing through sharp rises and plateaus.
- Among the LSE models, the standard and GTC specifications follow a smooth, almost linear upward trend throughout the period. In contrast, the quadratic and piecewise models remain broadly stable until approximately 2015, and then increase rapidly, generally overtaking the standard and GTC specifications after 2020. Some piecewise specifications peak in 2022 and decline in 2023, broadly reflecting the latest movement in the OPFP results. These models follow the main time-profile of OPFP but with higher efficiency scores, similar to the BC95 model.

Figure 4.4.2.1: AGD BC95 vs OPFP Trends – Cobb-Douglas Models

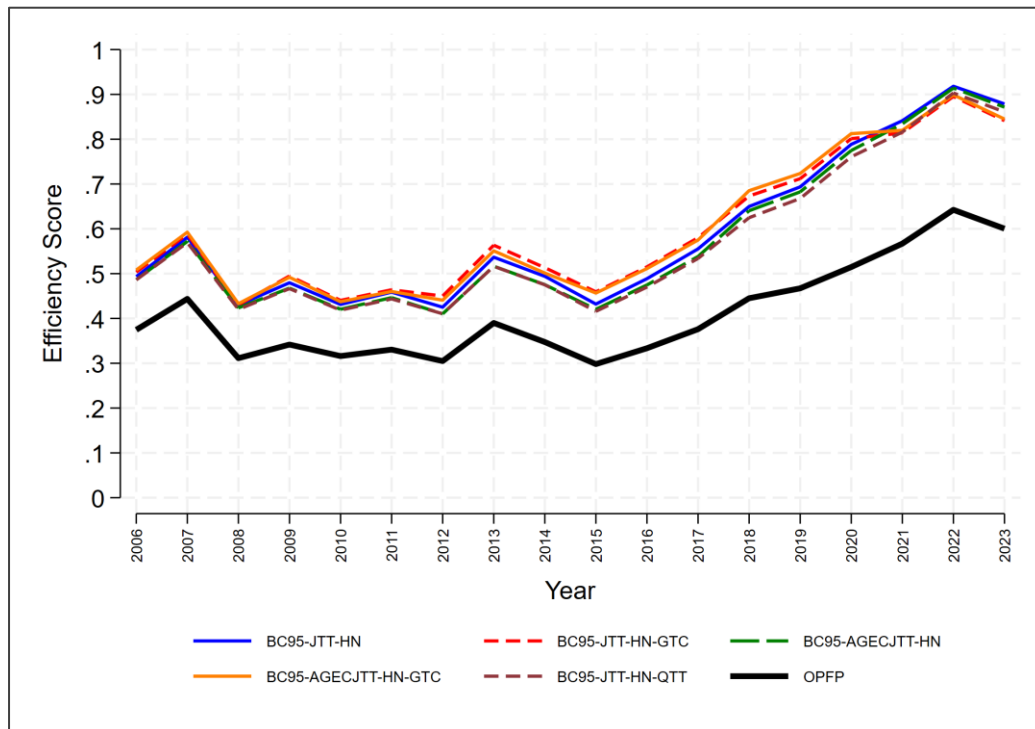


Figure 4.4.2.2: AGD BC95 vs OPFP Trend – Translog Models

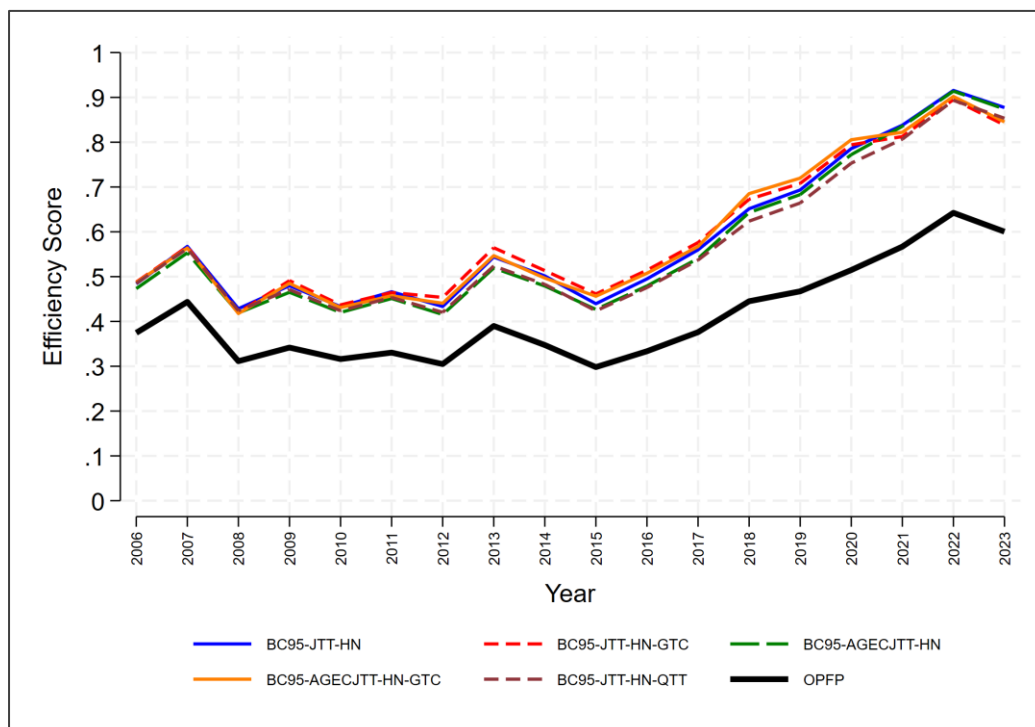


Figure 4.4.2.3: AGD BC95 vs OPFP Trend – Parsimonious Translog Models

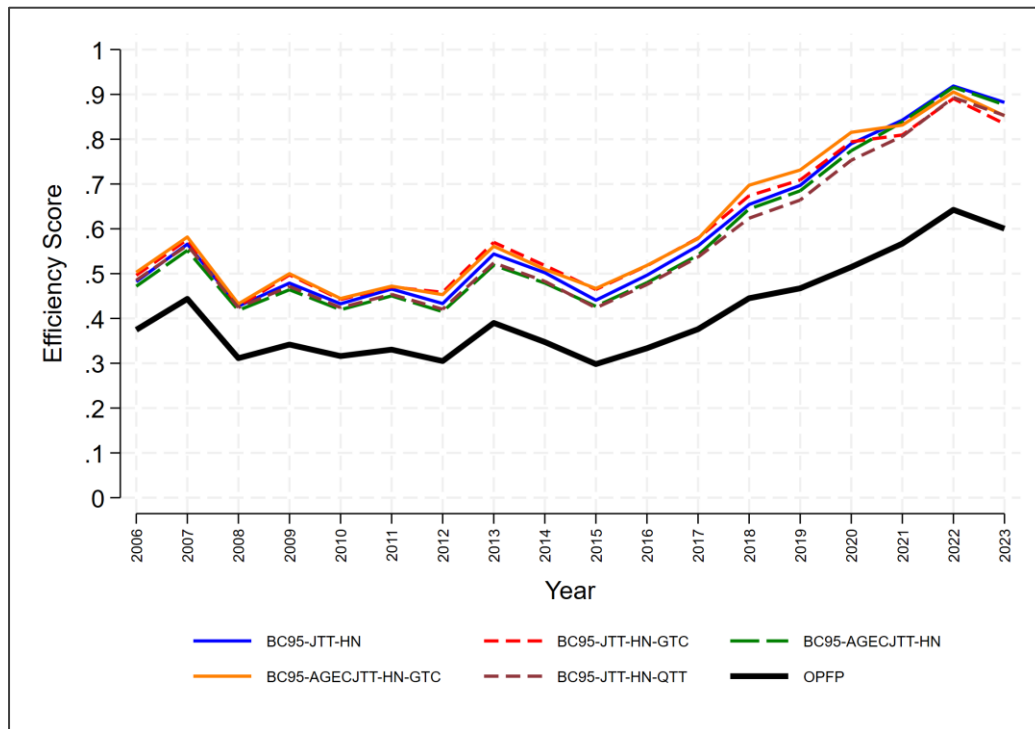


Figure 4.4.2.4: AGD Kumb90 vs OPFP Trends – Cobb-Douglas Models

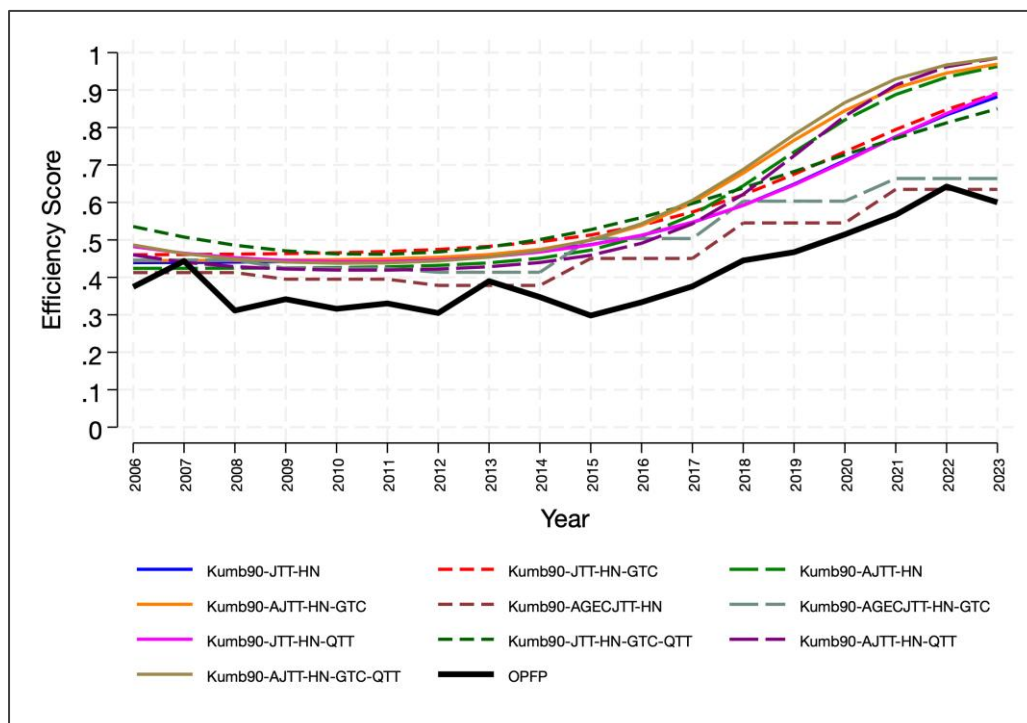


Figure 4.4.2.5: AGD Kumb90 vs OPFP Trends – Translog Models

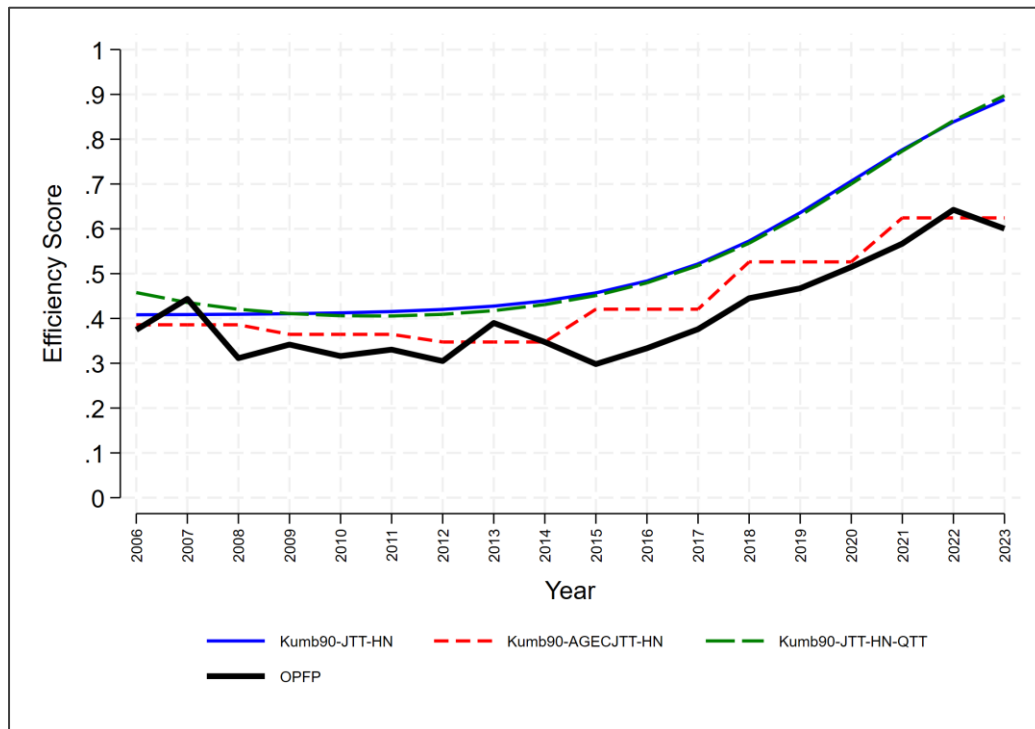


Figure 4.4.2.6: AGD Kumb90 vs OPFP Trends – Parsimonious Translog Models

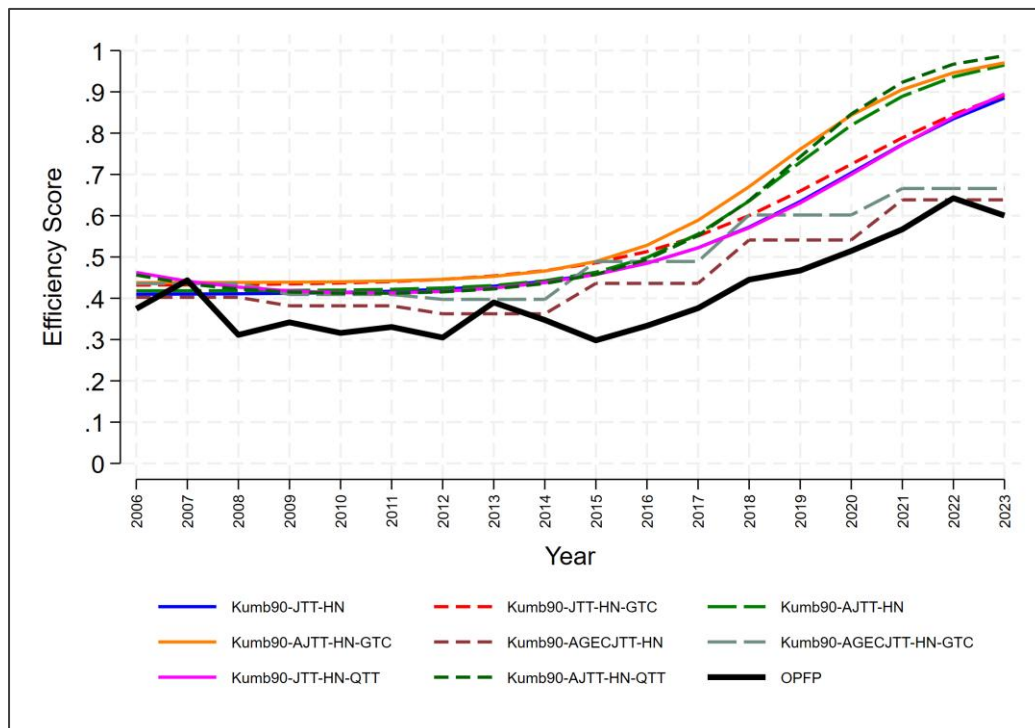


Figure 4.4.2.7: AGD LSE vs OPFP Trends – Cobb-Douglas Models

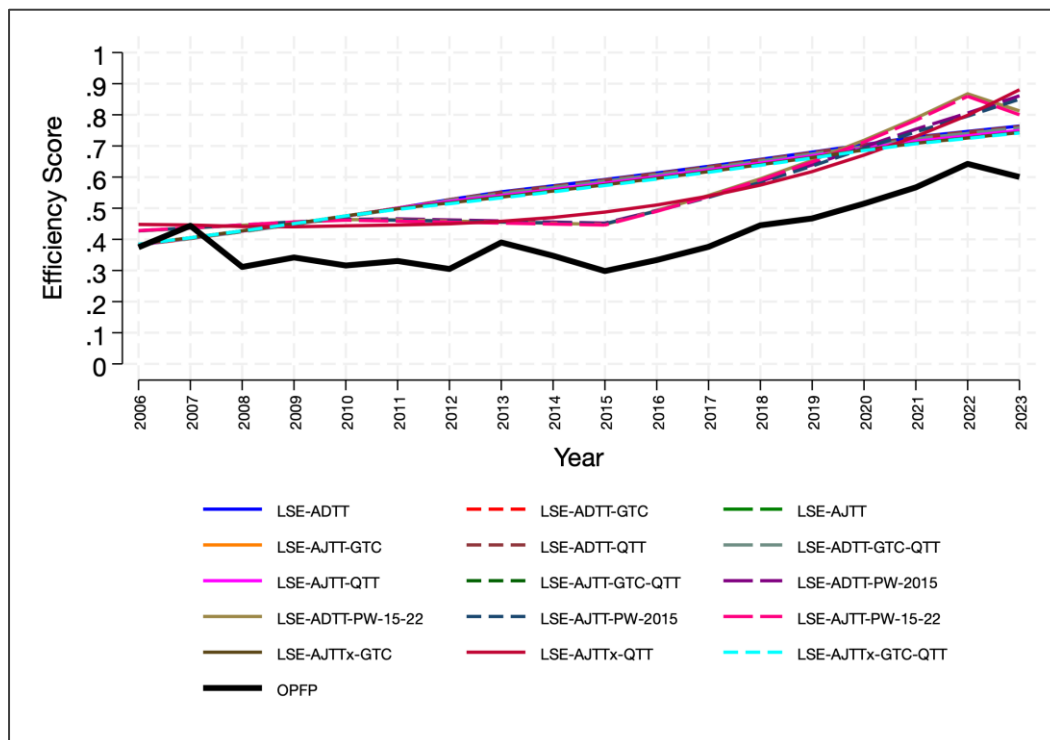


Figure 4.4.2.8: AGD LSE vs OPFP Trends – Translog Models

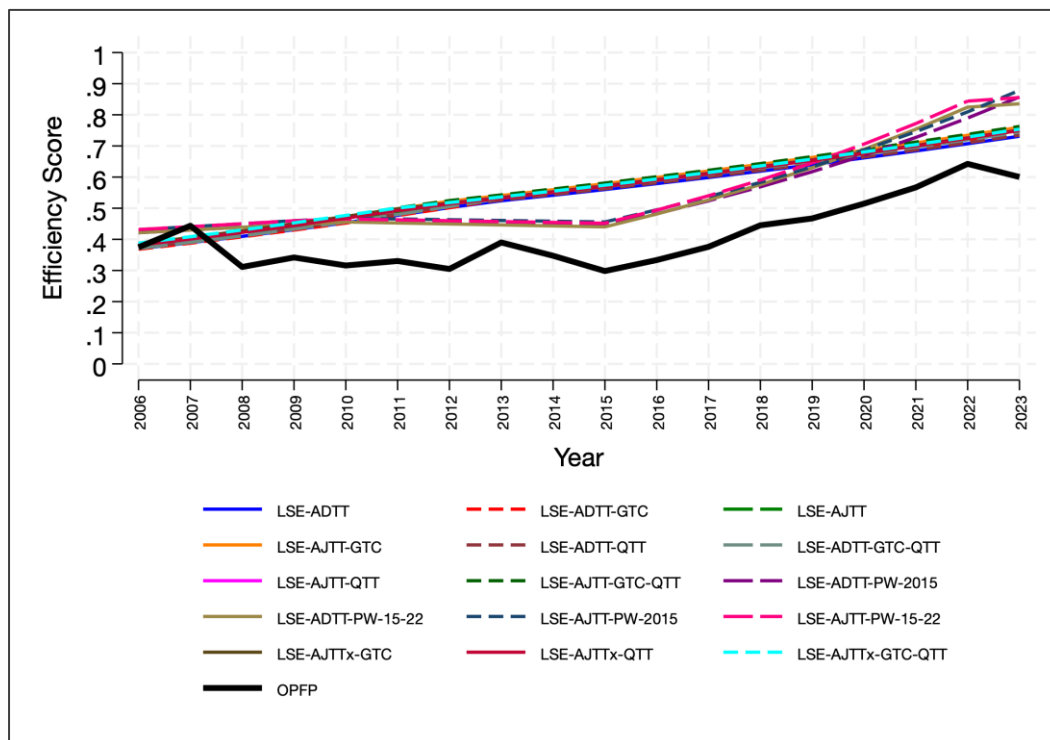
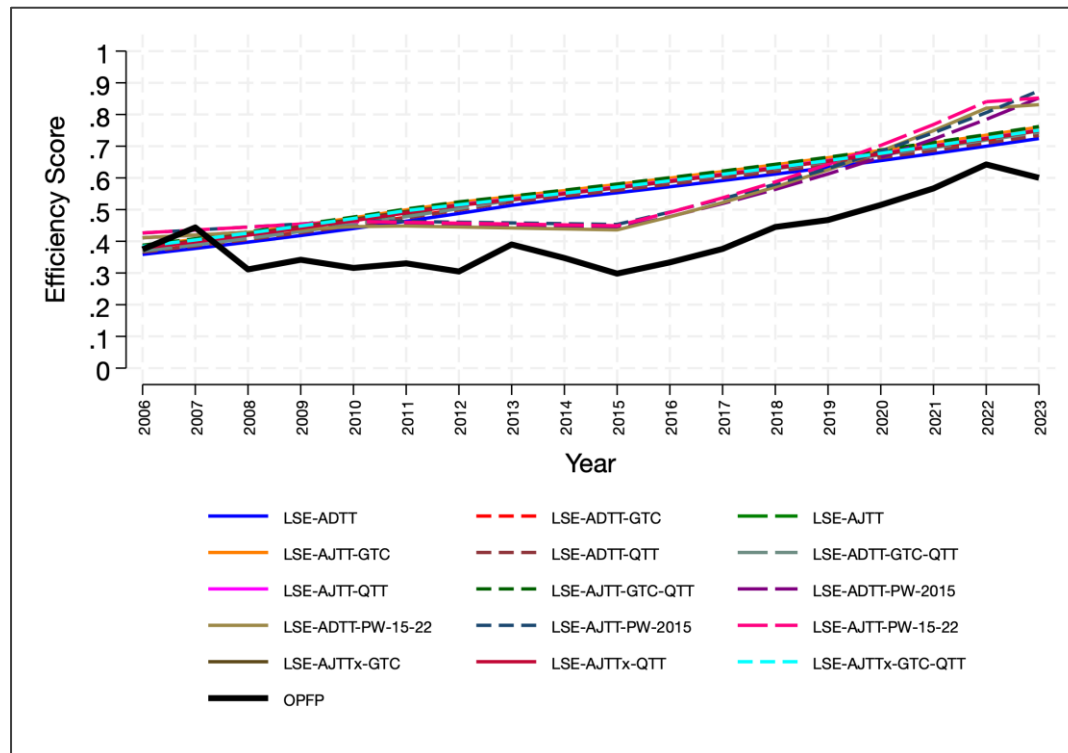


Figure 4.4.2.9: AGD LSE vs OPFP Trends – Parsimonious Translog Models



4.4.3 CIT

For CIT, Figures 4.4.3.1 to 4.4.3.9 show that:

- OPFP efficiency increased slightly between 2006 and 2008, decreased through 2010, temporarily increased in 2011, and reached its minimum in 2014. It then generally increased, despite decreases in 2019 and after 2021, before decreasing further in 2023. The efficiency score trends for the BC95 specifications closely reproduce these movements but generate consistently higher scores.
- Most Kumb90 specifications show a smooth U-shaped pattern, decreasing until 2010–2012, remaining relatively stable through 2014, and increasing thereafter. The AJTT paths initially increase at an accelerating rate until around 2018–2020, after which the rate of increase slows; some approach one by 2023. The AGEJTT specifications show a slight decrease until 2012–2014, followed by stepwise increases separated by plateaus.
- LSE scores are comparatively stable. The standard and GTC specifications increase gradually until around 2012 and then remain stable or decrease slightly. The QTT and piecewise specifications show a shallow U-shaped pattern, decreasing until 2014–2015 and increasing thereafter. The QTT and piecewise specifications broadly capture the decreases and subsequent increases in the other models, unlike the standard and GTC specifications.

Figure 4.4.3.1: CIT BC95 vs OPFP Trends – Cobb-Douglas Models

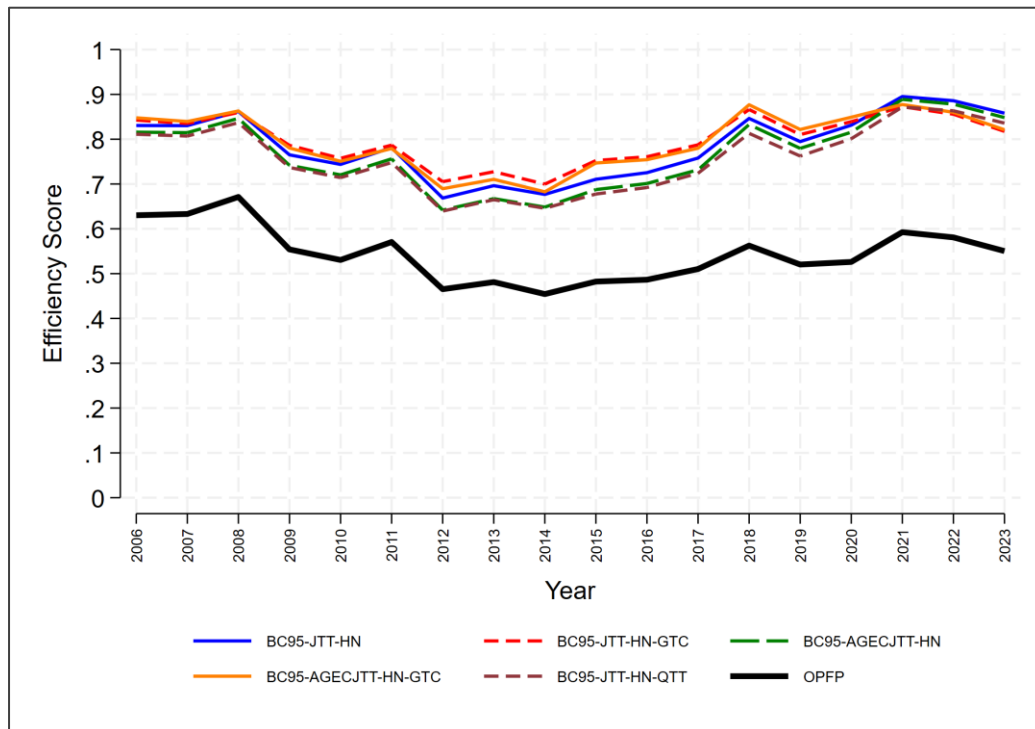


Figure 4.4.3.2: CIT BC95 vs OPFP Trend – Translog Models

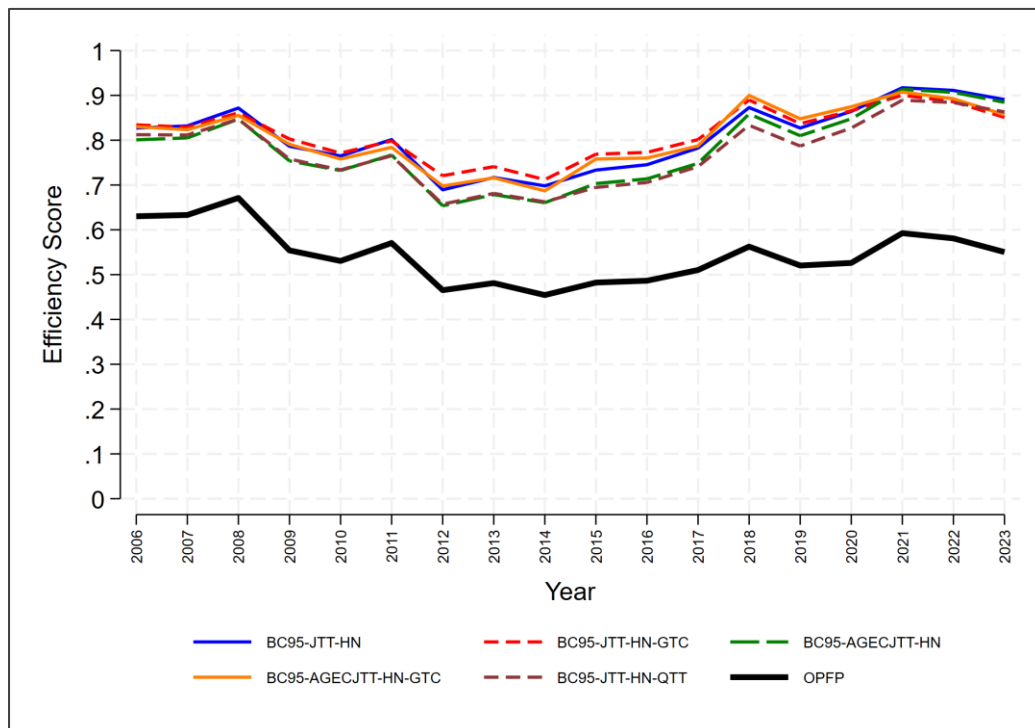


Figure 4.4.3.3: CIT BC95 vs OPFP Trend – Parsimonious Translog Models

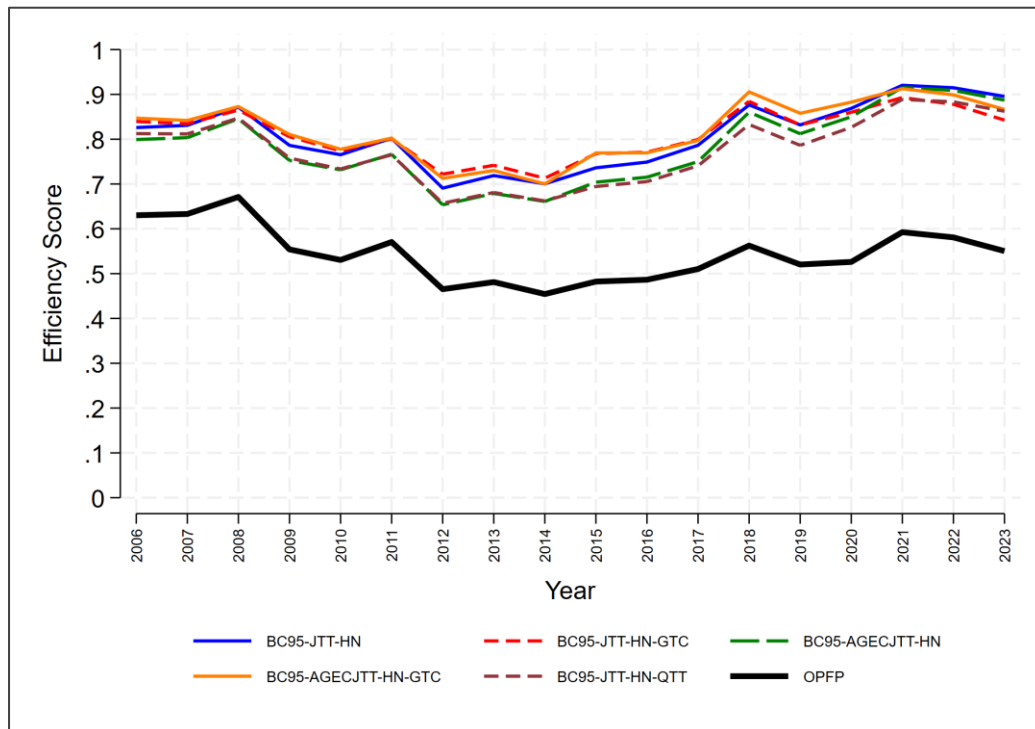


Figure 4.4.3.4: CIT Kumb90 vs OPFP Trends – Cobb-Douglas Models

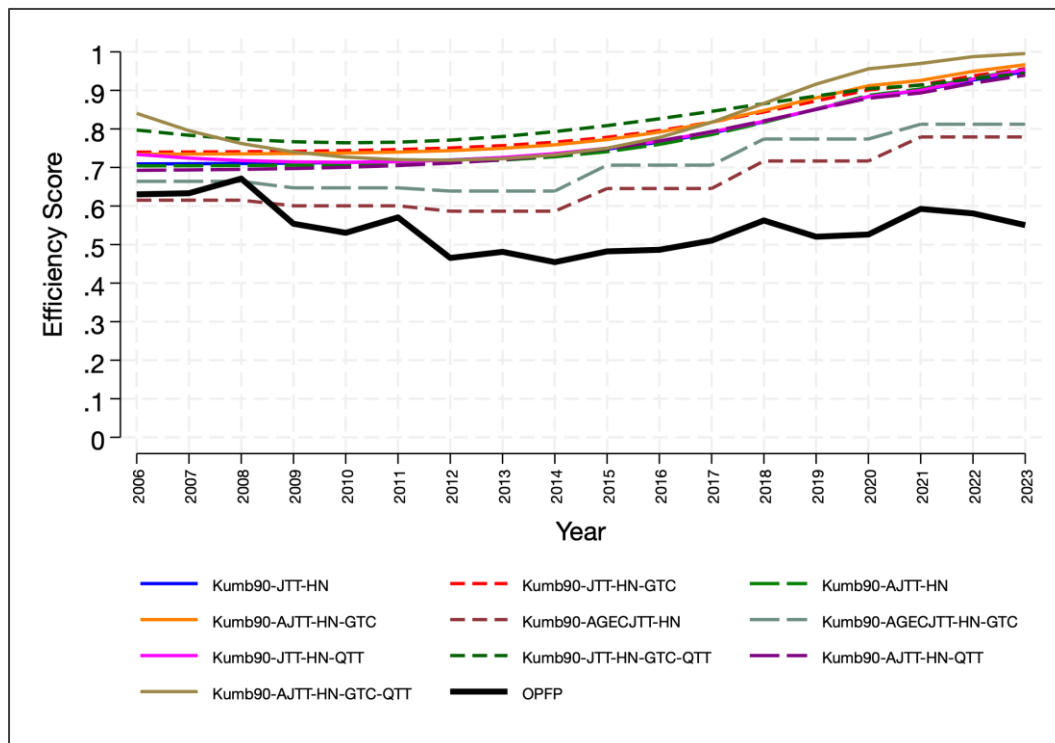


Figure 4.4.3.5: CIT Kumb90 vs OPFP Trends – Translog Models

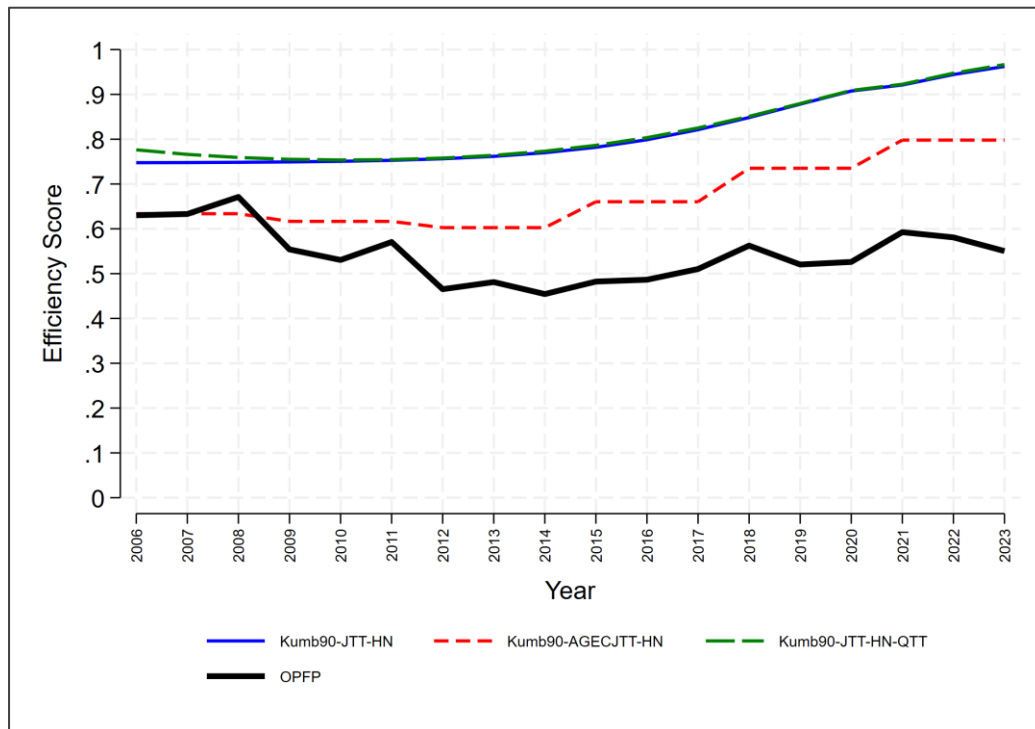


Figure 4.4.3.6: CIT Kumb90 vs OPFP Trends – Parsimonious Translog Models

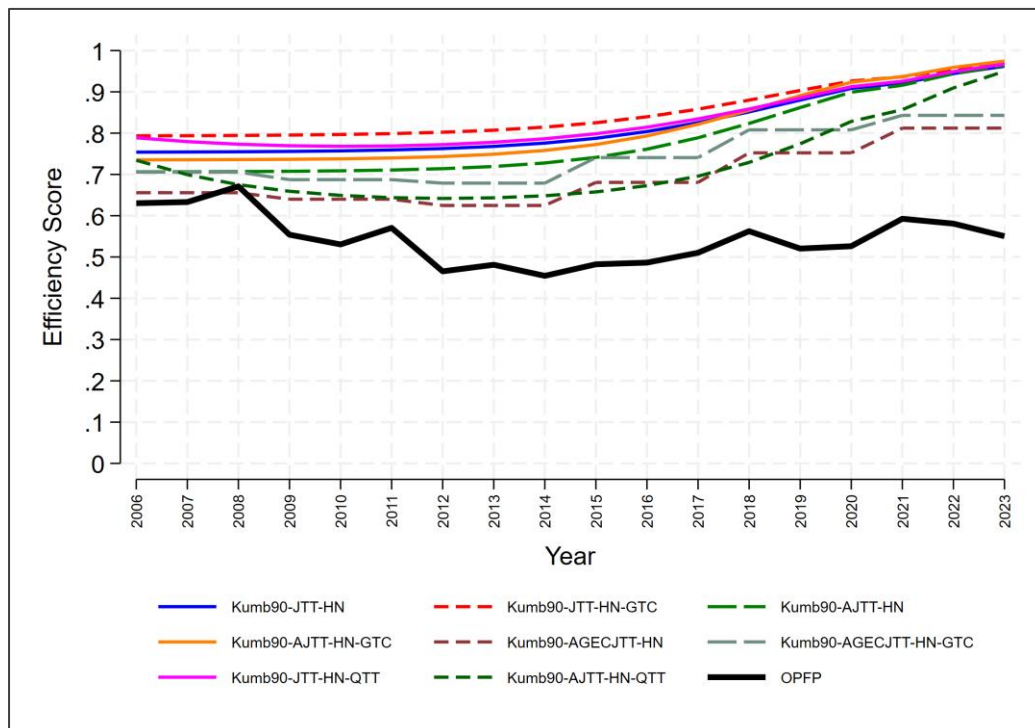


Figure 4.4.3.7: CIT LSE vs OPFP Trends – Cobb-Douglas Models

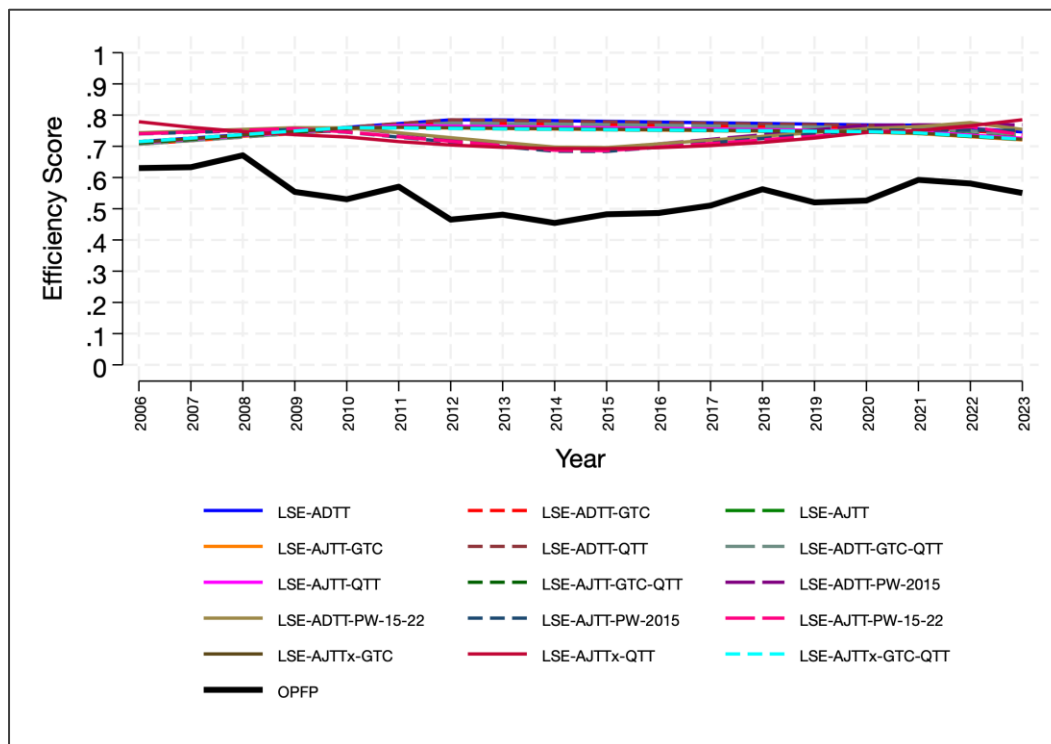


Figure 4.4.3.8: CIT LSE vs OPFP Trends – Translog Models

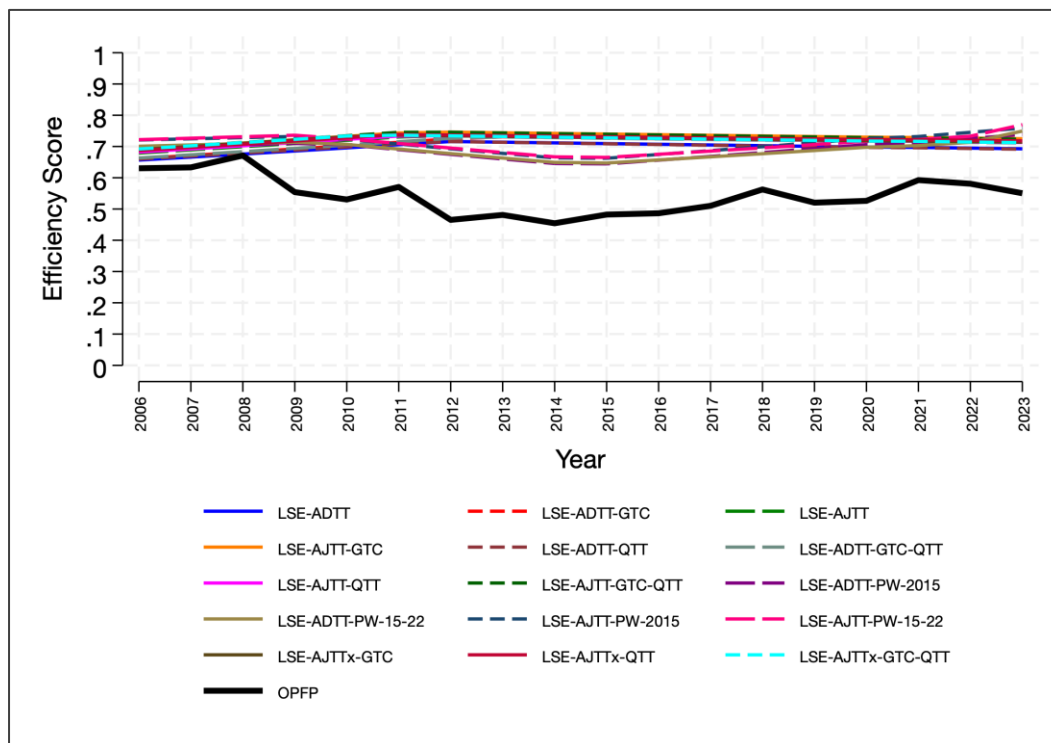
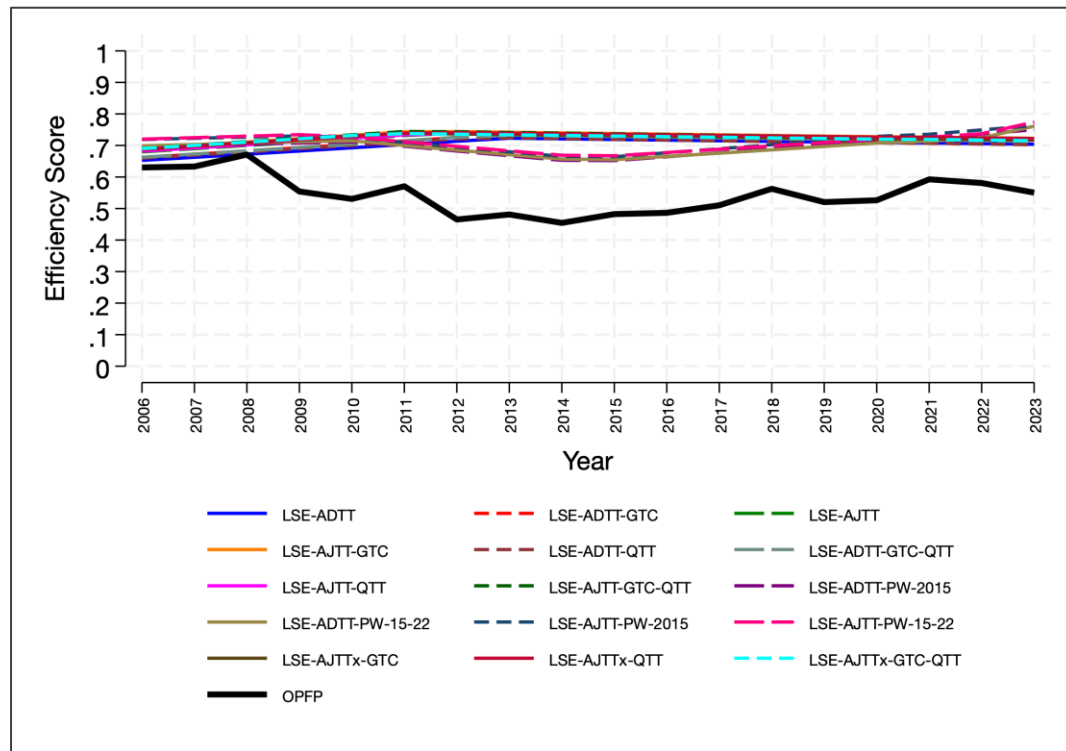


Figure 4.4.3.9: CIT LSE vs OPFP Trends – Parsimonious Translog Models



4.4.4 END

For END, Figures 4.4.4.1 to 4.4.4.9 show that:

- OPFP efficiency fluctuated between 2006 and 2016, reaching its minimum in 2008 and further low points in 2012 and 2014–2016. It then generally increased strongly through 2021, despite a slight decrease in 2019, before decreasing in 2022 and increasing again in 2023. The efficiency score trends for the BC95 specifications closely reproduce these movements but generate consistently higher scores. After 2020, they diverge slightly, with some increasing and others stabilising or decreasing.
- Most Kumb90 specifications decreased slightly or remained stable until approximately 2009–2012 and increased thereafter, particularly after 2015. The AJTT paths initially increase at an accelerating rate until around 2019–2020, after which the rate of increase slows. The AGE CJTT specifications show a stepwise pattern, decreasing slightly until 2012–2014 and then increasing through successive adjustments and plateaus.
- The standard and GTC LSE specifications show a smooth, almost linear increase throughout the period, with growth slightly decelerating over time. In contrast, the QTT and piecewise specifications remaining relatively stable until approximately 2015 and then increase at an accelerating rate. QTT and piecewise specifications broadly reflect the OPFP patterns, although they do not reproduce the short-term fluctuations.

Figure 4.4.4.1: END BC95 vs OPFP Trends – Cobb-Douglas Models

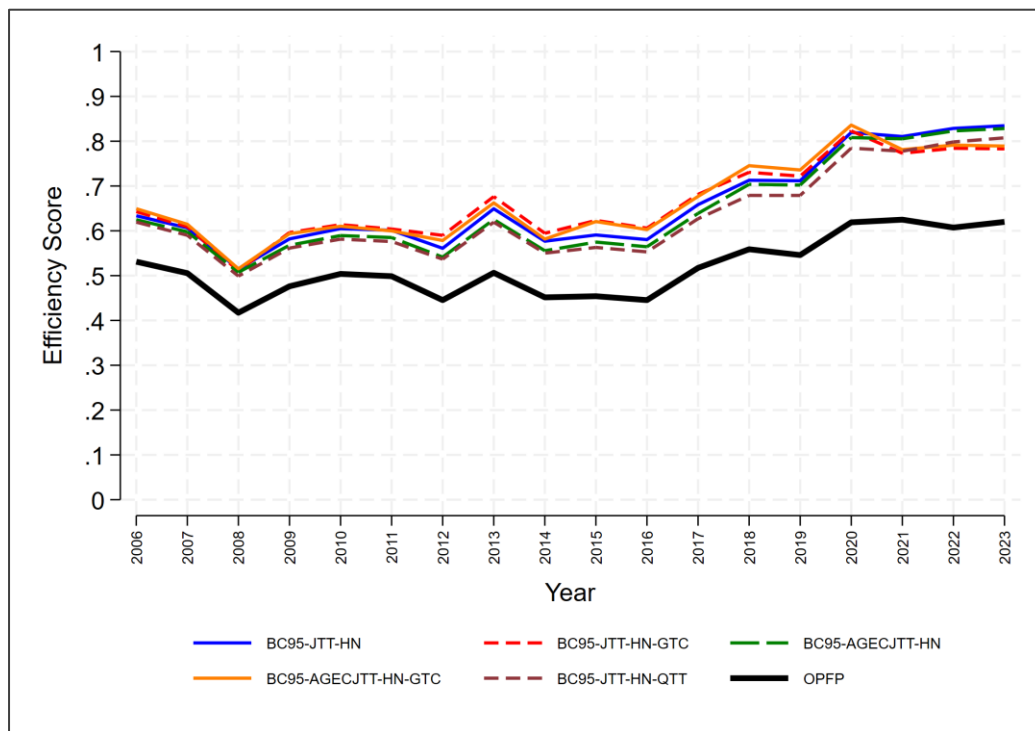


Figure 4.4.4.2: END BC95 vs OPFP Trend – Translog Models

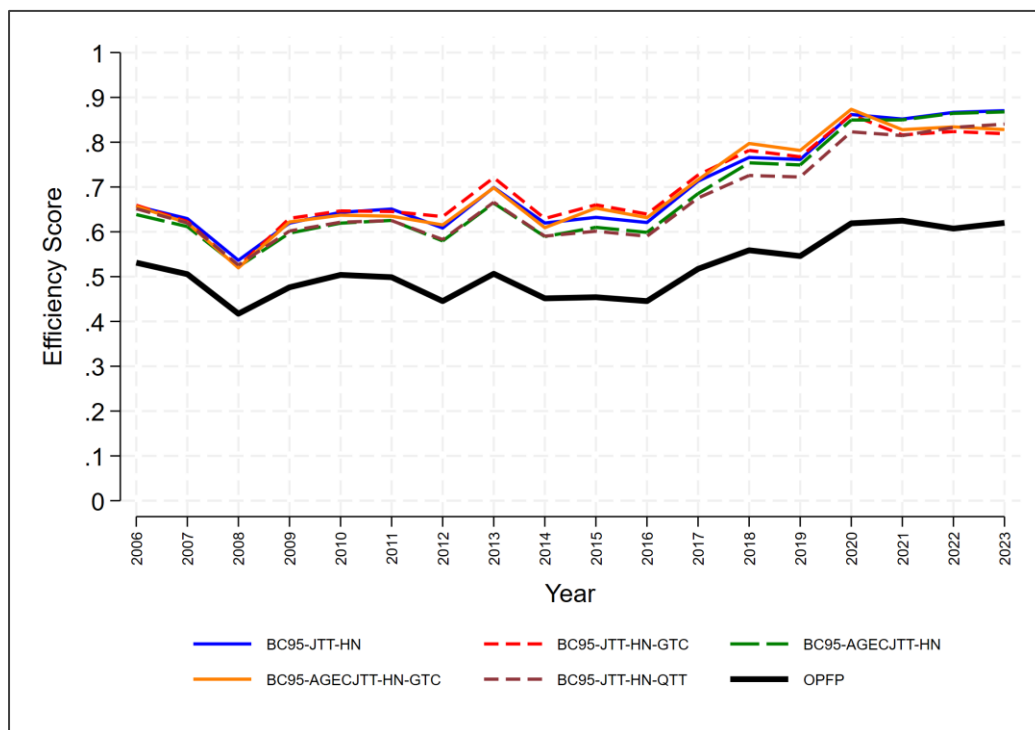


Figure 4.4.4.3: END BC95 vs OPFP Trend – Parsimonious Translog Models

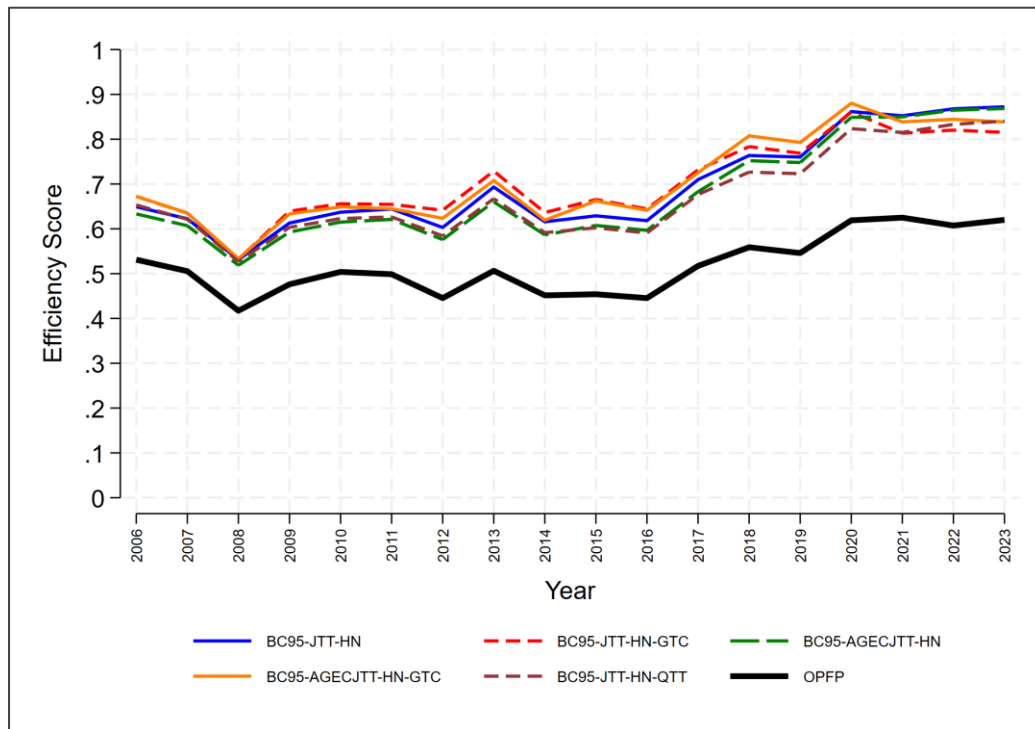


Figure 4.4.4.4: END Kumb90 vs OPFP Trends – Cobb-Douglas Models

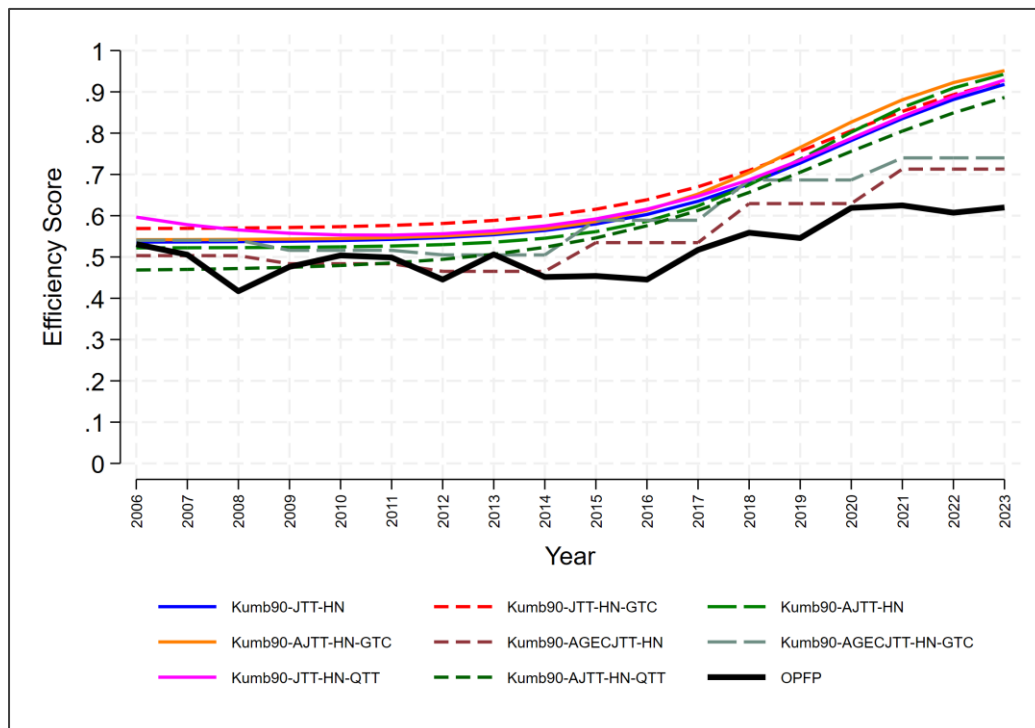


Figure 4.4.4.5: END Kumb90 vs OPFP Trends – Translog Models

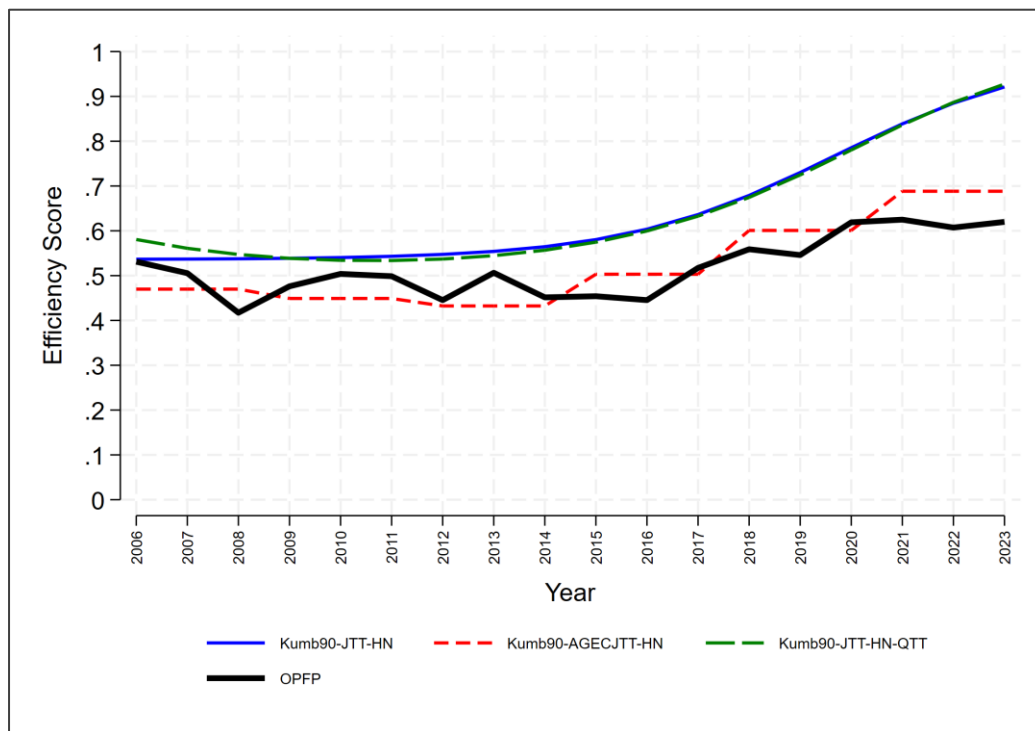


Figure 4.4.4.6: END Kumb90 vs OPFP Trends – Parsimonious Translog Models

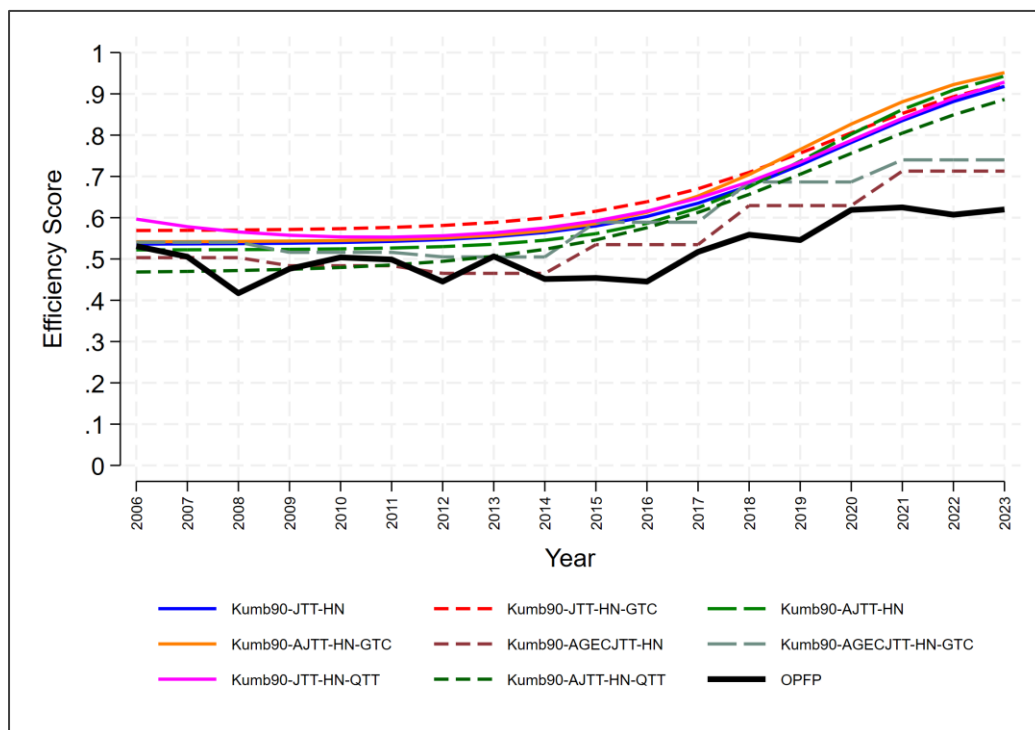


Figure 4.4.4.7: END LSE vs OPFP Trends – Cobb-Douglas Models

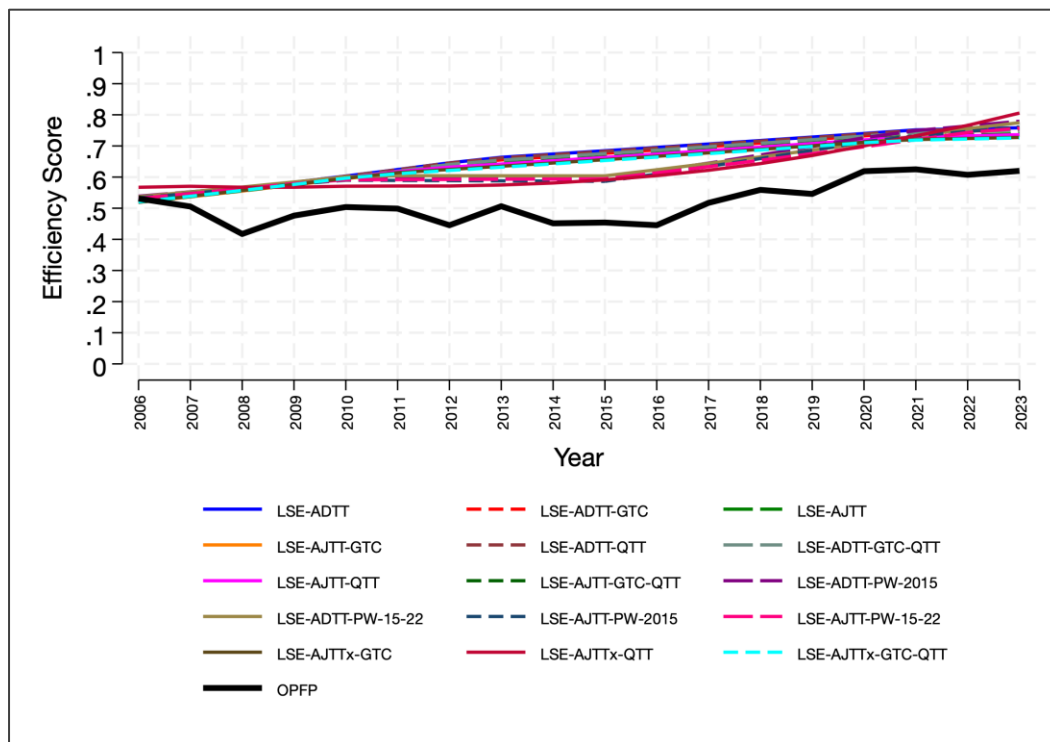


Figure 4.4.4.8: END LSE vs OPFP Trends – Translog Models

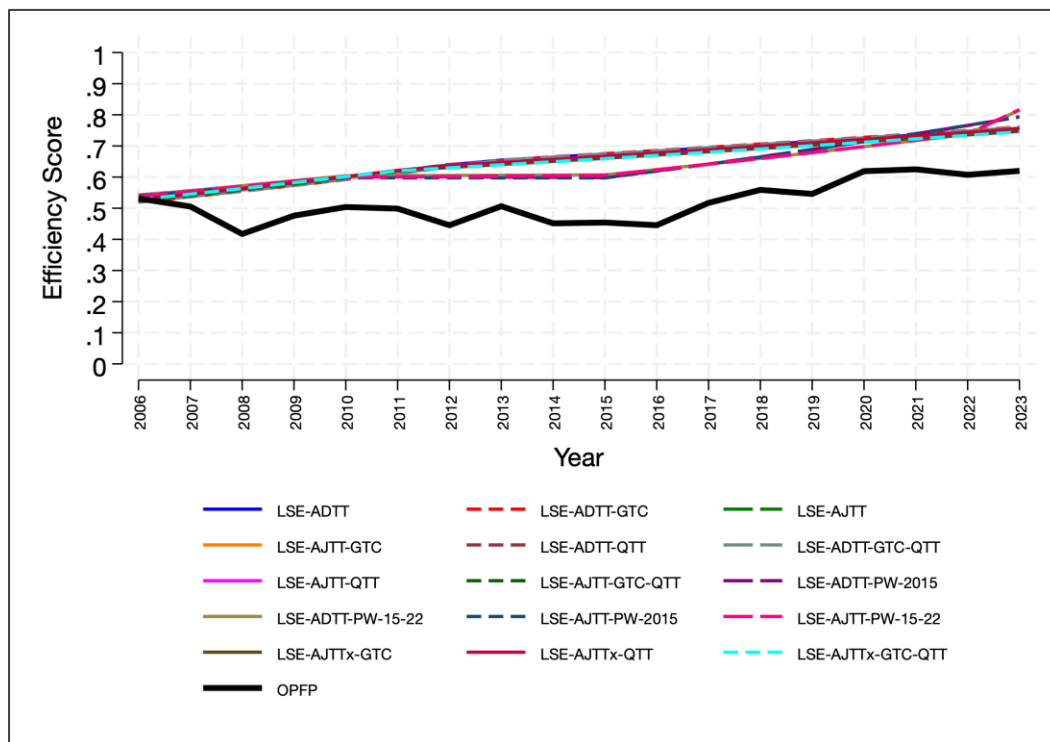
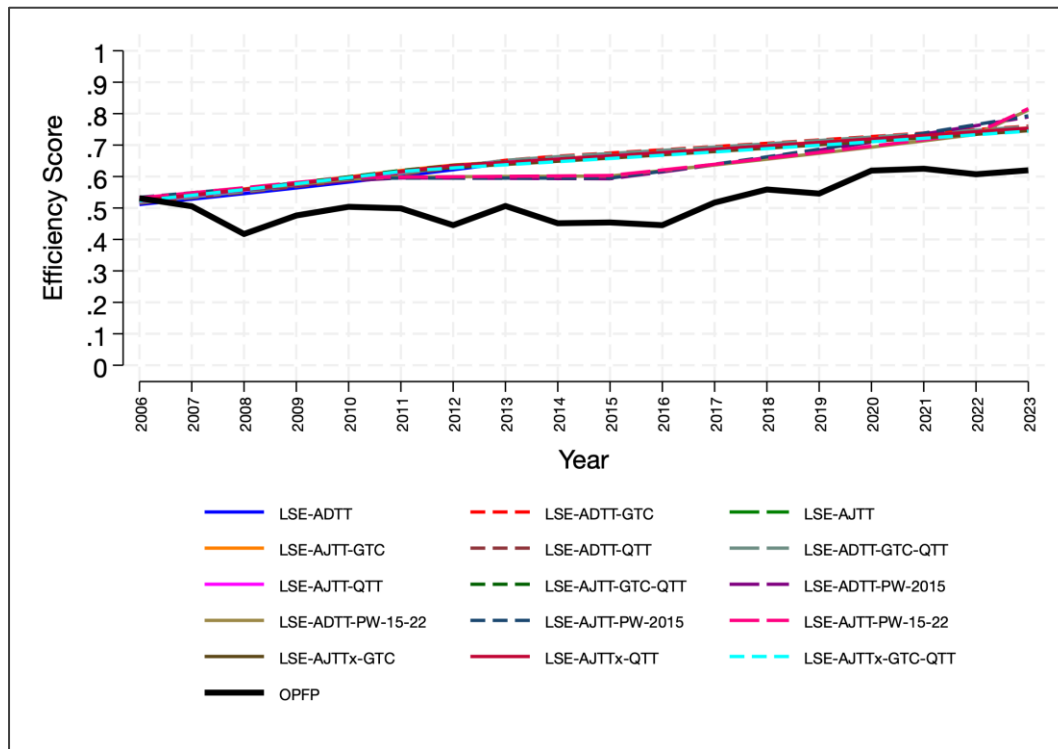


Figure 4.4.4.9: END LSE vs OPFP Trends – Parsimonious Translog Models



4.4.5 ENX

For ENX, Figures 4.4.5.1 to 4.4.5.9 show that:

- OPFP efficiency generally decreased from 2006 to 2013, despite temporary increases in 2009–2010, and reached a low point in 2013. After increasing in 2014 and decreasing again in 2015, efficiency increased strongly through 2020 and then decreased each year from 2021 to 2023. The efficiency score trends for the BC95 specifications closely reproduce these movements but generate consistently higher scores.
- Most Kumb90 specifications remained stable or decreased slightly until approximately 2009–2012 and increased thereafter, particularly after 2015. Their smooth curves initially increase at an accelerating rate until around 2019–2020, after which the rate of increase slows. The AGE CJTT specifications show a stepwise pattern, decreasing slightly until 2012–2014 and then increasing through successive adjustments and plateaus.
- The standard and GTC LSE specifications show smooth paths in which efficiency increases at a slowing rate until around 2013–2015, remains broadly stable through 2020, and then decreases. In contrast, the QTT and piecewise specifications show efficiency decreasing until around 2015 after small initial increases, before increasing at an accelerating rate.

Figure 4.4.5.1: ENX BC95 vs OPFP Trends – Cobb-Douglas Models

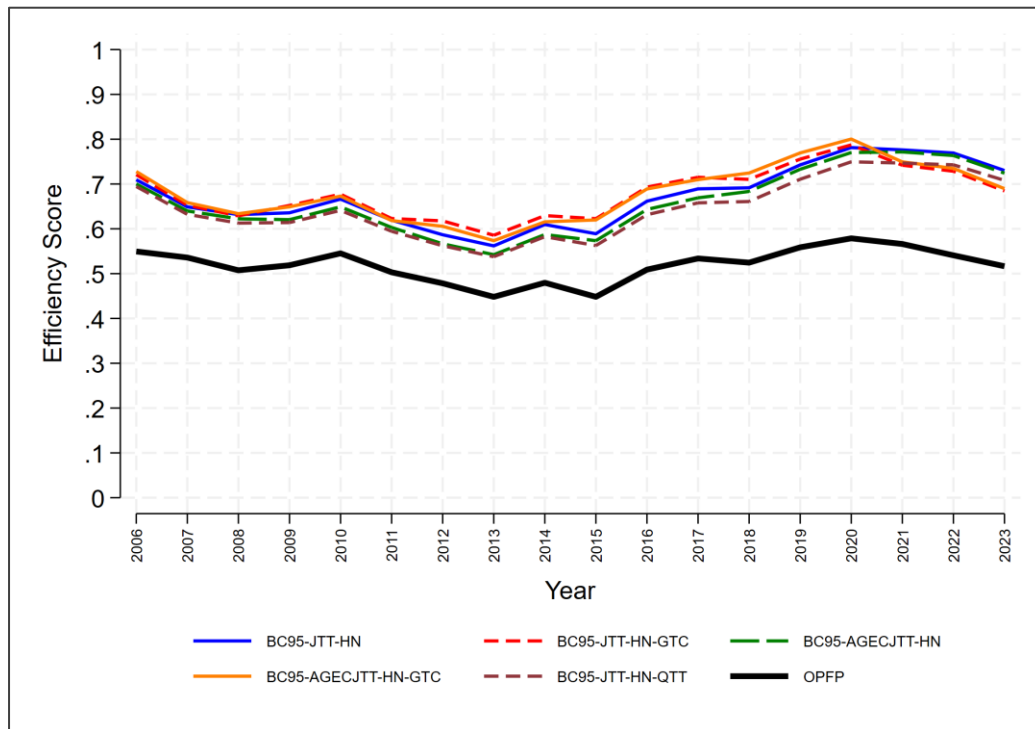


Figure 4.4.5.2: ENX BC95 vs OPFP Trend – Translog Models

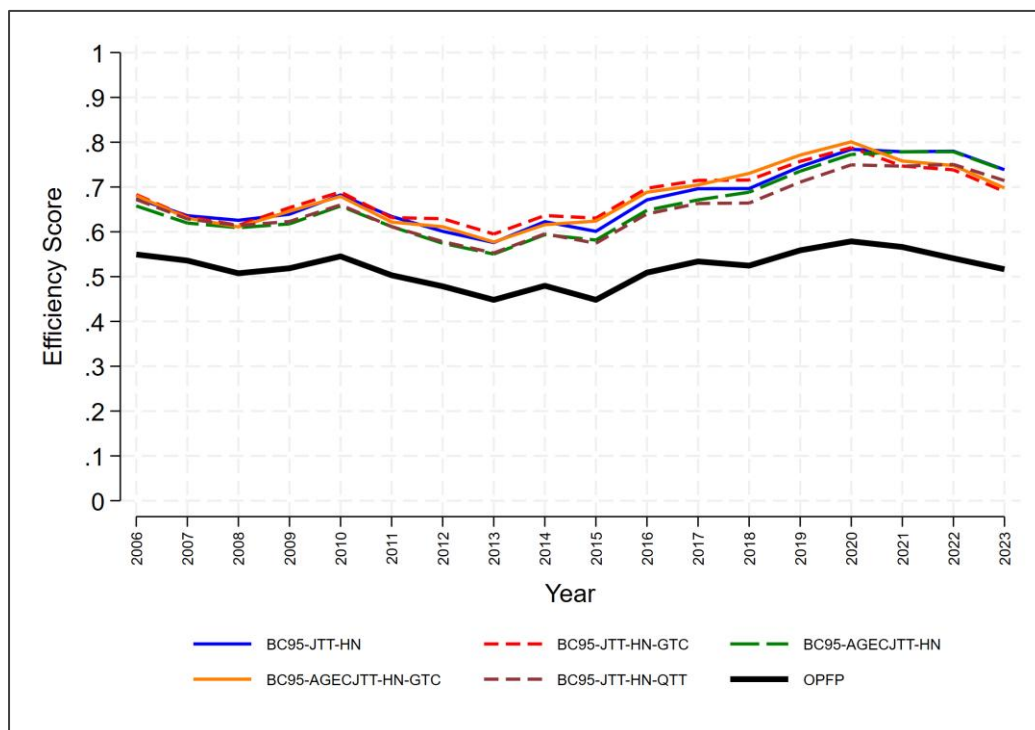


Figure 4.4.5.3: ENX BC95 vs OPFP Trend – Parsimonious Translog Models

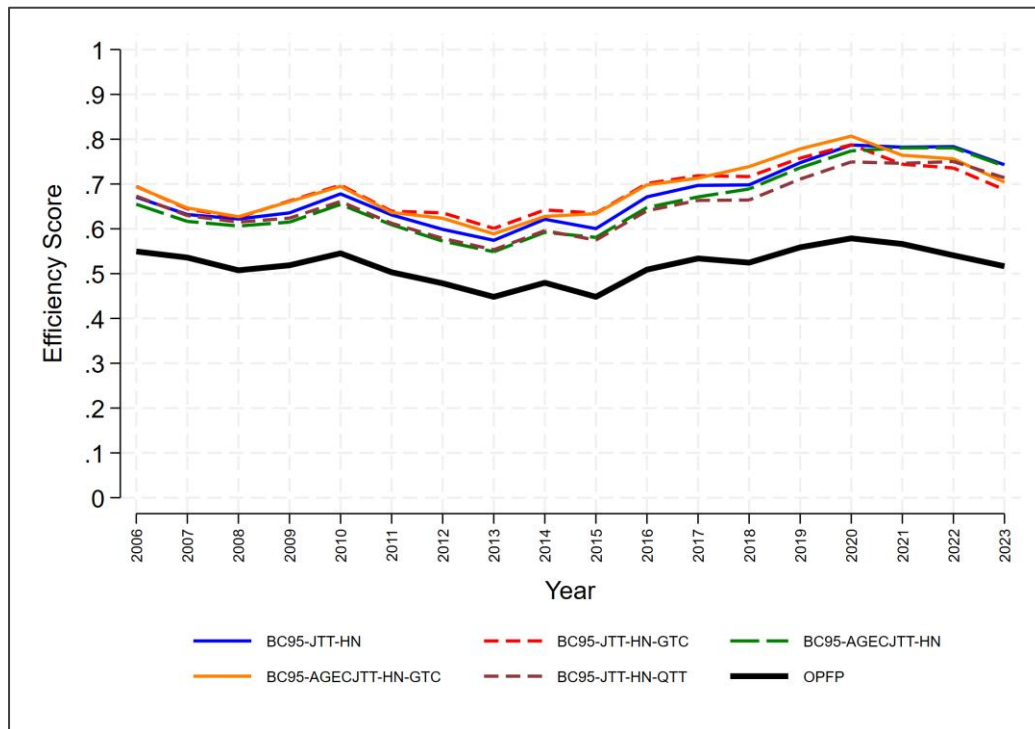


Figure 4.4.5.4: ENX Kumb90 vs OPFP Trends – Cobb-Douglas Models

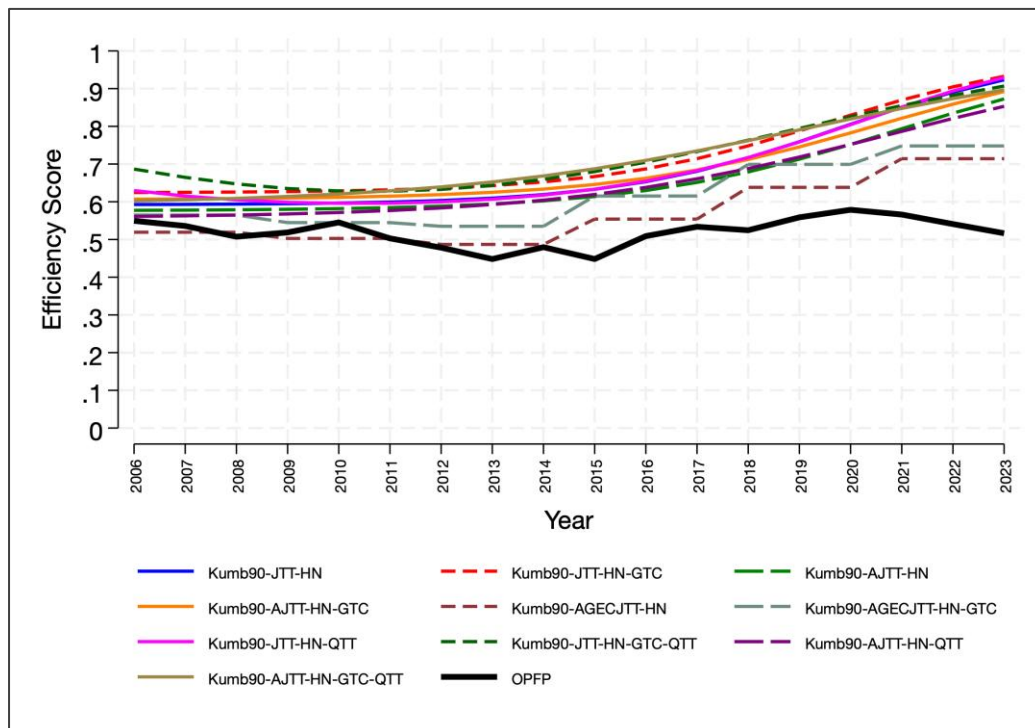


Figure 4.4.5.5: ENX Kumb90 vs OPFP Trends – Translog Models

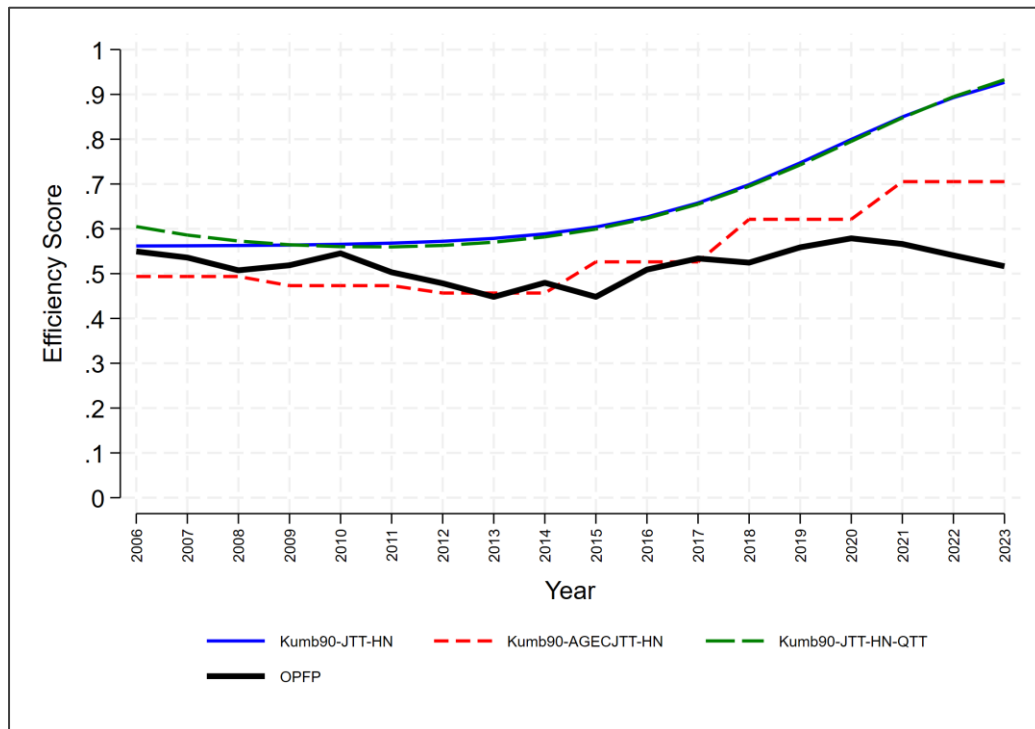


Figure 4.4.5.6: ENX Kumb90 vs OPFP Trends – Parsimonious Translog Models

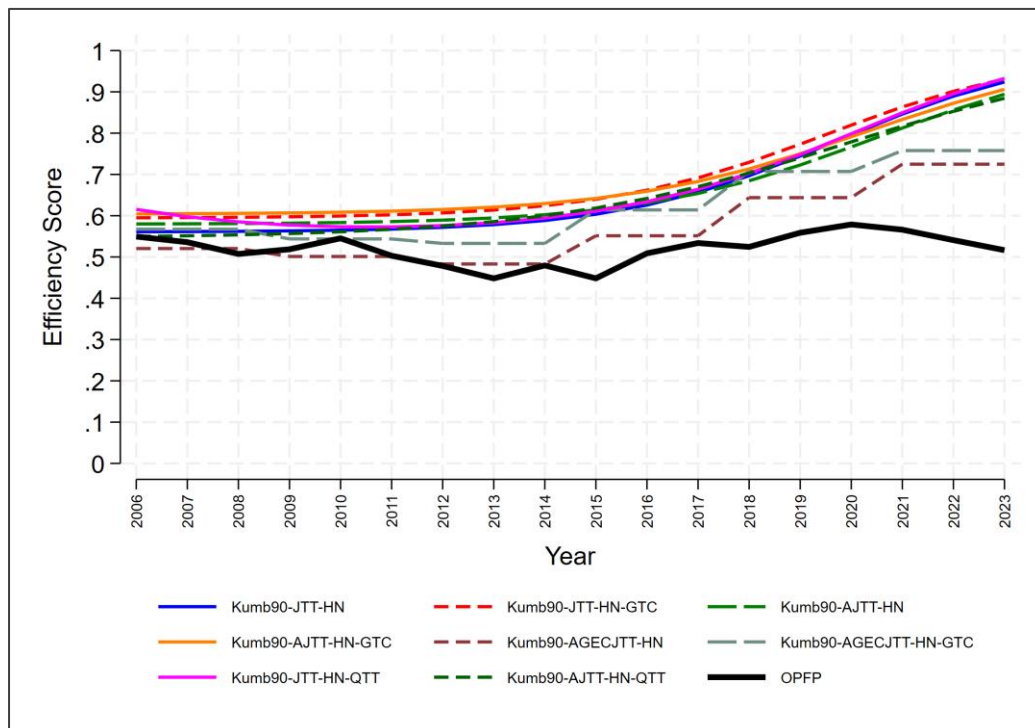


Figure 4.4.5.7: ENX LSE vs OPFP Trends – Cobb-Douglas Models

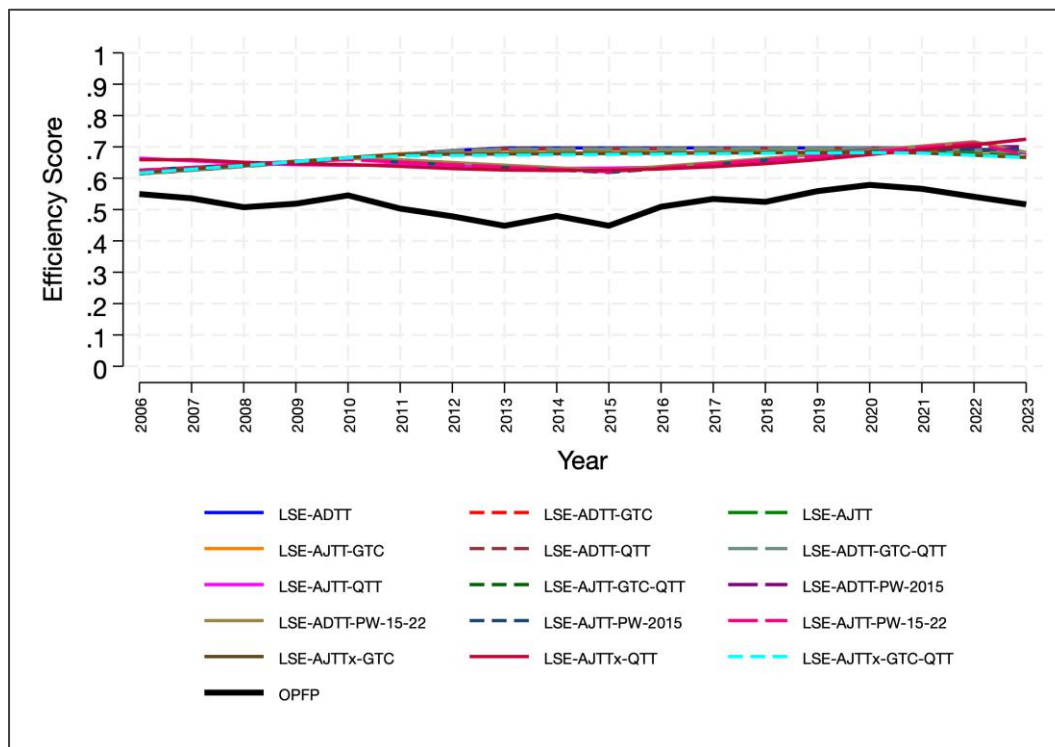


Figure 4.4.5.8: ENX LSE vs OPFP Trends – Translog Models

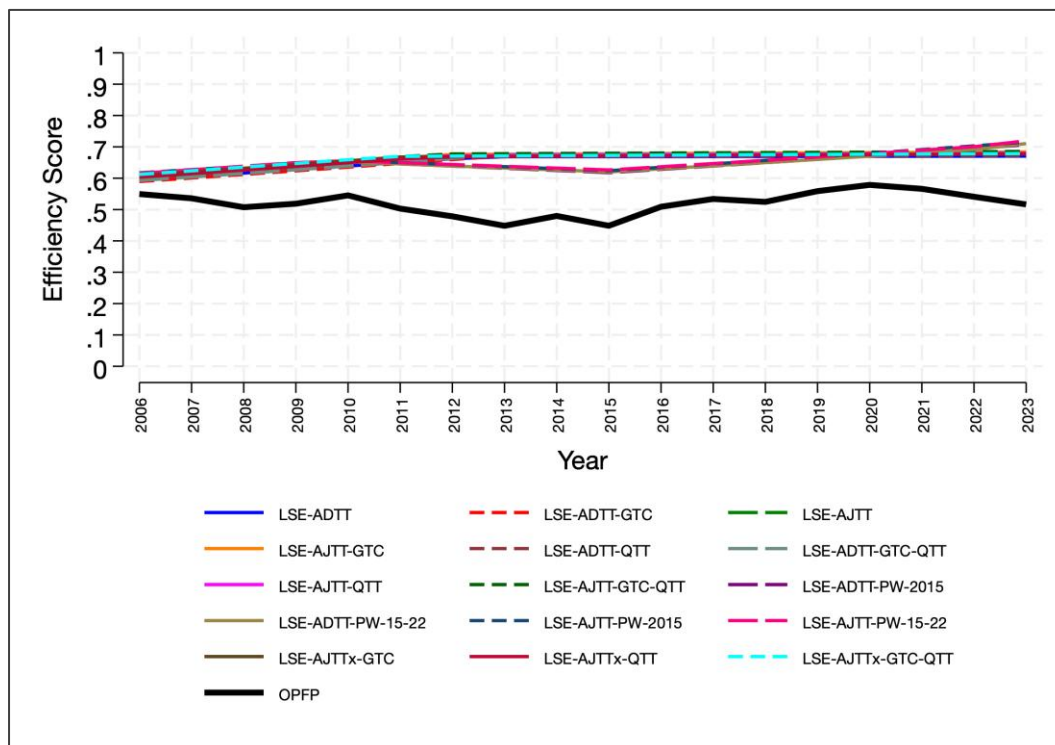
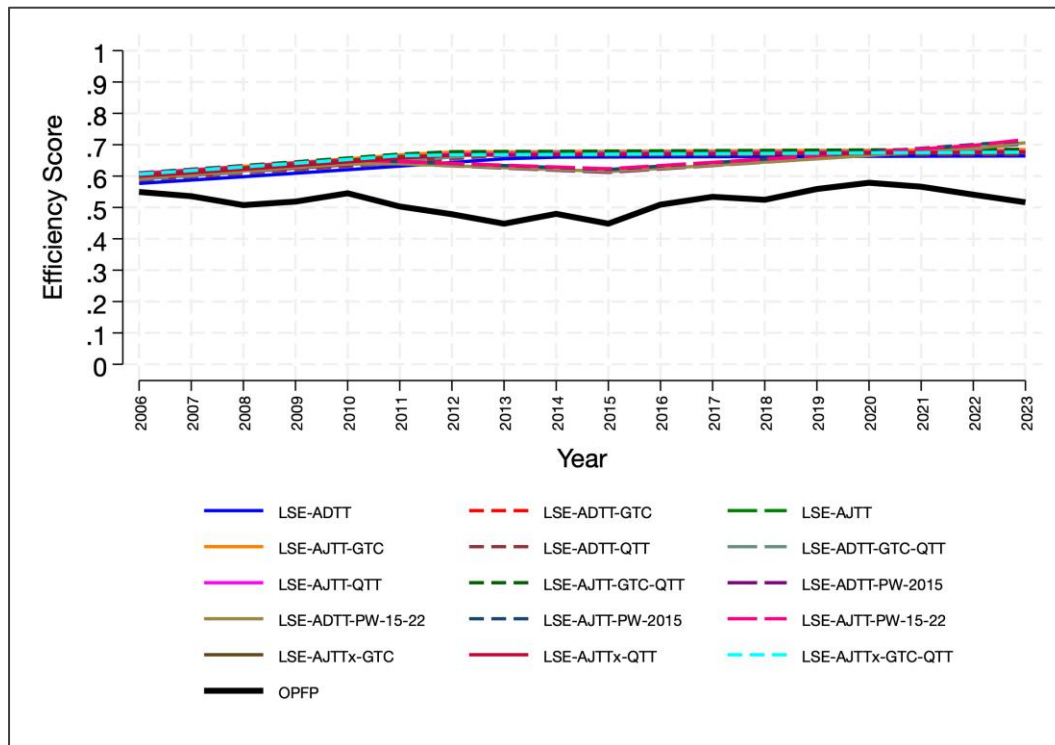


Figure 4.4.5.9: ENX LSE vs OPFP Trends – Parsimonious Translog Models



4.4.6 ERG

For ERG, Figures 4.4.6.1 to 4.4.6.9 show that:

- OPFP efficiency fluctuated considerably with an underlying upward trend. It increased in 2007, decreased through 2008, remained relatively stable until 2010, and decreased again through 2012. Efficiency increased in 2013–2014, decreased in 2015, increased through 2017, decreased until 2020, increased sharply in 2021, and decreased again in 2022–2023. The efficiency score trends for the BC95 specifications closely reproduce these movements but generate consistently higher scores.
- Most Kumb90 specifications show smooth, remaining stable or decreasing slightly until around 2010–2013 and increasing at an accelerating rate thereafter, particularly after 2015. The JTT specifications increase more strongly than the AJTT specifications, reaching approximately 0.84–0.89 by 2023, compared with approximately 0.65–0.77 for the AJTT specifications. The AGE CJTT specifications decrease until 2012–2014 and then increase through stepwise adjustments and plateaus.
- All LSE specifications show similar smooth curves, with efficiency increasing gradually at a slightly slowing rate from 2006 to 2022. The QTT and piecewise specifications are marginally lower during the middle years, but differences across LSE specifications are small.

Figure 4.4.6.1: ERG BC95 vs OPFP Trends – Cobb-Douglas Models

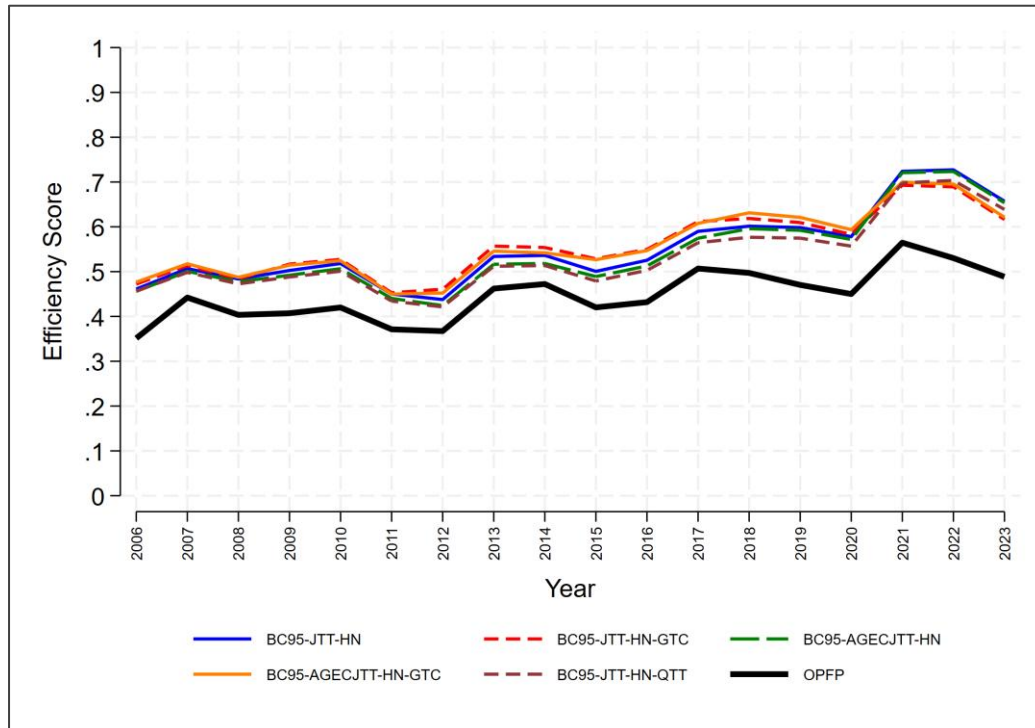


Figure 4.4.6.2: ERG BC95 vs OPFP Trend – Translog Models

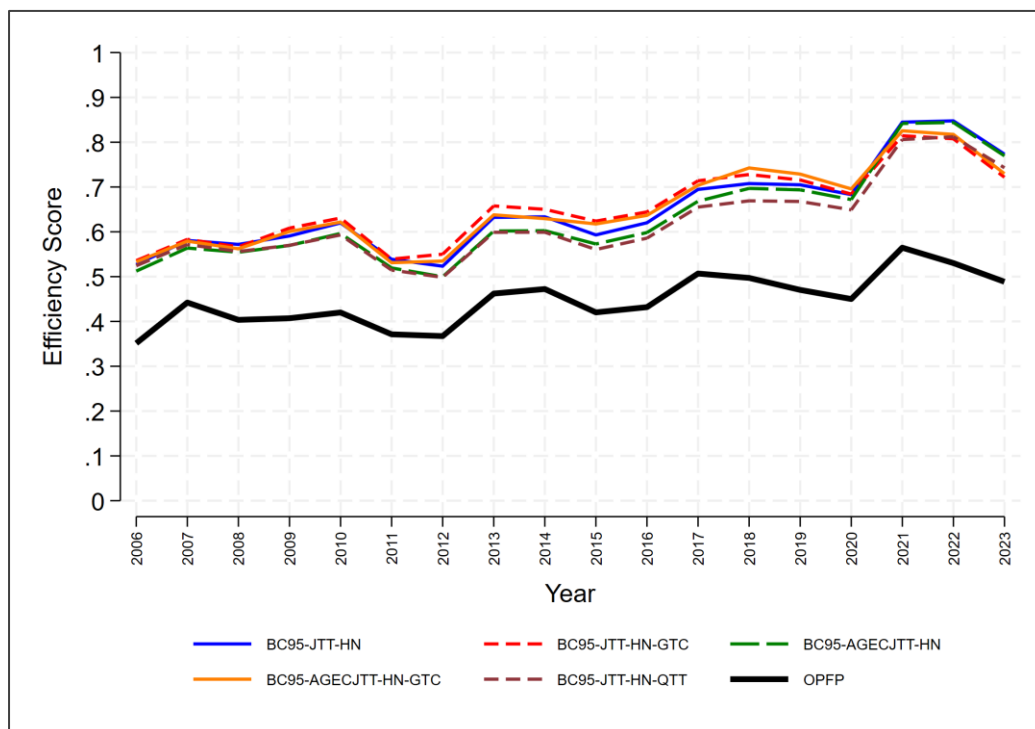


Figure 4.4.6.3: ERG BC95 vs OPFP Trend – Parsimonious Translog Models

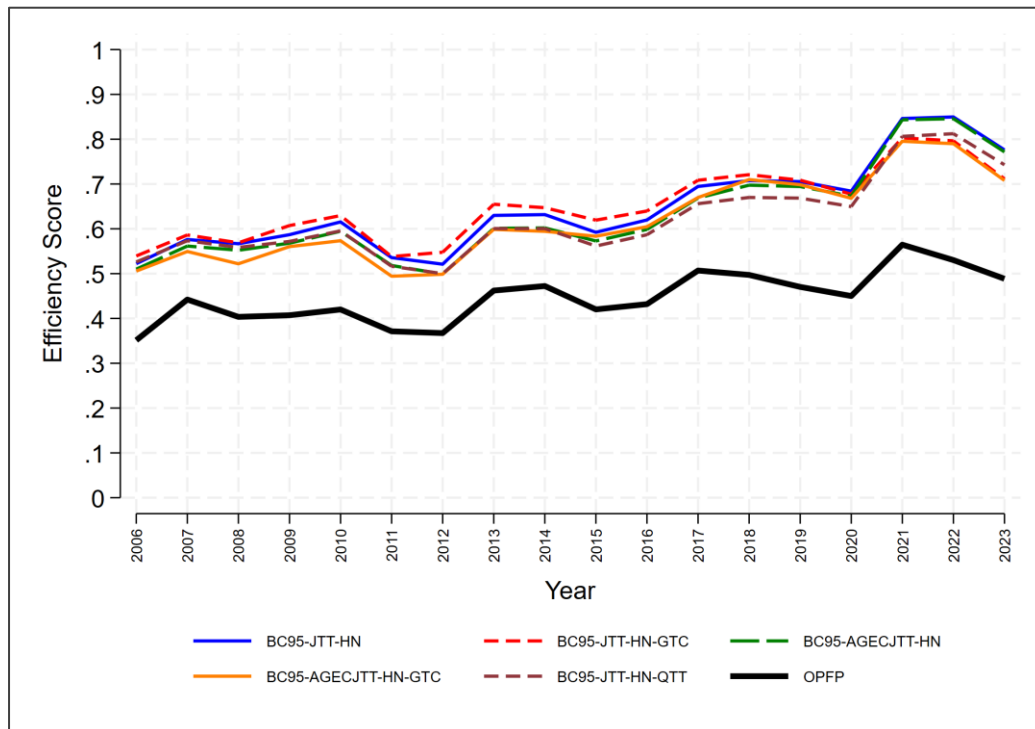


Figure 4.4.6.4: ERG Kumb90 vs OPFP Trends – Cobb-Douglas Models

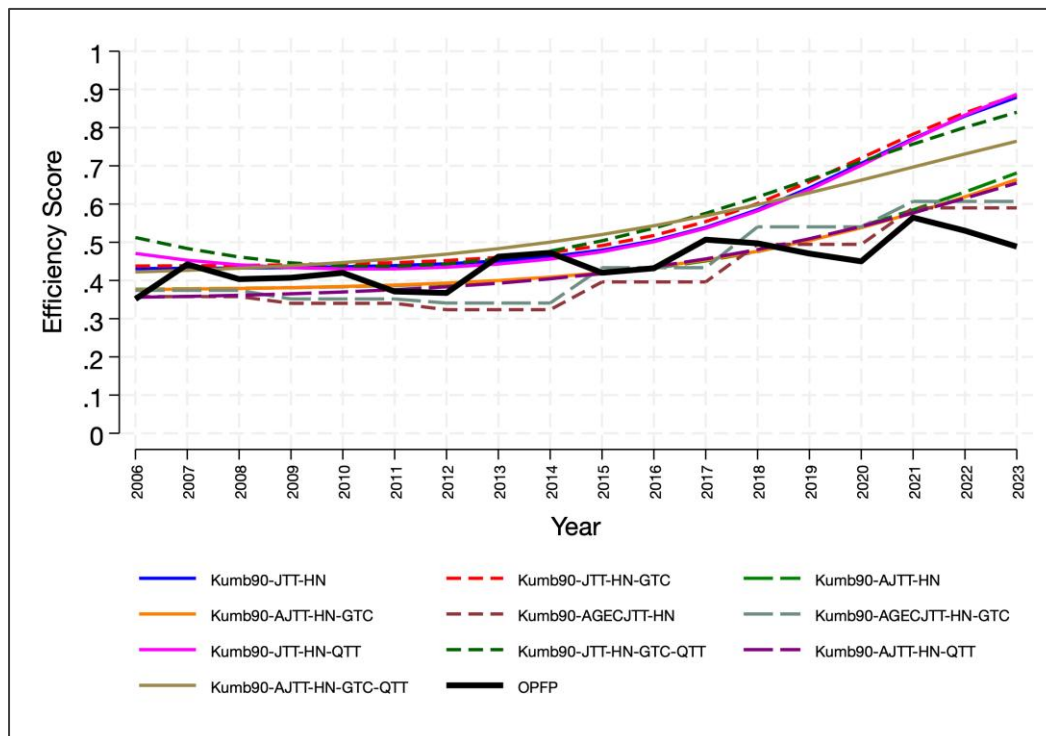


Figure 4.4.6.5: ERG Kumb90 vs OPFP Trends – Translog Models

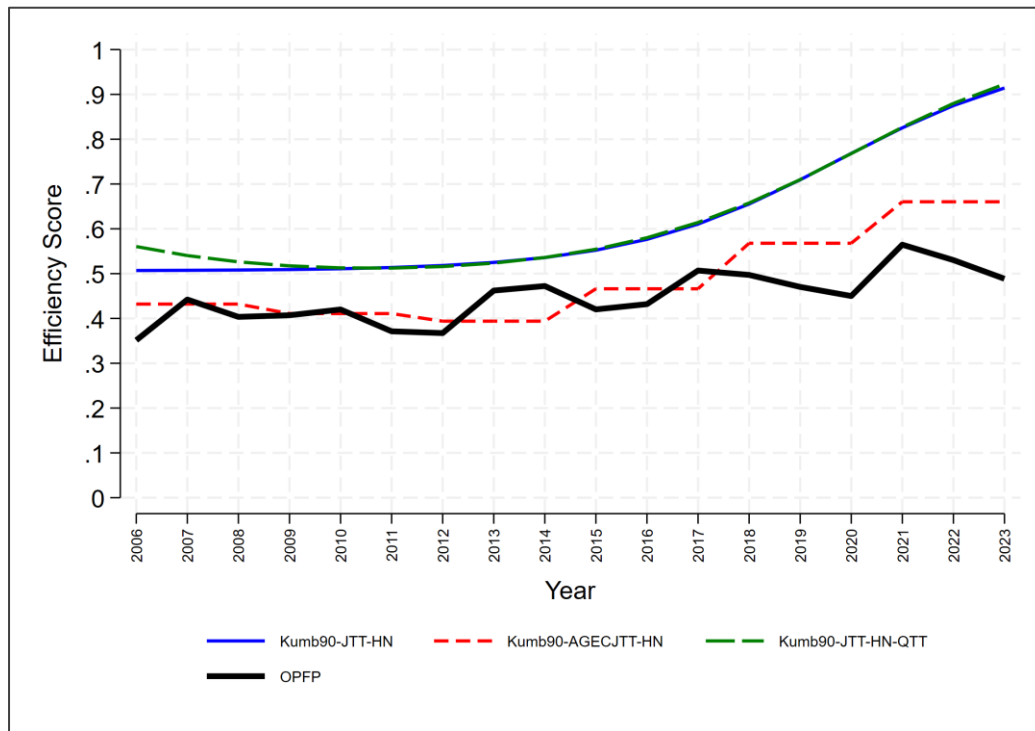


Figure 4.4.6.6: ERG Kumb90 vs OPFP Trends – Parsimonious Translog Models

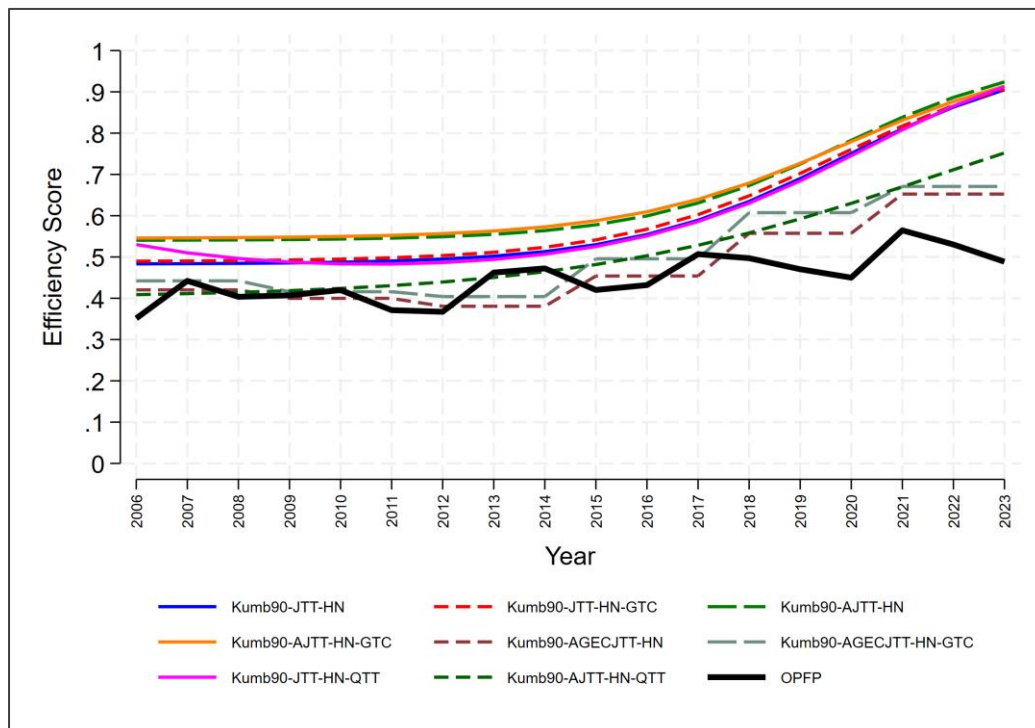


Figure 4.4.6.7: ERG LSE vs OPFP Trends – Cobb-Douglas Models

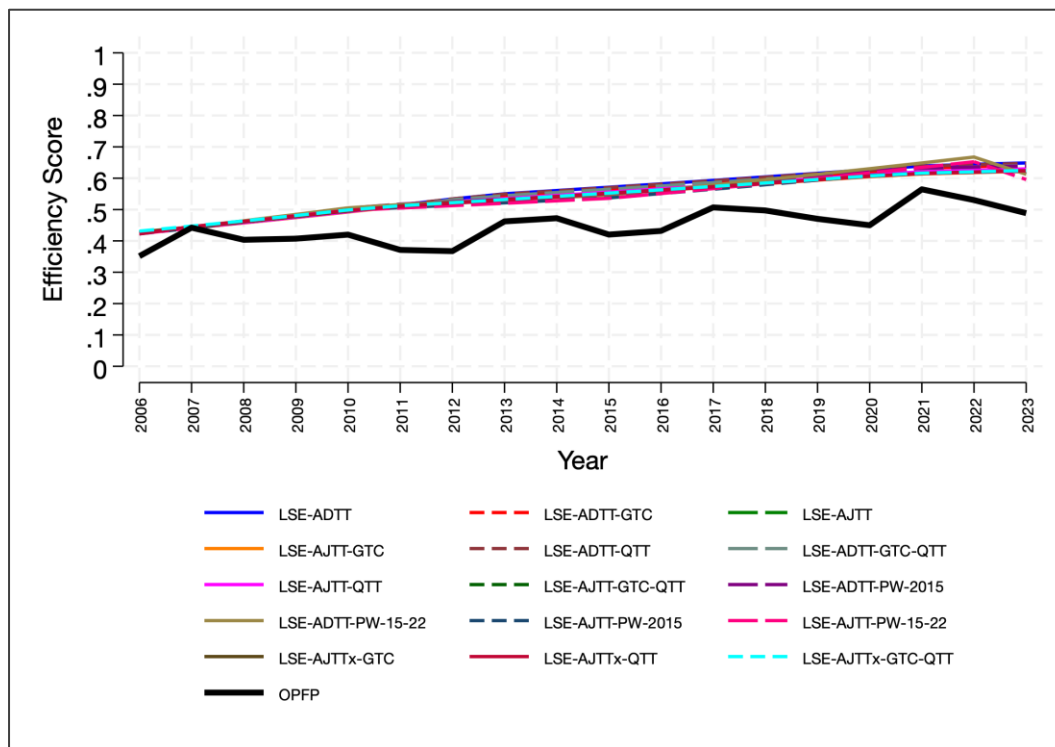


Figure 4.4.6.8: ERG LSE vs OPFP Trends – Translog Models

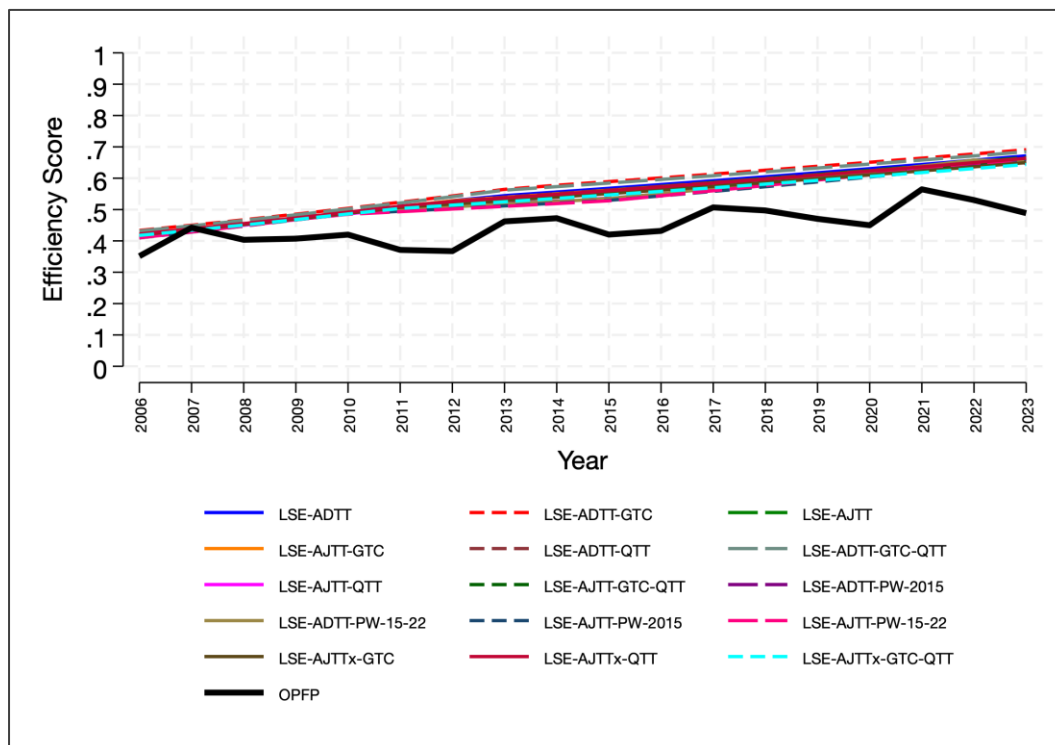
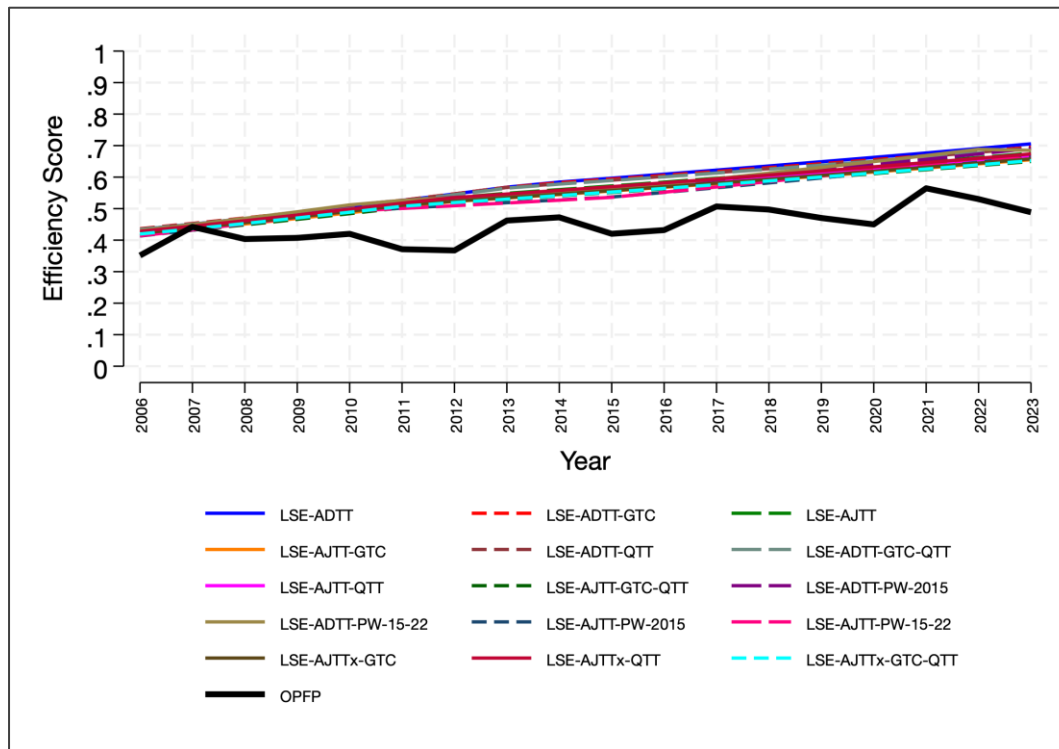


Figure 4.4.6.9: ERG LSE vs OPFP Trends – Parsimonious Translog Models



4.4.7 ESS

For ESS, Figures 4.4.7.1 to 4.4.7.9 show that:

- OPFP efficiency decreased from 2006 to 2008, temporarily increased in 2009, and then decreased to its minimum in 2012. It increased strongly through 2017, decreased in 2018–2019, increased again through 2022, and decreased in 2023. The efficiency score trends for the BC95 specifications closely reproduce these movements but generate consistently higher scores.
- The Kumb90 results show three distinct patterns. The JTT specifications remain broadly stable until approximately 2012 and then increase at an accelerating rate with particularly strong increases after 2017 and scores of approximately 0.86–0.90 by 2023. The AJTT specifications without QTT remain relatively stable until around 2014 and then increase at an accelerating rate, reaching approximately 0.78–0.80 in 2023. In contrast, the AJTT-QTT and AJTT-GTC-QTT specifications increase strongly between 2014 and 2020 but decrease thereafter. The AGE CJTT specifications decrease slightly until 2012–2014 and then increase through stepwise adjustments separated by plateaus.
- The standard and GTC LSE specifications show smooth increases at a slowing rate, flattening in the latest years. In contrast, the QTT and piecewise specifications remain relatively stable from approximately 2010 to 2015 and then increase at an accelerating rate. The spline 2015-2022 specifications show a decrease in 2023.

Figure 4.4.7.1: ESS BC95 vs OPFP Trends – Cobb-Douglas Models

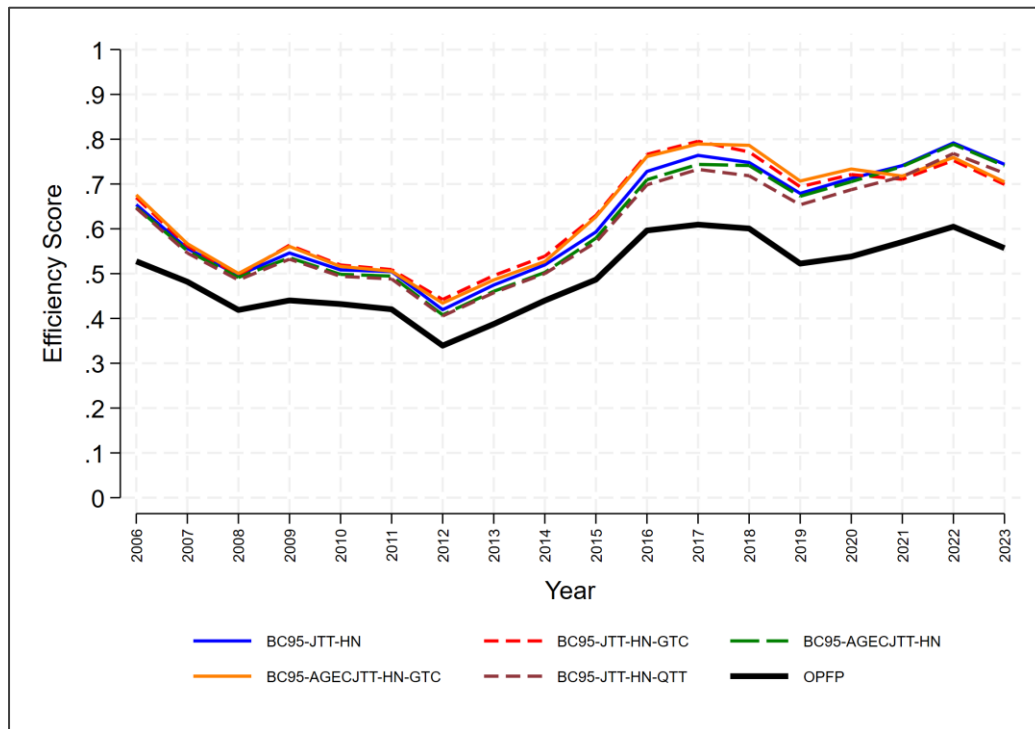


Figure 4.4.7.2: ESS BC95 vs OPFP Trend – Translog Models

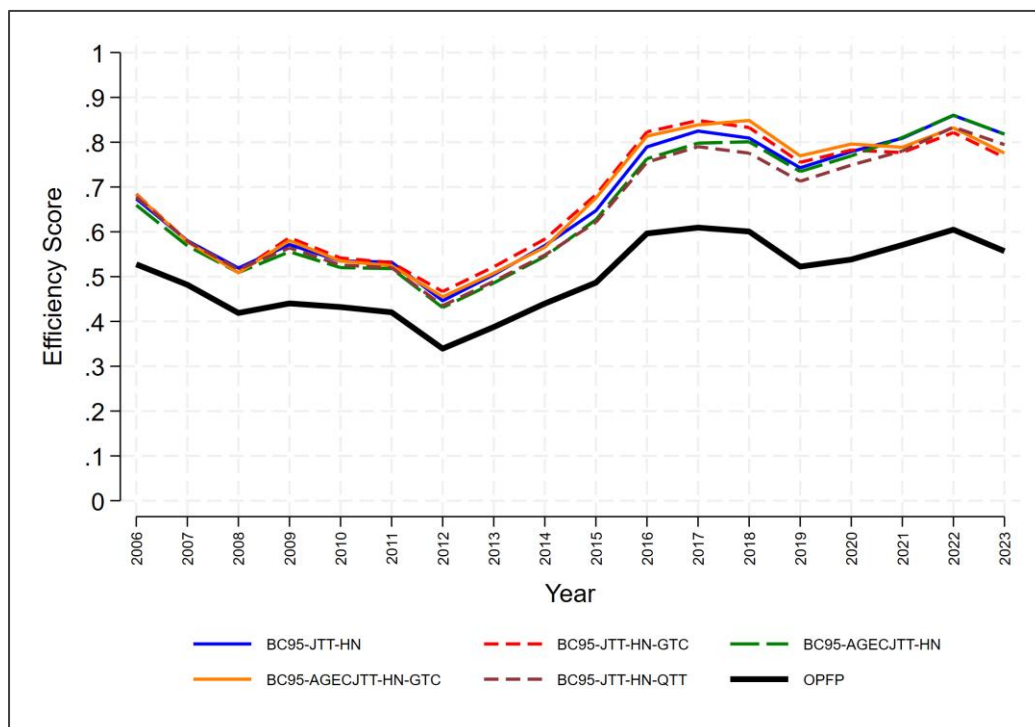


Figure 4.4.7.3: ESS BC95 vs OPFP Trend – Parsimonious Translog Models

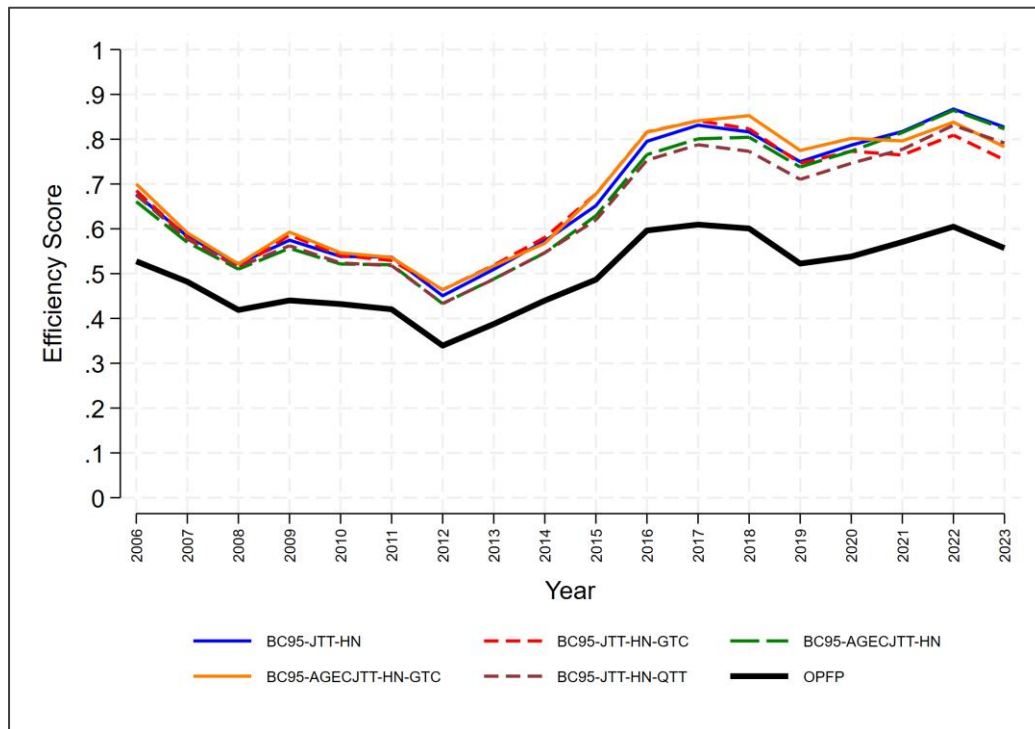


Figure 4.4.7.4: ESS Kumb90 vs OPFP Trends – Cobb-Douglas Models

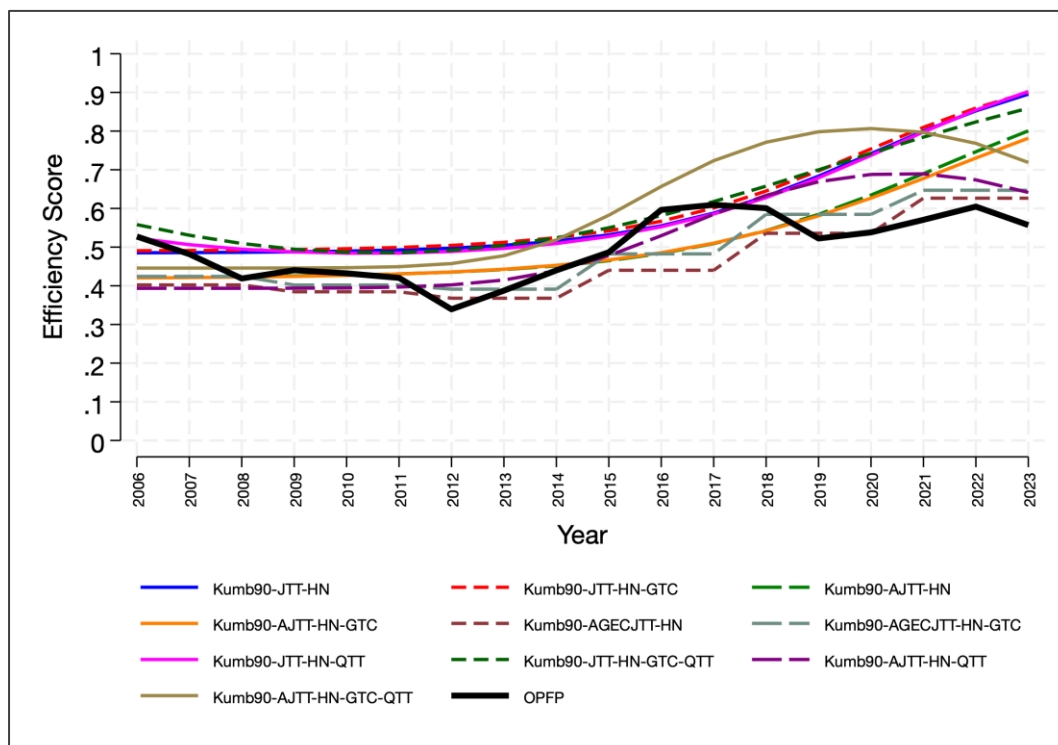


Figure 4.4.7.5: ESS Kumb90 vs OPFP Trends – Translog Models

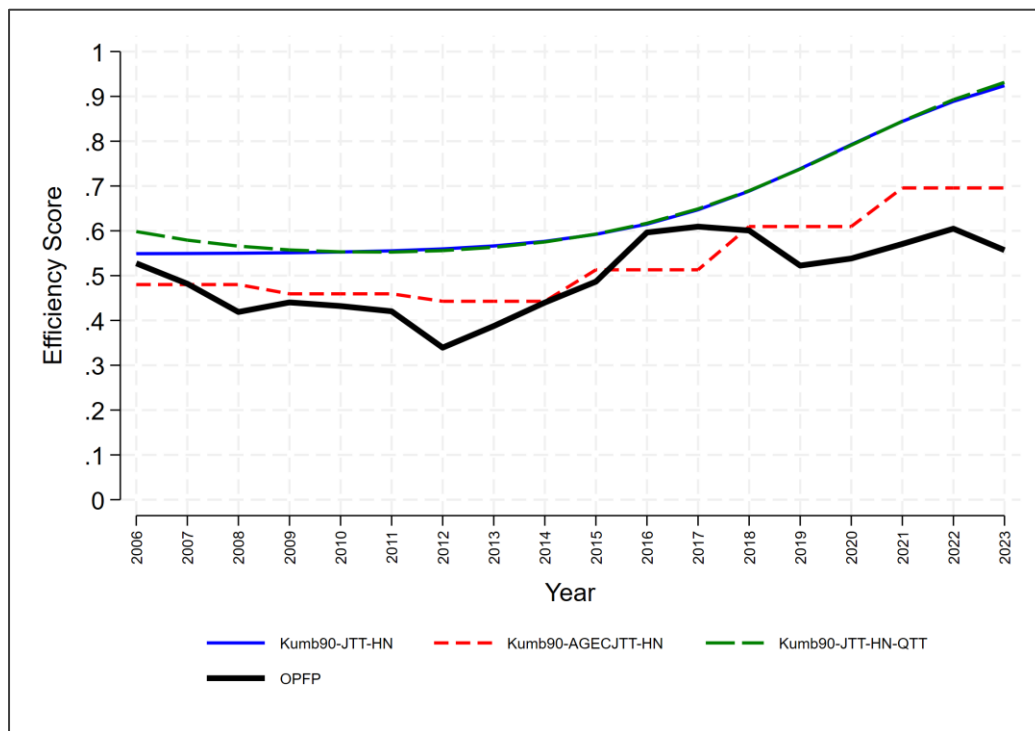


Figure 4.4.7.6: ESS Kumb90 vs OPFP Trends – Parsimonious Translog Models

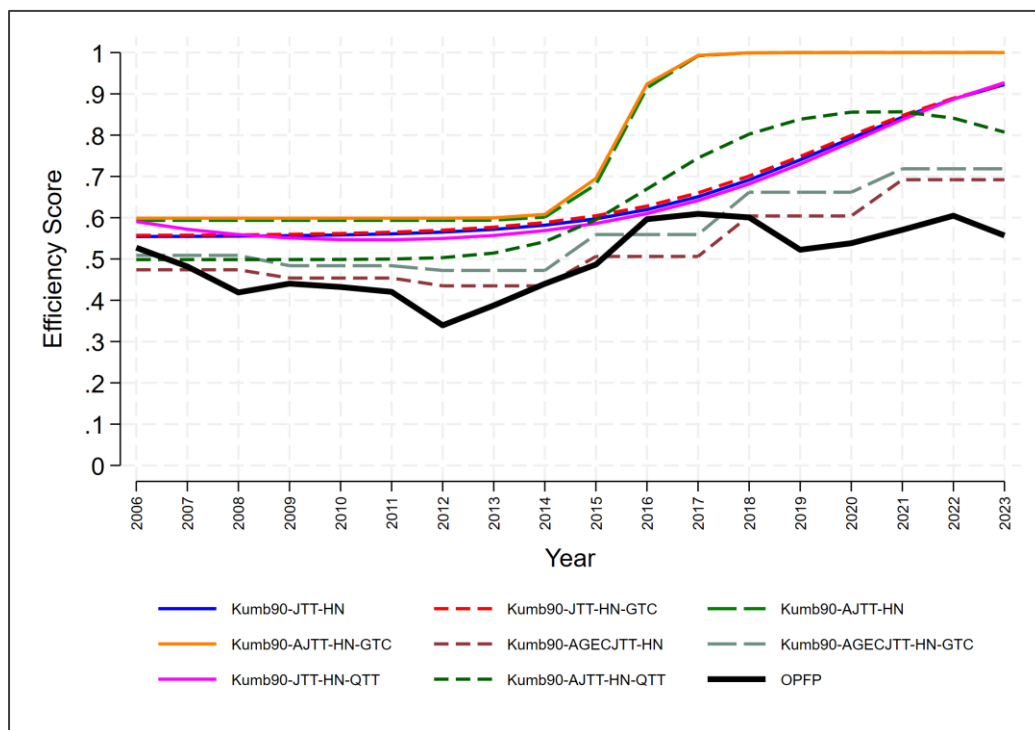


Figure 4.4.7.7: ESS LSE vs OPFP Trends – Cobb-Douglas Models

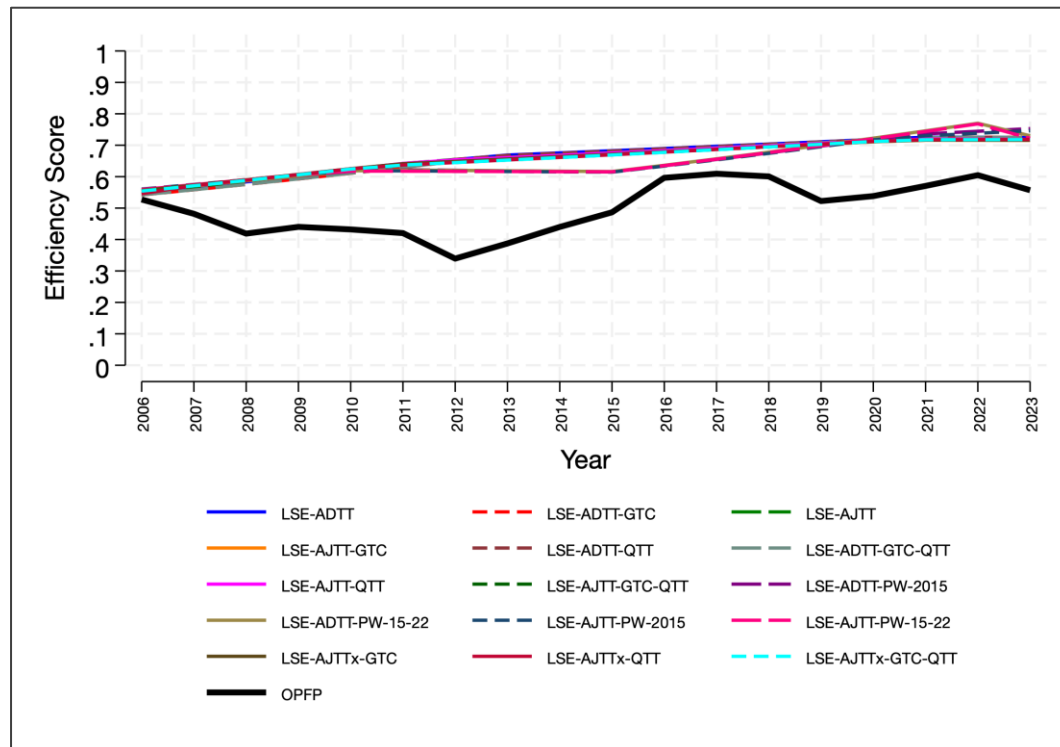


Figure 4.4.7.8: ESS LSE vs OPFP Trends – Translog Models

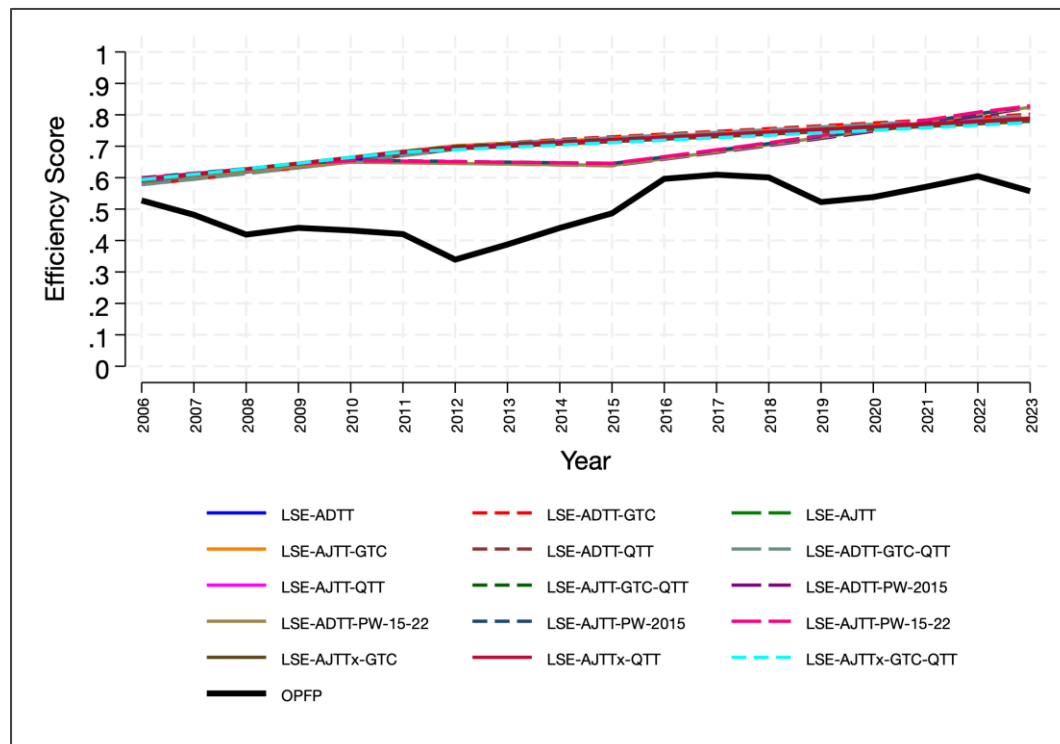
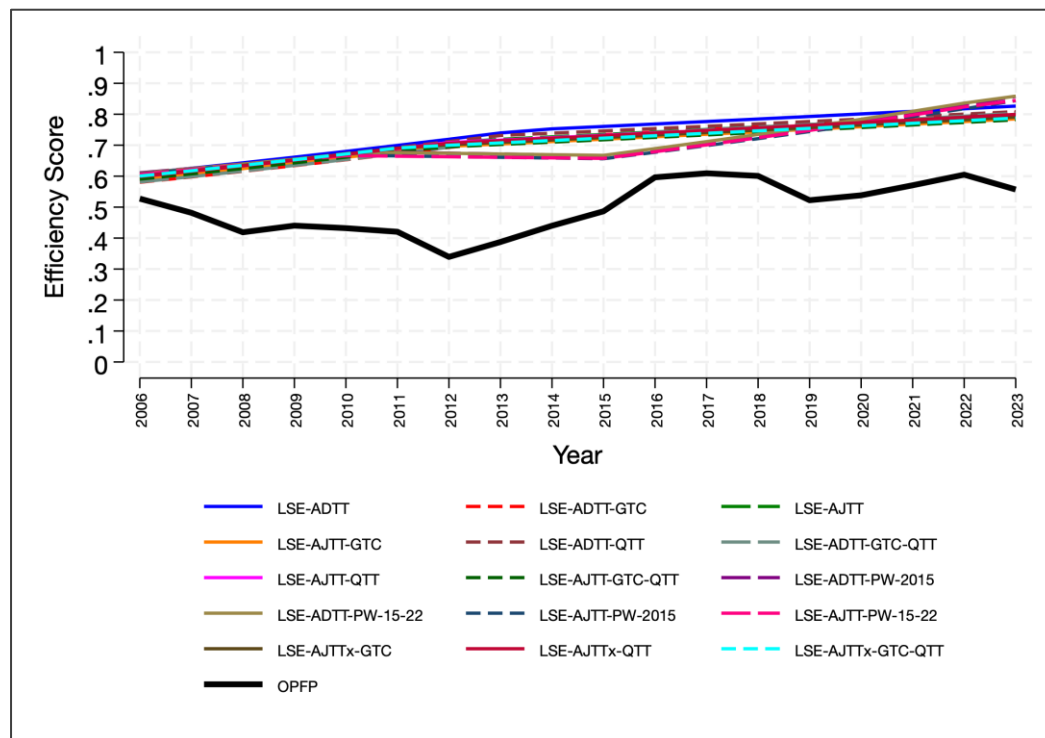


Figure 4.4.7.9: ESS LSE vs OPFP Trends – Parsimonious Translog Models



4.4.8 JEN

For JEN, Figures 4.4.8.1 to 4.4.8.9 show that:

- OPFP efficiency remained stable in 2006–2007, increased sharply in 2008, and then decreased continuously to its minimum in 2012. It remained relatively low and stable until 2017, generally increased through 2022, and decreased slightly in 2023. The efficiency score trends for the BC95 specifications closely reproduce these movements but generate consistently higher scores.
- Most JTT and AJTT Kumb90 specifications show smooth U-shaped curves, decreasing until approximately 2010–2013 and increasing thereafter, particularly after 2016. The increases accelerate over time, with most specifications converging to approximately 0.88–0.93 by 2023. The AGE CJTT specifications show a stepwise pattern, decreasing slightly until 2012–2014 and then increasing through successive adjustments separated by plateaus.
- The standard and GTC LSE specifications show smooth curves, increasing gradually at a slowing rate until around 2014–2016 and then remaining broadly stable or decreasing slightly. In contrast, the QTT and piecewise specifications decrease until approximately 2014–2015 and increase at an accelerating rate. QTT and piecewise specifications broadly reflect the OPFP, BC95 and Kumb90 decrease-and-increase pattern, although they do not reproduce the short-term fluctuations.

Figure 4.4.8.1: JEN BC95 vs OPFP Trends – Cobb-Douglas Models

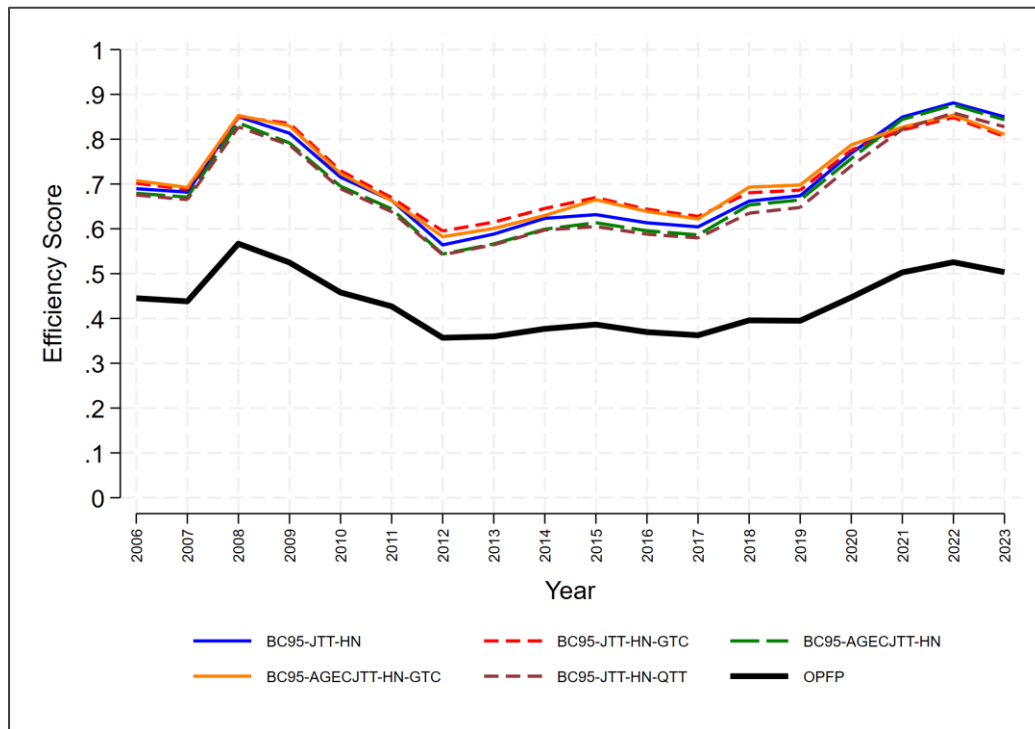


Figure 4.4.8.2: JEN BC95 vs OPFP Trend – Translog Models

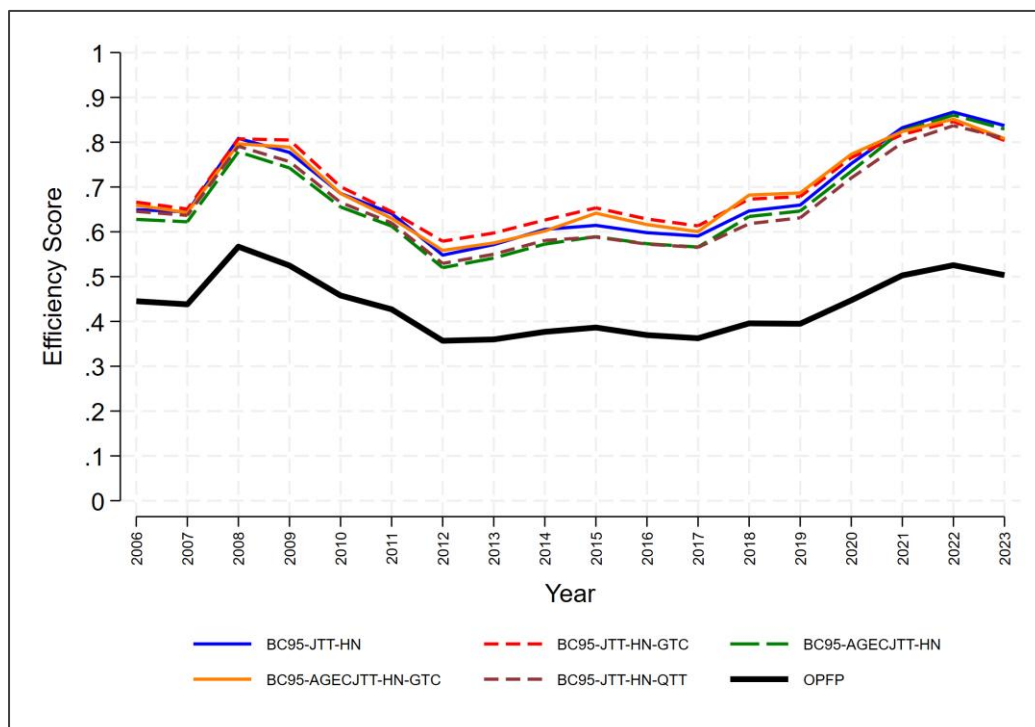


Figure 4.4.8.3: JEN BC95 vs OPFP Trend – Parsimonious Translog Models

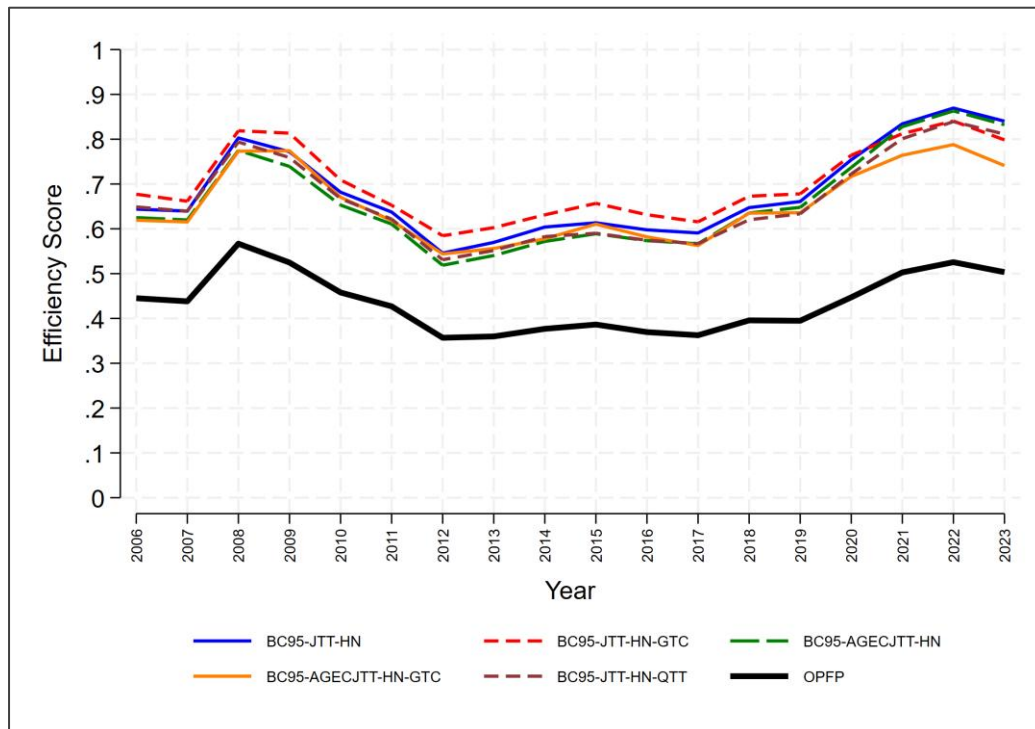


Figure 4.4.8.4: JEN Kumb90 vs OPFP Trends – Cobb-Douglas Models

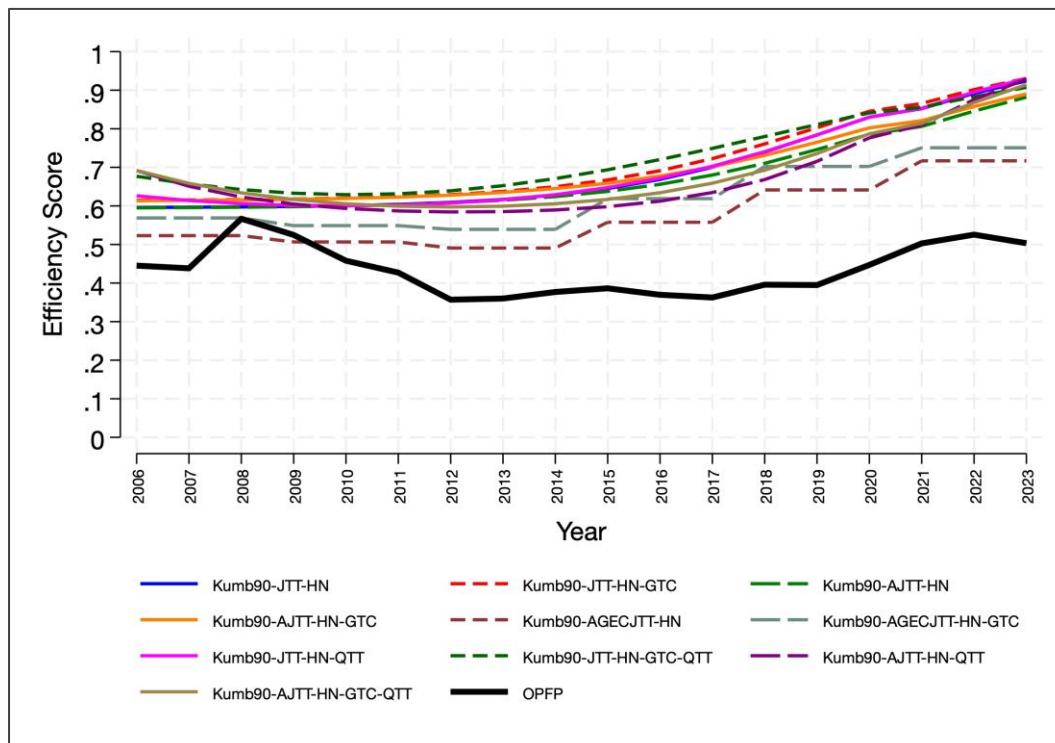


Figure 4.4.8.5: JEN Kumb90 vs OPFP Trends – Translog Models

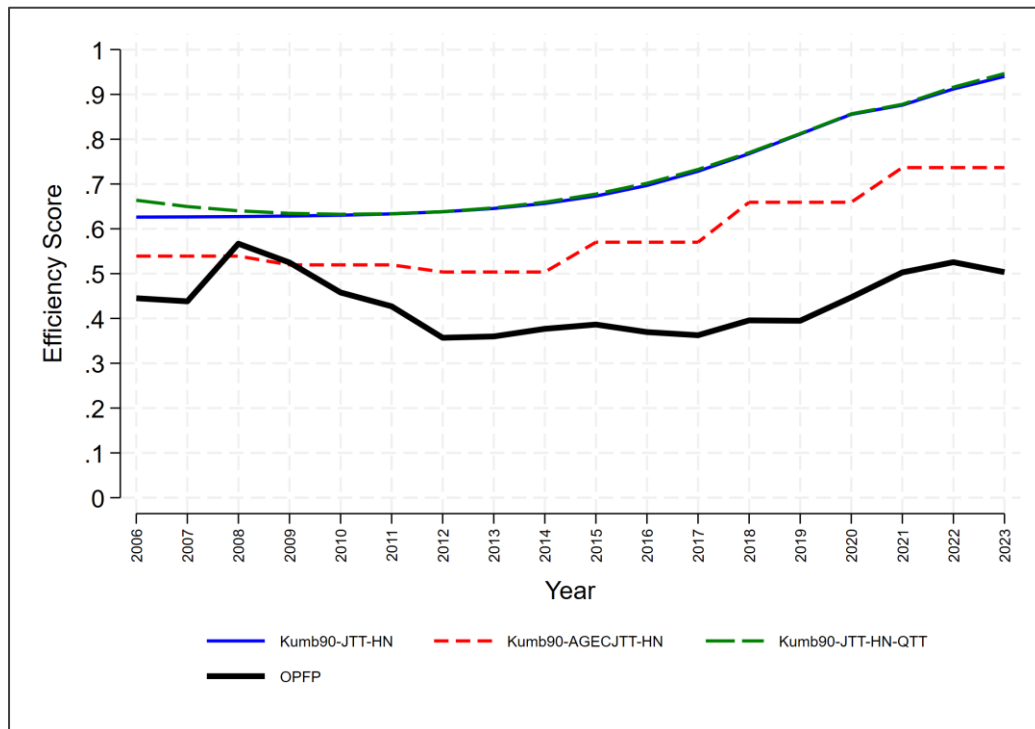


Figure 4.4.8.6: JEN Kumb90 vs OPFP Trends – Parsimonious Translog Models

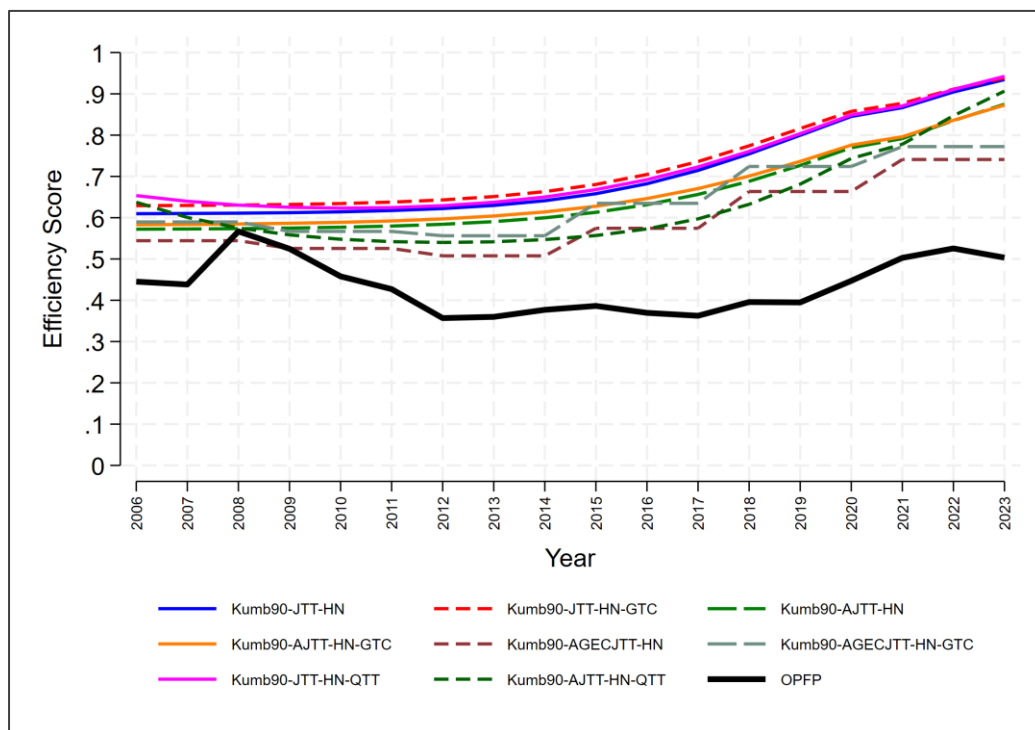


Figure 4.4.8.7: JEN LSE vs OPFP Trends – Cobb-Douglas Models

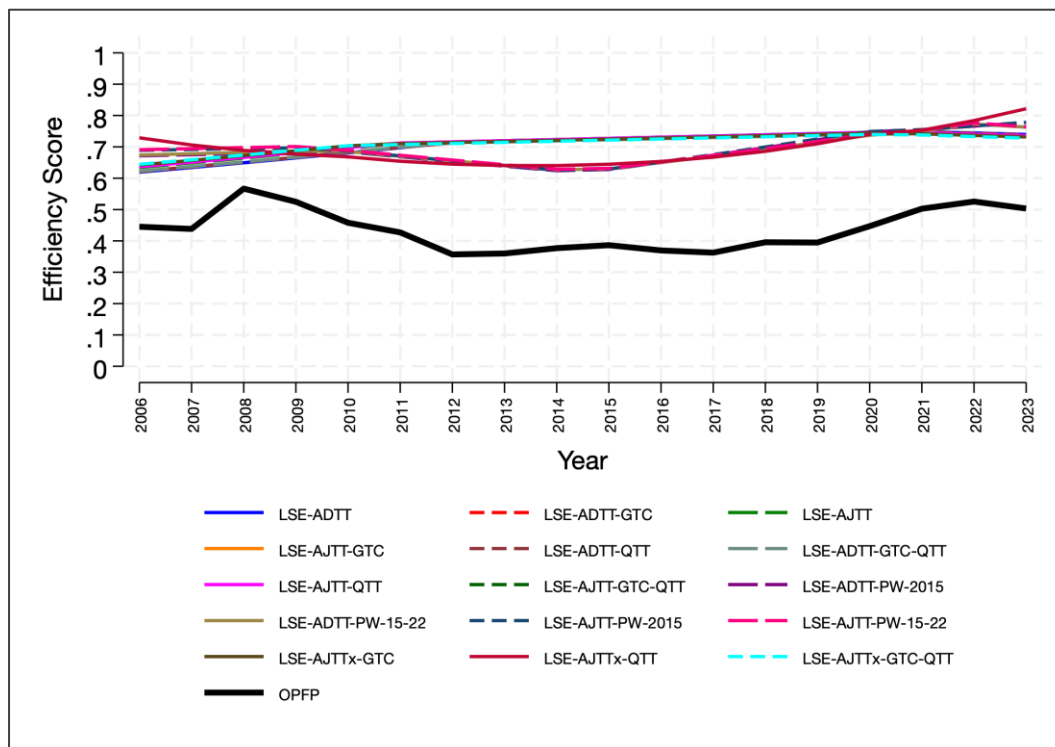


Figure 4.4.8.8: JEN LSE vs OPFP Trends – Translog Models

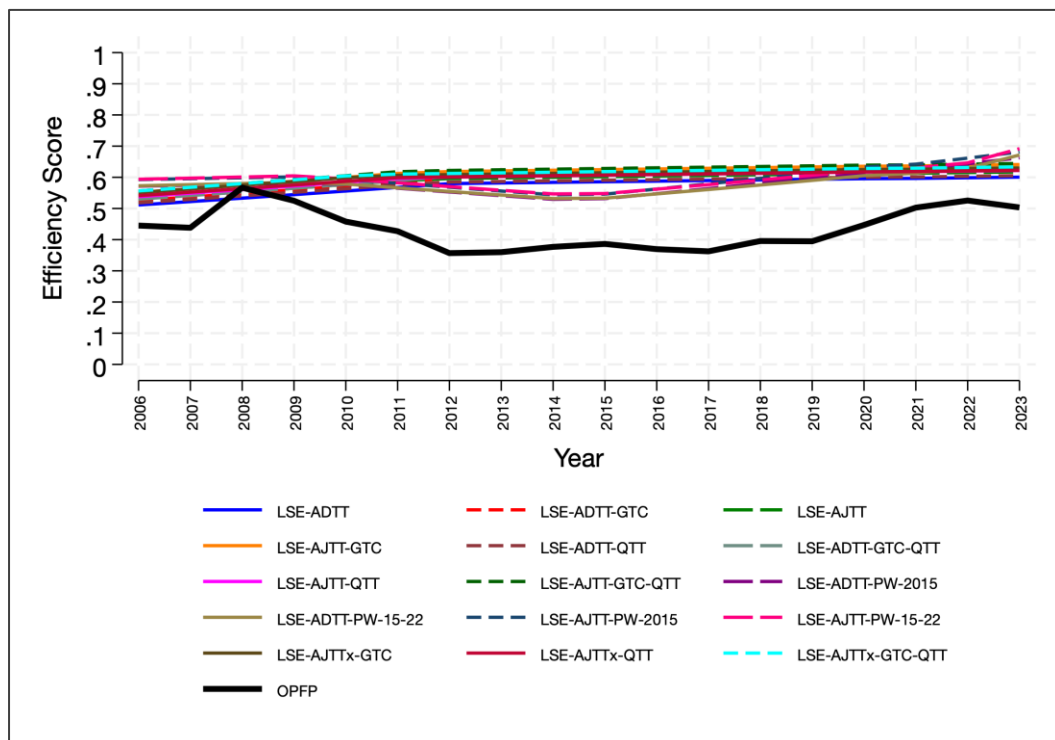
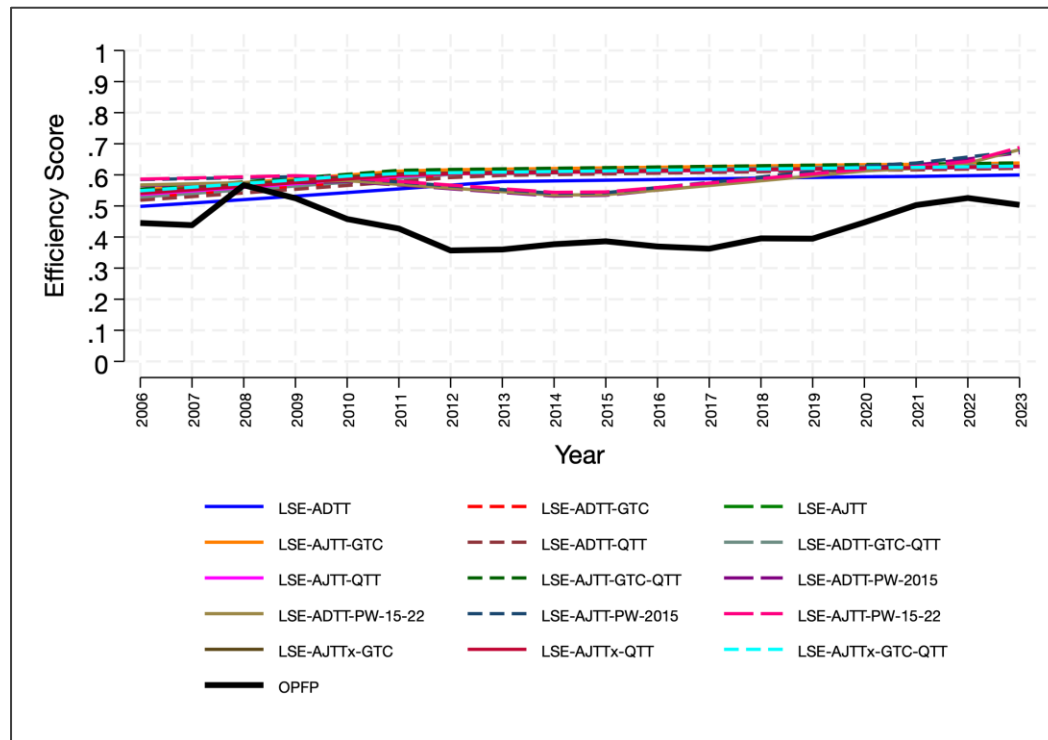


Figure 4.4.8.9: JEN LSE vs OPFP Trends – Parsimonious Translog Models



4.4.9 PCR

For PCR, Figures 4.4.9.1 to 4.4.9.9 show that:

- PCR's OPFP had little underlying trend. It increased from 2006 to 2008, decreased in 2009, increased through 2011, and then decreased sharply to its minimum in 2013. After remaining stable in 2014–2015, it increased sharply in 2016, decreased through 2018, increased until 2021, and decreased in 2022–2023. The efficiency score trends for the BC95 specifications closely reproduce these movements but generate consistently higher scores.
- The JTT Kumb90 specifications remain stable at approximately 0.88–0.93 until 2013–2014 and then increase towards 0.98–0.99. Most AJTT specifications start lower, remain stable until 2011–2012, and then increase at an accelerating rate, converging with the JTT results by 2023. The AJTT-GTC-QTT specification reaches one by 2017 and remains there. The AGE CJTT specifications decrease slightly until 2012–2014 and then increase through stepwise adjustments and plateaus.
- All LSE specifications show similar smooth paths, with efficiency increasing at a slowing rate towards one by approximately 2011–2012 and remaining at the efficiency frontier through 2020. Efficiency then decrease slightly thereafter, particularly in 2023.

Figure 4.4.9.1: PCR BC95 vs OPFP Trends – Cobb-Douglas Models

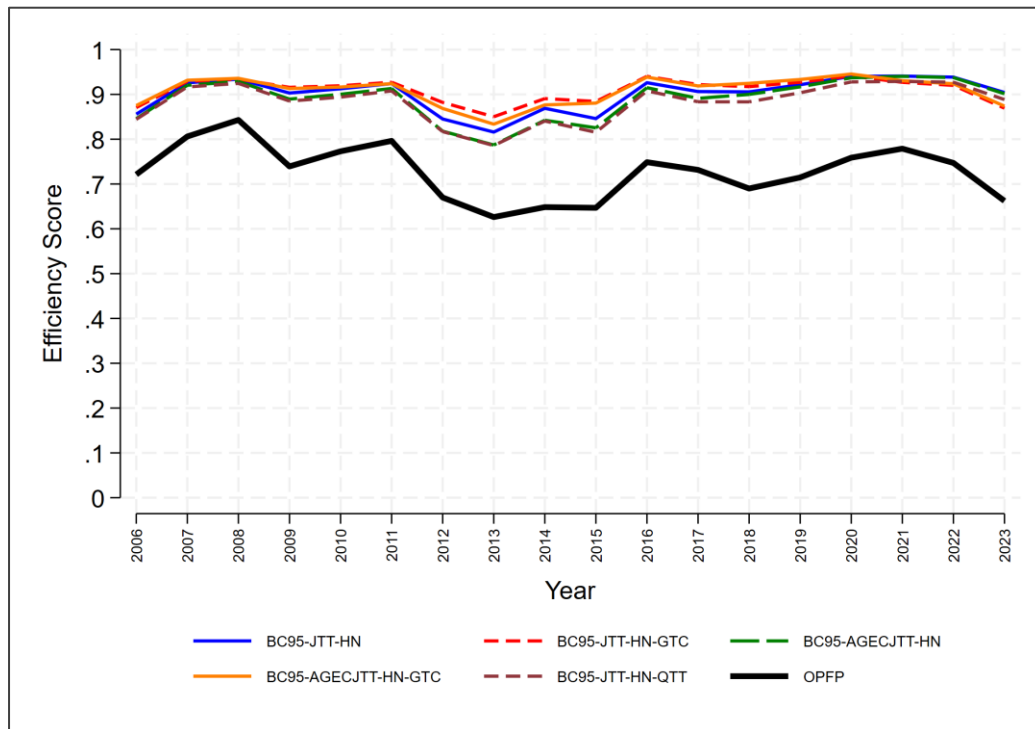


Figure 4.4.9.2: PCR BC95 vs OPFP Trend – Translog Models

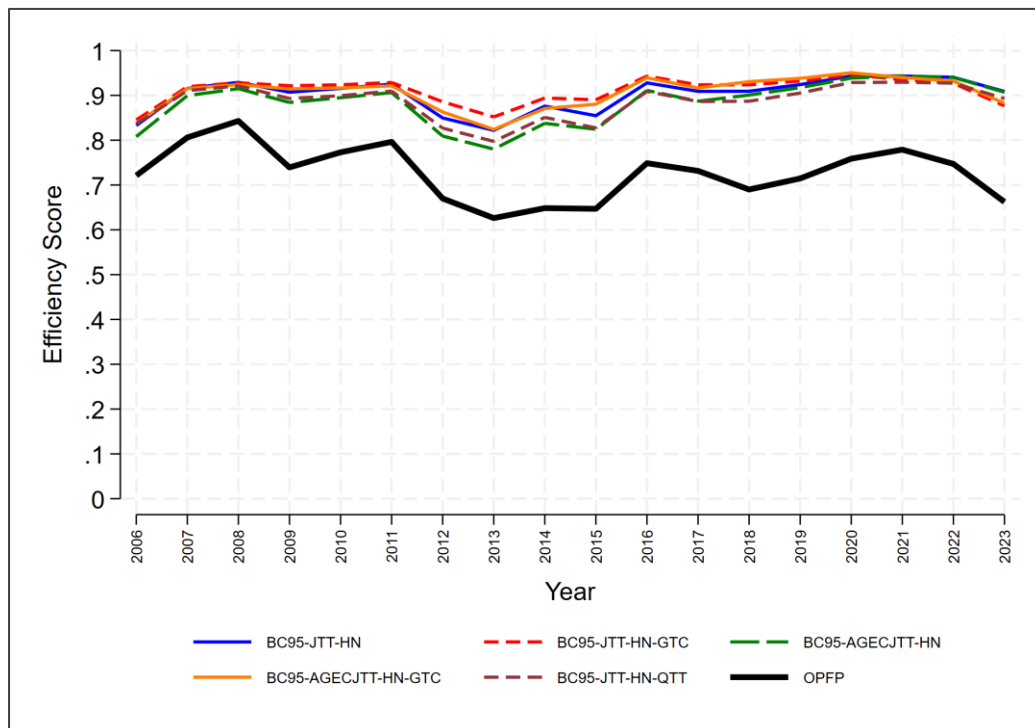


Figure 4.4.9.3: PCR BC95 vs OPFP Trend – Parsimonious Translog Models

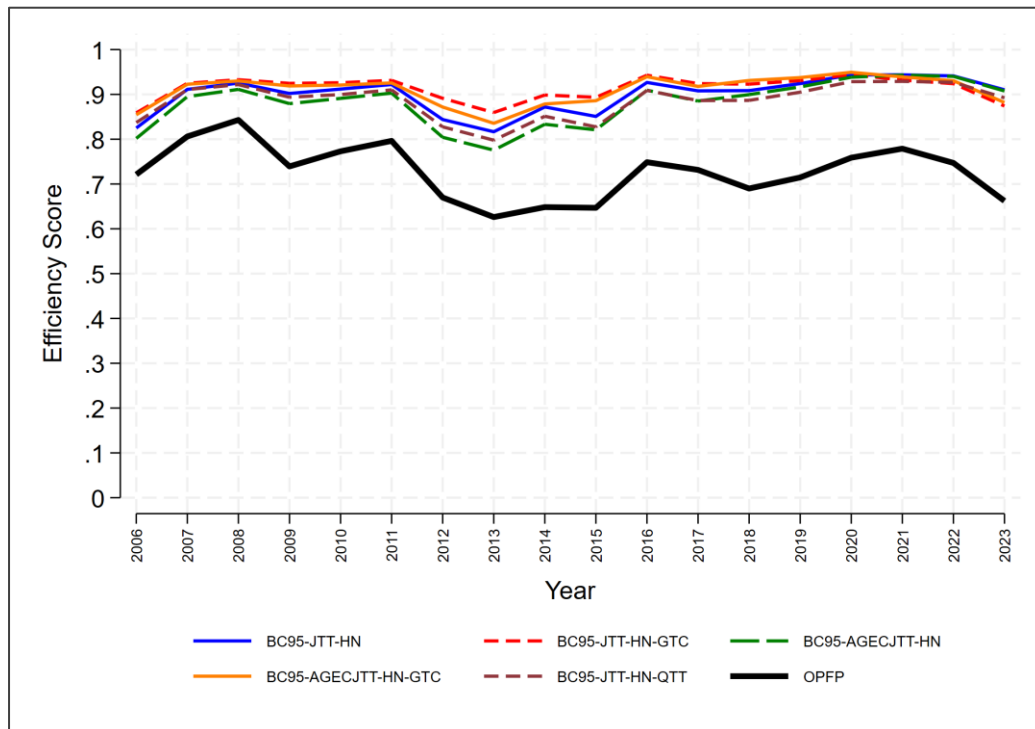


Figure 4.4.9.4: PCR Kumb90 vs OPFP Trends – Cobb-Douglas Models

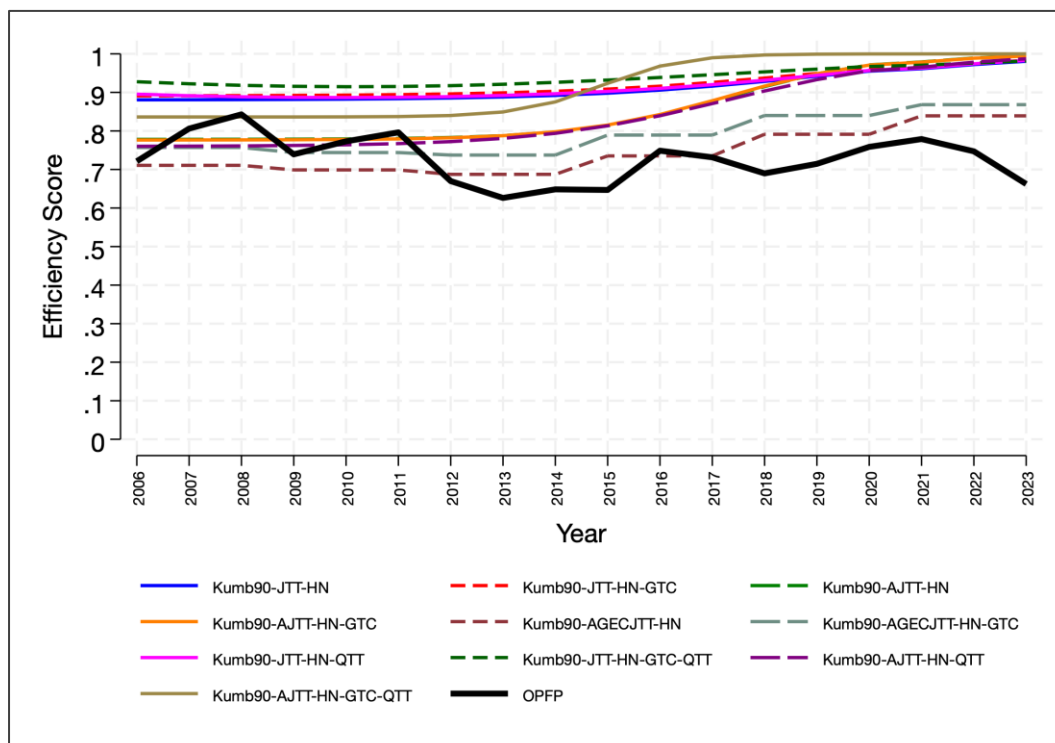


Figure 4.4.9.5: PCR Kumb90 vs OPFP Trends – Translog Models

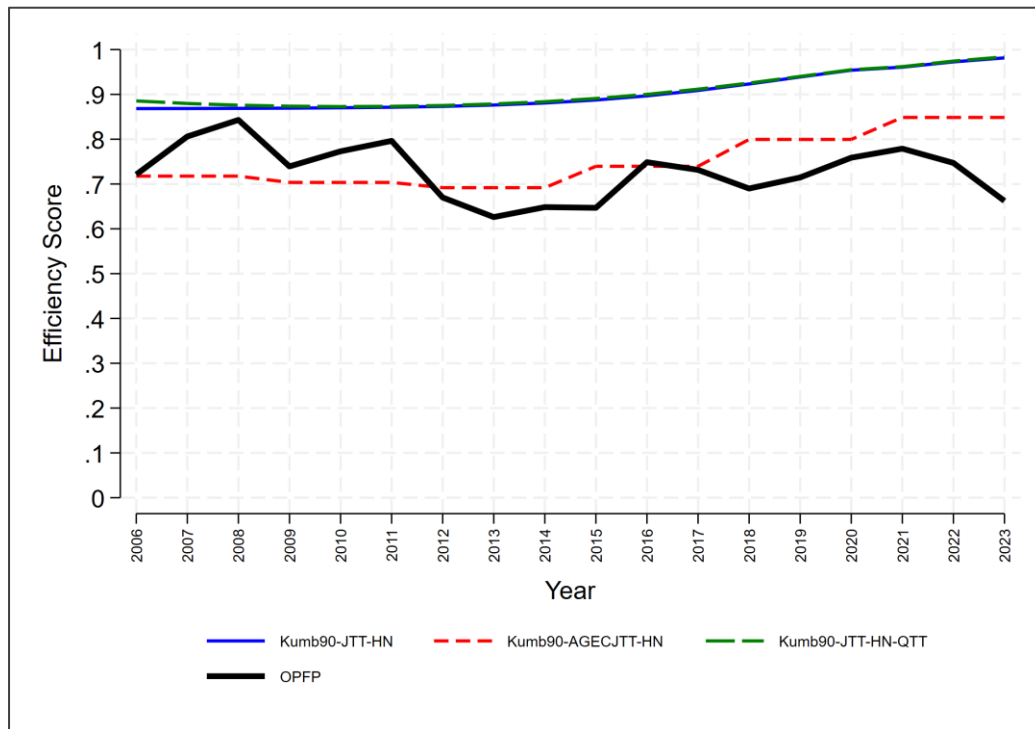


Figure 4.4.9.6: PCR Kumb90 vs OPFP Trends – Parsimonious Translog Models

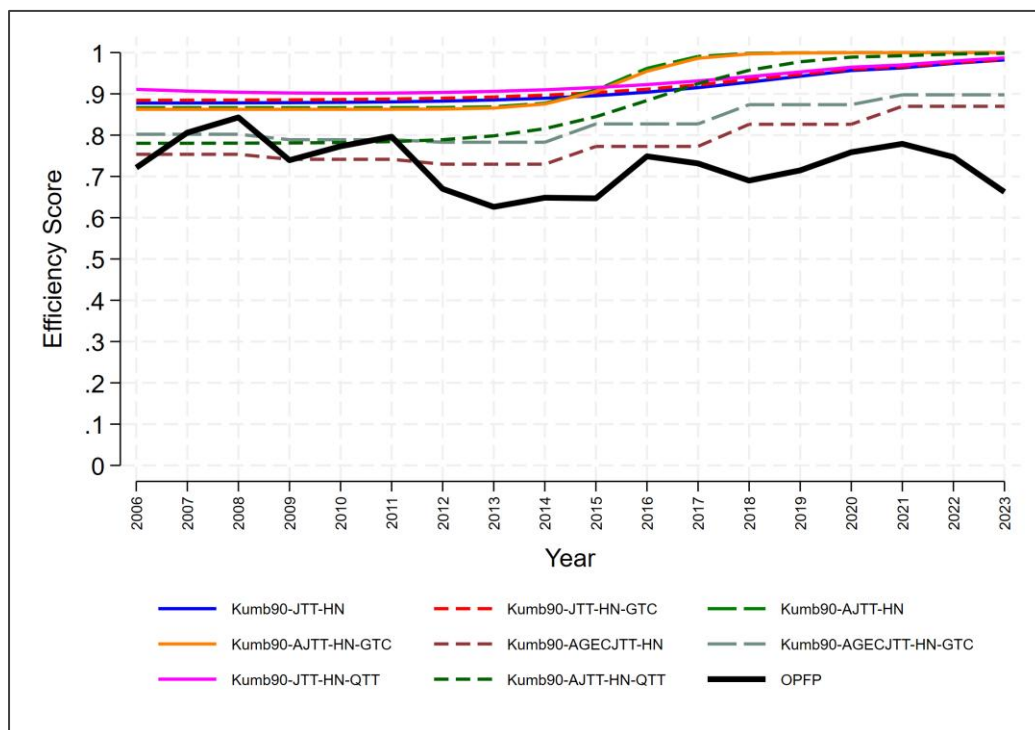


Figure 4.4.9.7: PCR LSE vs OPFP Trends – Cobb-Douglas Models

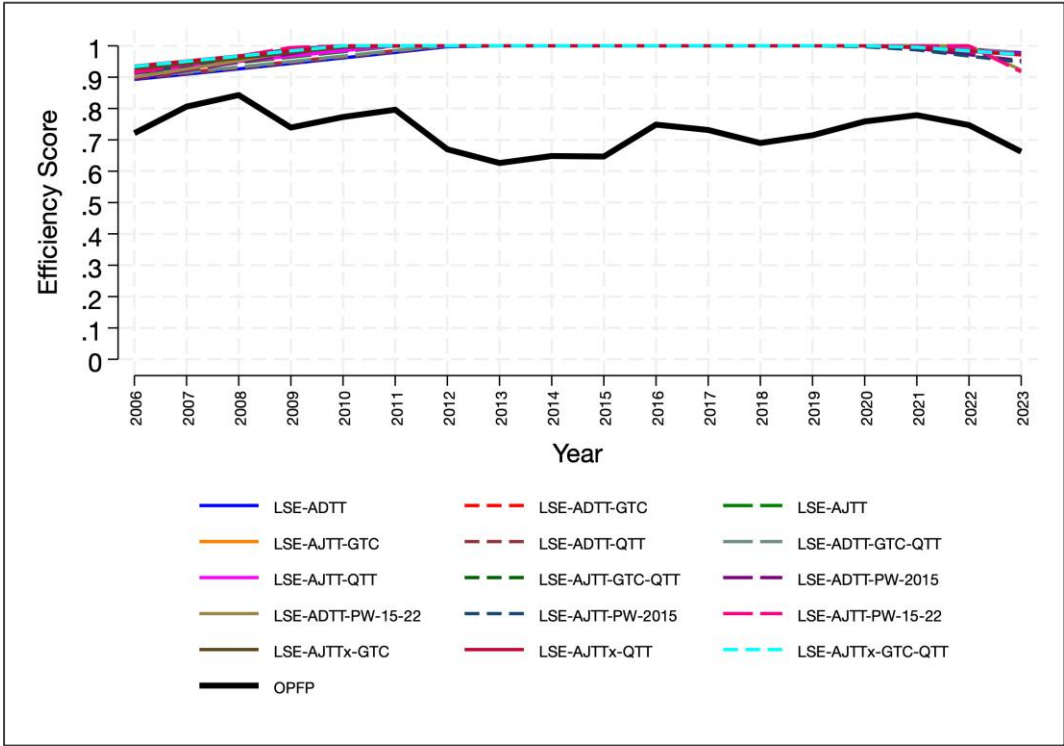


Figure 4.4.9.8: PCR LSE vs OPFP Trends – Translog Models

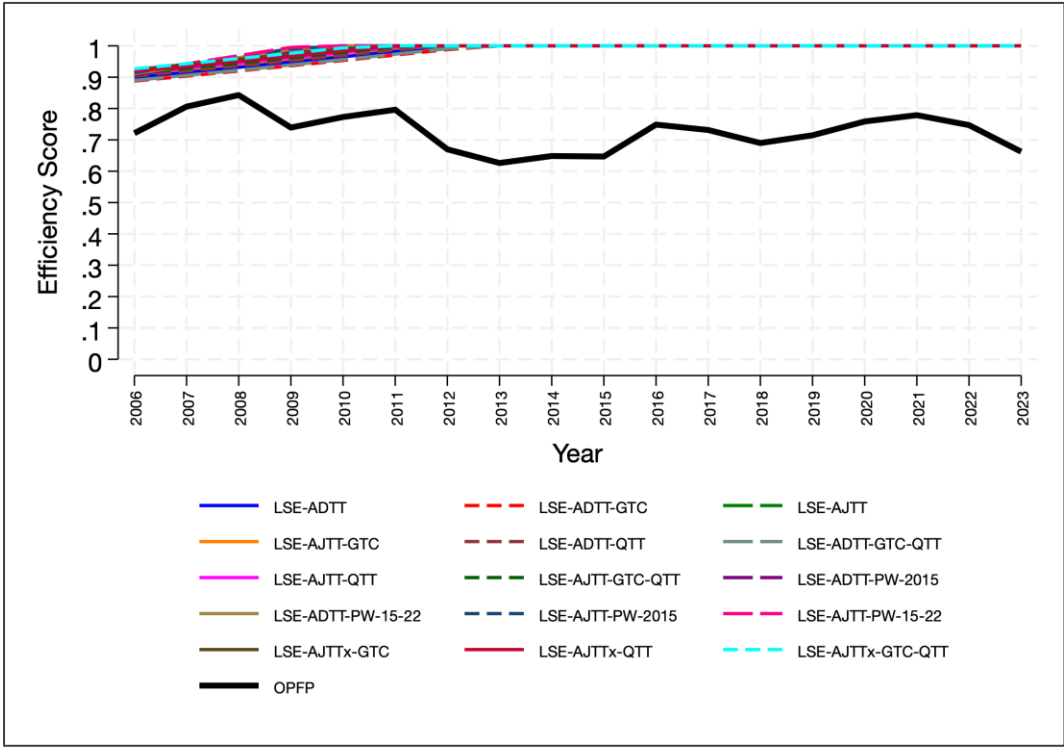
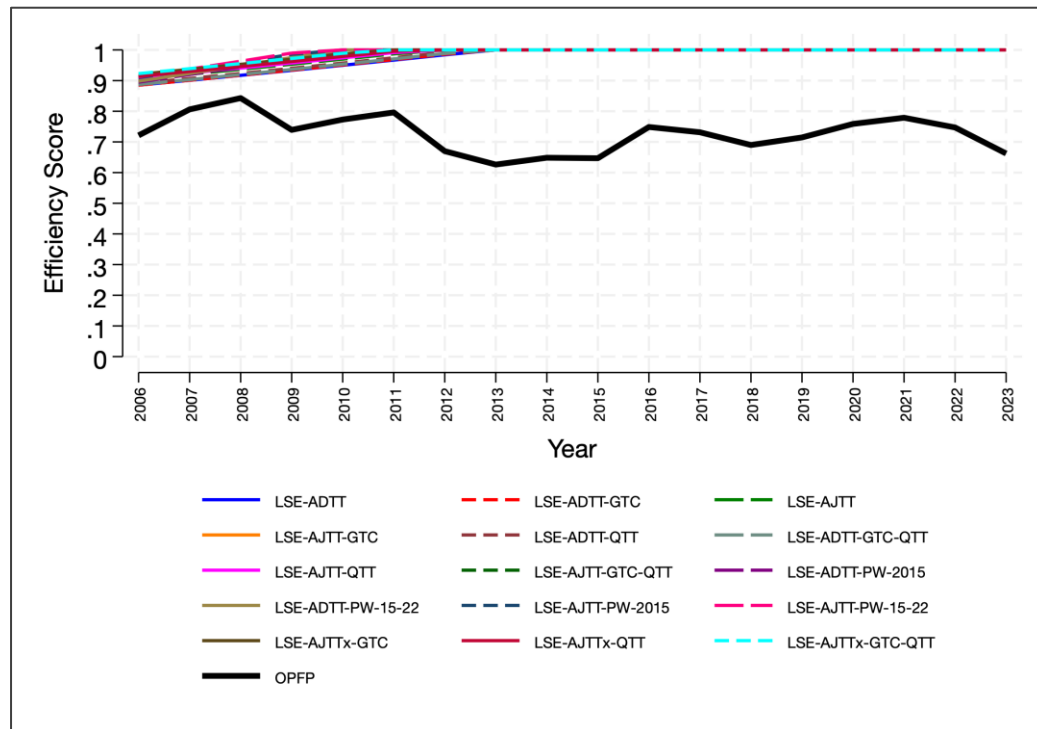


Figure 4.4.9.9: PCR LSE vs OPFP Trends – Parsimonious Translog Models



4.4.10 SAP

For SAP, Figures 4.4.10.1 to 4.4.10.9 show that:

- OPFP has an underlying downward trend over the sample period. Efficiency increased to its maximum in 2007, decreased strongly through 2011, and remained relatively stable until 2015. It increased sharply in 2016, decreased in 2017, fluctuated through 2020, and decreased thereafter, particularly in 2023. The efficiency score trends for the BC95 specifications broadly reproduce these movements and generally generate higher scores from 2010 onwards. Notably, they produce lower efficiency scores before 2010, unlike the other cases discussed so far. The downward trend is not as pronounced as with OPFP.
- The Kumb90 results for SAP's efficiency levels vary considerably across specifications. Trends also show differences. The JTT specifications remain high and relatively stable until around 2014 and then increase gradually to approximately 0.98–0.99. The AJTT and AJTT-GTC specifications decrease gradually, reaching approximately 0.80–0.83 by 2023, while AJTT-QTT increases modestly to around 0.86. In contrast, AJTT-GTC-QTT increases strongly after 2014 and reaches one by 2017. The AGE CJTT specifications decrease very slightly until 2012–2014 and then increase through stepwise adjustments and plateaus.
- The standard and GTC LSE specifications increase towards one during the early years and then decrease smoothly to approximately 0.79–0.81 by 2023. The QTT and

piecewise specifications decrease more quickly until around 2015, remain broadly stable through 2022, and decrease sharply in 2023. All LSE trajectories are substantially smoother than the OPFP and BC95 results, but they reflect the underlying trends more closely than the Kumb90 models.

Figure 4.4.10.1: SAP BC95 vs OPFP Trends – Cobb-Douglas Models

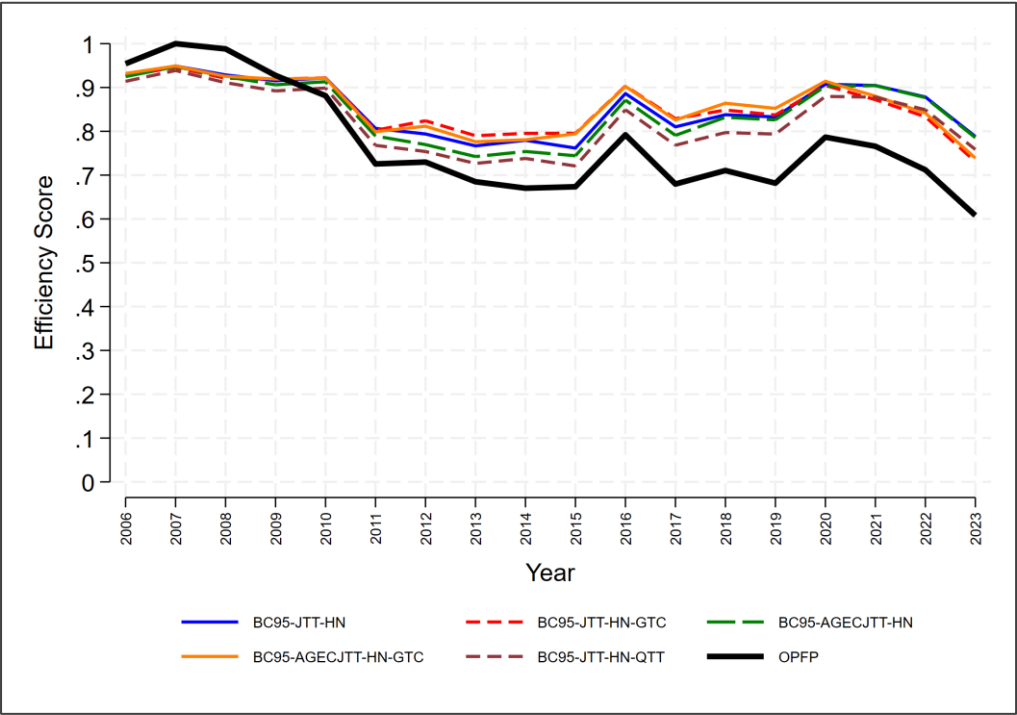


Figure 4.4.10.2: SAP BC95 vs OPFP Trend – Translog Models

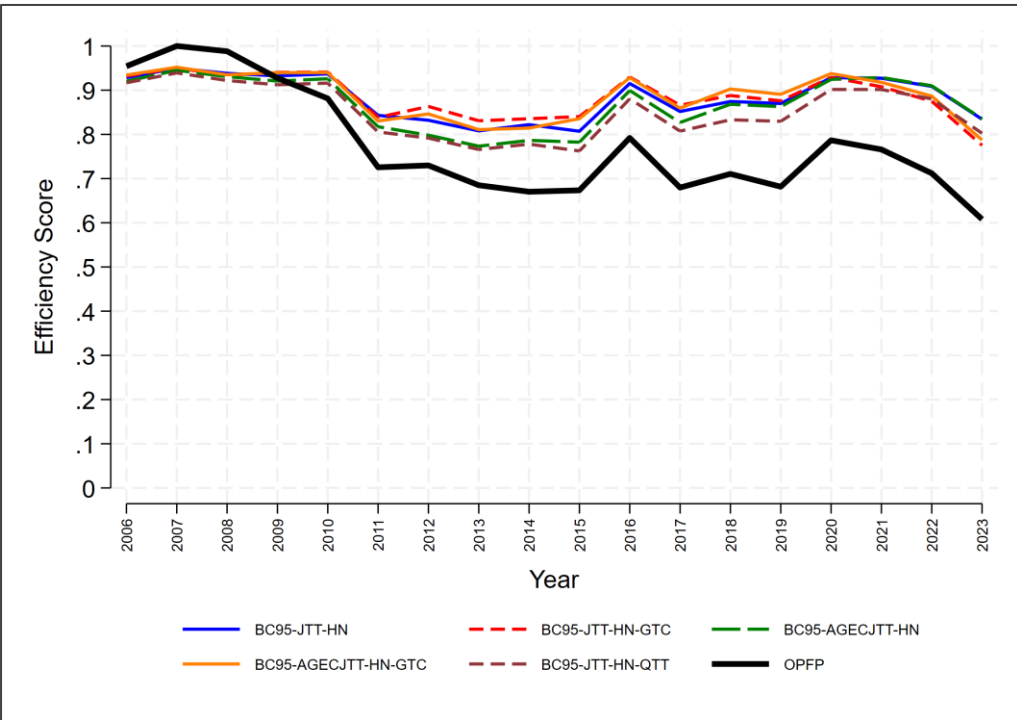


Figure 4.4.10.3: SAP BC95 vs OPFP Trend – Parsimonious Translog Models

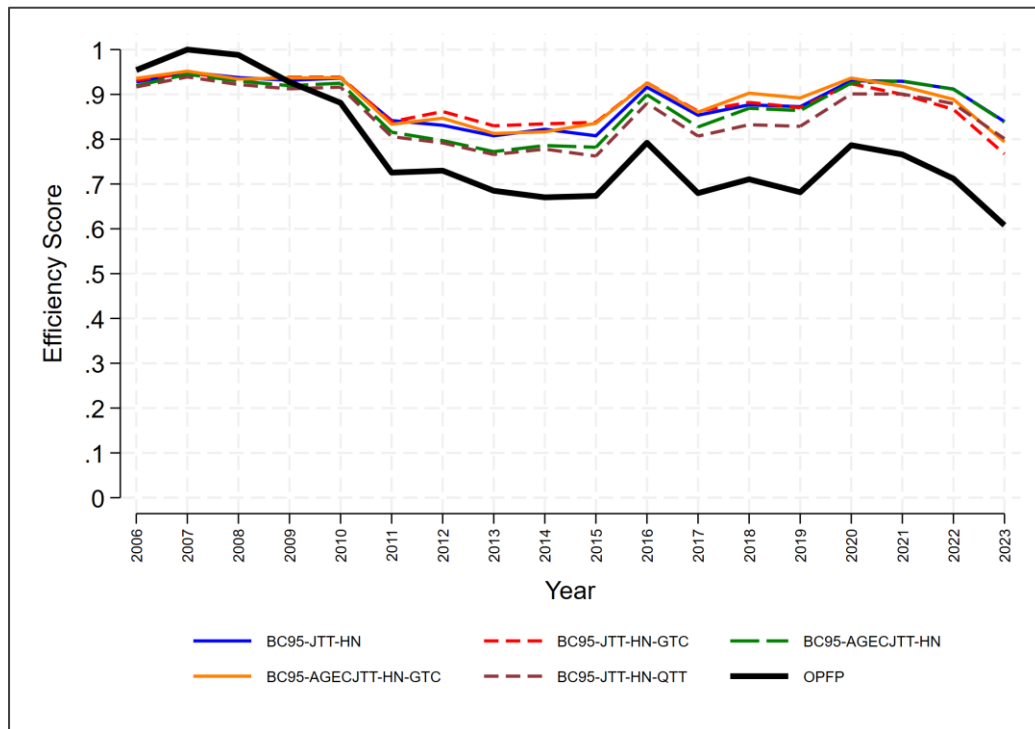


Figure 4.4.10.4: SAP Kumb90 vs OPFP Trends – Cobb-Douglas Models

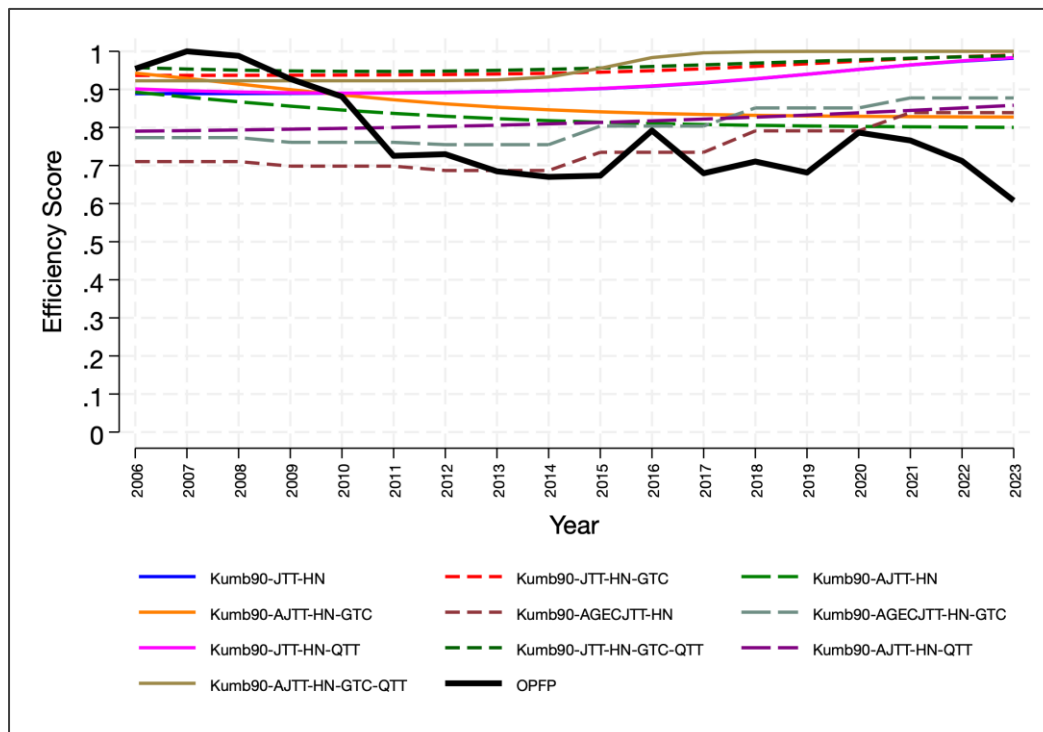


Figure 4.4.10.5: SAP Kumb90 vs OPFP Trends – Translog Models

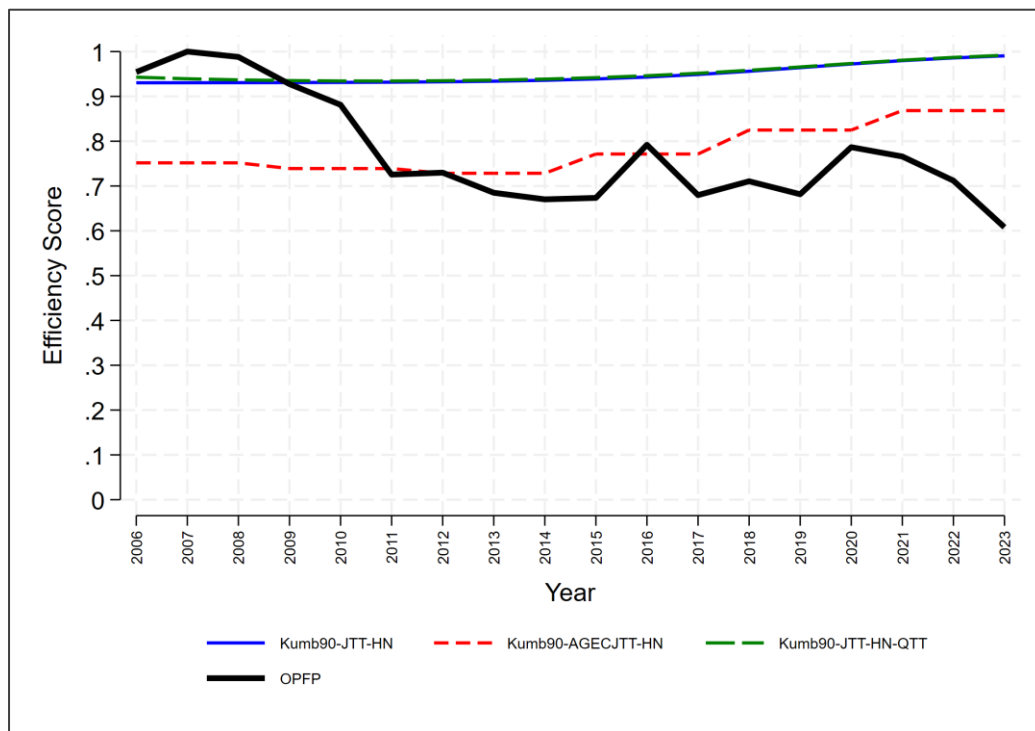


Figure 4.4.10.6: SAP Kumb90 vs OPFP Trends – Parsimonious Translog Models

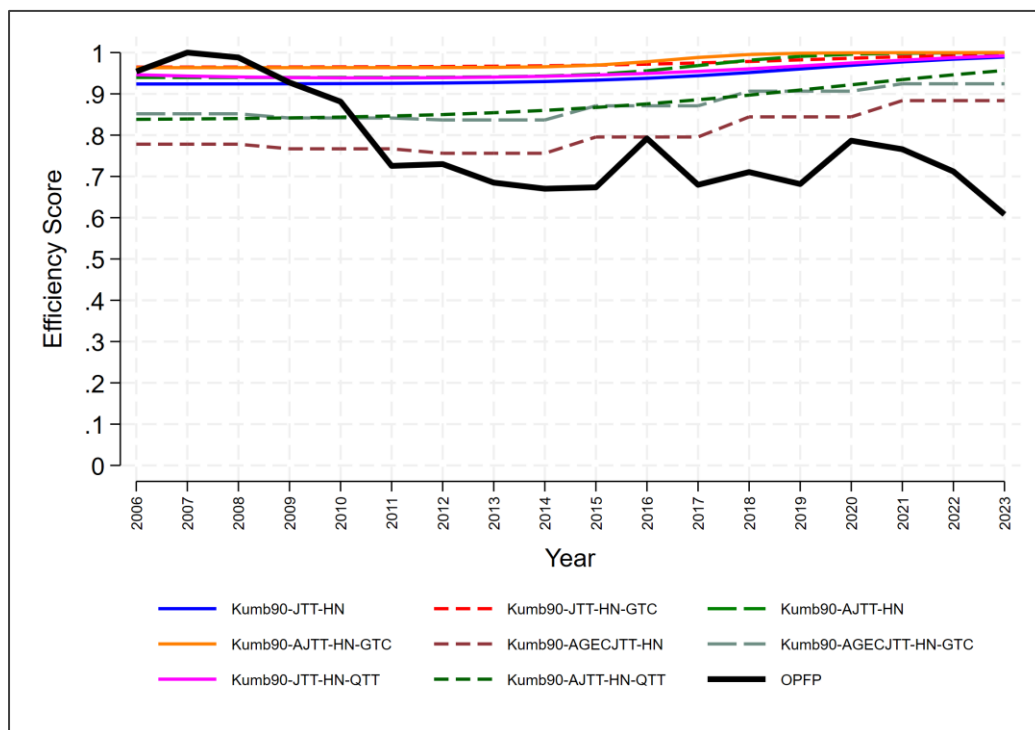


Figure 4.4.10.7: SAP LSE vs OPFP Trends – Cobb-Douglas Models

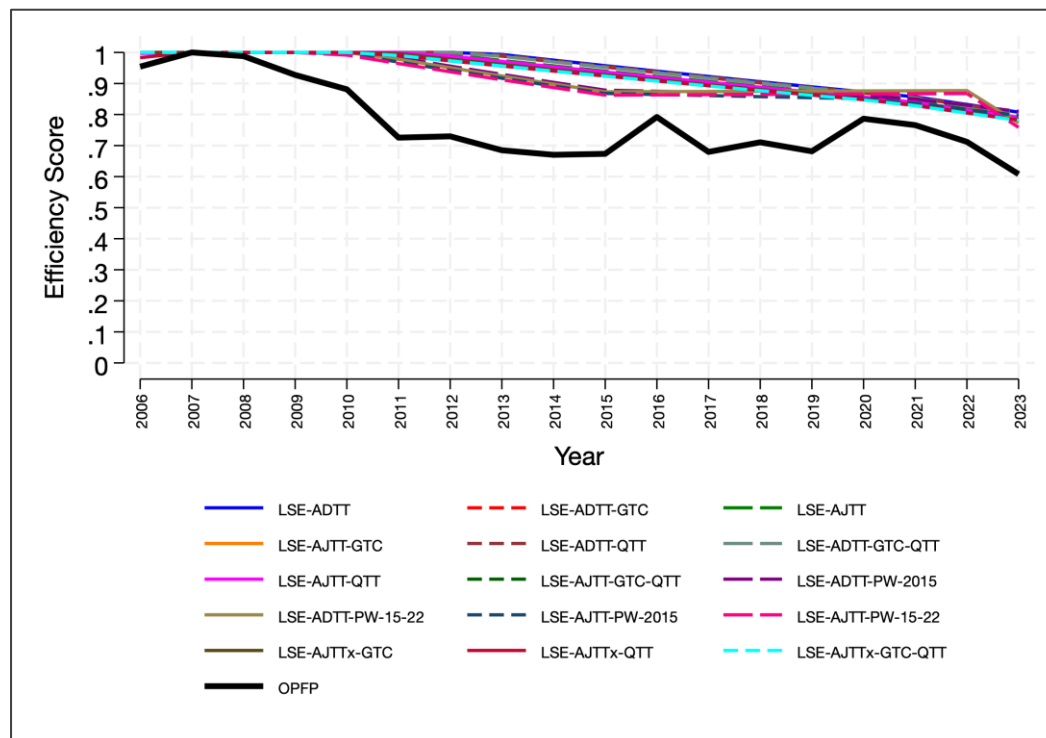


Figure 4.4.10.8: SAP LSE vs OPFP Trends – Translog Models

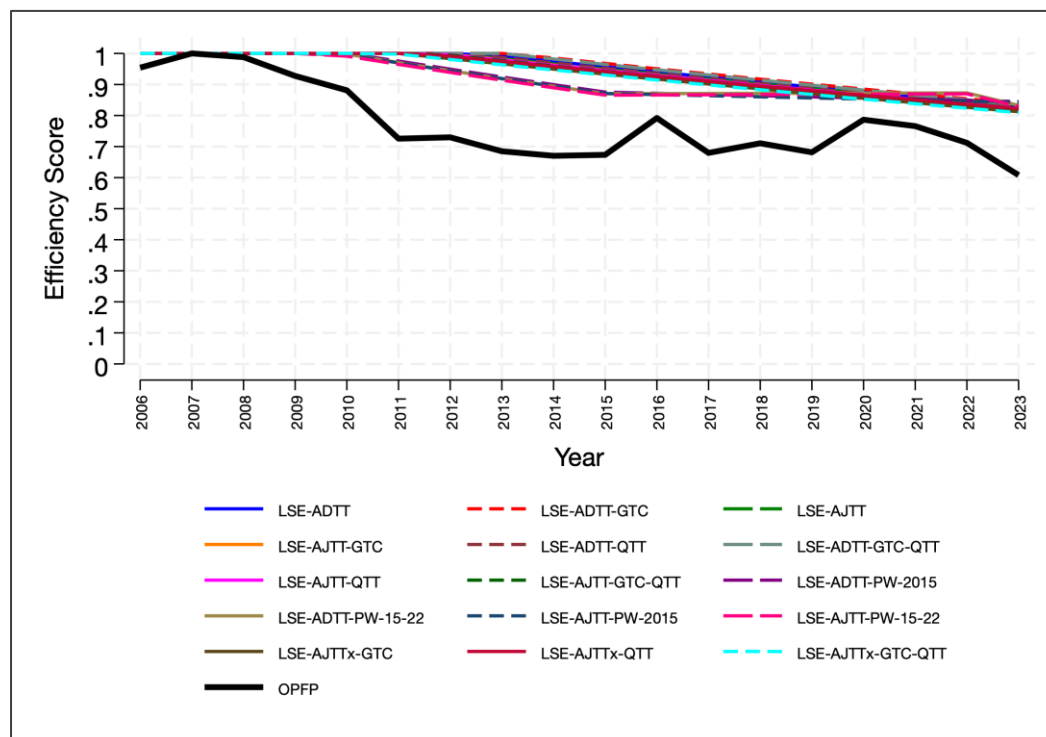
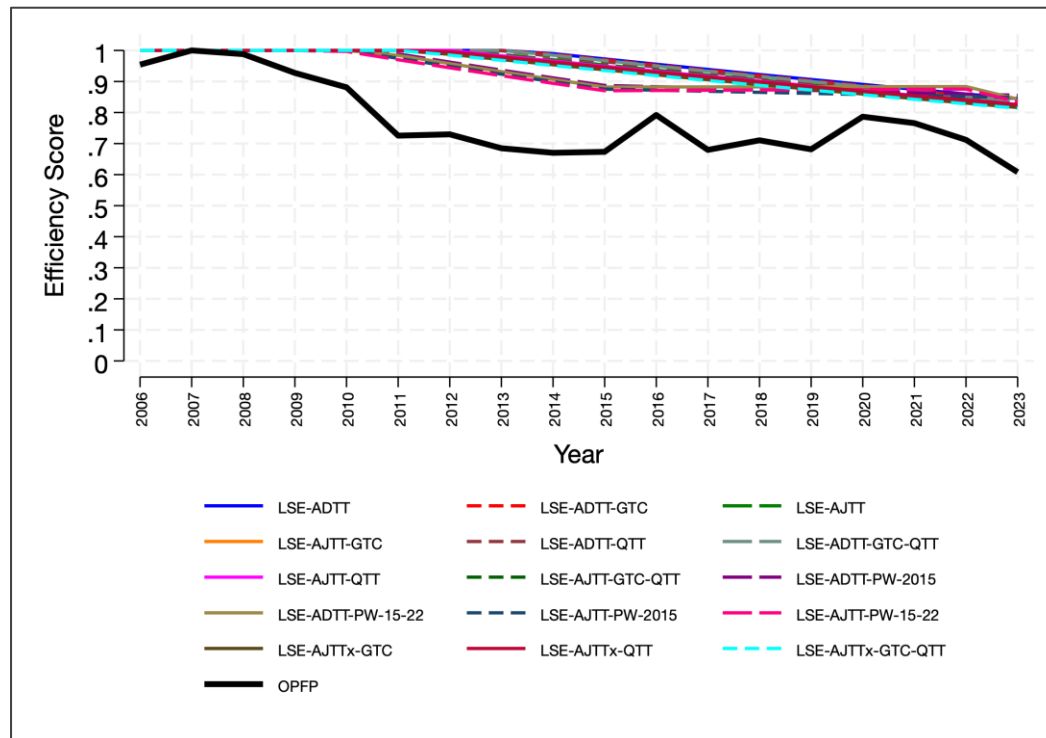


Figure 4.4.10.9: SAP LSE vs OPFP Trends – Parsimonious Translog Models



4.4.11 AND

For AND, Figures 4.4.11.1 to 4.4.11.9 show that:

- OPFP efficiency generally decreased from 2006 to 2016, despite temporary increases in 2008 and 2010, and reached its minimum in 2016. It increased sharply in 2017 and then remained relatively stable through 2023. The efficiency score trends for the BC95 specifications closely reproduce these movements but generate consistently higher scores. An improvement in 2022 and 2023 is more pronounced.
- The JTT Kumb90 specifications show shallow U-shaped curves, decreasing until 2010–2012 and then increasing strongly to approximately 0.94–0.95 by 2023. The AJTT specifications without QTT remain stable until 2013–2014 and then increase, while the AJTT-QTT specifications show deeper U-shaped curves. The AGE CJTT specifications decrease slightly until 2012–2014 and then increase through stepwise adjustments and plateaus. The efficiency trends in these models differ markedly from the OPFP series.
- The standard and GTC LSE specifications increase slightly until 2010–2011 and then decrease smoothly. In contrast, the QTT and piecewise specifications decrease until 2014–2015 and then increase at an accelerating rate. These specifications reflect the broad pattern in the OPFP and BC95 models but not their short-term fluctuations.

Figure 4.4.11.1: AND BC95 vs OPFP Trends – Cobb-Douglas Models

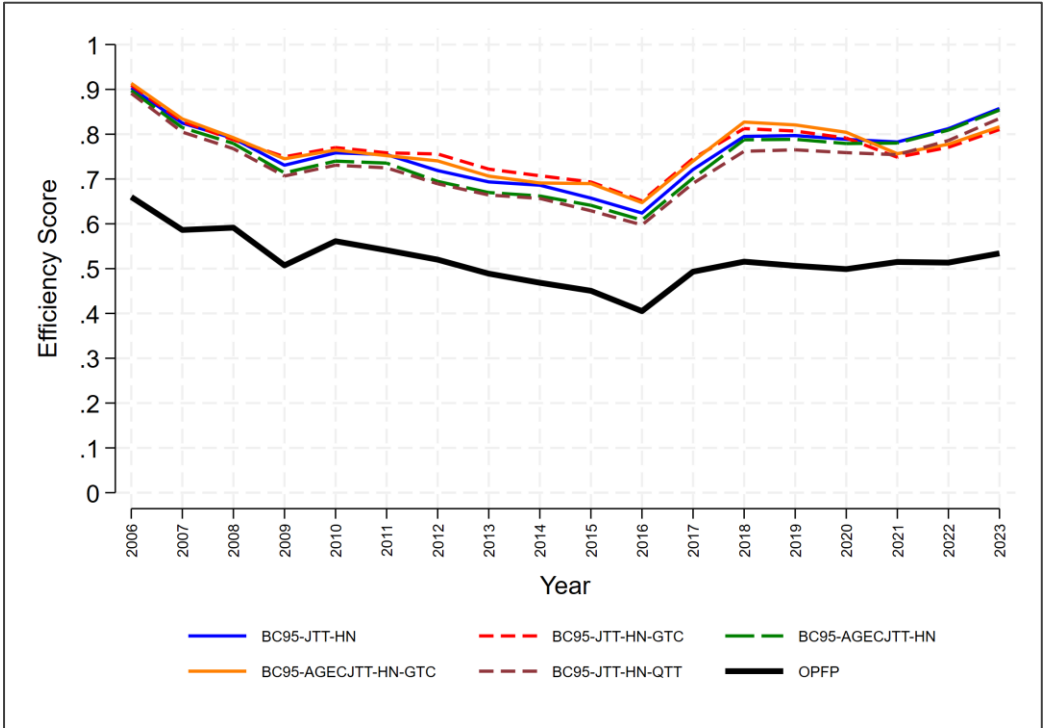


Figure 4.4.11.2: AND BC95 vs OPFP Trend – Translog Models

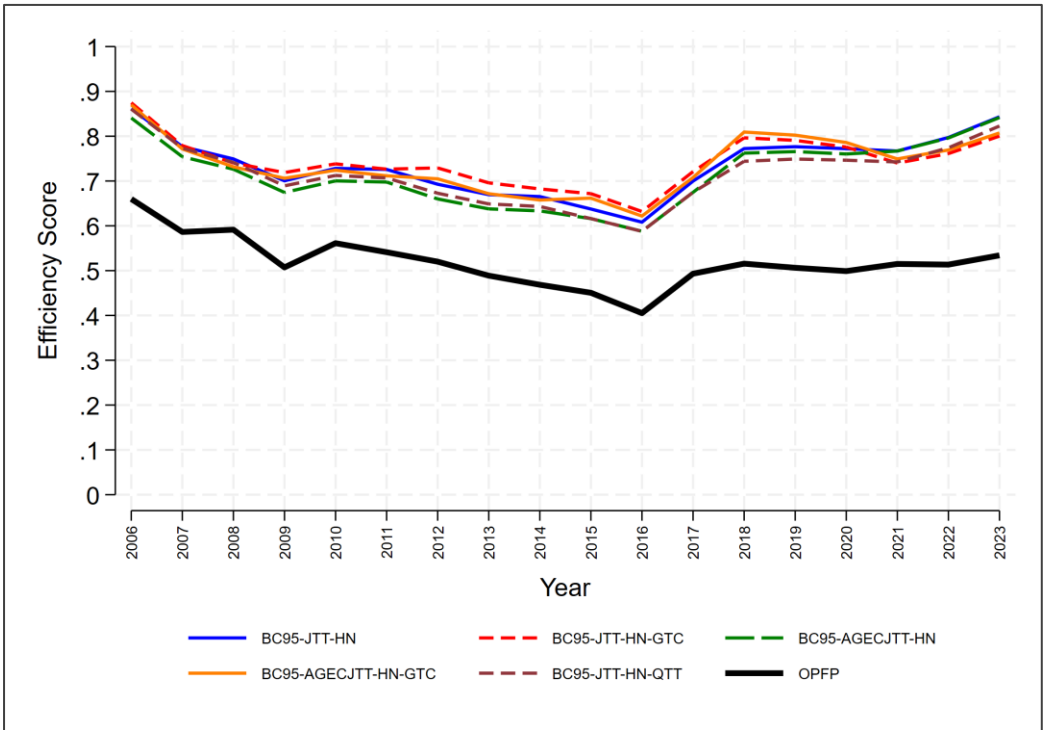


Figure 4.4.11.3: AND BC95 vs OPFP Trend – Parsimonious Translog Models

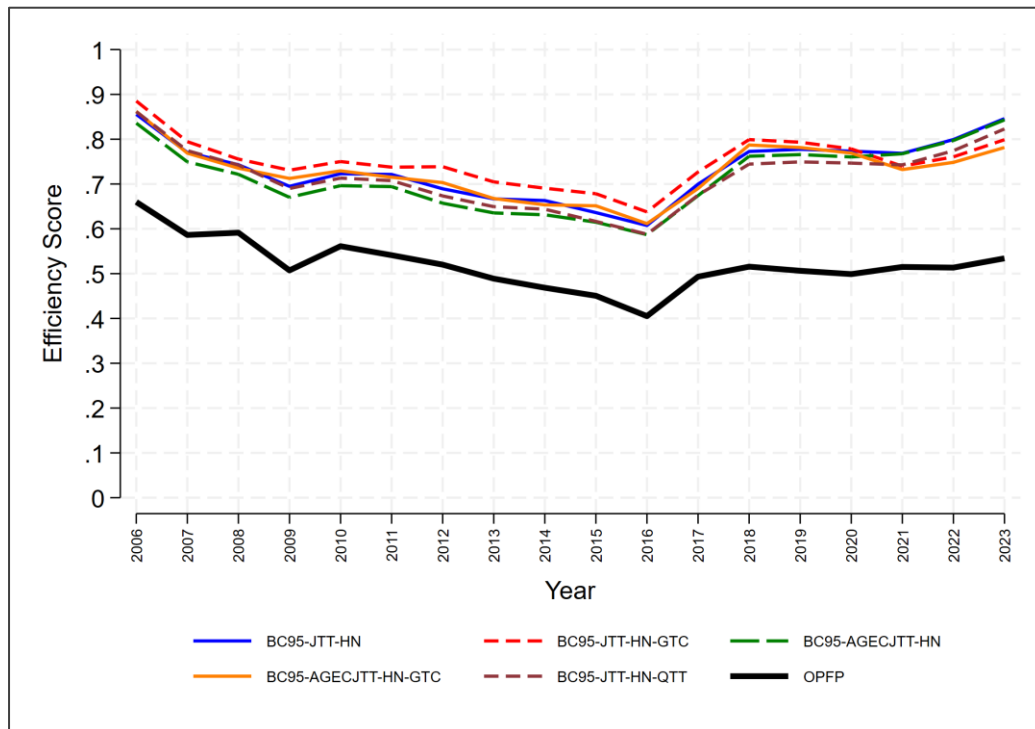


Figure 4.4.11.4: AND Kumb90 vs OPFP Trends – Cobb-Douglas Models

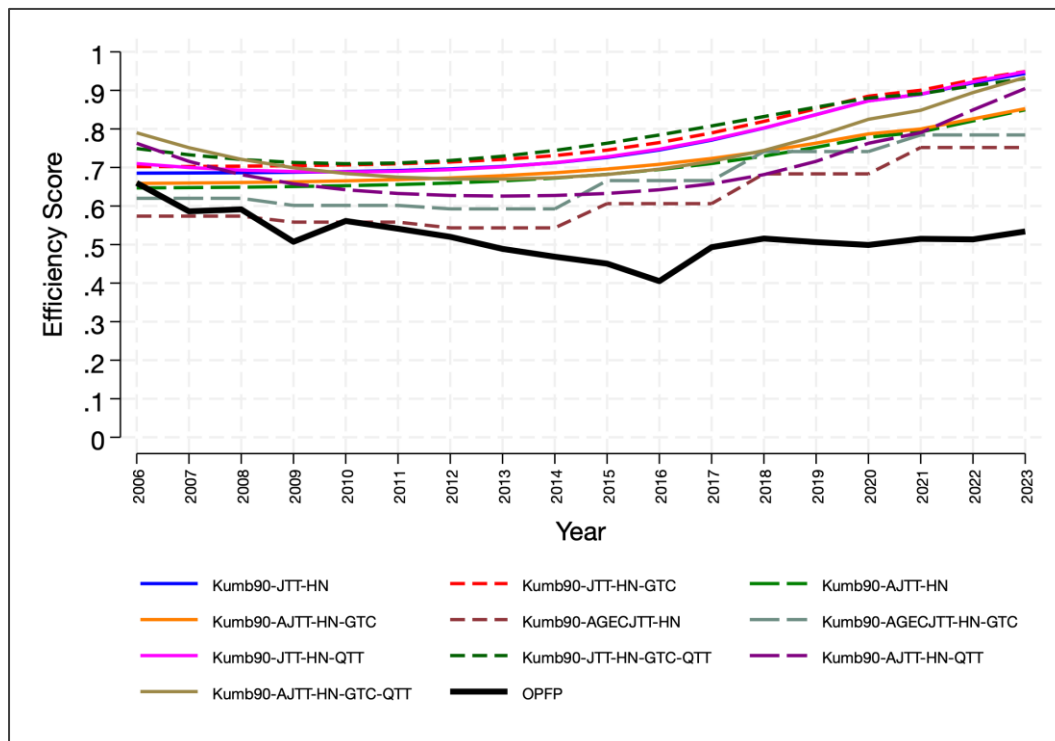


Figure 4.4.11.5: AND Kumb90 vs OPFP Trends – Translog Models

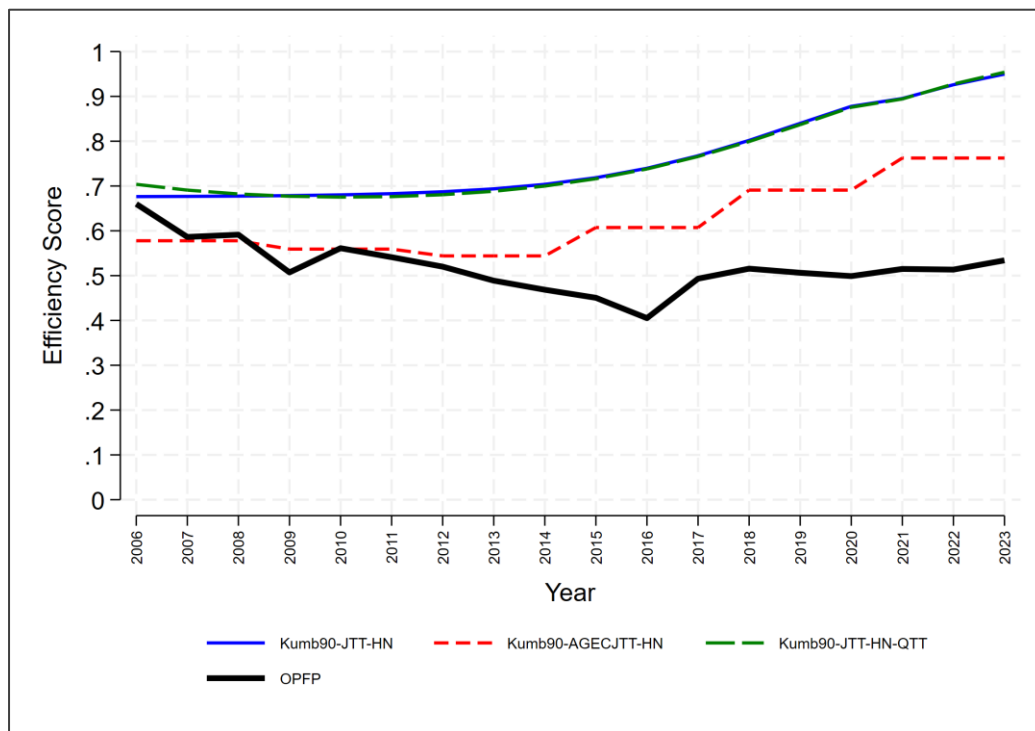


Figure 4.4.11.6: AND Kumb90 vs OPFP Trends – Parsimonious Translog Models

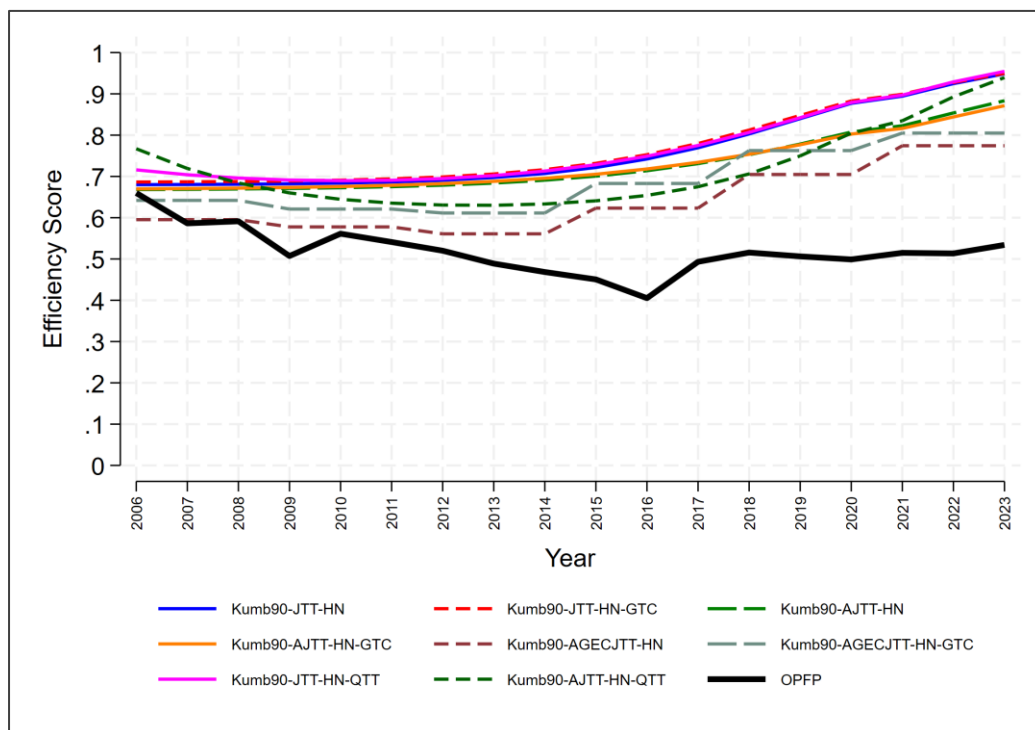


Figure 4.4.11.7: AND LSE vs OPFP Trends – Cobb-Douglas Models

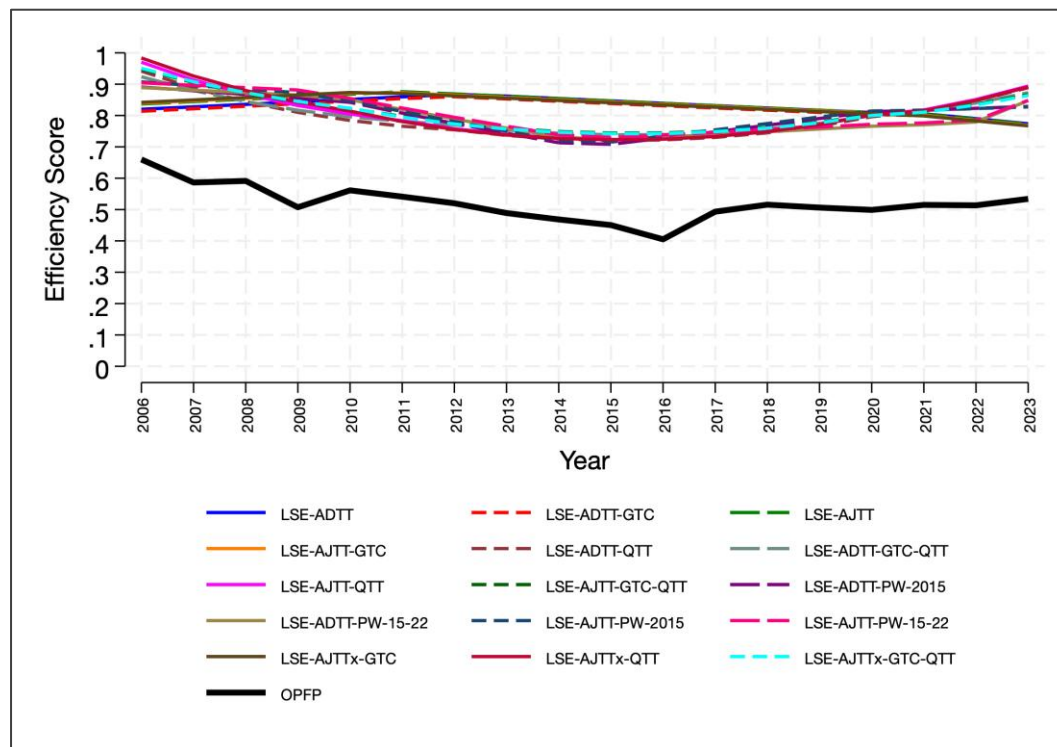


Figure 4.4.11.8: AND LSE vs OPFP Trends – Translog Models

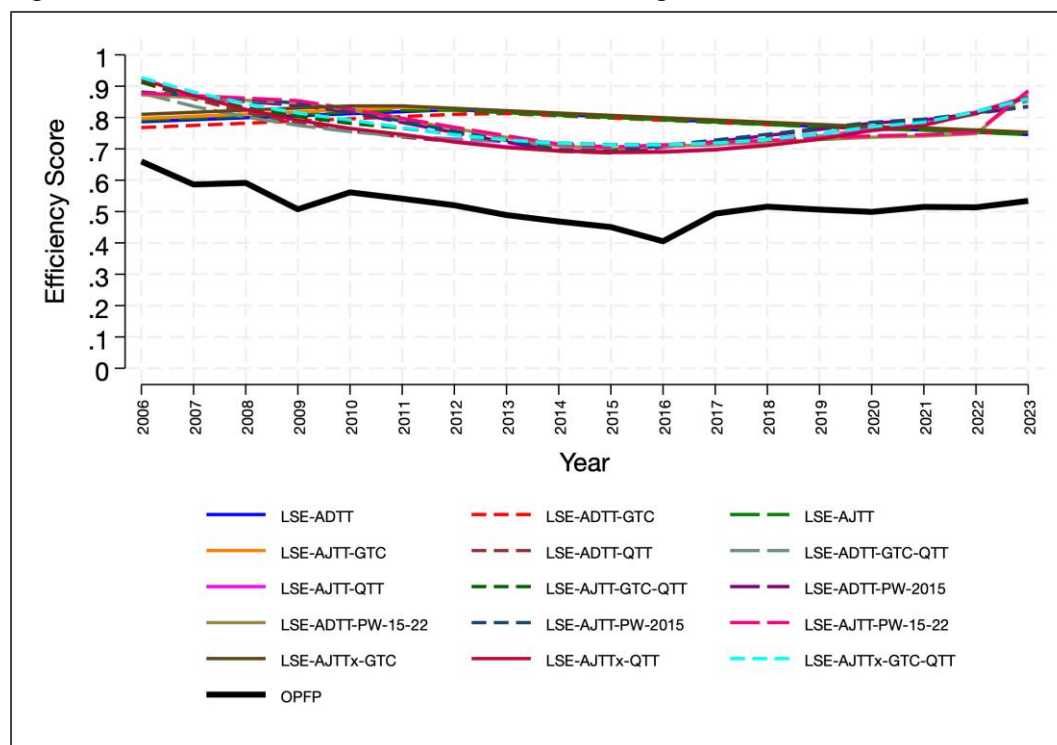
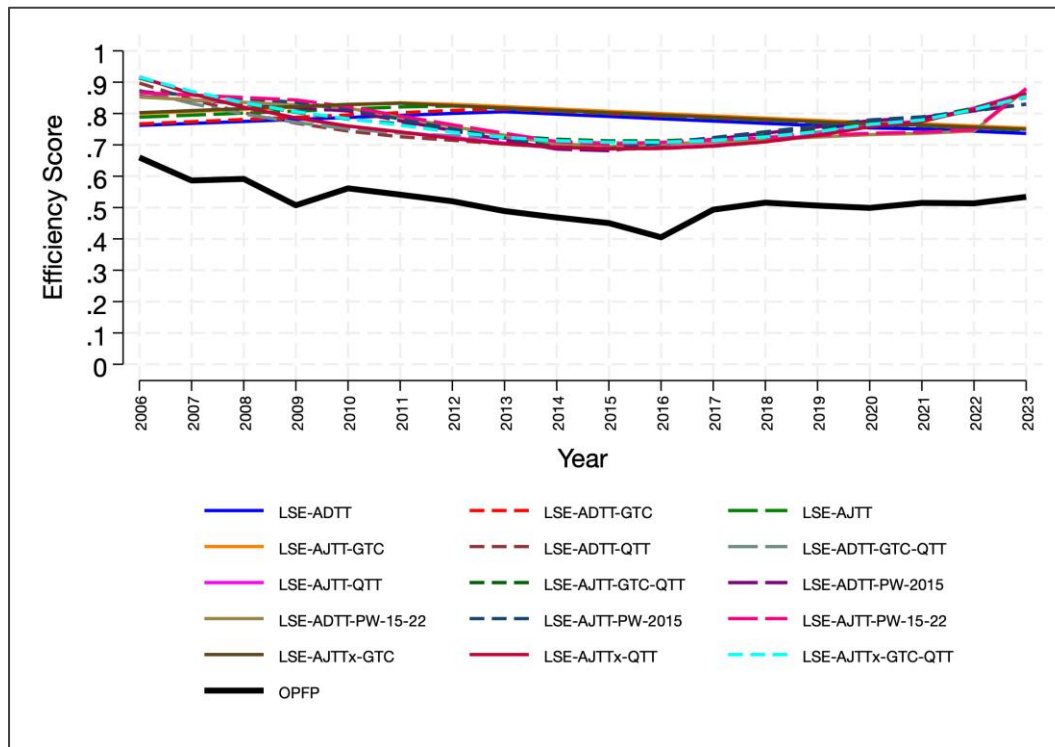


Figure 4.4.11.9: AND LSE vs OPFP Trends – Parsimonious Translog Models



4.4.12 TND

For TND, Figures 4.4.12.1 to 4.4.12.9 show that:

- OPFP efficiency fluctuated substantially, decreasing through 2010, increasing in 2011, decreasing in 2012, and increasing sharply in 2013 and 2015. It then decreased strongly through 2017, varied moderately until 2022, and increased in 2023. The efficiency score trends for the BC95 specifications closely reproduce these movements but generate consistently higher scores.
- The JTT Kumb90 specifications show shallow U-shaped curves, decreasing until approximately 2010–2012 and then increasing at an accelerating rate to around 0.94–0.96 by 2023. The AJTT specifications remain relatively stable until around 2014 and then increase more gradually to approximately 0.82–0.83. The AGE CJTT specifications decrease slightly until 2012–2014 and then increase through stepwise adjustments and plateaus.
- The standard and GTC LSE specifications increase smoothly until approximately 2012–2015 and then remain stable or decrease slightly. In contrast, the QTT and piecewise specifications show pronounced inverted U-shaped curves, increasing until 2015, decreasing through 2022, and generally increasing again in 2023. These specifications capture the broad OPFP 2015 peak but not the OPFP and BC95 dip in 2010 to 2012, nor the short-term fluctuations.

Figure 4.4.12.1: TND BC95 vs OPFP Trends – Cobb-Douglas Models

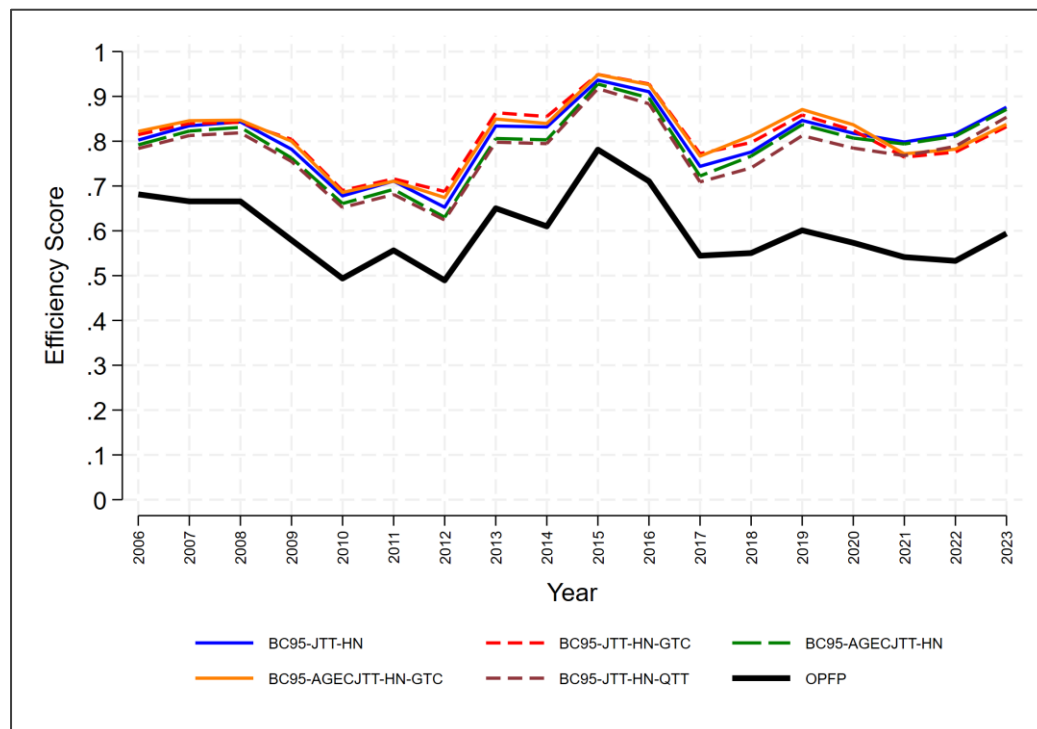


Figure 4.4.12.2: TND BC95 vs OPFP Trend – Translog Models

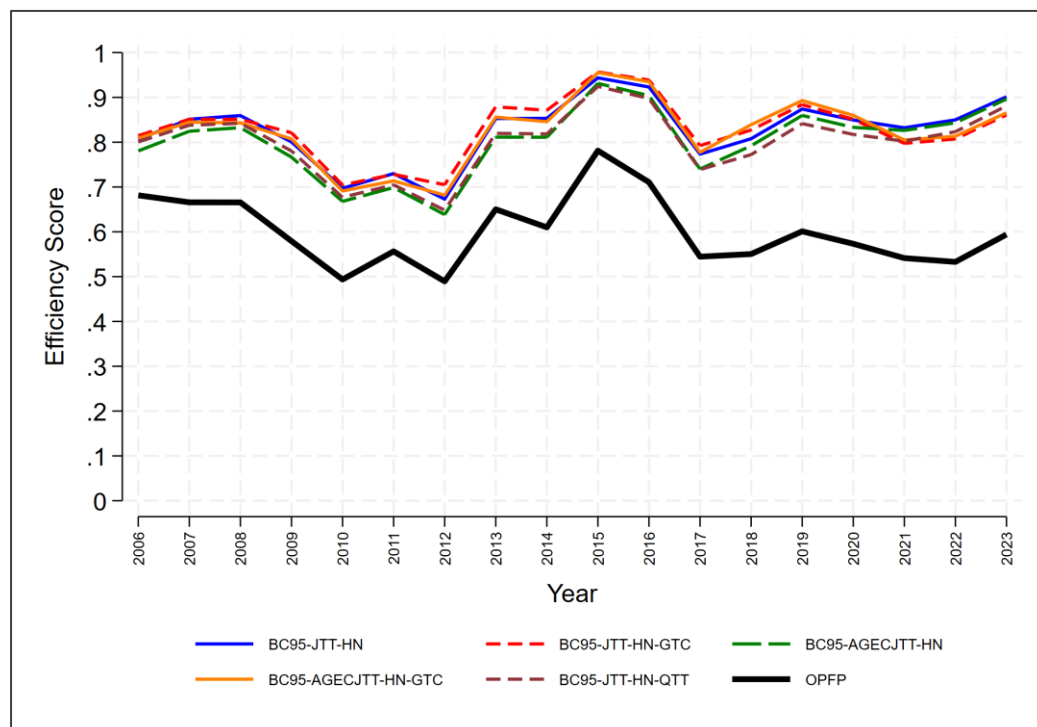


Figure 4.4.12.3: TND BC95 vs OPFP Trend – Parsimonious Translog Models

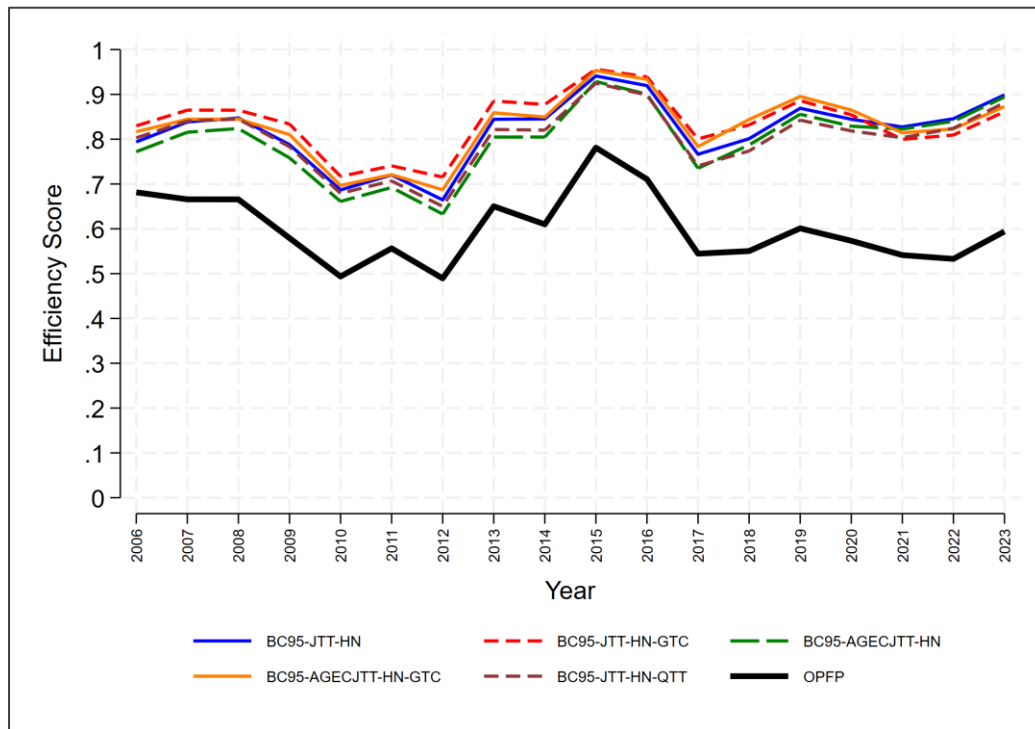


Figure 4.4.12.4: TND Kumb90 vs OPFP Trends – Cobb-Douglas Models

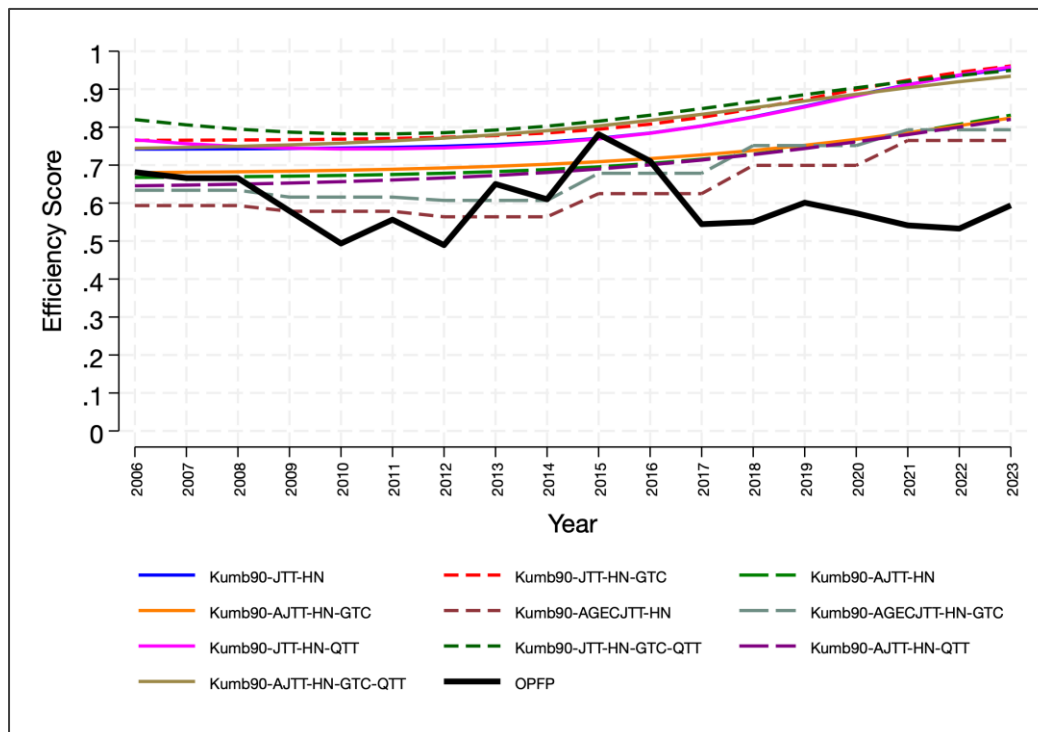


Figure 4.4.12.5: TND Kumb90 vs OPFP Trends – Translog Models

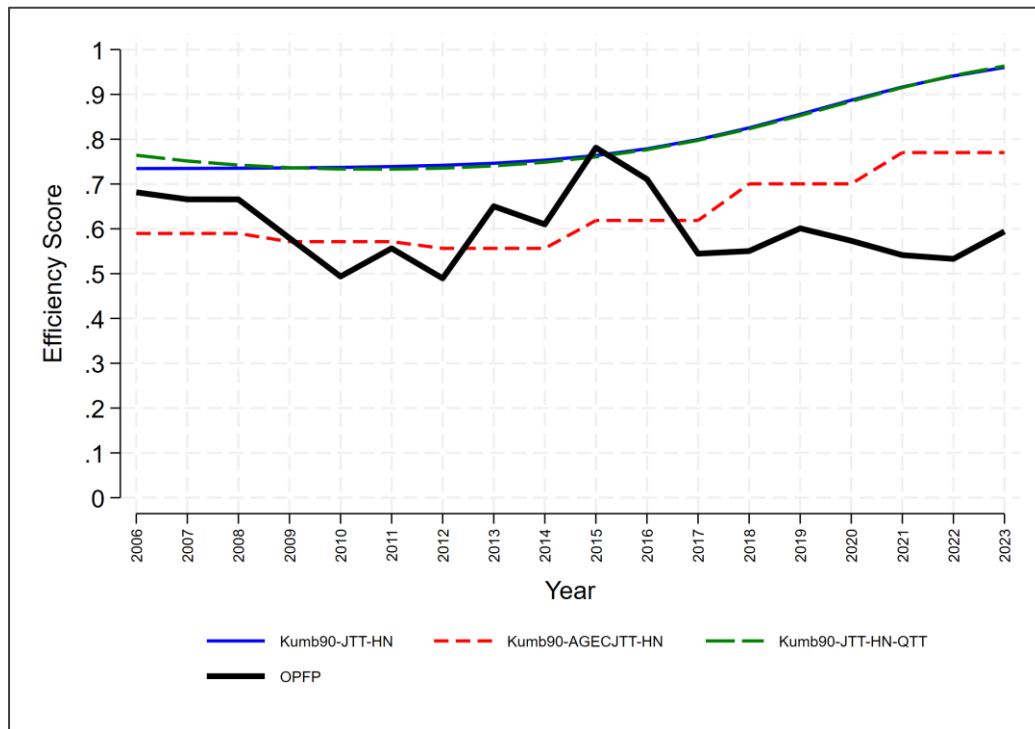


Figure 4.4.12.6: TND Kumb90 vs OPFP Trends – Parsimonious Translog Models

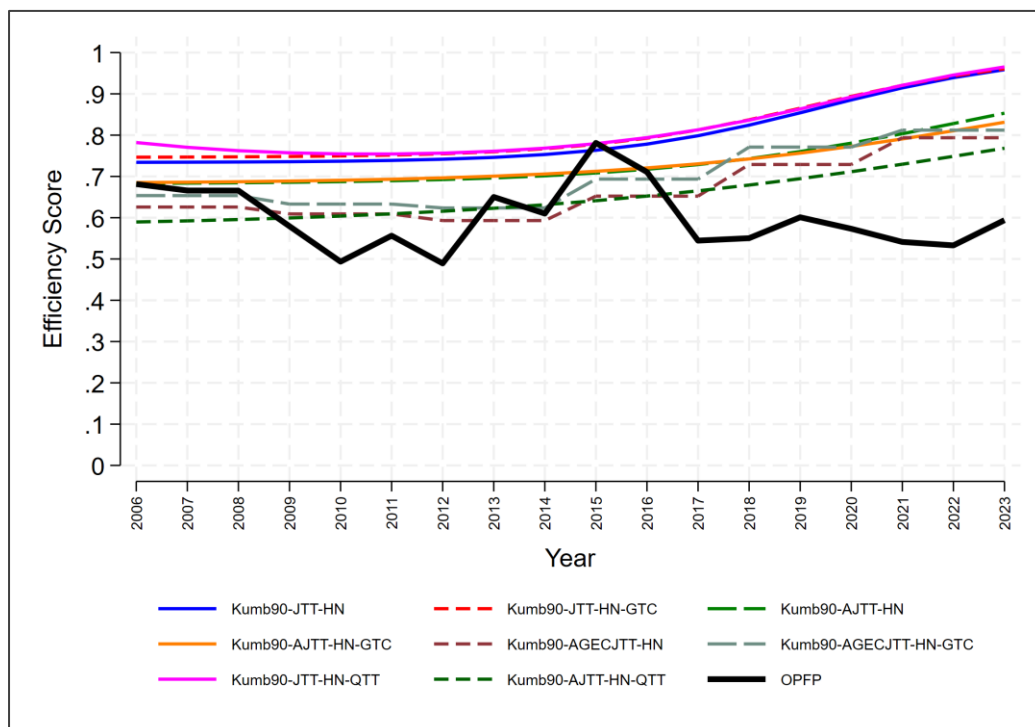


Figure 4.4.12.7: TND LSE vs OPFP Trends – Cobb-Douglas Models

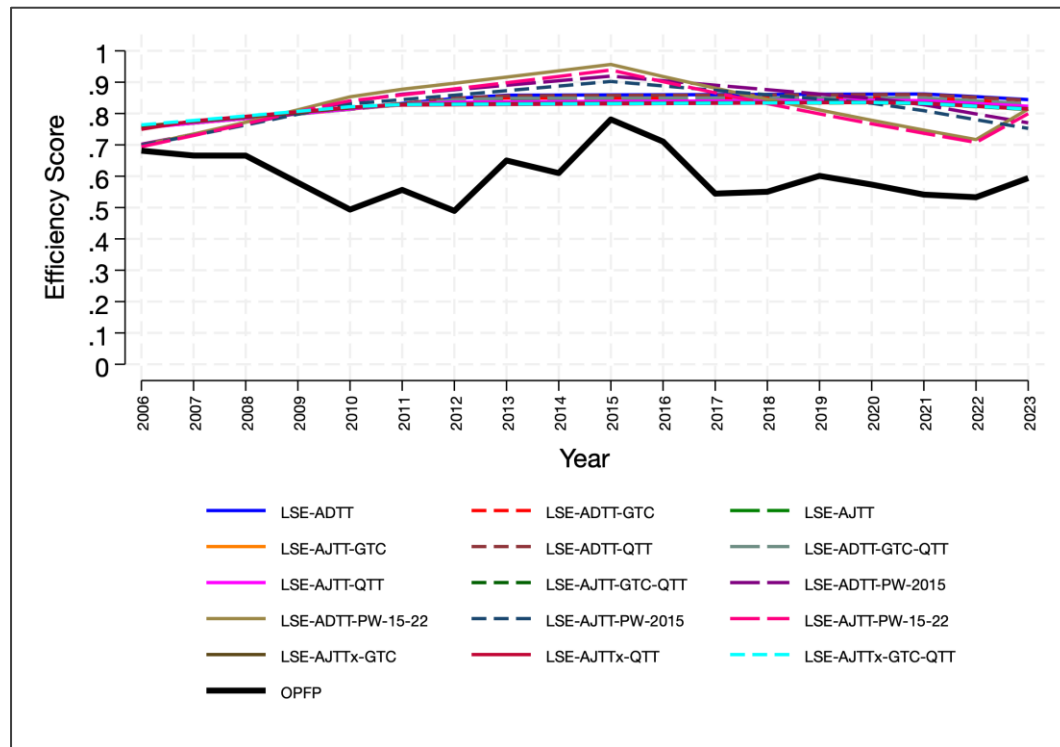


Figure 4.4.12.8: TND LSE vs OPFP Trends – Translog Models

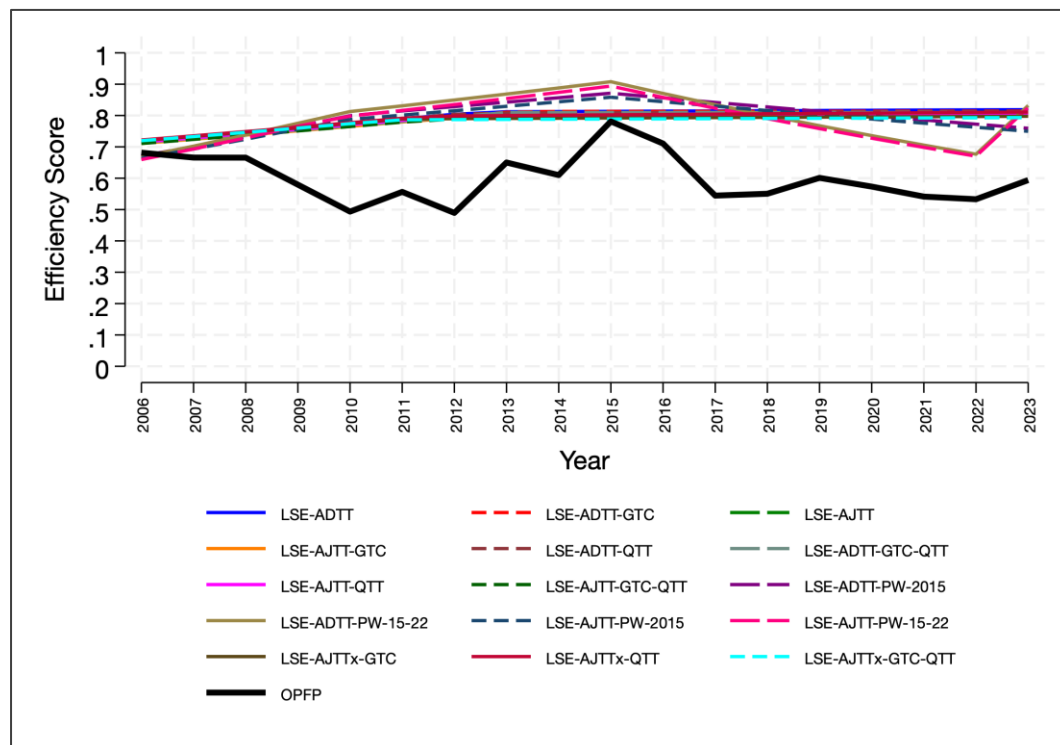
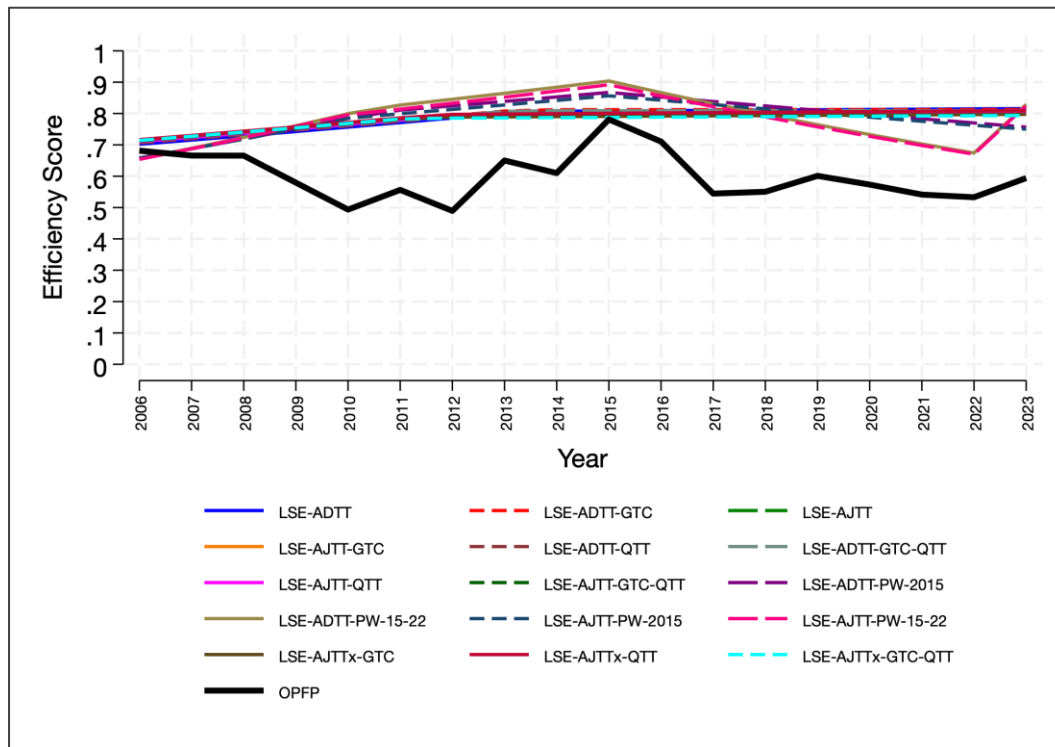


Figure 4.4.12.9: TND LSE vs OPFP Trends – Parsimonious Translog Models



4.4.13 UED

For UED, Figures 4.4.13.1 to 4.4.13.9 show that:

- OPFP efficiency increased gradually from 2006 to 2009 and then decreased to its minimum in 2012. It fluctuated until 2017, increased sharply in 2018, and remained relatively stable thereafter, with small decreases in 2020 and 2022. The efficiency score trends for the BC95 specifications closely reproduce these movements but generate consistently much higher scores.
- The JTT Kumb90 specifications remain relatively stable until approximately 2013–2014 and then increase smoothly to around 0.97 by 2023. The AJTT specifications increase more strongly after 2014, reach one between approximately 2018 and 2020, and remain there. The AGE CJTT specifications decrease slightly until 2012–2014 and then increase through stepwise adjustments and plateaus.
- The CD LSE specifications show similar smooth, with efficiency increasing at a slowing rate from approximately 0.75–0.79 in 2006 to the efficiency frontier by around 2020 and remaining there through 2023. In the TLG and TLGP LSE models, efficiency also steadily increases, but commencing at lower levels in 2006 and reaching 0.9 in 2023. The LSE models do not capture the dip in OPFP from 2011 to 2017.

Figure 4.4.13.1: UED BC95 vs OPFP Trends – Cobb-Douglas Models

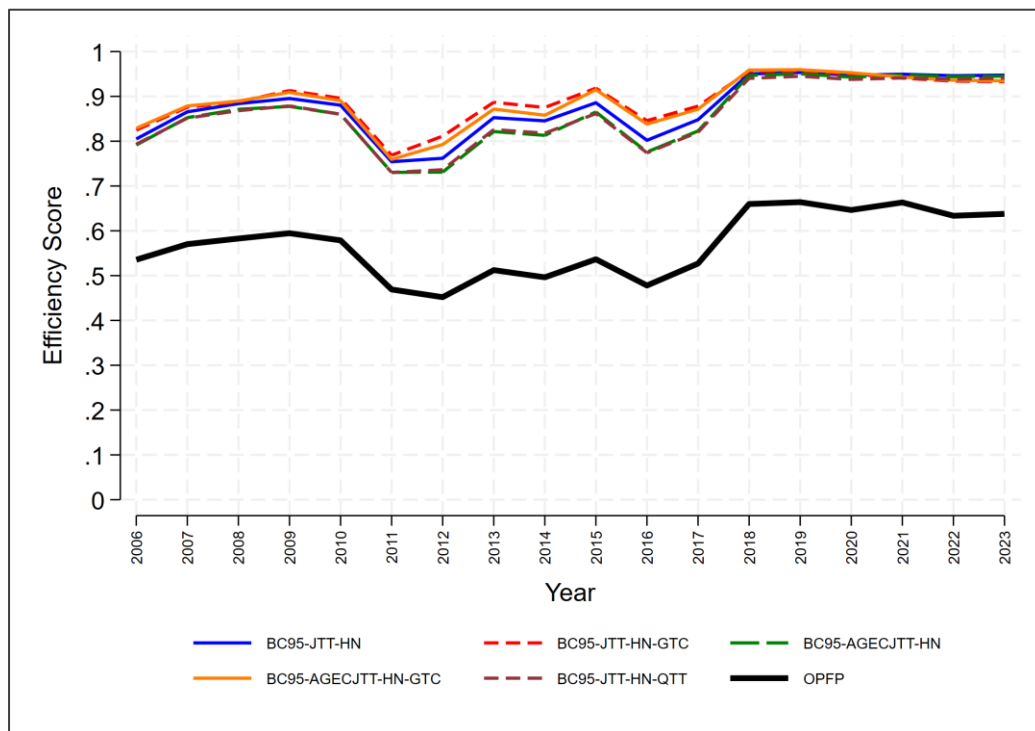


Figure 4.4.13.2: UED BC95 vs OPFP Trend – Translog Models

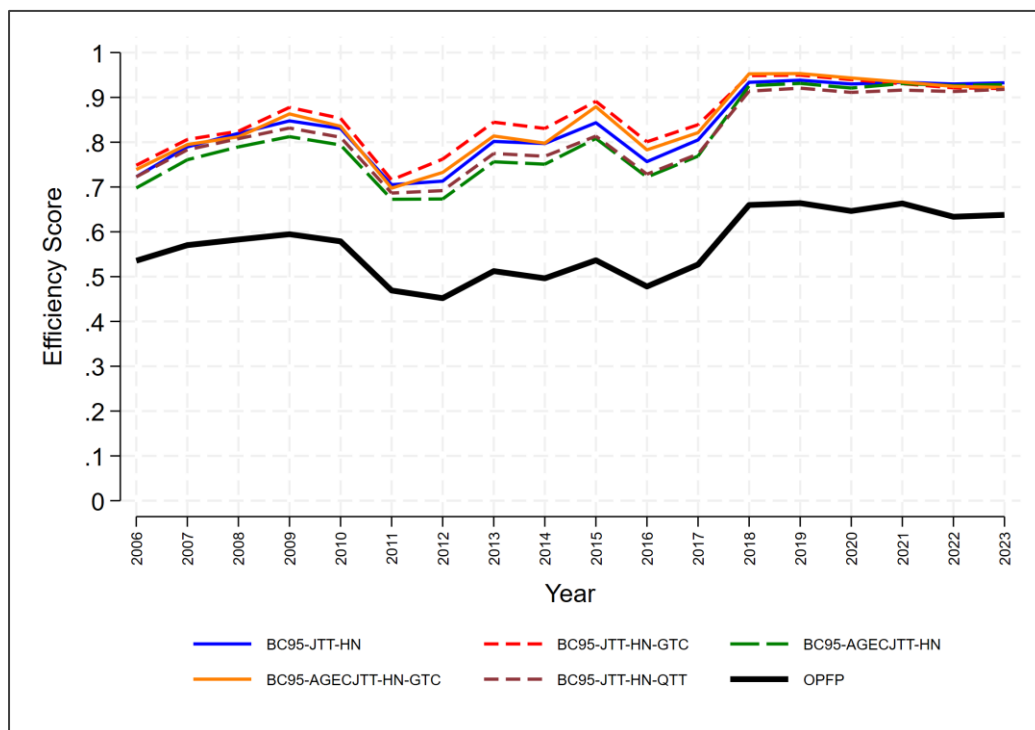


Figure 4.4.13.3: UED BC95 vs OPFP Trend – Parsimonious Translog Models

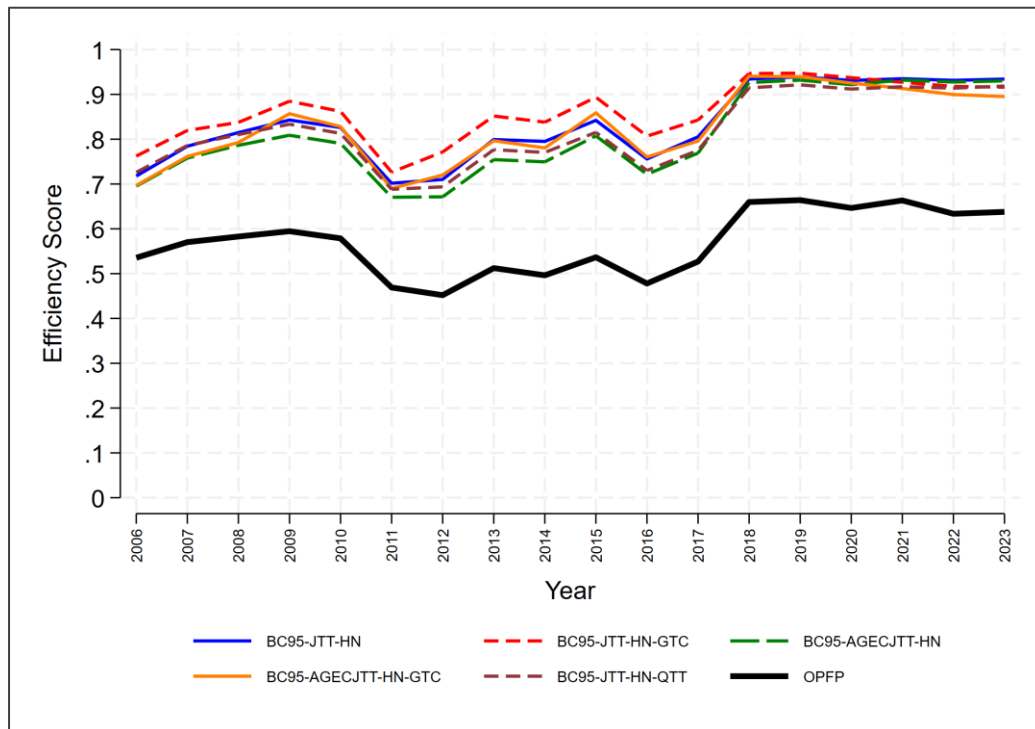


Figure 4.4.13.4: UED Kumb90 vs OPFP Trends – Cobb-Douglas Models

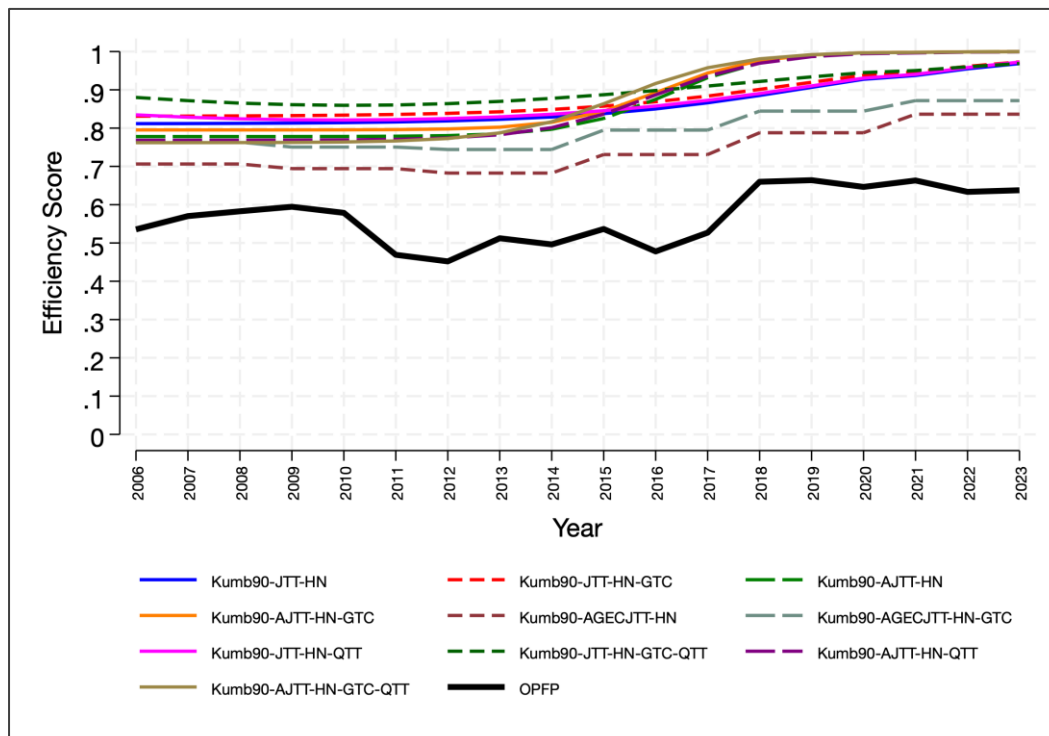


Figure 4.4.13.5: UED Kumb90 vs OPFP Trends – Translog Models

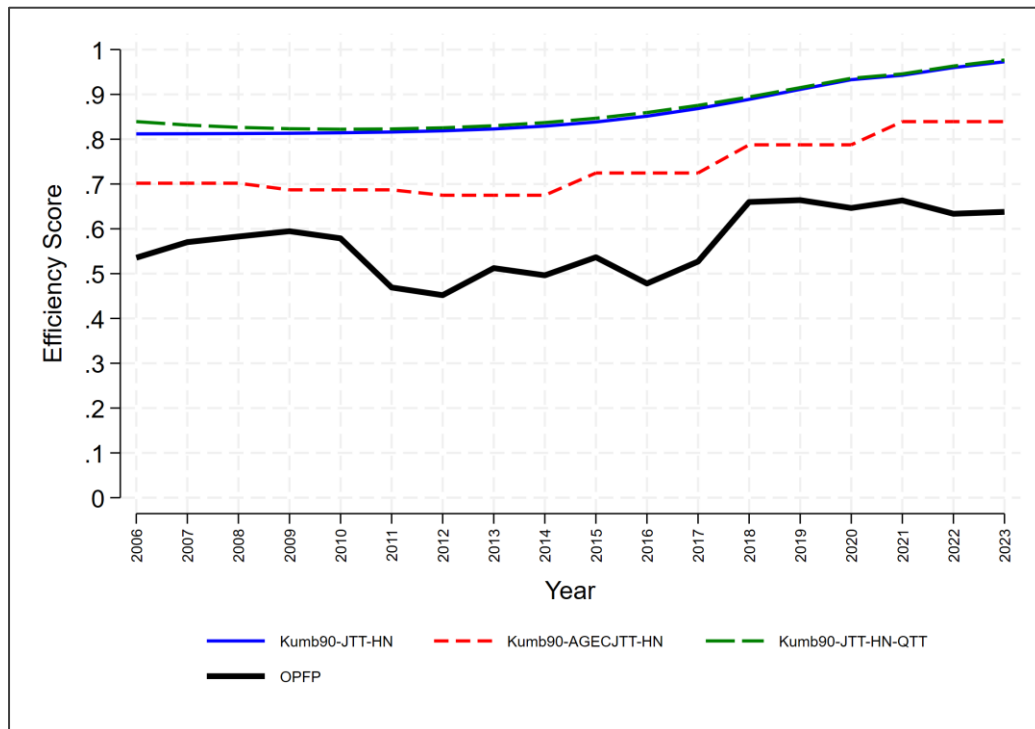


Figure 4.4.13.6: UED Kumb90 vs OPFP Trends – Parsimonious Translog Models

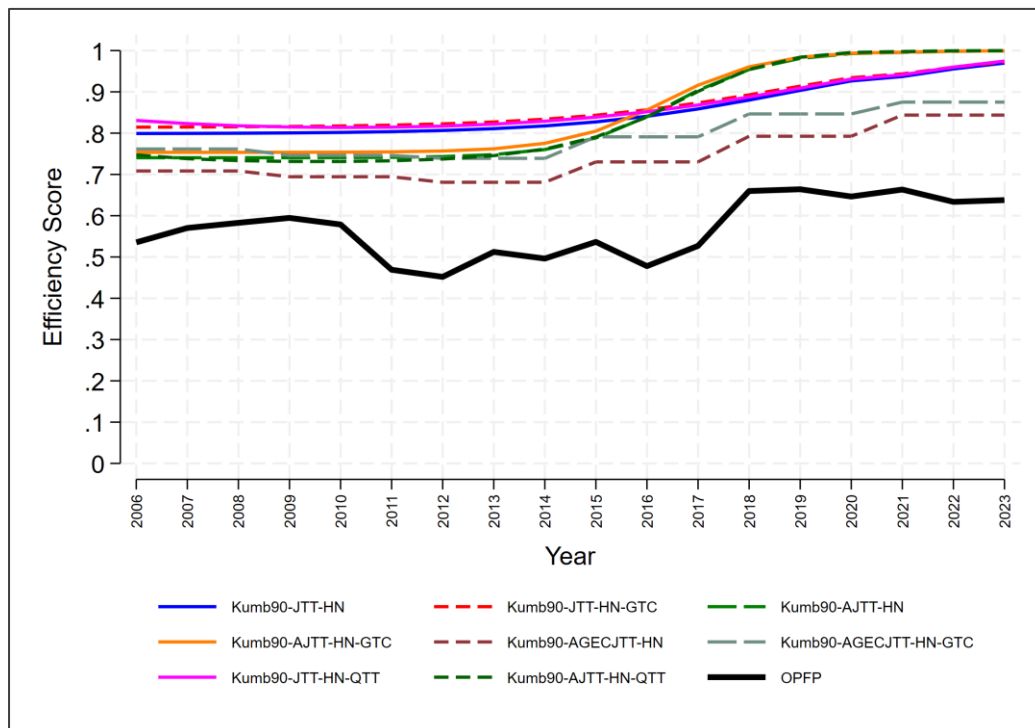


Figure 4.4.13.7: UED LSE vs OPFP Trends – Cobb-Douglas Models

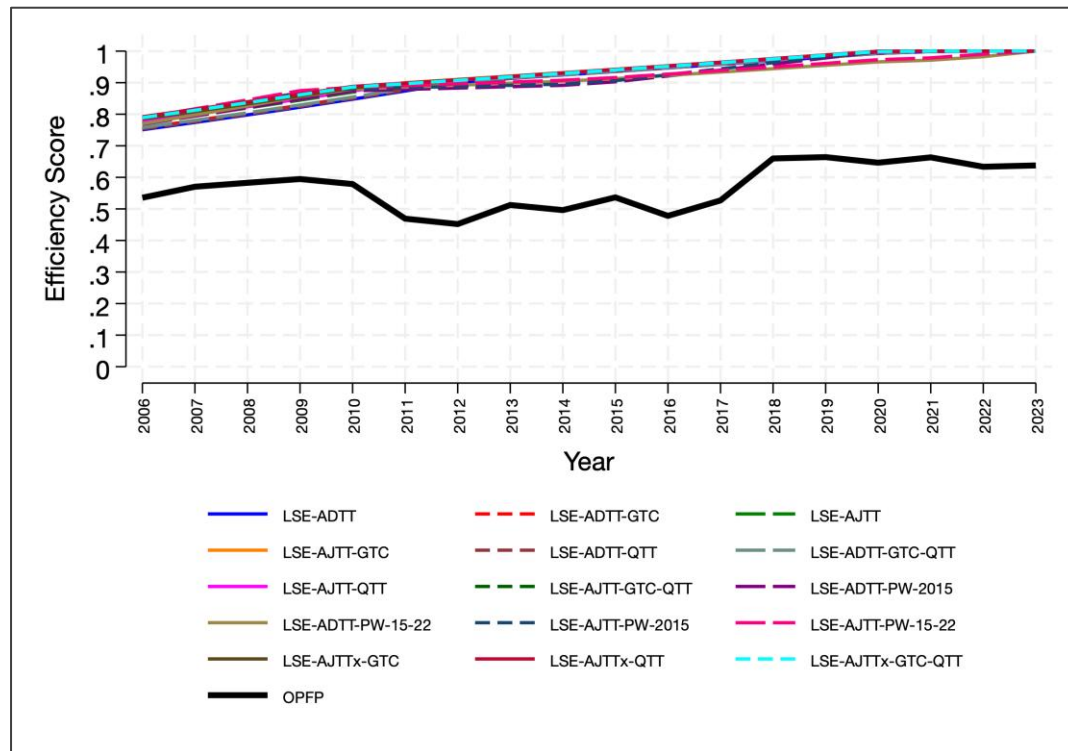


Figure 4.4.13.8: UED LSE vs OPFP Trends – Translog Models

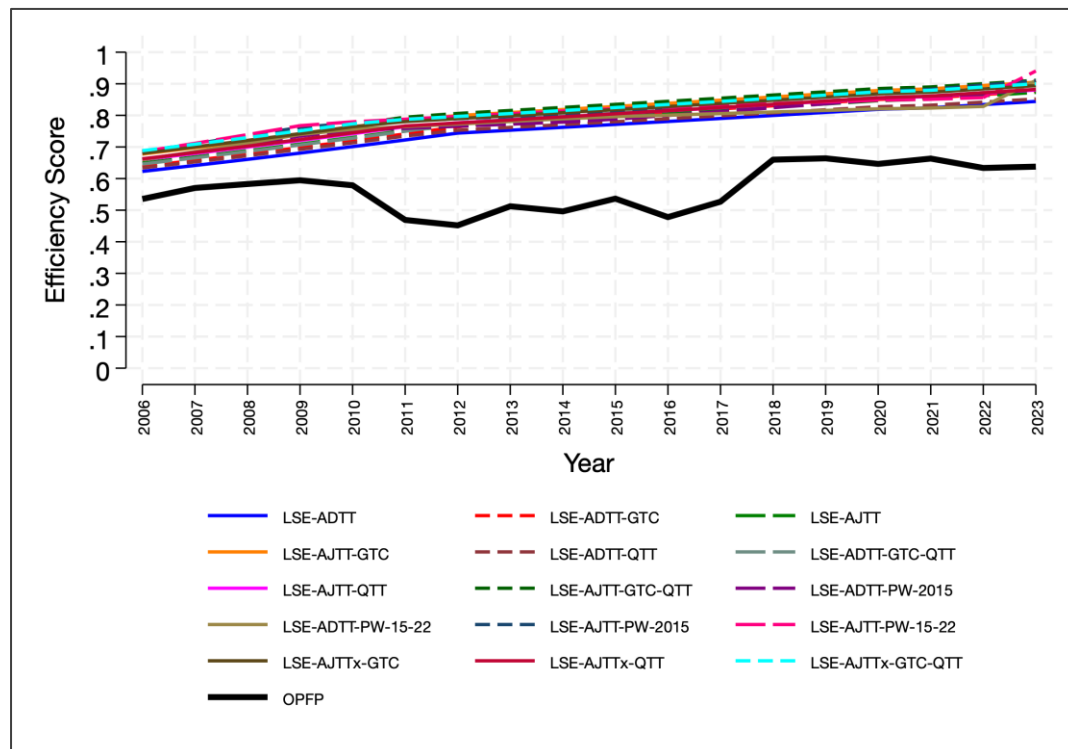
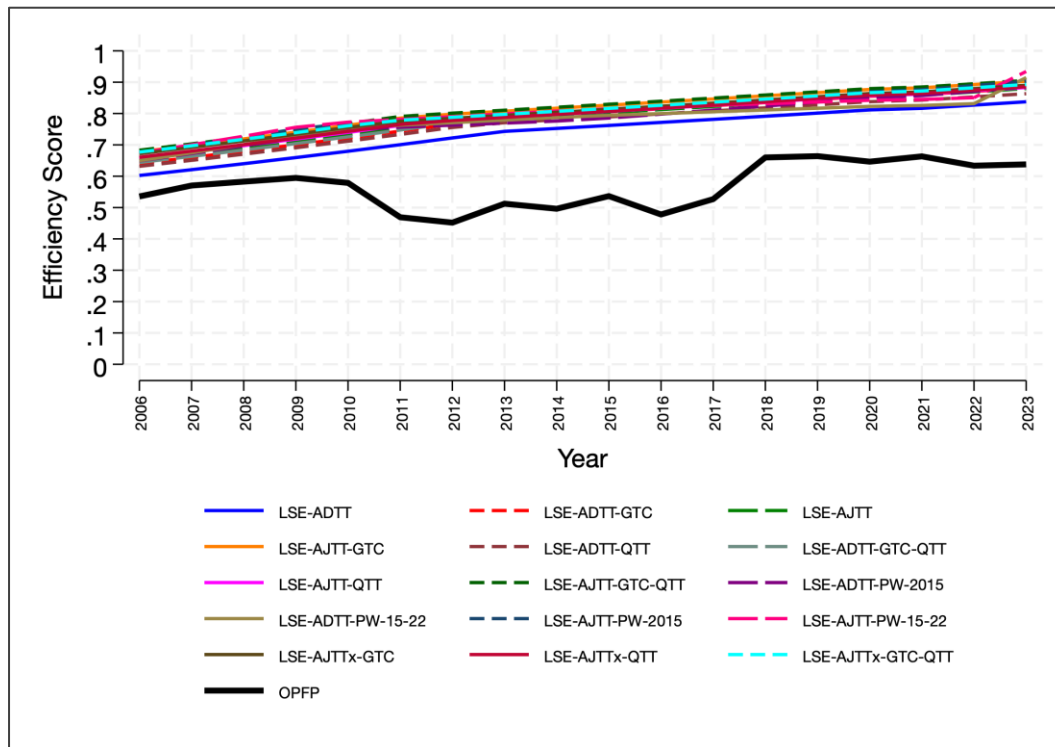


Figure 4.4.13.9: UED LSE vs OPFP Trends – Parsimonious Translog Models



4.4.14 Efficiency Profile Assessment

The main observations are:

- The shape of efficiency for the three functional forms (CD, TLG and TLGP) are generally similar.
- Across the CD, TLG, and TLGP functional forms, the BC95 efficiency trends closely track the annual movements in the OPFP series. In contrast, Kumb90 and LSE generally produce smoother trajectories with fewer short-term fluctuations.
- The BC95 estimates are tightly clustered for each DNSP. This indicates that the broad efficiency trends are robust to specification choice, although the estimated efficiency levels vary slightly.
- The additional flexibility of the AJTT specification does not usually change the broad trend relative to the corresponding JTT specification. However, differences in levels, curvature, and late-sample behaviour are material for some DNSPs, particularly within the Kumb90 models.
- Quadratic time trends and piecewise or spline terms introduce additional curvature. In the LSE models, the standard and GTC specifications often produce smooth, nearly linear or trends with a slowing rate of change, whereas the quadratic and piecewise specifications frequently produce trends that align better with the OPFP trends.

- For all DNSPs, the Kumb90 AGEJCJT specifications decrease slightly until around the middle of the sample and increase more strongly thereafter, with changes occurring as discrete adjustments separated by extended plateaus. This suggests that the AGEJC specification provides less temporal flexibility than expected because it does not capture year-to-year fluctuations within each subperiod.
- BC95 also reproduces late-sample turning points more consistently. Kumb90 and LSE often smooth or attenuate these changes, although the more flexible quadratic and piecewise specifications are more likely to capture them.

The efficiency scores produced by each model were regressed against the corresponding OPFP scores. The OPFP measure is not treated as a definitive or error-free measure of efficiency. However, the analysis suggests that it captures movements in productivity and efficiency trends that are not always reflected as clearly by some of the econometric models. The resulting R^2 provides a measure of the strength of the statistical association between the two sets of scores and is reported in Table 4.4.1. The main conclusions are:

- The BC95 models show the strongest alignment with the OPFP scores. The highest R^2 is recorded for the BC95-AGEJCJT-HN-GTC TLGP model followed closely by the BC95-JTT-HN-QTT TLG and TLGP models.
- TLG and TLGP generally produce higher R^2 values than CD, particularly for the BC95 and LSE models. Differences between TLG and TLGP are usually small.
- The LSE piecewise models perform better than the standard and quadratic LSE specifications.
- Among the quadratic LSE models, the AJTTx-QTT specification performs best, although it remains below the BC95 and piecewise LSE models.
- The Kumb90 models show the weakest overall alignment with OPFP.

Table 4.4.1: OPFP and Efficiency Scores Squared Correlation Coefficient*

<i>Model</i>	<i>Functional Form</i>	<i>R²</i>
<u><i>Revised Phase 2 Models</i></u>		
BC95-JTT-HN	CD	0.725
	TLG	0.773
	TLGP	0.767
BC95-JTT-HN-GTC	CD	0.713
	TLG	0.769
	TLGP	0.771
Kumb90-JTT-HN	CD	0.522
	TLG	0.521
	TLGP	0.531
Kumb90-JTT-HN-GTC	CD	0.536
	TLGP	0.544
Kumb90-AJTT-HN [#]	CD	0.468
	TLGP	0.568
Kumb90-AJTT-HN-GTC [#]	CD	0.463
	TLGP	0.572
LSE-ADTT	CD	0.578
	TLG	0.674
	TLGP	0.664
LSE-ADTT-GTC	CD	0.582
	TLG	0.669
	TLGP	0.666
LSE-AJTT	CD	0.573
	TLG	0.672
	TLGP	0.667
LSE-AJTT-GTC	CD	0.571
	TLG	0.659
	TLGP	0.661
LSE-AJTtx-GTC	CD	0.575
	TLG	0.667
	TLGP	0.667
<u><i>Generalised Efficiency Change</i></u>		
BC95-AGECJTT-HN	CD	0.731
	TLG	0.768
	TLGP	0.763
BC95-AGECJTT-HN-GTC	CD	0.717
	TLG	0.768
	TLGP	0.796
Kumb90-AGECJTT-HN [#]	CD	0.500
	TLG	0.532
	TLGP	0.567
Kumb90-AGECJTT-HN-GTC [#]	CD	0.497
	TLGP	0.574

Table 4.4.1: (cont.)

<i>Model</i>	<i>Functional Form</i>	<i>R²</i>
<i>Quadratic Time Trend Models</i>		
BC95-JTT-HN-QTT	CD	0.728
	TLG	0.791
	TLGP	0.790
Kumb90-JTT-HN-QTT	CD	0.531
	TLG	0.532
	TLGP	0.556
Kumb90-JTT-HN-GTC-QTT [#]	CD	0.567
Kumb90-AJTT-HN-QTT	CD	0.450
	TLGP	0.526
Kumb90-AJTT-HN-GTC-QTT	CD	0.570
LSE-ADTT-QTT	CD	0.602
	TLG	0.697
	TLGP	0.685
LSE-ADTT-GTC-QTT	CD	0.601
	TLG	0.686
	TLGP	0.684
LSE-AJTT-QTT	CD	0.596
	TLG	0.691
	TLGP	0.687
LSE-AJTT-GTC-QTT	CD	0.595
	TLG	0.674
	TLGP	0.678
LSE-AJTtx-QTT	CD	0.642
	TLG	0.700
	TLGP	0.695
LSE-AJTtx-GTC-QTT	CD	0.598
	TLG	0.683
	TLGP	0.685
<i>Piecewise Regression Models</i>		
LSE-ADTT-PW-2015	CD	0.662
	TLG	0.740
	TLGP	0.735
LSE-ADTT-PW-2015-2022	CD	0.671
	TLG	0.742
	TLGP	0.736
LSE-AJTT-PW-2015	CD	0.652
	TLG	0.730
	TLGP	0.732
LSE-AJTT-PW-2015-2022	CD	0.661
	TLG	0.731
	TLGP	0.733

Notes: * R² from Regressions of Efficiency Scores on OPFP Scores.

As noted in Section 4.1, the model may be subject to identification issues.

4.5 Parsimony

The principle of parsimony is applied when two or more models perform similarly. In such cases, the preferred model is the one that achieves comparable performance with a simpler specification. For this assessment, the number of estimated parameters is used as a practical measure of model complexity. Table 4.5.1 presents the number of parameters estimated in each model and ranks the models from the fewest to the most parameters.

Table 4.5.1: Parsimony Assessment

<i>Model</i>	<i>Functional Form</i>	<i># Parameters</i>
<u><i>Revised Phase 2 Models</i></u>		
BC95-JTT-HN	CD	13
	TLG	19
	TLGP	17
BC95-JTT-HN-GTC	CD	17
	TLG	23
	TLGP	21
Kumb90-JTT-HN	CD	14
	TLG	20
	TLGP	19
Kumb90-JTT-HN-GTC	CD	18
	TLGP	21
Kumb90-AJTT-HN	CD	26
	TLGP	29
Kumb90-AJTT-HN-GTC	CD	30
	TLGP	33
LSE-ADTT	CD	33
	TLG	39
	TLGP	37
LSE-ADTT-GTC	CD	37
	TLG	43
	TLGP	41
LSE-AJTT	CD	34
	TLG	40
	TLGP	38
LSE-AJTT-GTC	CD	38
	TLG	44
	TLGP	42
LSE-AJT Tx-GTC	CD	39
	TLG	45
	TLGP	43

Table 4.5.1: (cont.)

<i>Model</i>	<i>Functional Form</i>	<i># Parameters</i>
<u><i>Generalised Efficiency Change</i></u>		
BC95-AGECJTT-HN	CD	17
	TLG	23
	TLGP	21
BC95-AGECJTT-HN-GTC	CD	21
	TLG	27
	TLGP	25
Kumb90-AGECJTT-HN	CD	18
	TLG	24
	TLGP	20
Kumb90-AGECJTT-HN-GTC	CD	22
	TLGP	24
<u><i>Quadratic Time Trend Models</i></u>		
BC95-JTT-HN-QTT	CD	16
	TLG	22
	TLGP	20
Kumb90-JTT-HN-QTT	CD	16
	TLG	22
	TLGP	18
Kumb90-JTT-HN-GTC-QTT	CD	21
Kumb90-AJTT-HN-QTT	CD	32
	TLGP	39
Kumb90-AJTT-HN-GTC-QTT	CD	37
LSE-ADTT-QTT	CD	34
	TLG	40
	TLGP	38
LSE-ADTT-GTC-QTT	CD	38
	TLG	44
	TLGP	42
LSE-AJTT-QTT	CD	36
	TLG	41
	TLGP	39
LSE-AJTT-GTC-QTT	CD	40
	TLG	45
	TLGP	43
LSE-AJT Tx-QTT	CD	41
	TLG	43
	TLGP	41
LSE-AJT Tx-GTC-QTT	CD	41
	TLG	46
	TLGP	44

Table 4.5.1: (cont.)

<i>Model</i>	<i>Functional Form</i>	<i># Parameters</i>
<i>Piecewise Regression Models</i>		
LSE-ADTT-PW-2015	CD	46
	TLG	52
	TLGP	50
LSE-ADTT-PW-2015-2022	CD	59
	TLG	65
	TLGP	63
LSE-AJTT-PW-2015	CD	47
	TLG	53
	TLGP	51
LSE-AJTT-PW-2015-2022	CD	60
	TLG	66
	TLGP	64

4.6 Sensitivity analysis for alternative sample periods

This section assesses whether the 29 models with reported results remain robust when the estimation end year changes relative to the main 2006–2023 evaluation sample. The alternative periods are 2006–2019, 2006–2020, 2006–2021, 2006–2022 and 2006–2024. For each period, the assessment considers CD and TLG functional forms.²⁰ Appendix D reports the full assessment.

4.6.1 Computational Issues

The results showed the following patterns:

- *BC95 family*
 - BC95-JTT-HN and BC95-AGECJTT-HN recorded flat-region warnings in the earlier sample periods but became more computationally stable as the sample period expanded.
 - BC95-JTT-HN-GTC, BC95-AGECJTT-HN-GTC and BC95-JTT-HN-QTT exhibited computational issues in the first two sample periods. These issues were not observed in the intermediate periods but re-emerged when the sample was extended to 2006–2024. The issues included flat-region warnings and missing estimates.

²⁰ Estimating a separate parsimonious Translog specification for each alternative sample period would be considerably more time-consuming. For the same reason, the QTT specification selected for the 2006–2023 analysis was retained across the alternative sample periods, rather than sequentially removing statistically insignificant terms for each sample.

- *Kumb90 family*
 - Kumb90-JTT-HN-GTC, Kumb90-AJTT-HN-GTC and Kumb90-AGECJTT-HN-GTC were estimated without computational issues in all sample periods. For all flagged Kumb90 models, the computational issue was limited to a flat-region warning: Kumb90-JTT-HN recorded a flat-region warning only in the final sample period; Kumb90-AJTT-HN recorded a flat-region warning only in the first sample period; Kumb90-JTT-HN-QTT recorded a flat-region warning in every sample period except 2006–2020; Kumb90-AGECJTT-HN and Kumb90-AJTT-HN-QTT recorded flat-region warnings in every sample period.
- *LSE family*
 - LSE-AJTT-GTC-QTT also omitted variables because of collinearity in the first two sample periods.

Overall, this assessment indicates that the models of greatest concern are LSE-ADTT-PW-2015-2022, LSE-AJTT-PW-2015-2022 and BC95-AGECJTT-HN-GTC. For most other models, the identified issues were limited to flat-region warnings or were resolved when an additional year was added to the sample. By contrast, these three models exhibited more serious computational problems that persisted into, or re-emerged in, the final sample period.

4.6.2 Outputs and UGShare Coefficients

All BC95 and LSE models produced output coefficients with the expected signs and statistical significance across all sample periods and functional forms. For the Kumb90 models, all output coefficients had the expected signs across all models and sample periods. However, the criterion requiring all output coefficients to be statistically significant was not satisfied in the following cases:

- Kumb90-AGECJTT-HN-GTC under the TLG specification for the first sample period.
- Kumb90-AJTT-HN and Kumb90-AJTT-HN-GTC under the TLG specification for the sample periods ending in 2020, 2021 and 2022.
- Kumb90-AJTT-HN-QTT under the TLG specification for the sample period ending in 2021.

Regarding the UGShare coefficient, all BC95 and LSE models consistently produce a negative coefficient across all sample periods and functional forms. Within the Kumb90 family, the results were more mixed:

- Kumb90-JTT-HN, Kumb90-AJTT-HN, Kumb90-AGECJTT-HN and Kumb90-AJTT-HN-QTT produced negative UGShare coefficients across all sample periods and functional forms.

- Kumb90-JTT-HN-QTT produced a positive coefficient under the TLG specification in the first sample period but negative coefficients in all other cases. Kumb90-JTT-HN-GTC produced positive coefficients under the TLG specification in the first three sample periods, followed by negative coefficients in the final two periods.
- Kumb90-AJTT-HN-GTC produced positive coefficients under the TLG specification in the final three sample periods. Kumb90-AGECJTT-HN-GTC produced a positive coefficient under the TLG specification in the final sample period.

Overall, this assessment indicates that the TLG specifications of Kumb90-AJTT-HN-GTC and Kumb90-AGECJTT-HN-GTC are of greatest concern, as they failed the main outputs and UGShare coefficient assessments in the later sample periods.

4.6.3 Monotonicity Violations

The BC95 models exhibit a low incidence of monotonicity violations across all specifications and sample periods. For Australian DNSPs, the frequency of monotonicity violations remains below 10 per cent in every case. The main patterns are as follows:

- BC95-JTT-HN-GTC, BC95-AGECJTT-HN-GTC and BC95-JTT-HN-QTT exhibit relatively stable frequencies of monotonicity violations across the alternative sample periods. No DNSP records an excessive incidence of violations under any of these specifications.
- For BC95-JTT-HN and BC95-AGECJTT-HN, the frequency of total monotonicity violations generally increases slightly as additional years are included in the sample. For the Australian sample, the frequency of MV increases slightly with the inclusion of 2022 and 2024. No DNSP records an excessive incidence of violations for BC95-JTT-HN in samples ending up to 2022, or for BC95-AGECJTT-HN in samples ending up to 2024.

The pattern of monotonicity violations is more mixed across the Kumb90 models:

- Kumb90-JTT-HN, Kumb90-JTT-HN-GTC, Kumb90-AGECJTT-HN, Kumb90-AGECJTT-HN-GTC and Kumb90-JTT-HN-QTT improve as additional years are included in the sample and achieve a zero frequency of violations in the final sample period.
- Kumb90-AJTT-HN, Kumb90-AJTT-HN-GTC and Kumb90-AJTT-HN-QTT exhibit a less consistent pattern. Their monotonicity performance deteriorates over the intermediate sample periods before improving in the final period.

For the LSE models, the frequency of monotonicity violations remains relatively stable across the alternative sample periods, with limited variation over time. Most models record one DNSP with an excessive incidence of violations in each period. The AJTT piecewise

specifications exhibit the weakest monotonicity performance, recording two DNSPs with excessive violations in most sample periods, although this falls to one in the final period. By contrast, LSE-ADTT-GTC and LSE-ADTT-GTC-QTT provide the strongest and most consistent results, with no DNSPs recording an excessive incidence of monotonicity violations in any sample period.

Overall, the assessment indicates that the BC95 and LSE models exhibit the strongest monotonicity performance, with the frequency of violations for Australian DNSPs remaining below 25 per cent in every sample period. The Kumb90 models generally perform less well, with several specifications recording violation rates of 50 per cent or more in multiple periods. However, their performance tends to improve as additional years are included, with many specifications reaching zero violations in the final period. The model of greatest concern is Kumb90-AJTT-HN-GTC. By contrast, the strongest-performing models are BC95-JTT-HN-GTC, BC95-AGECJTT-HN-GTC, BC95-JTT-HN-QTT, LSE-ADTT-GTC and LSE-ADTT-GTC-QTT.

4.6.4 Concluding comments

The sensitivity analysis shows that robustness varies across models and assessment criteria. No model dominates under every criterion, and strong performance in one area does not necessarily imply robustness in another.

- The BC95 models generally provide the strongest overall results. However, some BC95 specifications remain sensitive to the estimation period from a computational perspective, particularly in the shorter samples and when the 2024 observations are included. BC95-AGECJTT-HN-GTC illustrates this trade-off, as it performs strongly on monotonicity but presents more serious computational concerns.
- The LSE models also exhibit robust output and UGShare coefficients and generally stable monotonicity performance. LSE-ADTT-GTC and LSE-ADTT-GTC-QTT provide the strongest results within this family, with no DNSPs recording excessive monotonicity violations across the alternative samples. The main concerns relate to collinearity in LSE piecewise models. The AJTT piecewise models also exhibit comparatively weaker monotonicity performance.
- The Kumb90 models show the greatest sensitivity to the selected sample period. These concerns are most pronounced for Kumb90-AJTT-HN-GTC, which performs inconsistently across several assessment dimensions.

Taken together, the results support excluding the following models from further consideration:

- Kumb90-AJTT-HN-GTC, Kumb90-AGECJTT-HN-GTC, Kumb90-AJTT-HN-QTT, LSE-ADTT-PW-2015-2022, LSE-AJTT-PW-2015-2022.

- The results tend to favour the BC95 and LSE specifications over the Kumb90 specifications, notwithstanding the computational-performance concerns identified for some BC95 models.

4.7 Summary Assessment

The final assessment and selection of the preferred models is not straightforward. Despite the application of clearly defined criteria, distinguishing among models with broadly similar performance inevitably involves some degree of judgement. The results also reveal trade-offs across the assessment criteria that cannot always be resolved unambiguously.

The assessment began with 33 models, of which only 5 were excluded before the final selection stage. This indicates that the results are broadly robust across a relatively large set of specifications, irrespective of the models ultimately selected as preferred.

The first step in the final selection process is therefore to identify a shortlist of preferred models within each model family. The preferred BC95 models are ranked as follows:

1. BC95-JTT-HN
2. BC95-JTT-HN-GTC
3. BC95-JTT-HN-QTT.

These models exhibit relatively similar overall performance. BC95-JTT-HN is preferred to BC95-JTT-HN-GTC and BC95-JTT-HN-QTT because the latter two specifications encounter some computational difficulties when 2024 is added to the sample (see Section 4.6.1 and Appendix D). BC95-JTT-HN-GTC is ranked ahead of BC95-JTT-HN-QTT on parsimony grounds. The inclusion of the quadratic time-trend terms does not materially improve the efficiency profile, while the monotonicity performance of the two models is broadly similar. The BC95-AGECJTT models are not preferred because they produce similar or weaker monotonicity results and do not materially improve the estimated efficiency profiles.

The preferred Kumb90 models are ranked as follows:

1. Kumb90-JTT-HN
2. Kumb90-JTT-HN-GTC, considering only the CD and TLGP specifications
3. Kumb90-JTT-HN-QTT.

These models recorded no monotonicity violations in the final sample period and did not exhibit serious computational problems. Given their broadly similar performance, the ranking primarily reflects parsimony, with simpler specifications preferred where the additional terms do not provide a material improvement.

The Kumb90-AJTT models were excluded because their results suggest potential identification issues and they performed less strongly in the sensitivity analysis. The Kumb90-

AGECJTT models were also excluded because some specifications produced large standard errors, raising concerns about the precision and reliability of the estimated coefficients.

The preferred LSE models are ranked as follows:

1. LSE-ADTT-GTC-QTT
2. LSE-ADTT-GTC
3. LSE-AJTT-QTT.

The shortlisted models perform relatively similarly overall. However, the ADTT specifications are ranked above the AJTT specification because they consistently produce fewer monotonicity violations and are more parsimonious. LSE-ADTT-GTC-QTT is ranked first because the quadratic time trend can provide additional flexibility in the estimated efficiency profile while retaining the strong monotonicity performance of the ADTT-GTC specification.

To consolidate the shortlisted models into an overall ranking we consider that the criteria-based assessment generally favours the BC95 models. These specifications produce coefficients that are consistent with theoretical expectations, low rates of monotonicity violations and relatively flexible efficiency profiles that can capture changes over time. However, the sensitivity analysis indicates that some BC95 specifications may be prone to computational instability when the estimation period changes.

The LSE models also provide stable and satisfactory results. However, their efficiency profiles are generally less flexible than those of the BC95 models. The Kumb90 models also frequently produce low rates of monotonicity violations. However, their efficiency profiles appear to be highly sensitive to the specification and normalisation of the time trend. They also produce less robust results across alternative sample periods. For these reasons, our final model rankings are:

1. BC95-JTT-HN
2. BC95-JTT-HN-GTC
3. BC95-JTT-HN-QTT
4. LSE-ADTT-GTC-QTT
5. LSE-ADTT-GTC
6. LSE-AJTT-QTT
7. Kumb90-JTT-HN
8. Kumb90-JTT-HN-GTC, considering only the CD and TLGP specifications
9. Kumb90-JTT-HN-QTT.

Regarding the functional form, the assessment generally favours the TLGP specification over the full TLG specification. The TLGP form retains the statistically relevant second-order terms while reducing unnecessary complexity, multicollinearity and computational burden, with little loss of model fit. It is therefore preferred where both specifications produce otherwise comparable results. Further, for these highest ranked models, the TLGP functional form is generally superior to the CD functional form.

To select among the short-listed models, it may be desirable to choose from among the two highest-ranked of the model families (BC95, LSE and Kumb90). The two preferred models would then be: (i) BC95-JTT-HN-GTC; (ii) LSE-ADTT-GTC-QTT. If more diversity is desirable, one could also include the second- or third-ranked BC95 and LSE models or include a Kumb90 model. In part this choice depends on whether the average efficiency scores are to be used in the BRFM or the end-of-period efficiency scores, since the Kumb90 models tend to produce less satisfactory time-profiles of efficiency than the BC95 and LSE models. The question of how efficiency scores are used in the benchmarking roll-forward model is discussed in section 6.

5 Decomposition of Productivity Changes

As outlined in the Phase 2 report, a cost function with time-varying inefficiency can decompose changes in productivity into four components: cost efficiency change (CEC), technical change (TC), scale efficiency change (SEC), and changes in operating environment factors (OEF).

5.1 Outline of the Method

The method presented here broadly follows Pacific Economics Group (2004), Lawrence (2007) and Economic Insights (2012). A decomposition of changes in opex partial factor productivity can be derived from the opex model, the general form of which, *at a given observation*, is:

$$\ln C = \alpha_0 + \sum_k \beta_k \ln q_k + \sum_n \gamma_n z_n + \tau(t|\lambda) + u + v \quad (5.1)$$

where C is the real opex; q_k is the quantity of output k ; z_n is the quantity (which may or may not be in logs) of the n th OEF (eg, jurisdictional dummies, the share of undergrounding); $\tau(t|\lambda)$ is the function modelling technical change (through its derivative w.r.t. time, $\tau'(t|\lambda)$), and λ is a set of parameters; u is the inefficiency term; and v is the idiosyncratic error term.

The first step is differentiating (5.1) with respect to time (t) and denoting the growth rate of a variable (eg, $\dot{q} = \partial \ln q / \partial t$) by a dot over the variable (eg, $\dot{q}$) yields:²¹

$$\dot{C} = \sum_k \varepsilon_{q_k} \dot{q}_k + \sum_n \gamma_n \frac{\partial z_n}{\partial t} + \tau'(t|\lambda) + \frac{\partial u}{\partial t} \quad (5.2)$$

where $\varepsilon_{q_k} = \beta_k$ represents the elasticity of real opex with respect to output k . Next, define the rate of change in an aggregate output index using the output elasticities as weights:

$$\dot{Q} \equiv \sum_k \left(\frac{\varepsilon_{q_k}}{\varepsilon_Q} \right) \dot{q}_k \quad \text{where: } \varepsilon_Q = \sum_k \varepsilon_{q_k} \quad (5.3)$$

The rate of partial factor productivity growth for opex inputs is: $P\dot{F}P_{opex} = \dot{Q} - \dot{C}$. Expanding this expression using equations (5.2), (5.3), gives the trend in the partial productivity of opex:

$$P\dot{F}P_{opex} = (1 - \varepsilon_Q) \dot{Q} - \sum_n \gamma_n \frac{\partial z_n}{\partial t} - \tau'(t|\lambda) - \frac{\partial u}{\partial t} \quad (5.4)$$

In this expression:

²¹ The derivative of the white noise term v with respect to time is itself a symmetrical random distribution with an expected value of zero.

- (a) $(1 - \varepsilon_Q)\dot{Q}$: The effects of changes in scale efficiency; ie, SEC.
- (b) $-\sum_n \gamma_n \frac{z_n}{\partial t}$: The effects on the sum of the partial productivity of changes from OEFs; ie, OEFC.
- (c) $-\tau'(t|\lambda)$: The rate of technical change or ‘frontier shift’; ie, TC.
- (d) $-\frac{\partial u}{\partial t}$: The change in cost efficiency or ‘catch-up’; ie, CEC.

Noting that $\dot{C} = \dot{Q} - P\dot{F}P_{opex}$, the signs of the effects of these components on the rate of change of real opex are opposite to the signs of their effects on PFP.

In practice, the decomposition between the efficiency ‘catch-up’ effect and the rate of ‘frontier shift’ may not always be reliable. The ability to estimate historical productivity accurately can be constrained by incomplete data, uncertainty in how inputs and outputs are measured, and limited capacity to properly control for operating environment factors that are unrelated to productivity changes.

5.2 Results

These components, together with the change in Opex partial factor productivity (Opex PFP), were calculated for each time-varying model. Figures 5.1 to 5.30 present the average contribution of SEC, TC, OEFC and CEC to Opex PFP growth over the full 2006 to 2023 sample period. The results cover the nine shortlisted models presented in section 4.7, each estimated using the CD and TLGP functional forms.

Key observations include:

- *Scale efficiency change (SEC)* is small under all specifications, indicating that movements toward or away from the optimal operating scale make only a limited contribution to productivity change. CD estimates are uniformly positive, implying modest improvements in scale efficiency. TLGP results are more variable: LSE generally indicates a deterioration, BC95 produces mixed results across DNSPs, and Kumb90 continues to indicate improvement.
- *Technical change (TC)* is negative across all models, thus making a negative contribution to opex productivity. Given that genuine technical regress is unlikely, the estimated increases in opex due to TC may instead reflect the influence of unmodelled operating environment factors, such as increasing regulatory and other obligations.
- *Operating environment factor change (OEFC)* is positive across the models, indicating that a positive contribution to opex productivity (ie, lowering opex over time). This is consistent with the observed increase in undergrounding and its expected effect in reducing opex requirements.

- *Cost efficiency change (CEC)* differs considerably across DNSPs. Nevertheless, average CEC is positive for every model group, suggesting an overall improvement in cost efficiency over 2006–2023.
- *Opex PFP change (OPFPC)* is positive on average under every specification, indicating overall productivity improvement, although some individual DNSPs record declines.

Figure 5.1 Scale Efficiency Change (SEC) - Average 2006-2023 – BC95 Models - CD

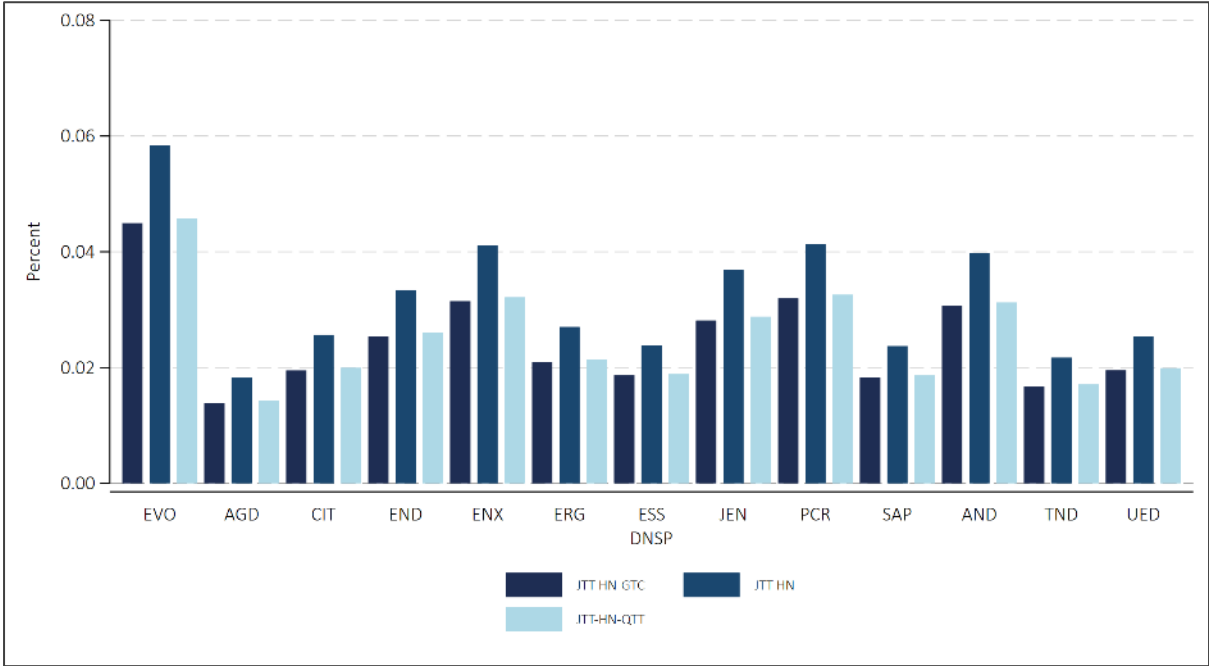


Figure 5.2 Scale Efficiency Change (SEC) - Average 2006-2023 – BC95 Models - TLGP

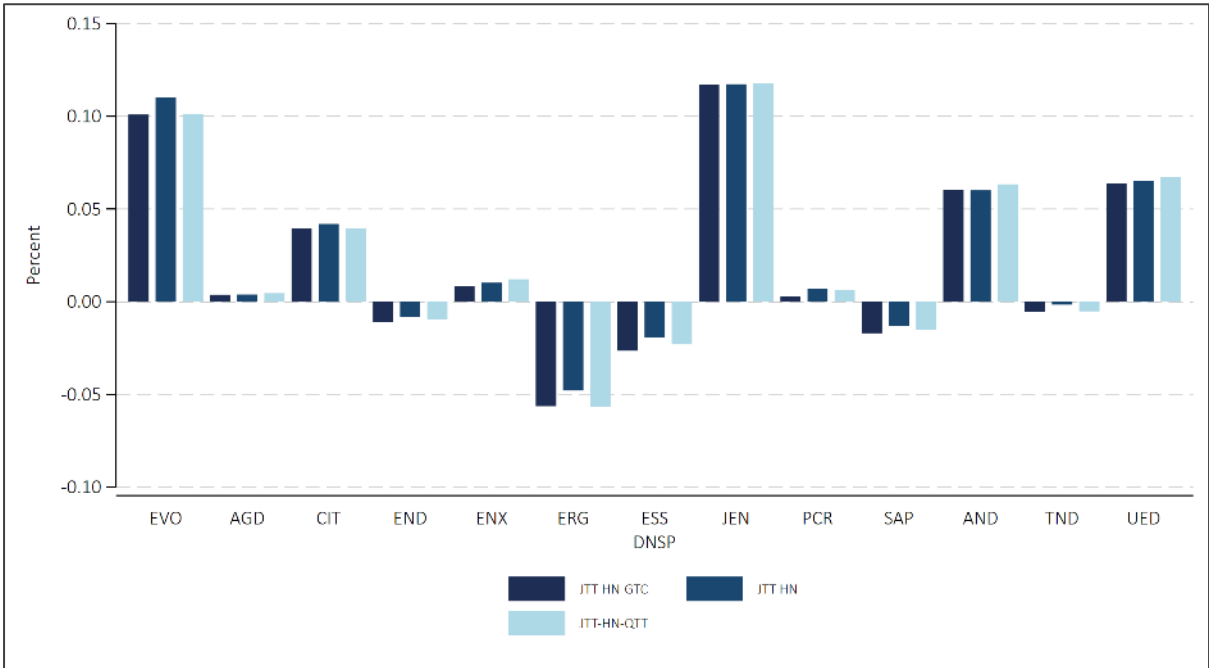


Figure 5.3 Scale Efficiency Change (SEC) - Average 2006-2023 – Kumb90 Models - CD

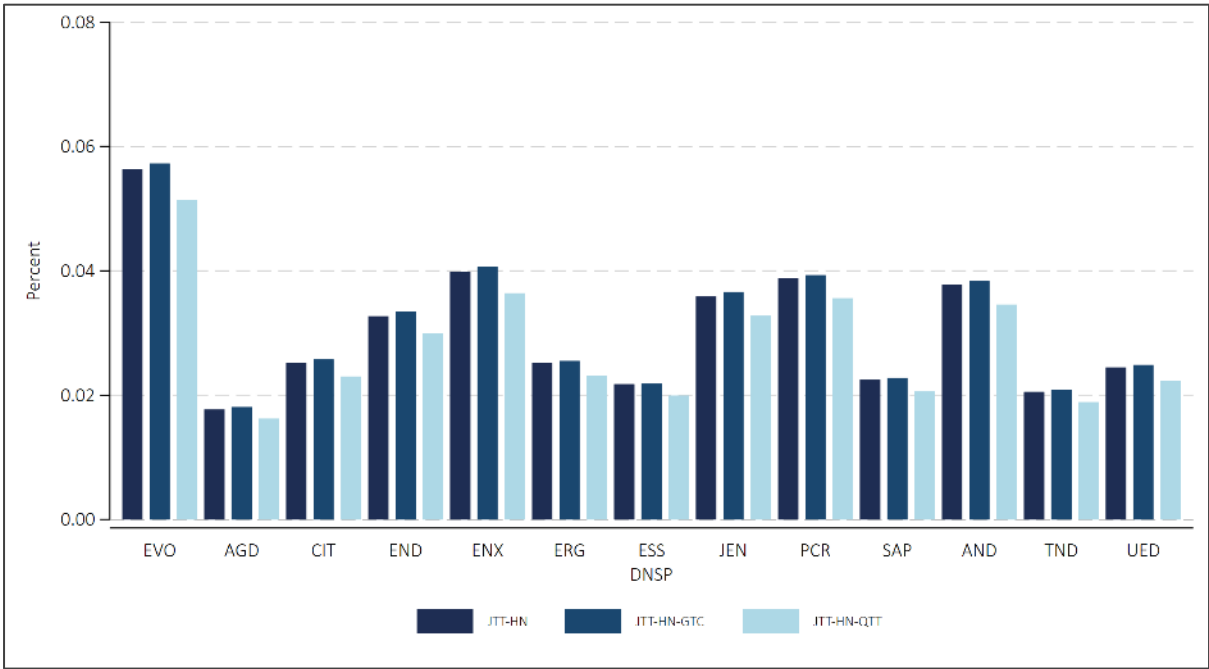


Figure 5.4 Scale Efficiency Change (SEC) - Average 2006-2023 – Kumb90 Models – TLGP

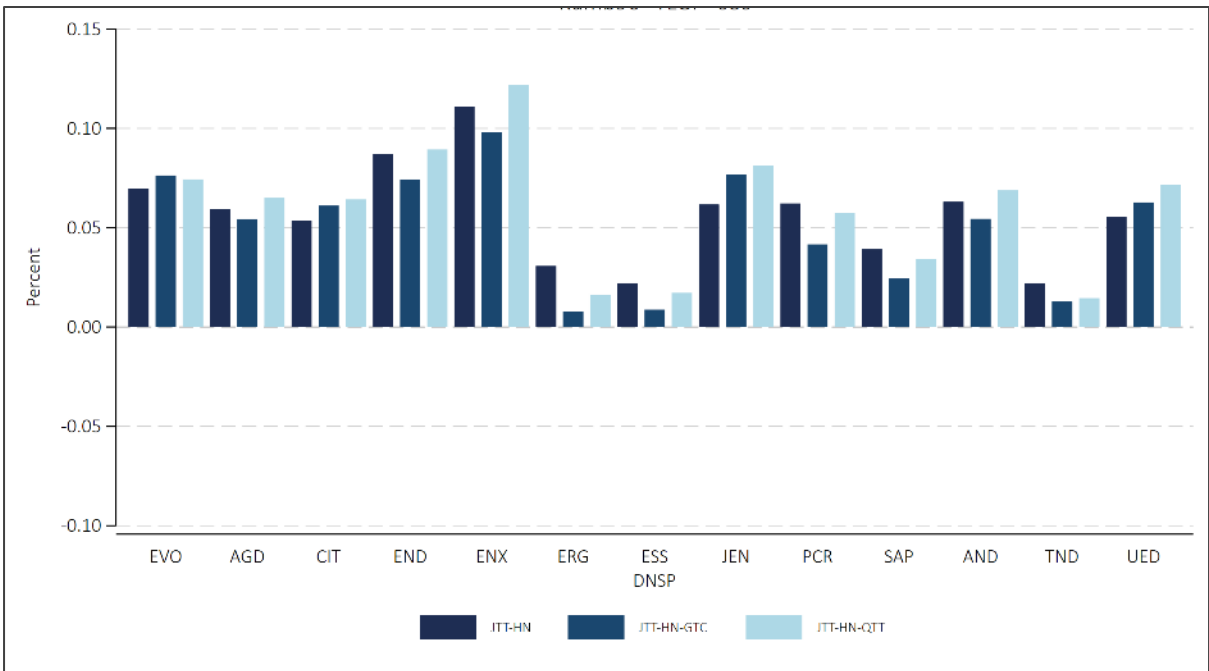


Figure 5.5 Scale Efficiency Change (SEC) - Average 2006-2023 – LSE Models - CD

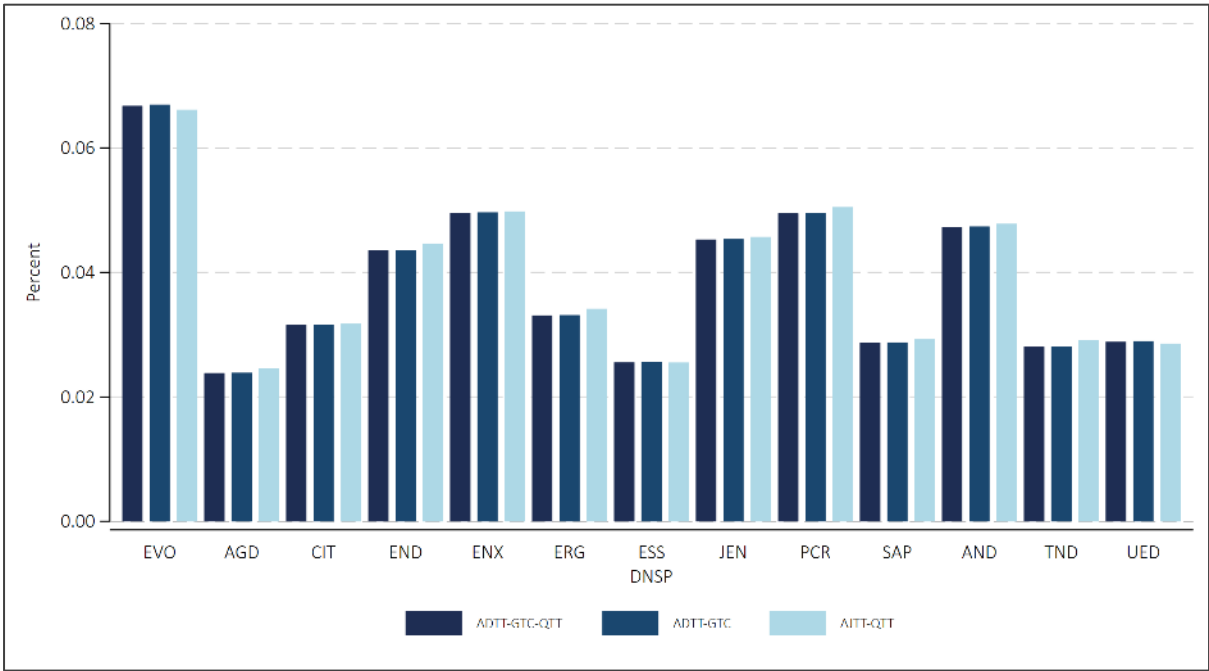


Figure 5.6 Scale Efficiency Change (SEC) - Average 2006-2023 – LSE Models – TLGP

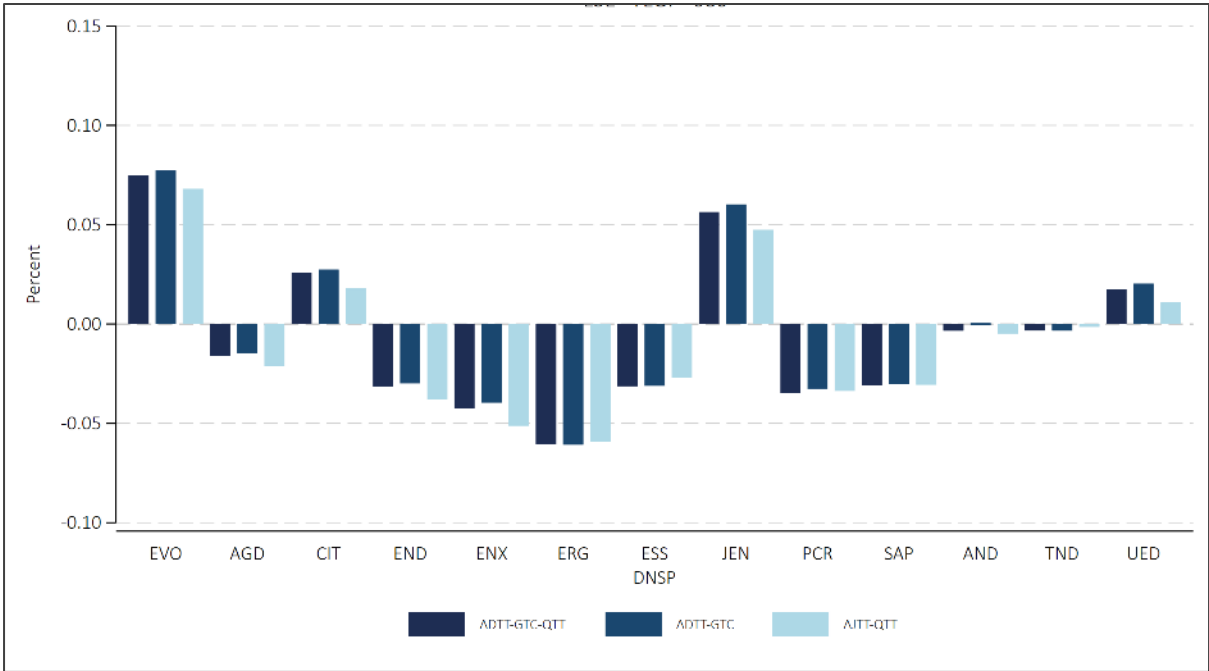


Figure 5.7 Technical Change (TC) - Average 2006-2023 – BC95 Models - CD

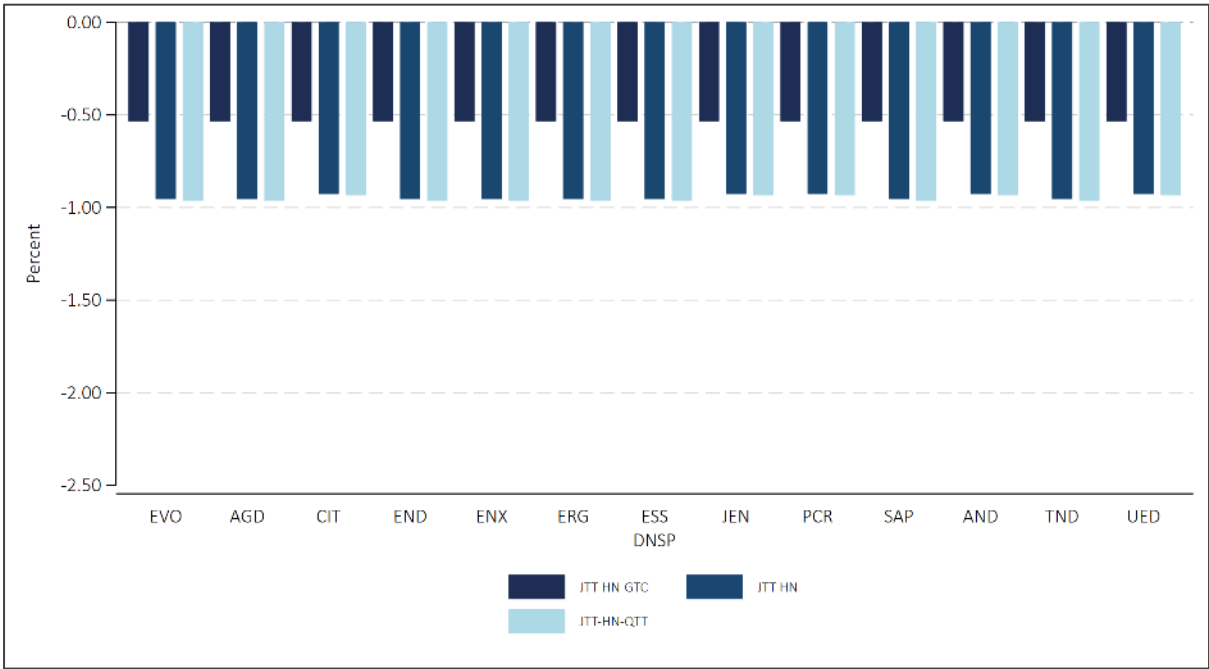


Figure 5.8 Technical Change (TC) - Average 2006-2023 – BC95 Models - TLGP

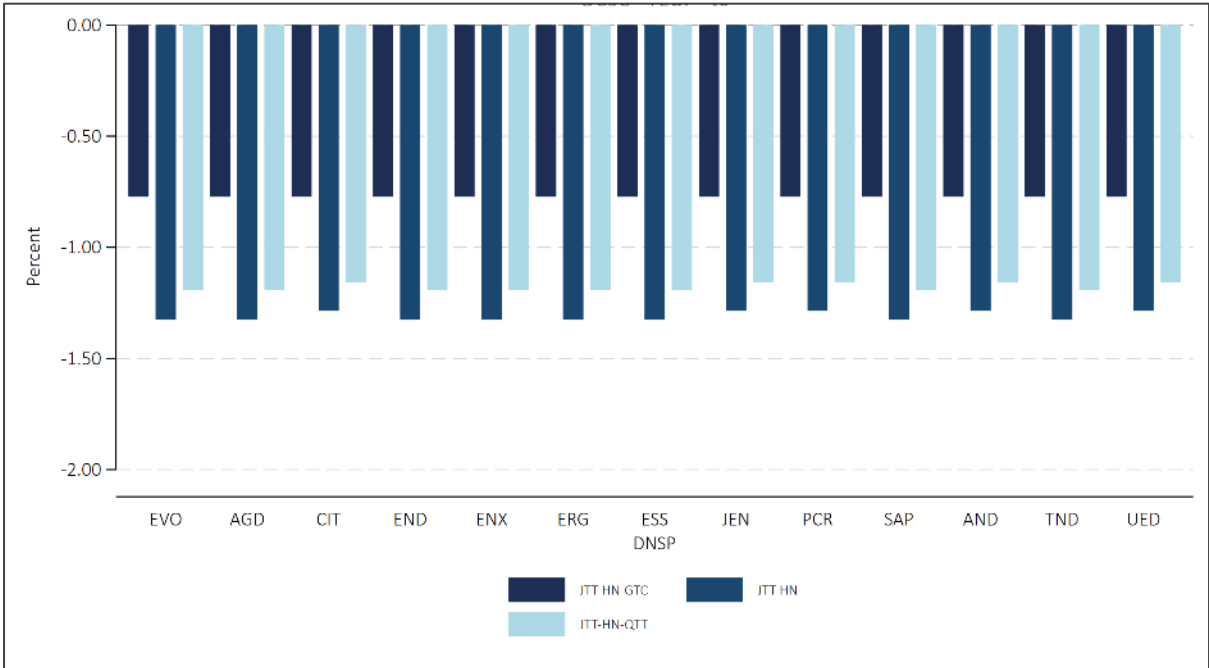


Figure 5.9 Technical Change (TC) - Average 2006-2023 – Kumb90 Models - CD

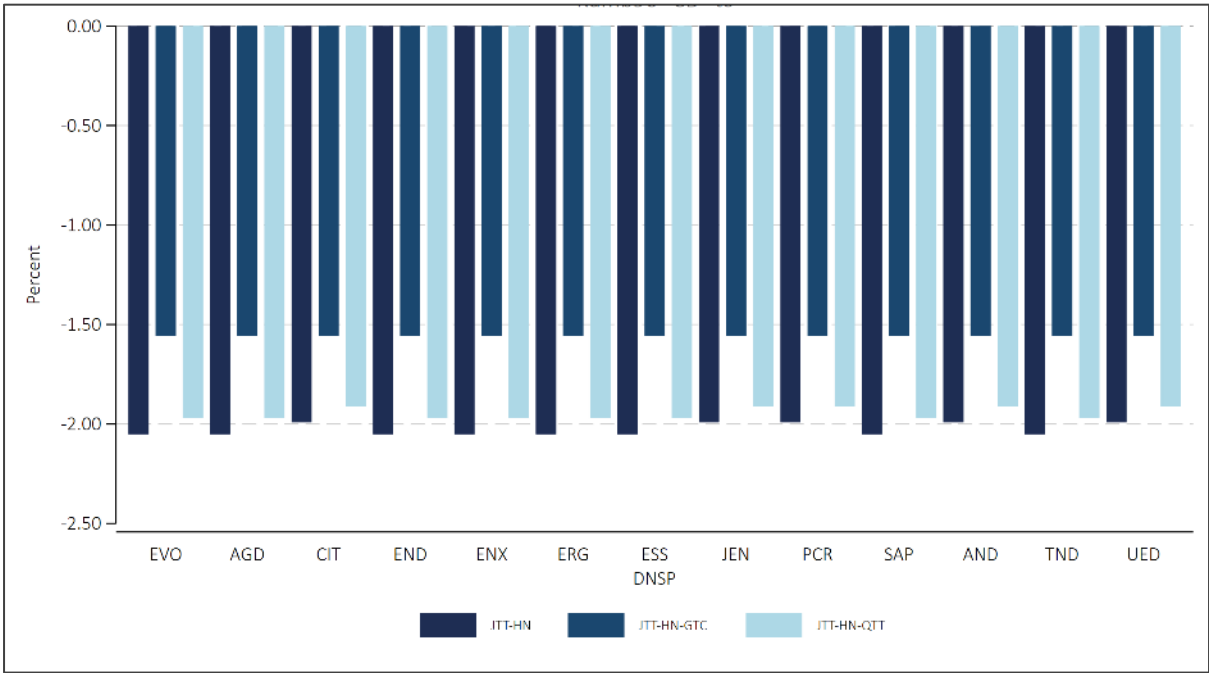


Figure 5.10 Technical Change (TC) - Average 2006-2023 – Kumb90 Models - TLGP

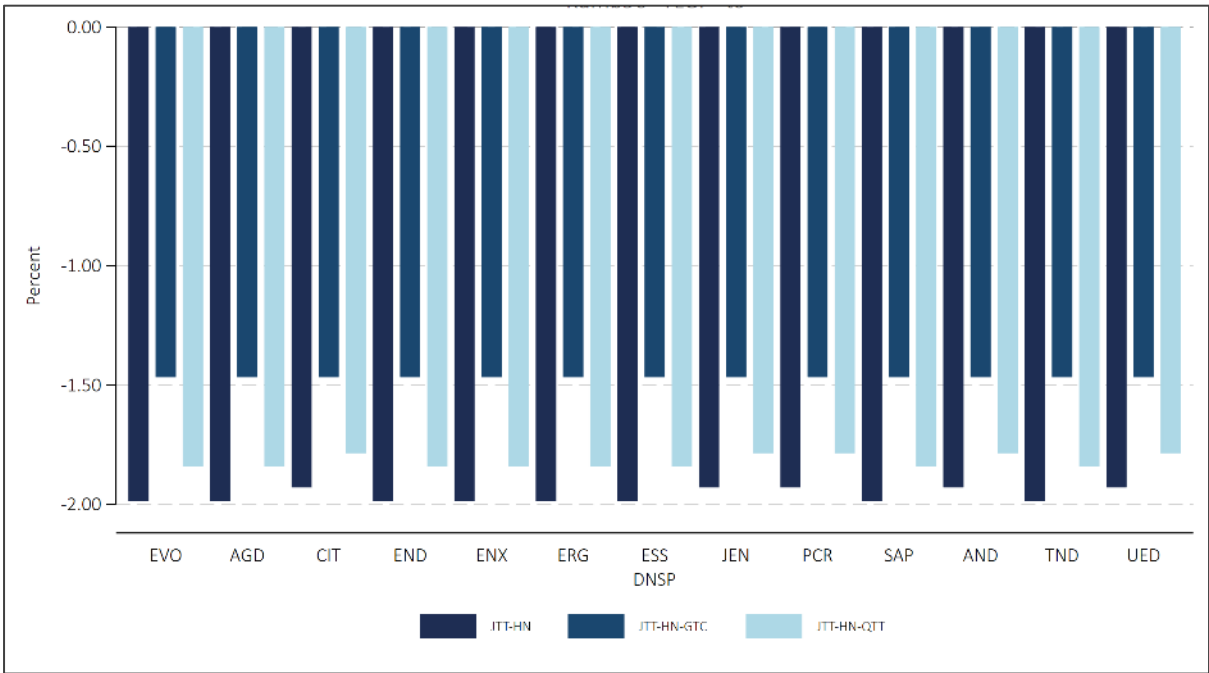


Figure 5.11 Technical Change (TC) - Average 2006-2023 – LSE Models - CD

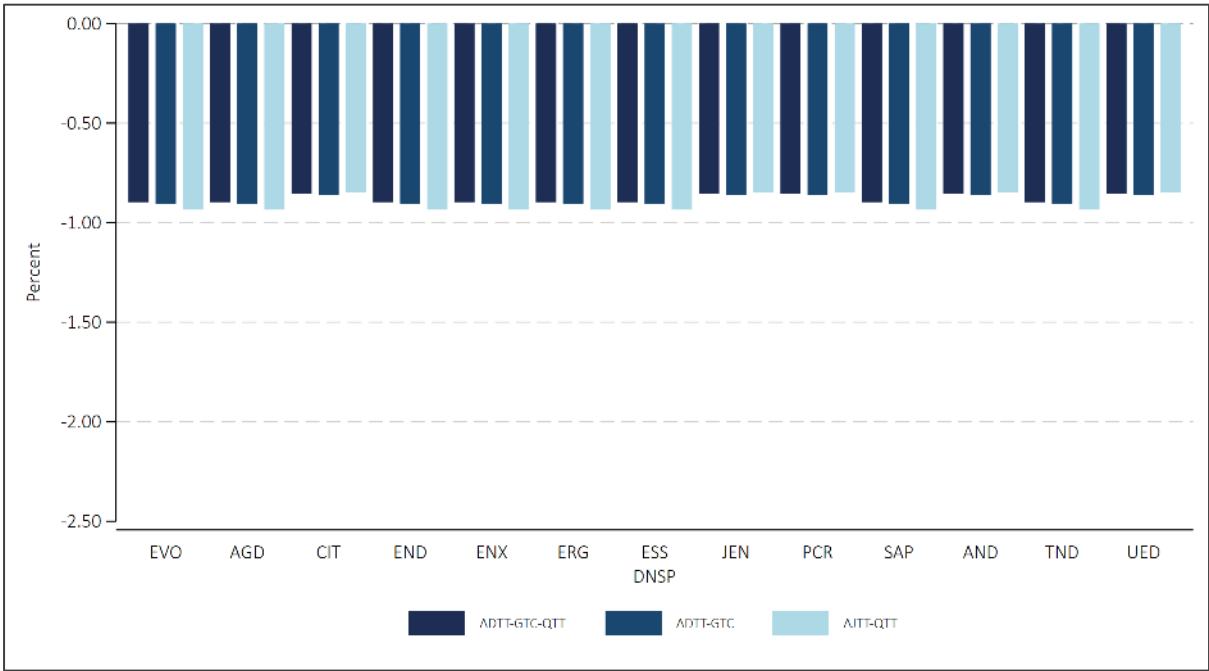


Figure 5.12 Technical Change (TC) - Average 2006-2023 – LSE Models - TLGP

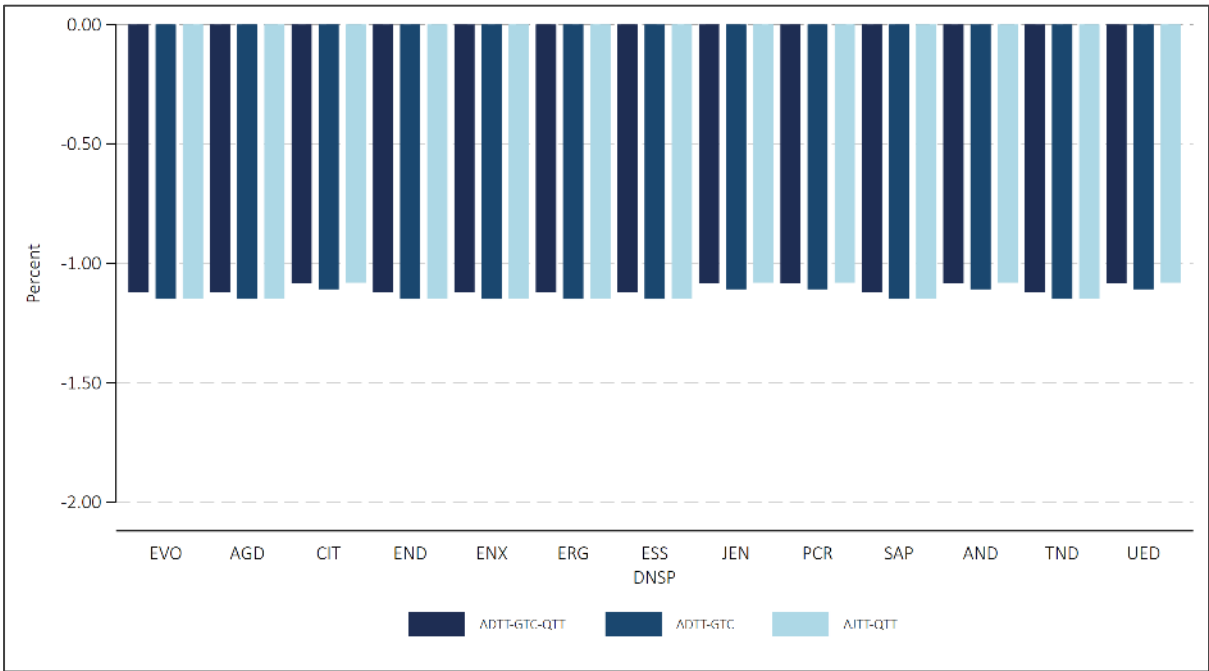


Figure 5.13 Effect of Change in the UG variable (OEFC) - Average 2006-2023 – BC95 Models - CD

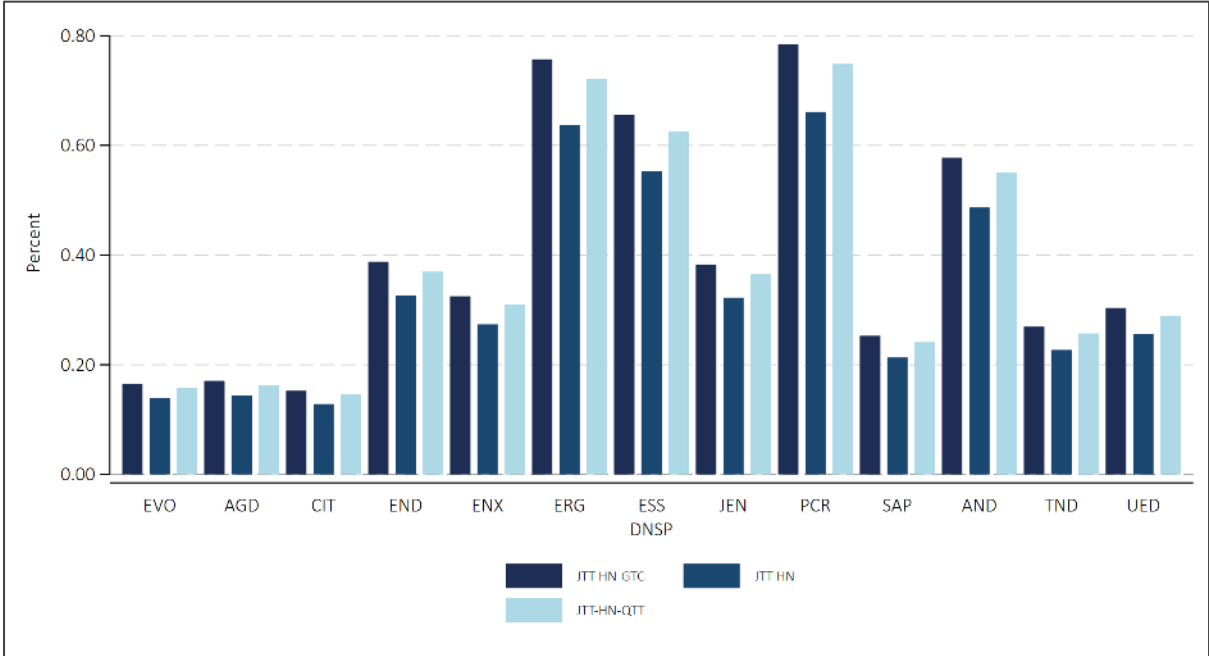


Figure 5.14 Effect of Change in the UG variable (OEFC) - Average 2006-2023 – BC95 Models - TLGP

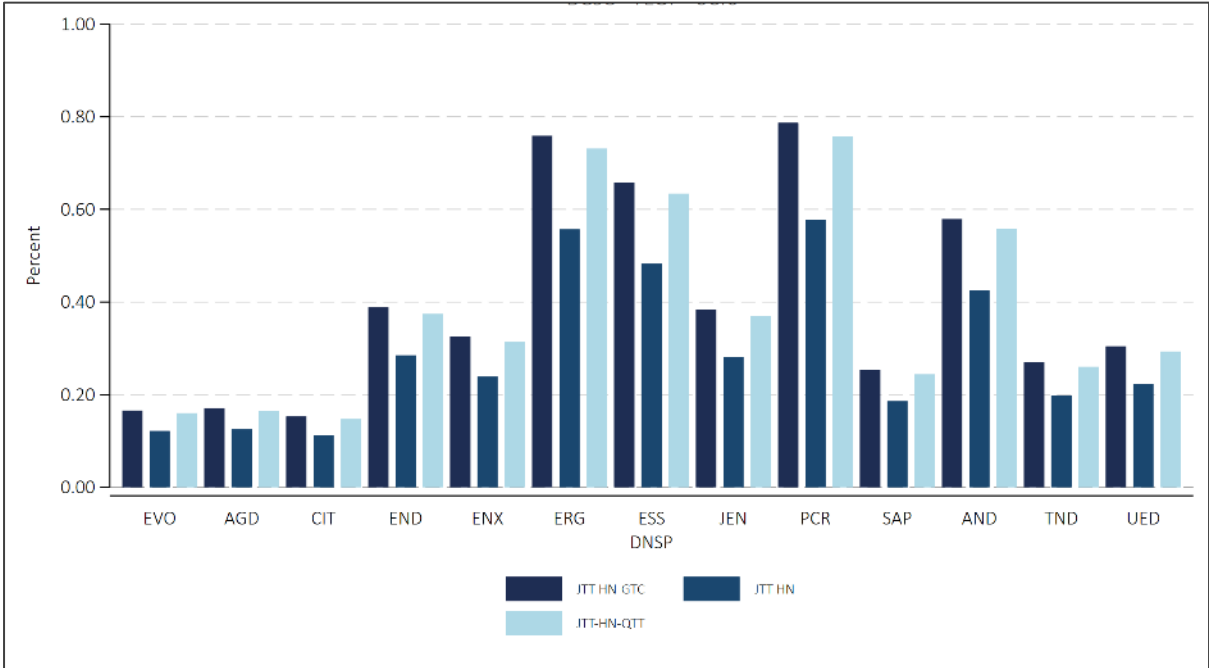


Figure 5.15 Effect of Change in the UG variable (OEFC) - Average 2006-2023 – Kumb90 Models - CD

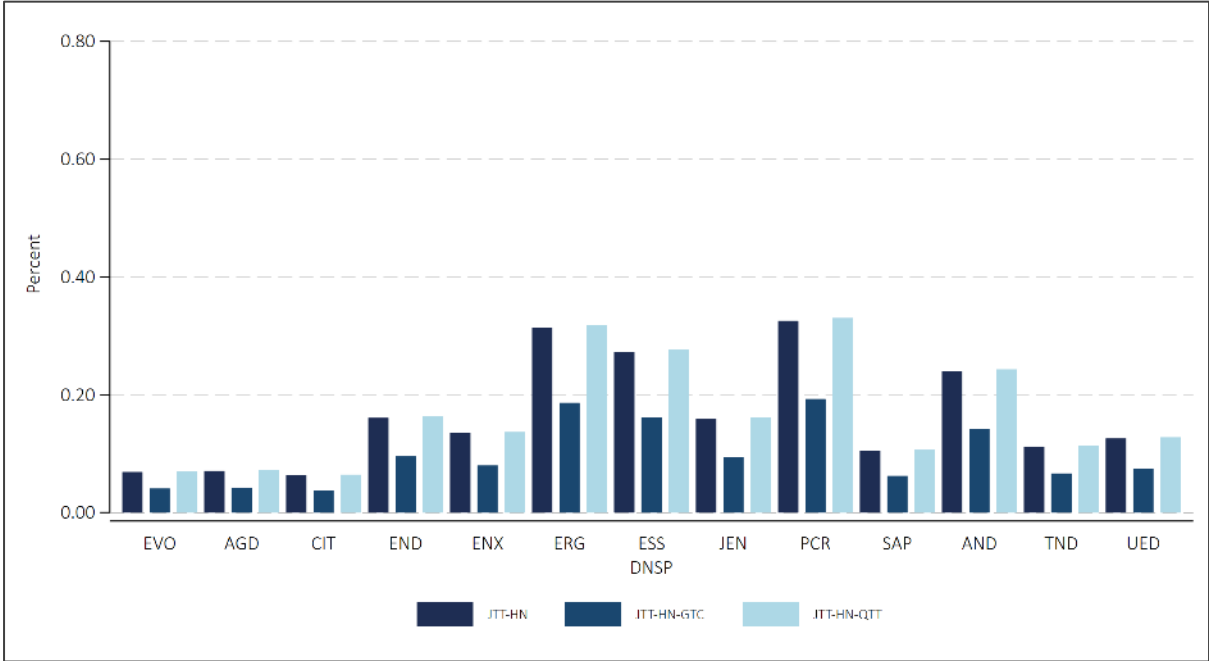


Figure 5.16 Effect of Change in the UG variable (OEFC) - Average 2006-2023 – Kumb90 Models - TLGP

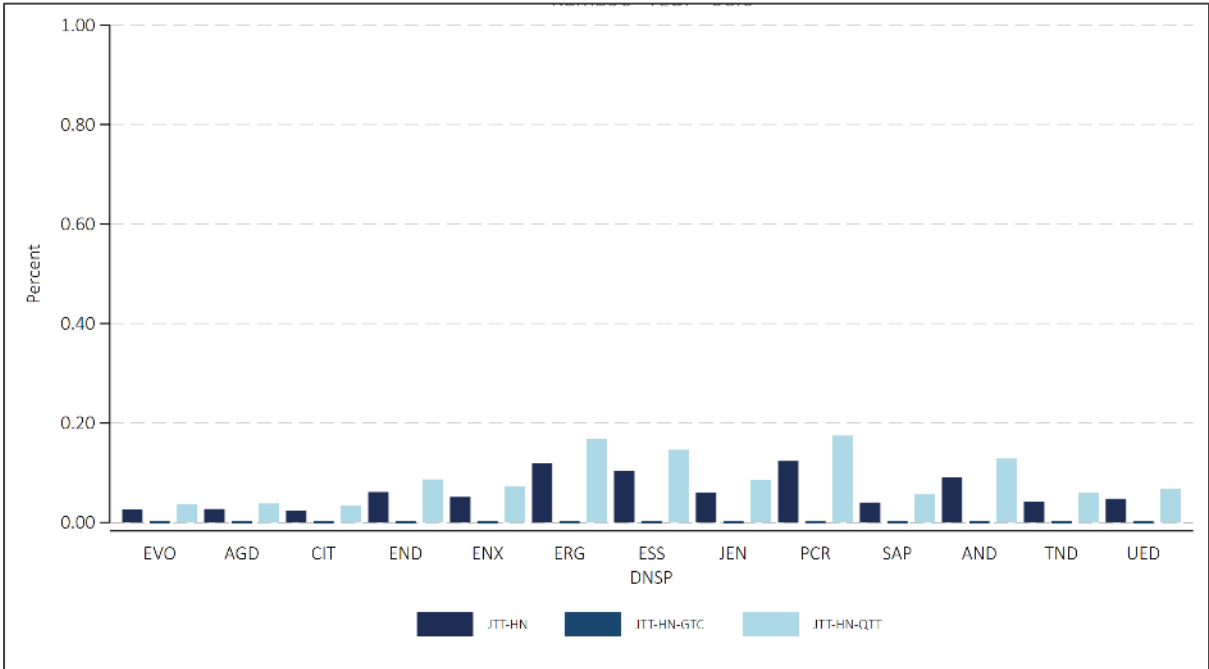


Figure 5.17 Effect of Change in the UG variable (OEFC) - Average 2006-2023 – LSE Models - CD

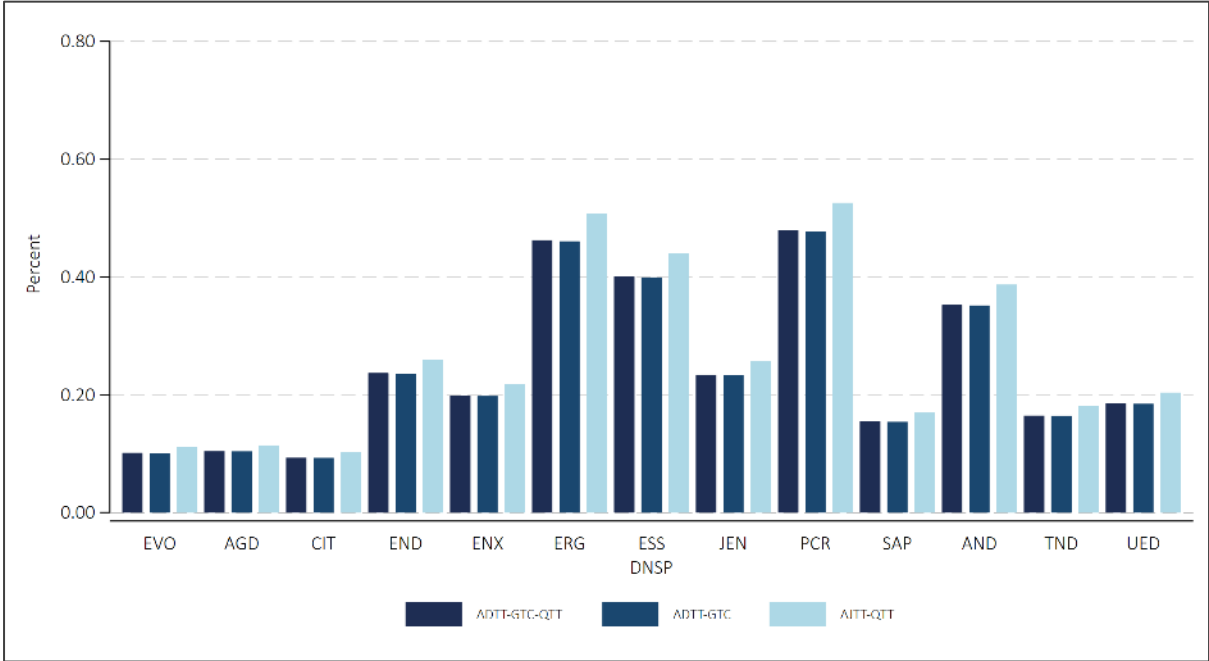


Figure 5.18 Effect of Change in the UG variable (OEFC) - Average 2006-2023 – LSE Models – TLGP

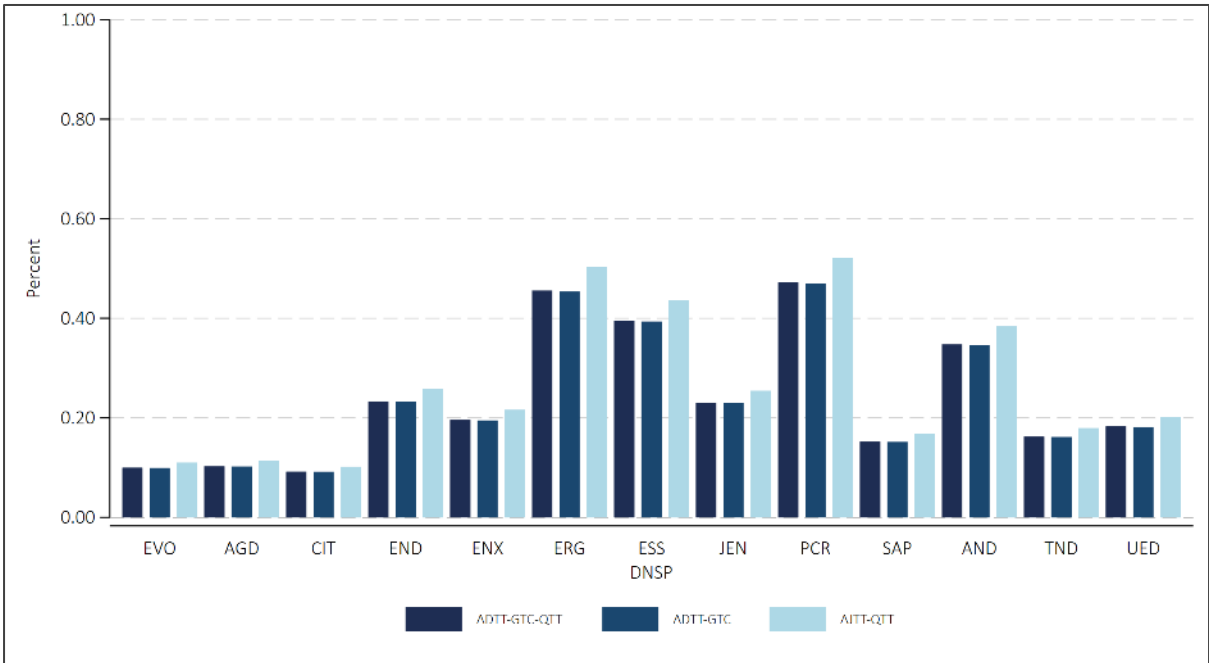


Figure 5.19 Cost Efficiency Change (CEC) - Average 2006-2023 – BC95 Models - CD

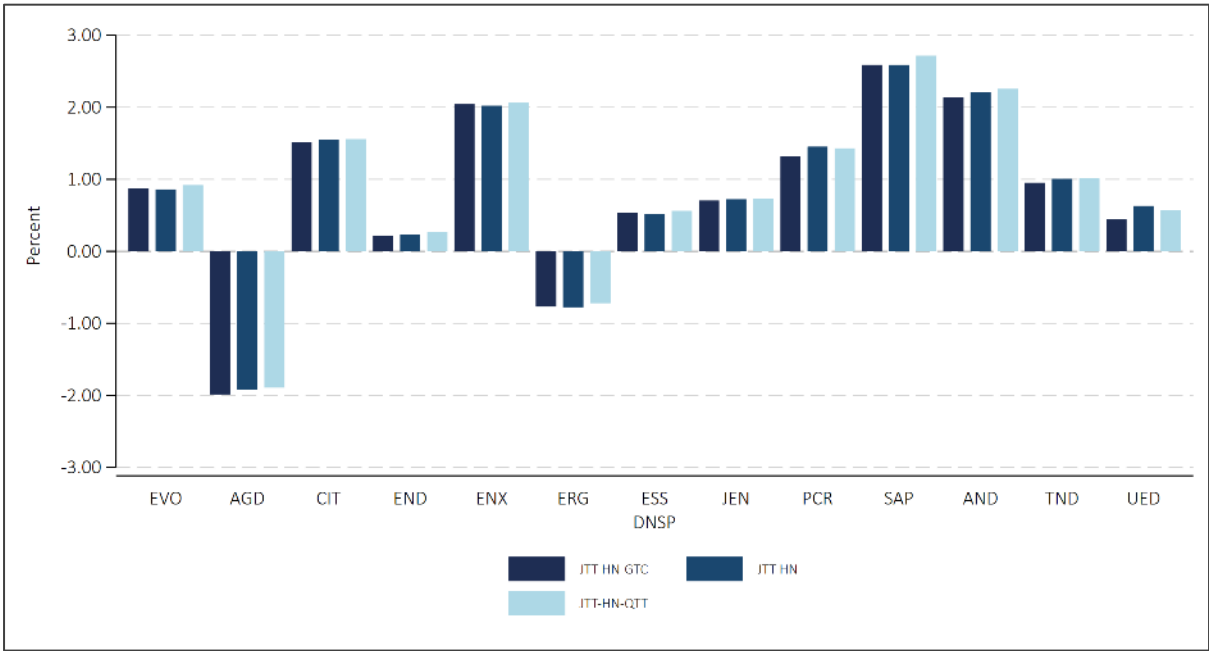


Figure 5.20 Cost Efficiency Change (CEC) - Average 2006-2023 – BC95 Models - TLGP

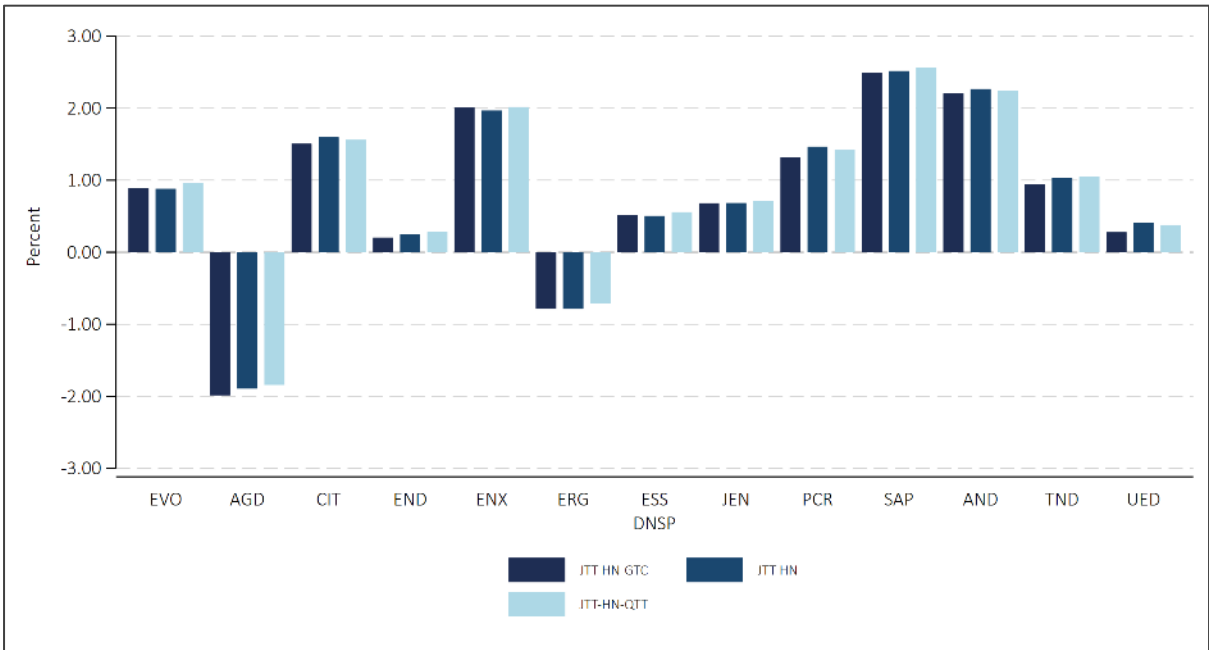


Figure 5.21 Cost Efficiency Change (CEC) - Average 2006-2023 – Kumb90 Models - CD

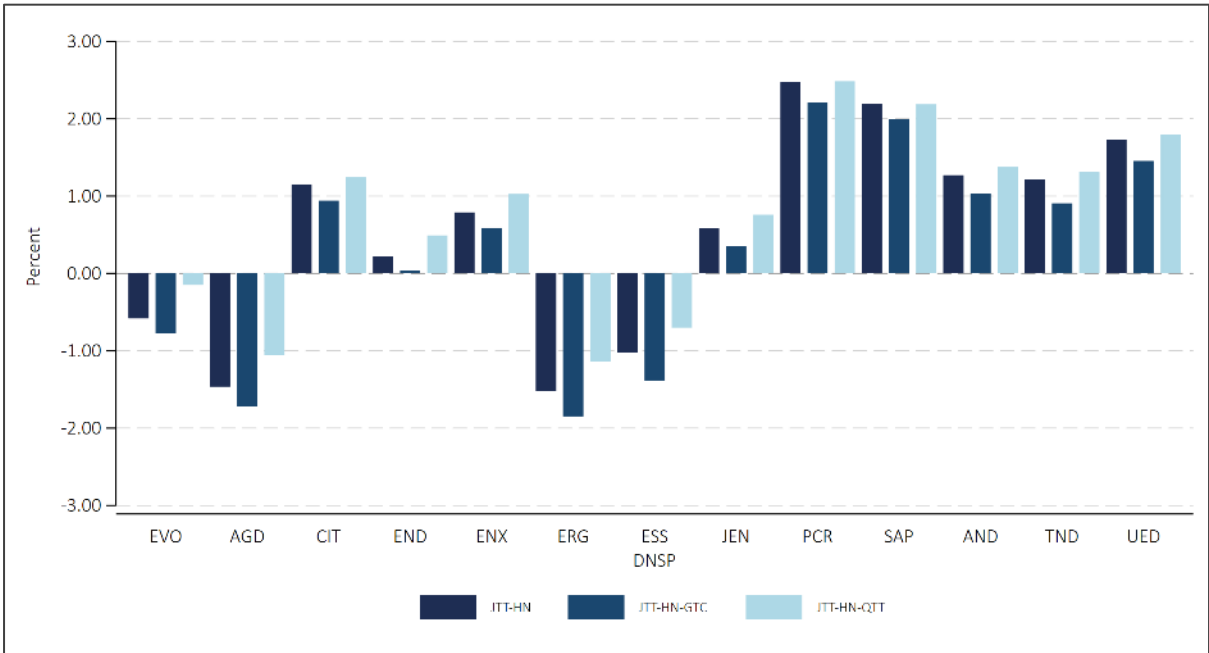


Figure 5.22 Cost Efficiency Change (CEC) - Average 2006-2023 – Kumb90 Models - TLGP

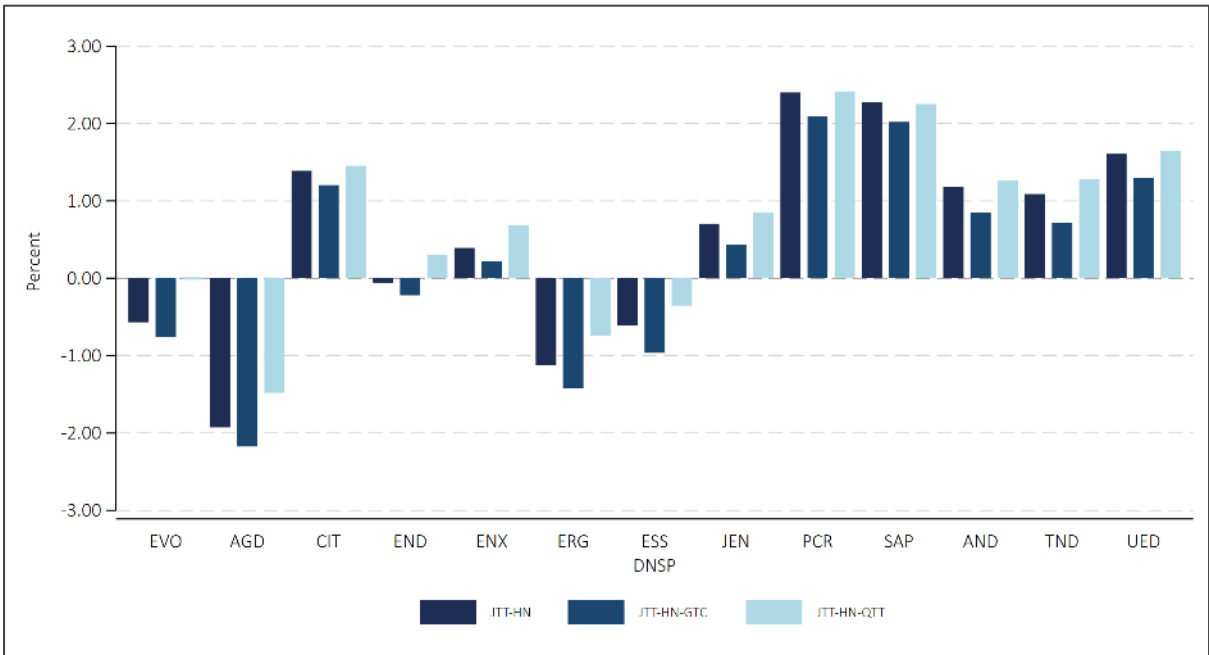


Figure 5.23 Cost Efficiency Change (CEC) - Average 2006-2023 – LSE Models - CD

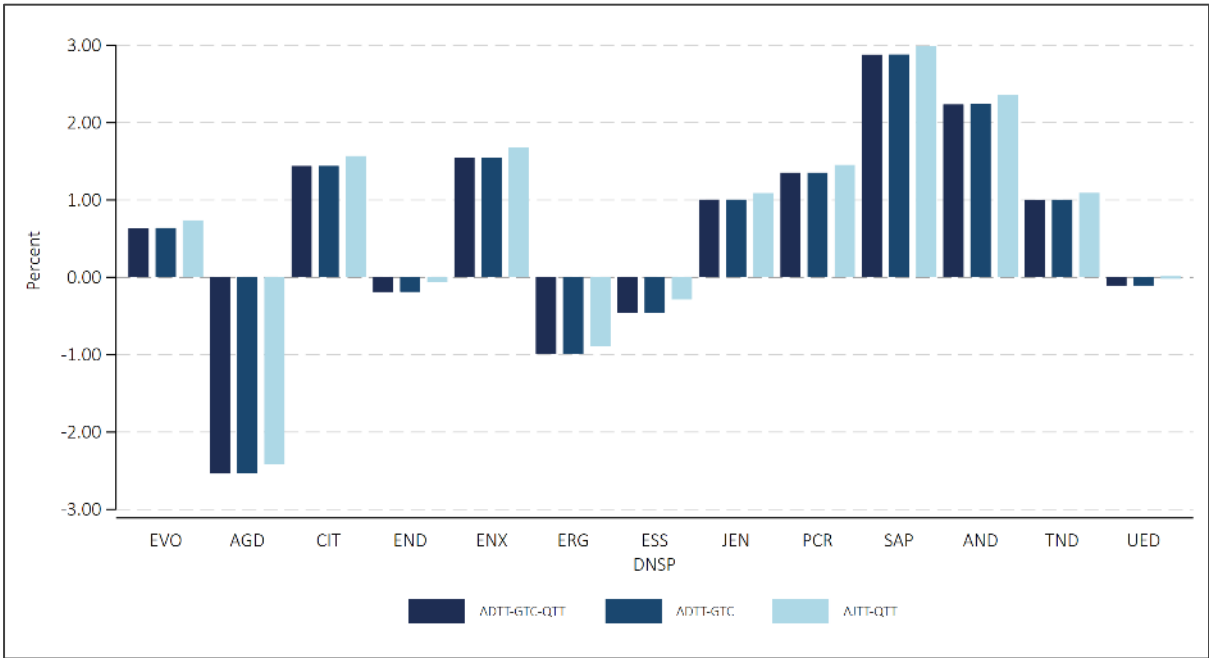


Figure 5.24 Cost Efficiency Change (CEC) - Average 2006-2023 – LSE Models - TLGP

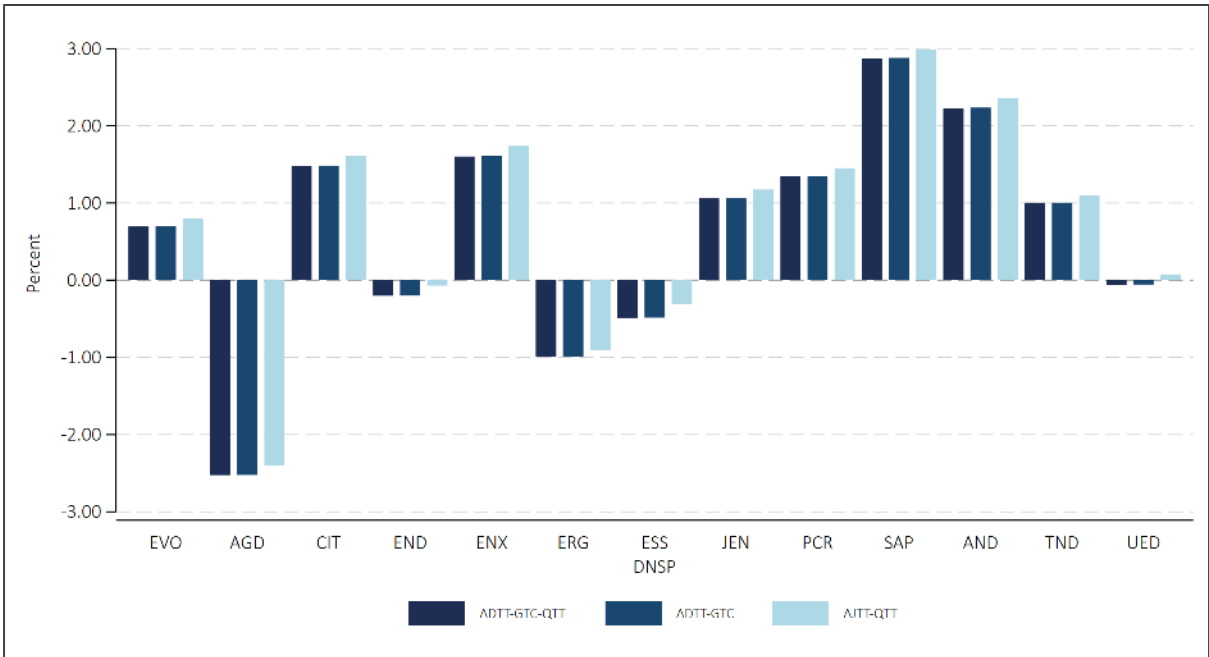


Figure 5.25 Opex PFP change - Average 2006-2023 – BC95 Models - CD

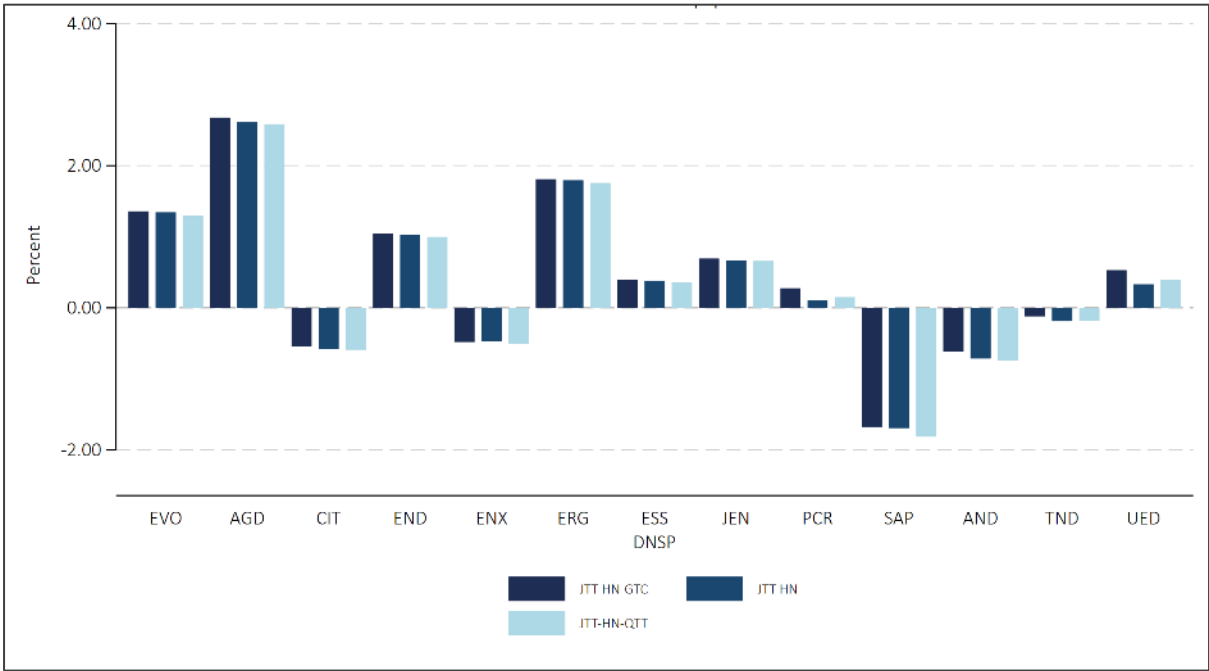


Figure 5.26 Opex PFP change - Average 2006-2023 – BC95 Models - TLGP

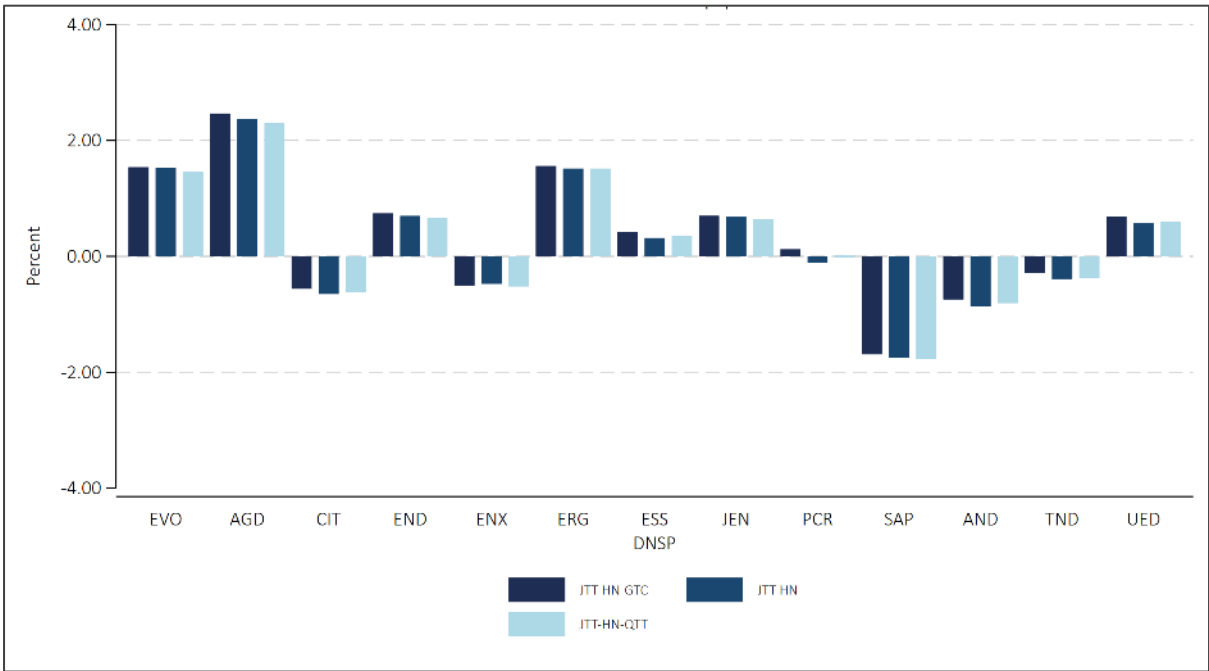


Figure 5.27 Opex PFP change - Average 2006-2023 – Kumb90 Models - CD

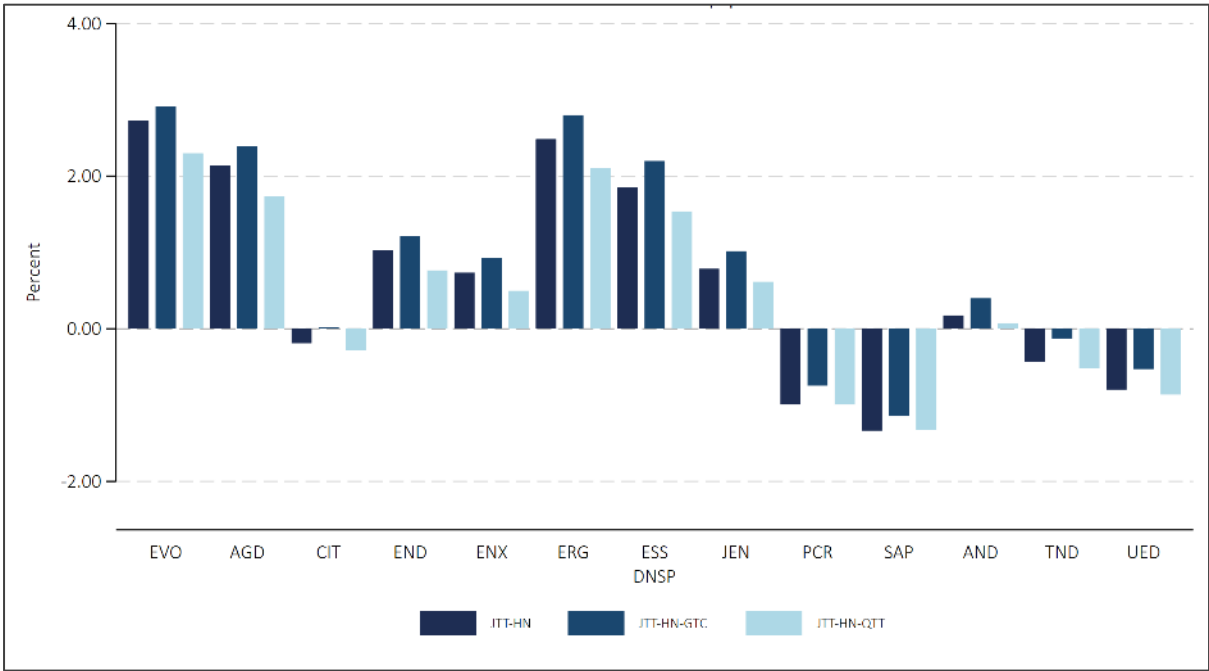


Figure 5.28 Opex PFP change - Average 2006-2023 – Kumb90 Models - TLGP

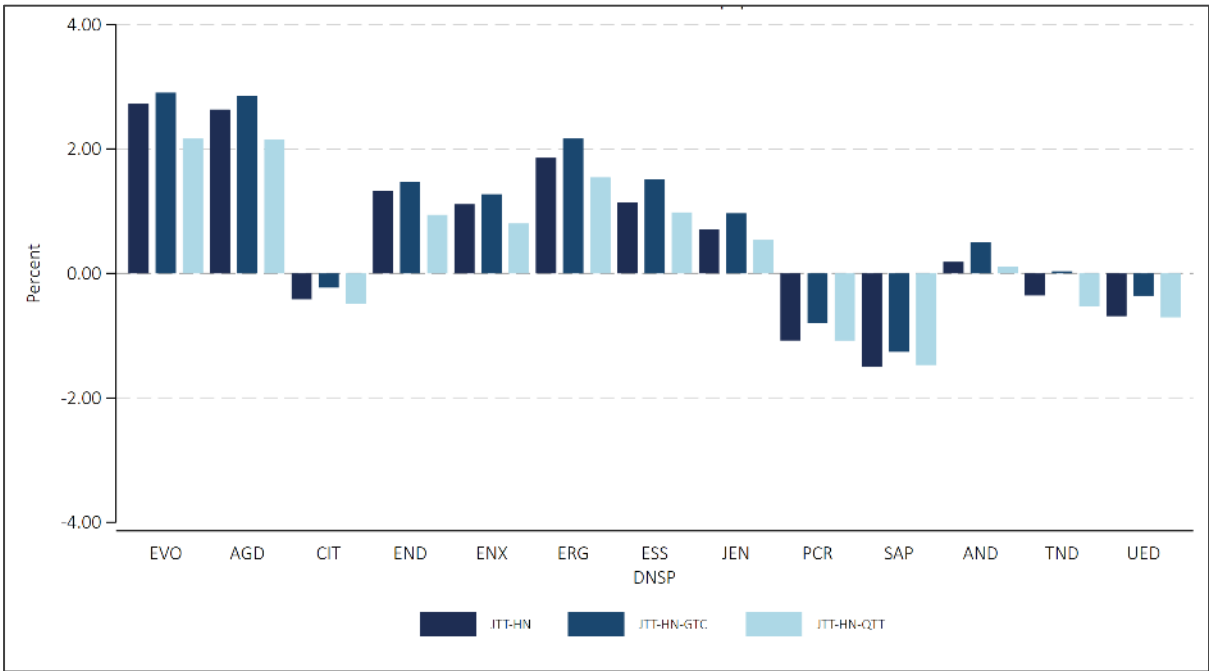


Figure 5.29 Opex PFP change - Average 2006-2023 – LSE Models - CD

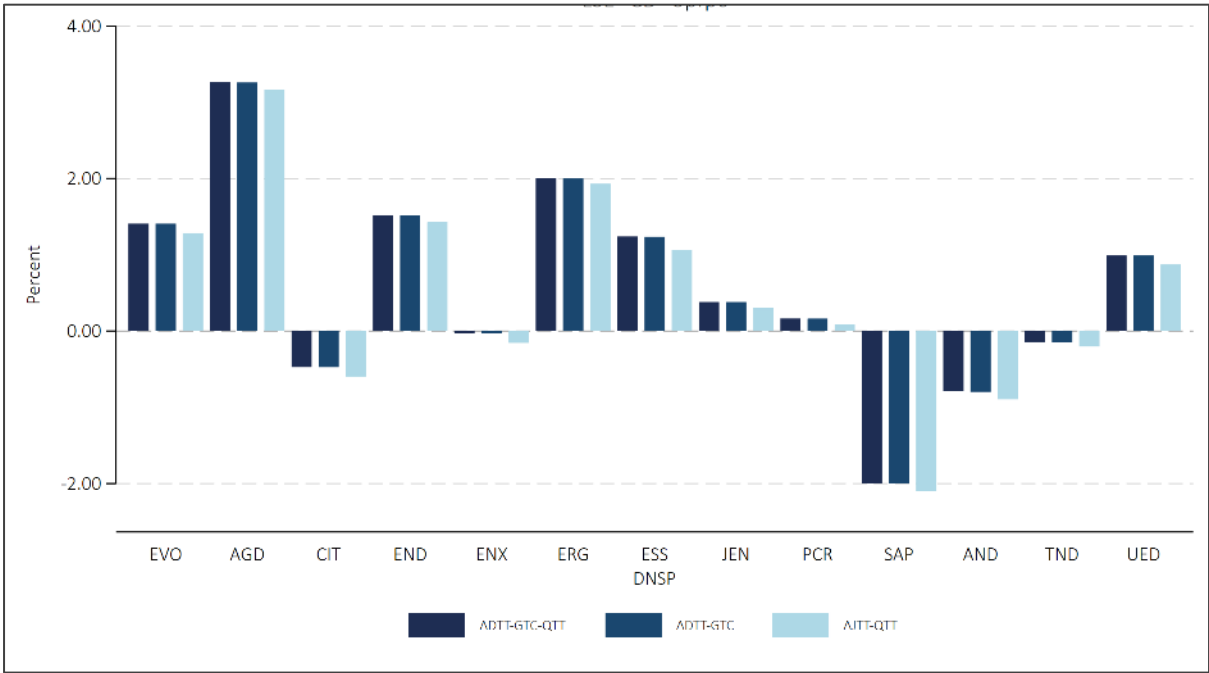
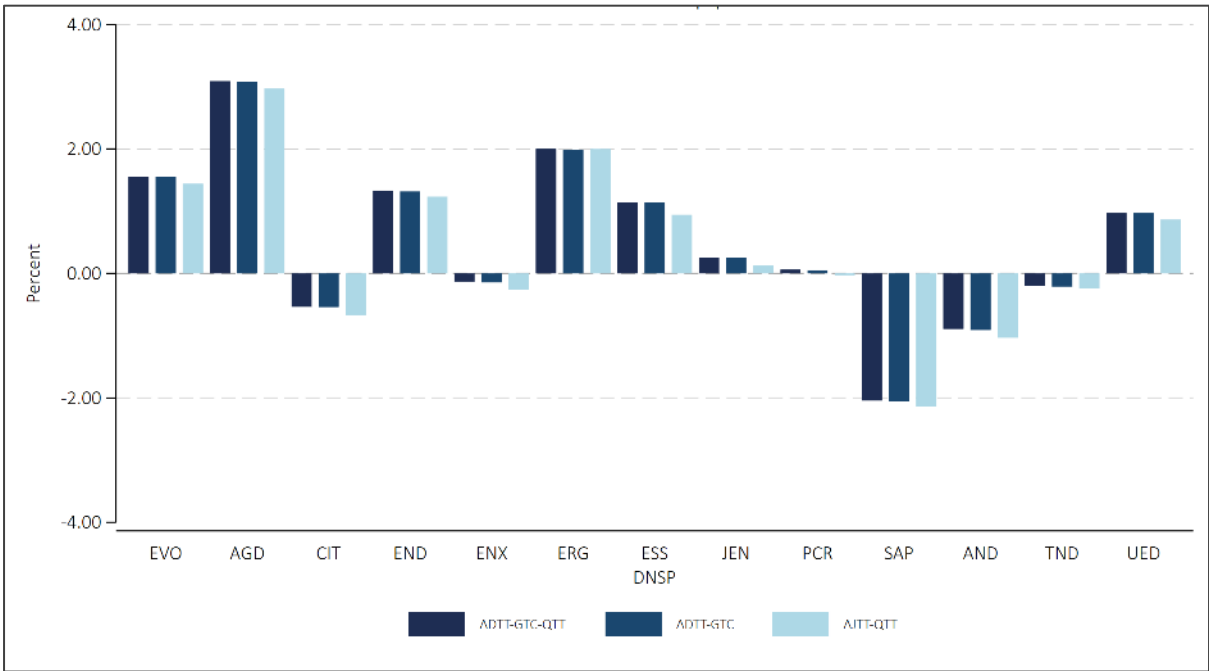


Figure 5.30 Opex PFP change - Average 2006-2023 – LSE Models - TLGP



6 Benchmarking Roll-forward Model (BRFM) Issues

6.1 Options and Submissions

6.1.1 Key options

If the opex econometric models incorporate time-varying efficiency, the available options for calculating base-year efficient opex include:

- (a) Averaging the estimated efficiency scores over the sample period and applying the BRFM as before.
- (b) Using the efficiency estimated at the end of the sample period (if the base year is the same as the last year of the sample period) or, if the base year is after the last year of the sample, using the BRFM to bridge the shorter gap between the last year of the sample period and the base year. The latter would most often be the case. Therefore, this option is henceforth assumed to involve applying the BRFM over a single year from the last sample year to the base year.
- (c) Both options could be combined. This would increase the number of estimates of efficient base-year opex. Since the averaging of several valid base-year efficient opex estimates should improve the reliability of the resulting (averaged) estimate, having more estimates to average over may be an advantage.²²

Appendix C sets out formally how each of option (a) and (b) are calculated. Most submissions supported option (a), though views were not unanimous. This section discusses the various issues that remain to be considered under each option.

6.1.2 Specific Issues Raised in Submissions

AusNet is concerned that option (a) “may introduce systematic bias. For DNSPs exhibiting improving efficiency over time, averaging efficiency scores across the sample period may mask recent performance improvements which are most relevant for regulatory decision-making. Conversely, DNSPs whose efficiency has deteriorated over time may be inadvertently advantaged, as sample-average efficiency scores may be overstated by earlier, less relevant performance levels.” Option (b) would allow the efficiency benchmark to reflect the most recent efficiency performance. This is discussed in section 6.2.4 on the choice of benchmark comparison point.

²² Consider a set of N independent, randomly distributed parameter estimates: $p_1, \dots, p_N$, with a set of corresponding standard errors: $s_1, \dots, s_N$. The mean parameter estimate is: $\bar{p} = (\sum_{i=1}^N p_i)/N$. The standard error of $\bar{p}$ is: $\bar{s} = \sqrt{(\sum_{i=1}^N s_i^2)/N^2}$. Furthermore, if all s_i have the same value, then $\bar{s} = \sqrt{s_i^2/N}$, which is decreasing as N increases.

Frontier Economics (2026) noted that one advantage of option (b) over option (a) is that it does not rely as much on the estimated output weights, which it regards as unreliable. This is because it is only necessary to roll forward the efficient opex from the end of the sample period to the base year, typically one year. Under option (a), when the full sample period is used, the roll forward will be over about 10 years based on the data sample available in 2026. On the other hand, Frontier Economics (2026, 74) suggests there are “a number of complex issues that the AER would need to take into account, and address, before changing its current approach of rolling forward from the middle of the benchmarking period.” Several criticisms were made of option (b):

- (1) The efficiency score estimate in the last period of the sample is likely to be more uncertain than the average of all the efficiency scores over the sample period. It would increase the potential for inaccuracy if the assessment of efficient opex is based on the efficiency score in the last year of the sample. This is because averaging of *actual* opex (expressed in constant dollars), against which the efficient opex is compared, effectively removes the effect of the random disturbance, which is assumed to impact opex in the econometric model.
- (2) Option (b) would require that the *ex post* adjustments for operating environment factors (OEFs) would need to be specific to a single year, namely the last year of the sample, rather than an average effect over the whole sample period. This raises the question of whether it would be practical to do so. This issue is discussed in section 6.2.3 on OEF calculation.
- (3) Specific questions are raised relating to the effects of option (b) on businesses’ incentives to make cost efficiencies. Some of these questions relate to the interaction between base year efficiency adjustments and the Efficiency Benefit Sharing Scheme (EBSS). This issue is discussed in section 6.3.

Frontier Economics recommends that the AER should (i) retain its current BRFM approach, and (ii) before making any changes to its BRFM approach, the AER should clarify these issues and consult with stakeholders on them. This section discusses the options for applying the BRFM.

6.2 Discussion of Selected Issues

6.2.1 Method of Averaging Efficiency Scores

In option (a), the averaging of each DNSP’s efficiency scores across the sample period will require a choice as to whether to use a simple average or a weighted average (with real opex in each year as the weight). A simple average would give equal weight to each year’s efficiency score, whereas a weighted average would give greater weight to years in which the DNSP

incurred higher real opex. This was not discussed in either the Quantonomics Phase 2 report or the submissions, but may need to be considered.

6.2.2 Frontier Economics: Inaccuracy of Efficiency Score Estimates in Sample End-Year

There are two observations regarding Frontier Economics' argument (1). Firstly, while the argument is valid, it needs to be qualified. It does not account for the fact that the BRFM is generally applied to several valid models. The estimated base-year efficient opex will be an average of this set of estimates. This averaging tends to iron out inaccuracies in individual estimates. Hence, even option (b) would have some averaging of different estimates, and option (c) more so. It may also be relevant that the actual base year opex, which the AER compares to the estimated efficiency base year opex, generally has the impact of abnormal one-off items removed by the AER. Hence, at least part of the random effects on opex which Frontier Economics refers to, may be removed through this procedure.

Secondly, Frontier Economics does not maintain this argument for all models. It is argued that basing an efficiency assessment on the last year of the sample would not be feasible with either the Kumb90 or LSE models because their restrictive time patterns for inefficiency are likely to result in inaccurate efficiency estimates in the final year of the sample. Because these models impose relatively restrictive patterns on how efficiency can change over time, a change in a DNSP's actual efficiency in the final year may not be fully reflected in its estimated efficiency for that year for these models. However, for the BC95 models, the efficiency assessment could be based on the final year of the sample, because its flexibility means that in the final year of the sample, there will be an observation-specific estimate of efficiency.

We consider these issues by applying the BRFM using the alternative estimated models and using both the period-average method (ie, option (a)) and the period-end method (option (b)) in section 6.4.

6.2.3 Calculating OEFs for a Single Year

The AER's approach to adjusting for material OEFs not included in the econometric models involves:

- Quantifying the effects of each OEF on each DNSP's *efficient* cost, on average, over the sample period.
- Expressing the effect of the OEF on opex in percentage terms relative to its effect on the opex of the reference DNSP (ie, the average of the group of efficient DNSPs).
- Aggregating these percentage effects over all OEFs.
- Adjusting the (period-average) efficiency target of each *inefficient* DNSP for the aggregate percentage effect of the OEFs.

Frontier Economics points out that when using the efficiency estimated at the end of the sample period and using the BRFM to bridge to the base year, the calculation of the OEF adjustment should be different. There may be some potential impracticalities in applying the OEF adjustment to the final year, rather than at the sample midpoint, because:

- Some OEFs may change considerably from year to year, such as the effects of cyclones. It may be more difficult to estimate the effect of cyclones in an individual year than their effect in an average year. “However, in some circumstances, a sample average may be taken as an estimate of the OEF in the final year due to challenges in estimating the OEF” (Frontier Economics 2026, 73).
- At present, the group of reference DNSPs is based on the estimated efficiency scores, representing averages for the sample period. This raises the question of whether the reference DNSP should be based on the final year of the sample also. This is discussed in section 6.2.4, on Benchmarking Comparison Point (BCP).

Regarding the first of these two points, Frontier Economics ignores possible improvements in practicality. In its Evoenergy decision, the AER included some new OEFs, such as for differences in workers’ compensation costs between jurisdictions. Calculating the impact of OEFs in the period-end timeframe would avoid the need to back-cast their effect over the whole sample period. That is one example of an improvement in practicality.

6.2.4 Efficient Peers and the Benchmark Comparison Point

The AER originally formulated the BCP based on the marginal efficiency of a group of the five most efficient DNSPs at the mid-sample period. Based on the latest efficiency estimates in the 2025 benchmarking study (Quantonomics 2025a, 34), the BCP of 0.75 at the mid-sample point is similar to the average efficiency score of Australian DNSPs for the period 2006 to 2024 (average of four models). Six DNSPs had a higher efficiency score.

The BRFM is only applied to those DNSPs that have efficiencies, after adjusting for OEFs, that are below 0.75. However, as noted by AusNet, with time-varying efficiency measurement, the group of most-efficient DNSPs at the mid-sample period need not be the same as the group of most-efficient DNSPs at the end of the sample period. Hence, some DNSPs may be deemed efficient when they are no longer among the five most efficient DNSPs. By using the end-of-sample efficiency in the BRFM, it would be possible to define the group of most-efficient DNSPs based on their efficiency in a year or period close to the base year. This would arguably be more relevant for determining those that should be regarded as among the group of most-efficient peers, for whom the BRFM would not be applied.

The AER has stated that the BCP is intended to be equivalent, when the sample midpoint is used, to the efficiency score for the business at the bottom of the upper third of companies in the benchmark sample (ie, the 5th most efficient DNSP). At present, the application of the BCP at the sample mid-point, together with the trend to improved opex efficiency, may result in

most DNSPs being found efficient in the base year. If the end-of-sample efficiency is used in the BRFM, the BCP could be more dynamic as the data sample is extended. Since there has been a general increase in opex efficiency in the latter half of the sample period, the efficiency of the fifth most efficient DNSP at the end of the sample period may be higher than 0.75. The BCP could be adjusted to be equal to a dynamic standard, such as the marginal efficiency of the top third of DNSPs.

6.3 Efficiency Incentives

6.3.1 Efficiency Incentives and Asymmetry of Efficiency Adjustment

Frontier Economics argues that the AER's current approach to adjusting base-year opex to ensure it reflects efficient opex is asymmetric. Thus, DNSPs with efficiency below the target have their sample average opex adjusted downwards, whereas DNSPs with efficiency above the target do not have their sample average opex adjusted upwards. Frontier Economics states that this asymmetry is not problematic when the BRFM is based on the mid-sample efficiency because "DNSPs with an average efficiency score over the benchmarking period above 75% is deemed to be efficient" (Frontier Economics 2026, 69). The asymmetry is seen as a problem when the sample end-year efficiency score is used in the BRFM, although their argument relies on assuming the efficiency score is applied "without exempting DNSPs that are at least 75% efficient in the final year" (Frontier Economics 2026, 69).

As has been discussed in previous consultations, the threshold of 0.75 effectively allows for uncertainty or inaccuracy in the estimation of the precise efficiency of a DNSP. The AER can have a high degree of confidence that DNSPs with an efficiency score below this target (after accounting for OEF impact) are materially inefficient. And the efficiency score divided by this target provides a conservative measure of inefficiency. Asymmetry will arise due to the upper bound on efficiency scores being 1.0. In equation (C.1) in Appendix C, we show how this may be applied in the period-average model, although at present, the AER does not calculate the BRFM for DNSPs that have efficiency scores above the BCP. When the period-end BRFM is used, the application of the constraint that the maximum efficiency score is 1.0 is applied at the last year of the sample (see section C.2 of Appendix C).

Regarding Frontier Economics' argument, firstly, as noted above, it assumes that when the BRFM is applied from the last year of the sample, DNSPs that are at least 75 per cent efficient in that year are not "exempted". This appears to mean these DNSPs are not deemed to be efficient, as is the practice when the BRFM is based on the period average. The grounds for that assumption are not made out.

Second, Frontier Economics argues that a firm with opex efficiency of around 0.75 is at risk of having an efficiency adjustment depending on its base year opex outturn (which has a random component). It is then argued that this produces an incentive for the firm to delay making cost efficiencies until the base year. The argument is unclear, and the conclusion is

counter-intuitive. If the firm makes efficiencies before the base year, it retains the benefit of those efficiencies for six years in total. However, if efficiencies are imposed on the firm through a base-year efficiency adjustment, then it does not retain any of the benefits. They flow fully to consumers. Hence, there appears to be a strong incentive for the firm to make efficiencies, rather than wait to have a base-year adjustment imposed on it.

Third, Frontier Economics states:

... a DNSP that reduces the final [sample] year opex beyond the target efficiency level will lower the estimate of efficient base year opex and therefore the opex allowance (*assuming the DNSP is unable to improve efficiency and therefore is subject to a base year reduction*). However, in such a circumstance the EBSS would (under the AER's existing framework) no longer be applied (due to the base year reduction), and the DNSP would not receive the EBSS benefit of the efficiency improvements in the final year. (Frontier Economics 2026, 70 italics added)

The assumption highlighted in italics suggests that the DNSPs' assessed efficiency in the final sample year is unconnected to its opex outturn in that year, which we consider an unreasonable assumption. If a DNSP reduces opex below the efficiency target, then it *will not* have an efficiency adjustment imposed on it. An efficiency adjustment depends on base year opex being above the efficiency target, and the amount of the adjustment relates to that excess.

In light of these three apparent shortcomings in Frontier Economics' argument, we recommend that the AER should disregard its proposal that the adjustment of efficiency scores for the Benchmark Comparison Point should apply symmetrically such that DNSPs with efficiencies above 0.75 would receive an upward adjustment to their base year opex.

6.3.2 Other issues relating to the BRFM

Frontier Economics also raised some complex issues regarding the interaction between the BRFM and the Efficiency Benefit Sharing Scheme (EBSS), not previously considered in this consultation. The operations of the EBSS are outside the scope of this report.

6.4 Analysis of BRFM Options

The purpose of this analysis is to compare the operations of the benchmarking roll-forward model (BRFM) with alternative time-varying inefficiency models, under the following scenarios:

- Simple averaging the efficiency scores for a DNSP over the sample period and using the BRFM as currently used;
- Using the DNSP's efficiency score estimated for the final year of the benchmarking period and rolling forward only to bridge between that year and the base year used to forecast efficient opex over the next regulatory period.

The data sample period used is 2006-2023 and the assumed base year is 2024. The BRFM method is set out in Appendix C. The Benchmark Comparison Point (BCP) of 0.75 is imposed in all cases.

The results for the estimated base-year opex efficiency adjustment for each DNSP are presented in this section for eight models considered across only two functional forms, CD and TLG. The nine models are:

- BC95-JTT-HN, BC95-JTT-HN-GTC, BC95-JTT-HN-QTT,
- LSE-ADTT-GTC-QTT, LSE-ADTT-GTC, LSE-AJTT-QTT
- Kumb90-JTT-HN, and Kumb90-JTT-HN-QTT.

Tables 6.1 to 6.8 in this section present the estimated opex efficiency percentage adjustment implied by each model for each DNSP in the base year (2024). The results include the application of the BRFM from the period-average efficiency score to the base year, and from the period-end (2023) to the base year. These tables also show:

- the cost adjustments when the results of the CD and TLG models are averaged
- the cost adjustments when the period-average and period-end results are averaged, and
- the average of those two averages.

Figure 6.1 illustrates the estimated opex efficiency adjustment, averaged over all DNSPs, for each of the nine models using the period-average BRFM method. It includes these results for both the CD and TLG versions of these models. Figure 6.2 presents the equivalent information for the period-end BRFM method.

Figure 6.1 suggests that when the period-average BRFM method is used, there is considerable variation in the average base year efficiency adjustment, including:

- Differences between models of the same family. For BC95, JTT-HN-GTC produces the highest adjustment, followed by JTT-HN-QTT, with JTT-HN the lowest. For LSE, ADTT-GTC-QTT produces the highest adjustment, followed by ADTT-GTC, with AJTT-QTT the lowest. For Kumb90, JTT-HN-QTT produces the highest adjustment and JTT-HN the lowest.
- Differences between different model families, with Kumb90 models producing, on average, lower efficiency adjustments and LSE models producing, on average, much larger efficiency adjustments (with BC95 models mostly in between);
- Differences between CD and TLG models, with TLG models producing, on average, lower efficiency adjustments (especially for the BC95 models).

Table 6.1: BRFM Results: BC95-JTT-HN-GTC (%)

<i>DNSP</i>	<i>Cobb-Douglas</i>			<i>Translog</i>			<i>Average of CD & TLG</i>		
	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>
EVO	7.5	10.3	8.9	0.6	3.2	1.9	4.0	6.8	5.4
AGD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CIT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
END	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ENX	16.6	16.1	16.4	13.9	15.0	14.4	15.3	15.6	15.4
ERG	27.0	27.8	27.4	12.3	15.4	13.9	19.6	21.6	20.6
ESS	11.6	13.7	12.6	2.5	5.3	3.9	7.1	9.5	8.3
JEN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PCR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SAP	3.5	0.0	1.6	0.0	0.0	0.0	0.6	0.0	0.0
AND	14.3	13.1	13.7	14.9	14.3	14.6	14.6	13.7	14.2
TND	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
UED	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<i>Average</i>	<i>6.2</i>	<i>6.2</i>	<i>6.2</i>	<i>3.4</i>	<i>4.1</i>	<i>3.7</i>	<i>4.7</i>	<i>5.2</i>	<i>4.9</i>

Table 6.2: BRFM Results: BC95-JTT-HN (%)

<i>DNSP</i>	<i>Cobb-Douglas</i>			<i>Translog</i>			<i>Average of CD & TLG</i>		
	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>
EVO	0.0	3.2	0.5	0.0	0.0	0.0	0.0	0.0	0.0
AGD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CIT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
END	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ENX	10.1	10.6	10.3	6.5	8.9	7.7	8.3	9.8	9.0
ERG	21.5	23.0	22.2	5.7	9.4	7.5	13.6	16.2	14.9
ESS	5.1	8.1	6.6	0.0	0.0	0.0	0.0	3.5	1.6
JEN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PCR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SAP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AND	7.2	8.1	7.6	8.1	9.7	8.9	7.7	8.9	8.3
TND	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
UED	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<i>Average</i>	<i>3.4</i>	<i>4.1</i>	<i>3.6</i>	<i>1.6</i>	<i>2.2</i>	<i>1.9</i>	<i>2.3</i>	<i>2.9</i>	<i>2.6</i>

Table 6.3: BRFM Results: BC95-JTT-HN-QTT (%)

<i>DNSP</i>	<i>Cobb-Douglas</i>			<i>Translog</i>			<i>Average of CD & TLG</i>		
	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>
EVO	2.7	7.2	5.0	0.0	0.0	0.0	0.0	3.4	1.2
AGD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CIT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
END	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ENX	13.5	13.3	13.4	10.9	12.0	11.4	12.2	12.7	12.4
ERG	24.3	25.1	24.7	10.7	13.0	11.8	17.5	19.0	18.3
ESS	8.3	10.6	9.5	0.0	1.8	0.4	3.7	6.2	4.9
JEN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PCR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SAP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AND	10.7	10.5	10.6	11.6	11.9	11.8	11.2	11.2	11.2
TND	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
UED	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<i>Average</i>	4.6	5.1	4.9	2.6	3.0	2.7	3.4	4.0	3.7

Table 6.4: BRFM Results: LSE-ADTT-GTC-QTT (%)

<i>DNSP</i>	<i>Cobb-Douglas</i>			<i>Translog</i>			<i>Average of CD & TLG</i>		
	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>
EVO	9.3	20.9	15.1	11.9	24.2	18.0	10.6	22.5	16.6
AGD	0.0	3.9	0.0	0.0	5.7	0.0	0.0	4.8	0.0
CIT	0.0	2.2	0.0	0.9	5.9	3.4	0.0	4.0	1.6
END	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ENX	16.7	17.2	17.0	16.5	16.6	16.6	16.6	16.9	16.8
ERG	26.8	24.8	25.8	23.1	19.7	21.4	24.9	22.3	23.6
ESS	6.8	10.5	8.6	0.0	1.5	0.0	1.9	6.0	3.9
JEN	0.0	3.2	0.0	5.7	16.4	11.0	0.0	9.8	4.1
PCR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SAP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AND	9.0	6.2	7.6	11.4	8.6	10.0	10.2	7.4	8.8
TND	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
UED	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<i>Average</i>	5.3	6.8	5.7	5.3	7.6	6.2	4.9	7.2	5.8

Table 6.5: BRFM Results: LSE-ADTT-GTC (%)

<i>DNSP</i>	<i>Cobb-Douglas</i>			<i>Translog</i>			<i>Average of CD & TLG</i>		
	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>
EVO	9.2	20.8	15.0	11.4	23.8	17.6	10.3	22.3	16.3
AGD	0.0	3.9	0.0	0.0	6.0	0.0	0.0	5.0	0.0
CIT	0.0	2.1	0.0	0.7	5.7	3.2	0.0	3.9	1.5
END	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ENX	16.8	17.1	16.9	16.6	16.7	16.7	16.7	16.9	16.8
ERG	26.8	24.7	25.8	22.4	19.0	20.7	24.6	21.8	23.2
ESS	6.8	10.4	8.6	0.0	0.9	0.0	1.6	5.7	3.7
JEN	0.0	3.2	0.0	6.1	16.8	11.5	0.0	10.0	4.4
PCR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SAP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AND	6.2	17.4	11.8	9.1	19.9	14.5	7.6	18.7	13.1
TND	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
UED	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<i>Average</i>	5.1	7.7	6.0	5.1	8.4	6.5	4.7	8.0	6.1

Table 6.6: BRFM Results: LSE-AJTT-QTT (%)

<i>DNSP</i>	<i>Cobb-Douglas</i>			<i>Translog</i>			<i>Average of CD & TLG</i>		
	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>
EVO	7.7	22.4	15.0	9.5	24.8	17.1	8.6	23.6	16.1
AGD	0.0	4.9	0.0	0.0	4.8	0.0	0.0	4.9	0.0
CIT	0.0	3.9	0.0	0.0	5.6	1.1	0.0	4.7	0.3
END	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ENX	16.5	12.2	14.4	13.4	17.0	15.2	15.0	14.6	14.8
ERG	25.3	26.1	25.7	23.7	22.7	23.2	24.5	24.4	24.4
ESS	3.7	11.3	7.5	0.0	2.9	0.0	0.0	7.1	2.9
JEN	0.0	3.0	0.0	3.0	17.6	10.3	0.0	10.3	2.5
PCR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SAP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AND	5.5	3.9	4.7	8.1	6.9	7.5	6.8	5.4	6.1
TND	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
UED	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<i>Average</i>	4.5	6.7	5.2	4.4	7.9	5.7	4.2	7.3	5.2

Table 6.7: BRFM Results: Kumb90-JTT-HN (%)

<i>DNSP</i>	<i>Cobb-Douglas</i>			<i>Translog</i>			<i>Average of CD & TLG</i>		
	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>
EVO	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AGD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CIT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
END	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ENX	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0
ERG	12.5	0.0	4.6	1.4	0.0	0.0	7.0	0.0	0.8
ESS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
JEN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PCR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SAP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AND	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TND	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
UED	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<i>Average</i>	1.0	0.0	0.4	0.1	0.0	0.0	0.5	0.0	0.1

Table 6.8: BRFM Results: Kumb90-JTT-HN-QTT (%)

<i>DNSP</i>	<i>Cobb-Douglas</i>			<i>Translog</i>			<i>Average of CD & TLG</i>		
	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>	<i>Period Average BRFM</i>	<i>Period- End BRFM</i>	<i>Average of both methods</i>
EVO	2.7	7.2	5.0	0.0	0.0	0.0	0.0	3.4	1.2
AGD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CIT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
END	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ENX	13.5	13.3	13.4	10.9	12.0	11.4	12.2	12.7	12.4
ERG	24.3	25.1	24.7	10.7	13.0	11.8	17.5	19.0	18.3
ESS	8.3	10.6	9.5	0.0	1.8	0.4	3.7	6.2	4.9
JEN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PCR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SAP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AND	10.7	10.5	10.6	11.6	11.9	11.8	11.2	11.2	11.2
TND	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
UED	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<i>Average</i>	4.6	5.1	4.9	2.6	3.0	2.7	3.4	4.0	3.7

Figure 6.1: Average base-year efficiency adjustment: Period-average BRFM method

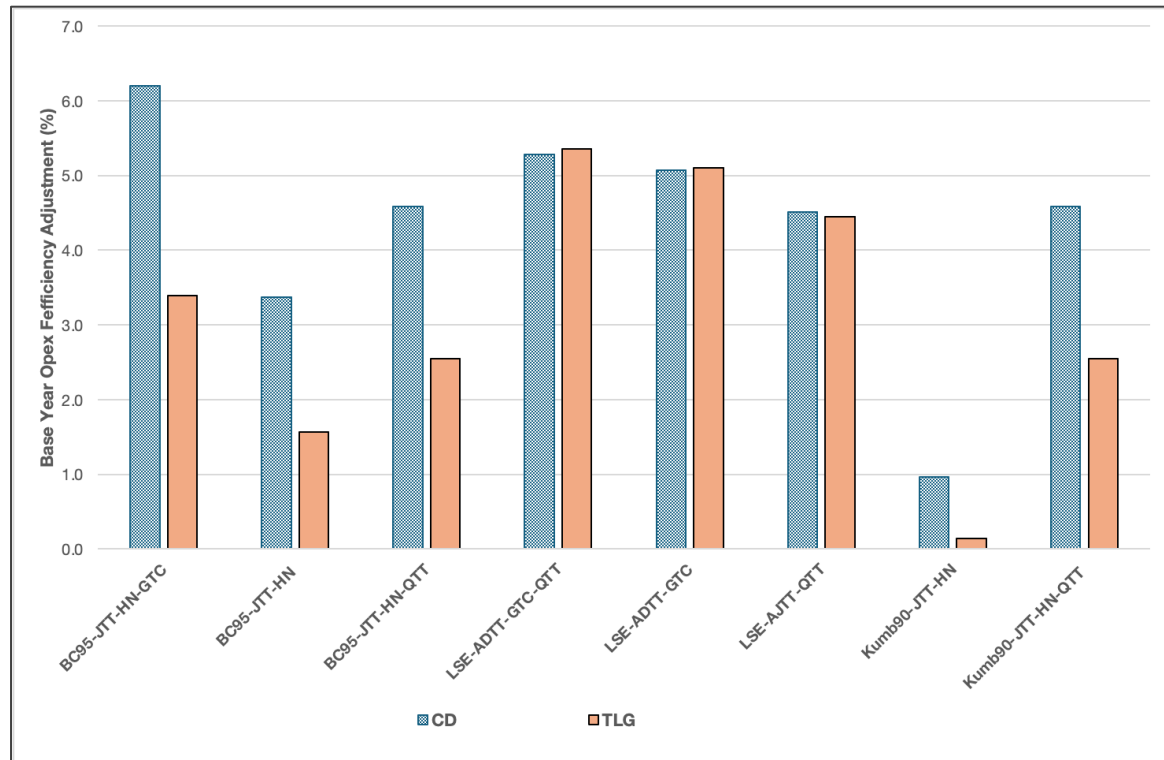
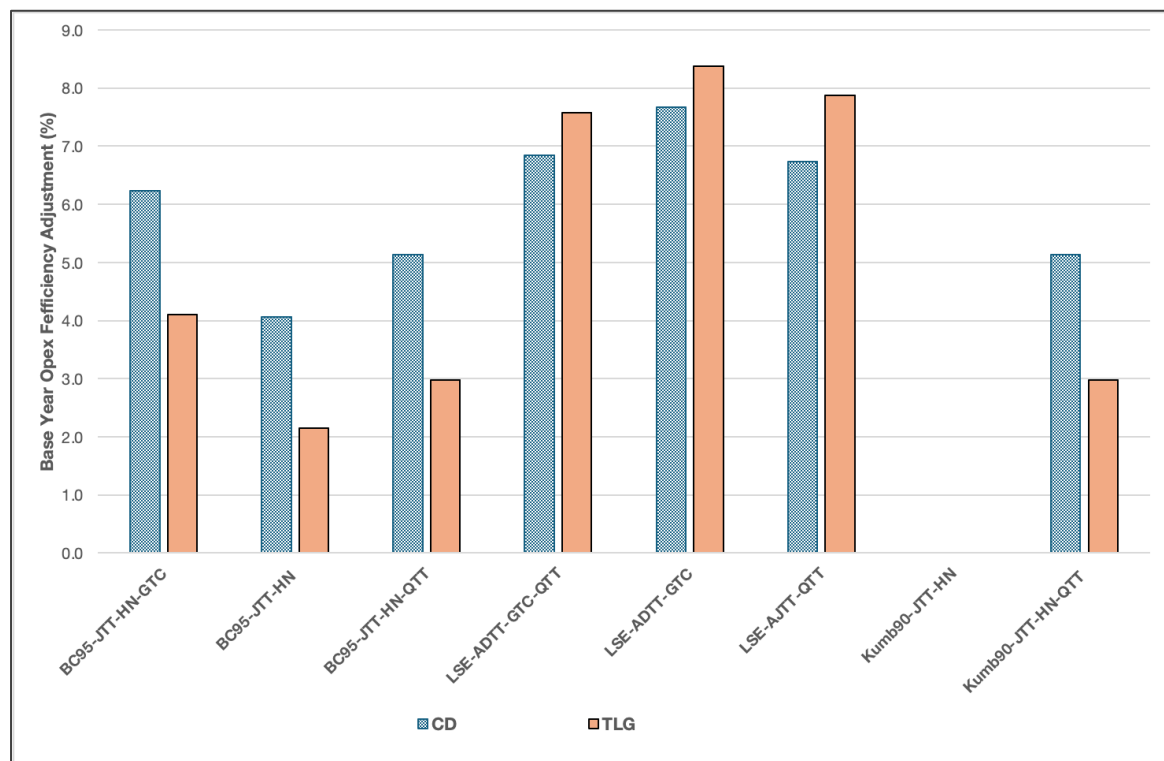


Figure 6.2: Average base-year efficiency adjustment: Period-end BRFM method



Using the TLG model, the average efficiency adjustment for all DNSPs using the period-average method ranges from 1.6 to 3.4 per cent in the BC95 models, 4.4 to 5.3 per cent in the LSE models, and 0.1 to 2.6 per cent in the Kumb90 models. On average across the three BC95 models, the CD average efficiency adjustment is 4.7 per cent compared to 2.5 per cent for the TLG models. Across the three LSE models, the average efficiency adjustment 5.0 per cent in both the CD and TLG versions. The average adjustments for the Kumb90 CD and TLG models are 2.8 and 1.3 per cent, respectively.

Using the period-average method, over all six BC95 models, the average efficiency adjustment is 3.6 per cent. Over all six LSE models, the average efficiency adjustment is 5.0 per cent. Over all four Kumb90 models, the average efficiency adjustment is just 2.1 per cent.

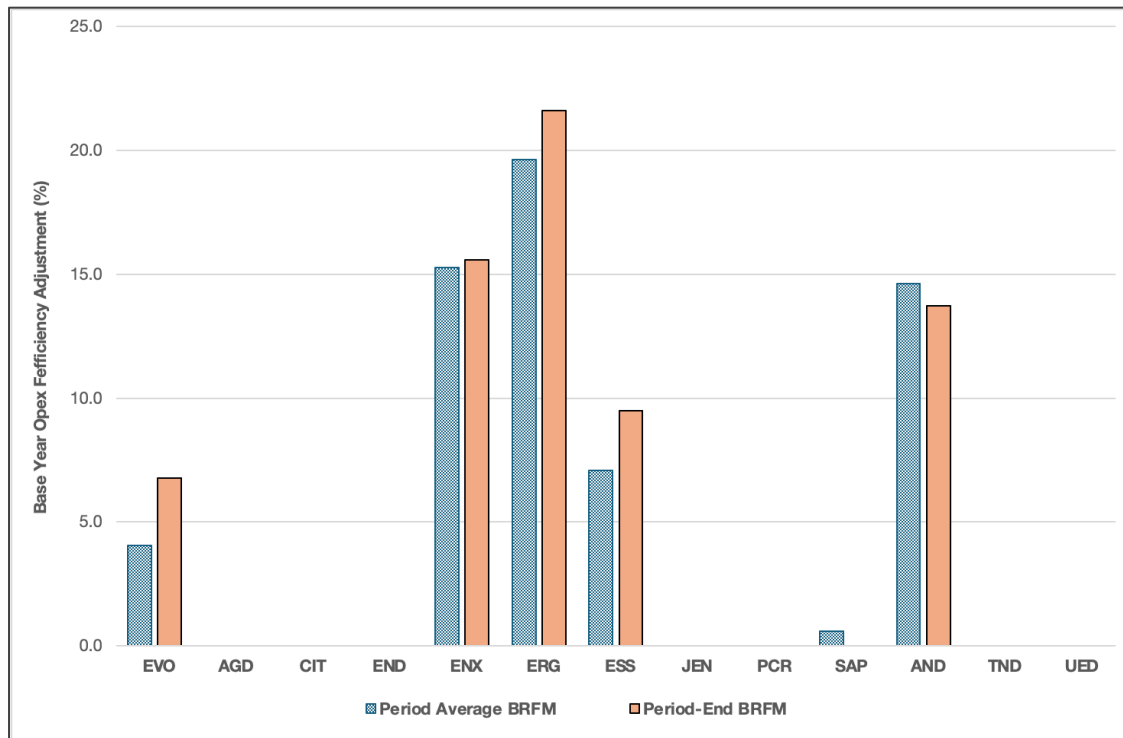
Figure 6.2 also suggests that when the period-end BRFM method is used, there is considerable variation in the average base year efficiency adjustment, including:

- Differences between models of the same family. In the BC95 family, BC95-JTT-HN-GTC produces the highest adjustment, followed by BC95-JTT-HN-QTT, with BC95-JTT-HN the lowest. In the LSE family, LSE-ADTT-GTC produces the highest adjustment. In the Kumb90 family, Kumb90-JTT-HN-QTT produces the highest adjustment and Kumb90-JTT-HN the lowest.
- Differences between different model families, with Kumb90 models producing, on average, lower efficiency adjustments and LSE models producing, on average, much larger efficiency adjustments (with BC95 models mostly in between);
- Differences between CD and TLG models, with TLG models producing, on average, lower efficiency adjustments.

Using the period-end method, over all six BC95 models, the average efficiency adjustment is 4.1 per cent (compared to 3.6 per cent using the period-average method). Over all six LSE models, the average efficiency adjustment is 7.5 per cent (compared to 5.0 per cent using the period-average method). Over all four Kumb90 models, the average efficiency adjustment is 2.0 per cent (compared to 2.1 per cent using the period-average method). On average, when the period-end BRFM method is used, Kumb90 models produced reasonably similar average efficiency adjustments, whereas for the BC95 and LSE model, efficiency adjustments are higher.

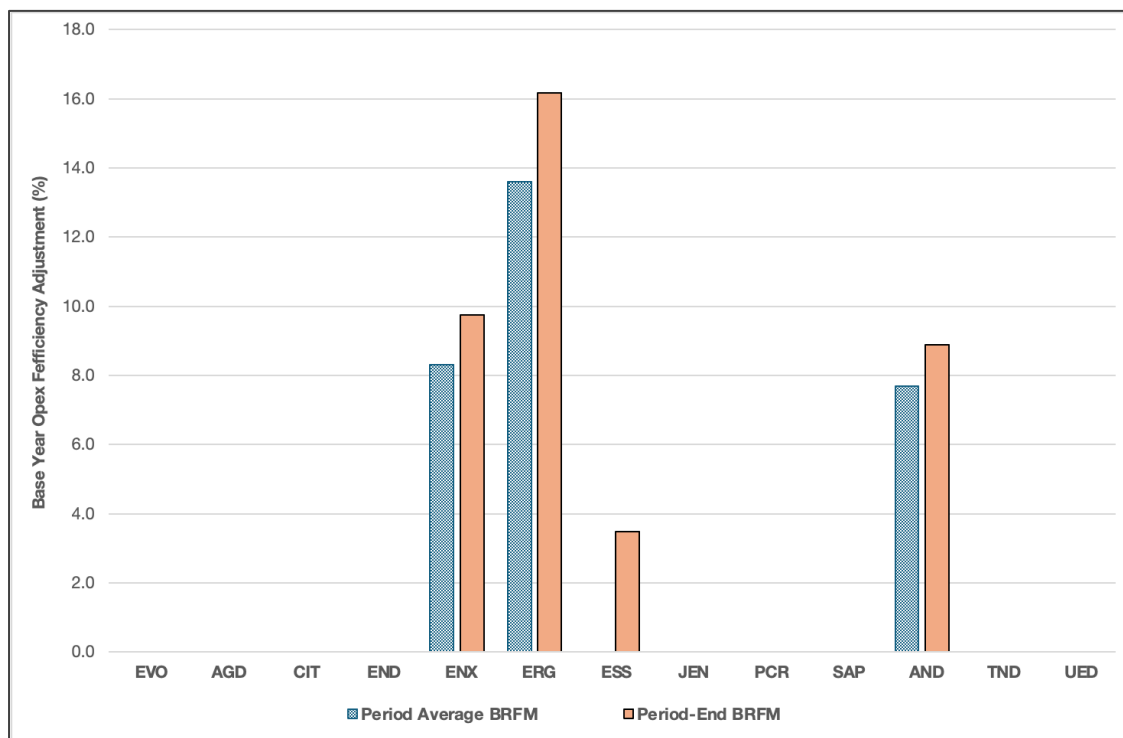
Figures 6.3 to 6.10 present, for each of the eight models, the efficiency adjustment for each DNSP, based on the average of the CD and TLG versions of the model. These charts also compare the estimated efficiency adjustments using both the period-average and period-end BRFM methods. Comparing these charts shows how the different models and different BRFM methods assess the required efficiency adjustments for each DNSP. Some DNSPs have very different assessed efficiency adjustments between different models.

Figure 6.3: Base-year efficiency adjustment by DNSP:* BC95-JTT-HN-GTC



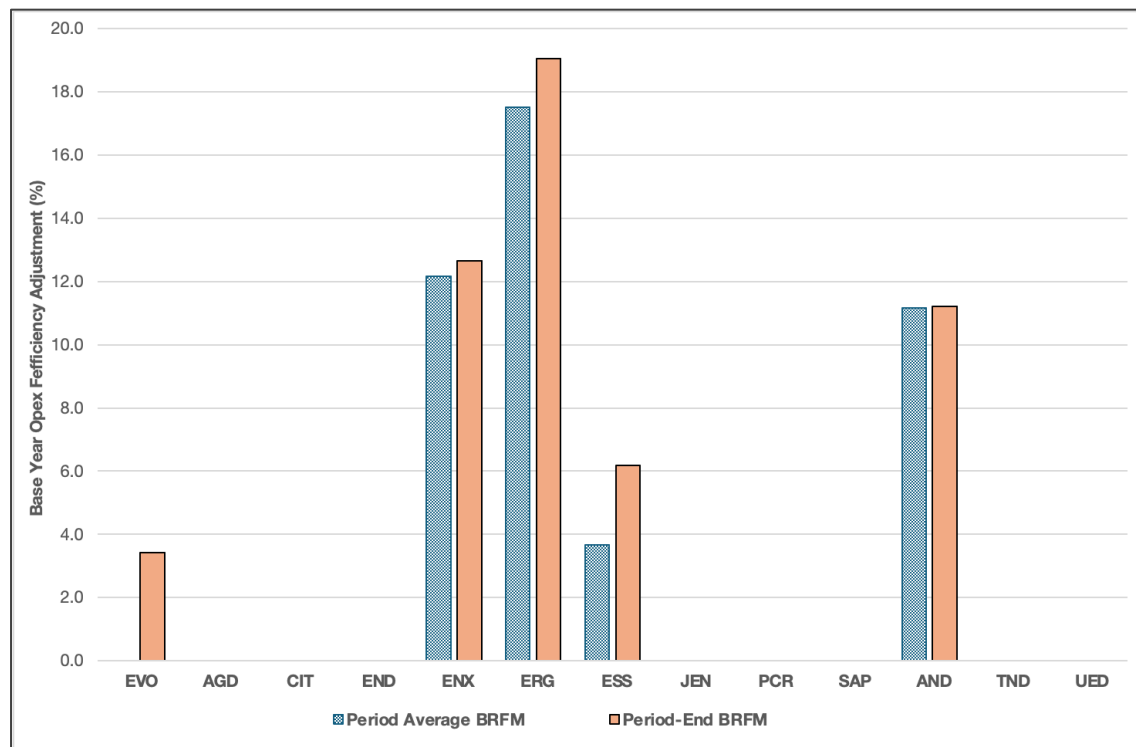
Note: * Averages of CD and TLG models.

Figure 6.4: Base-year efficiency adjustment by DNSP:* BC95-JTT-HN



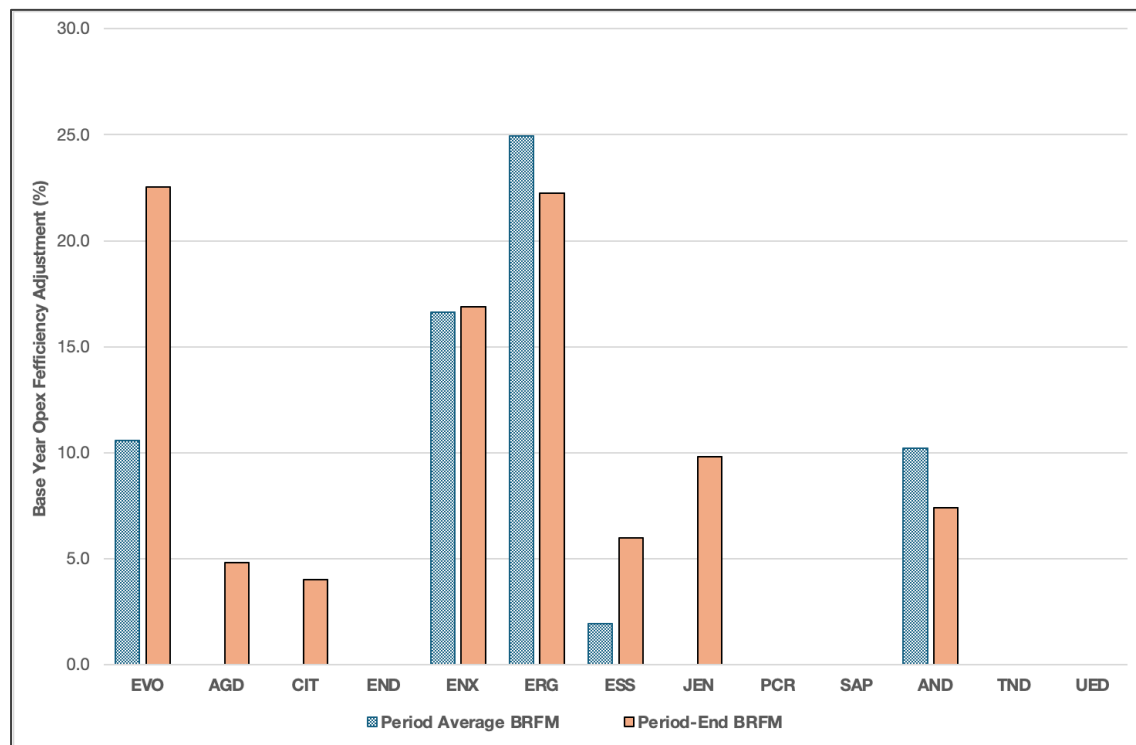
Note: * Averages of CD and TLG models.

Figure 6.5: Base-year efficiency adjustment by DNSP:* BC95-JTT-HN-QTT



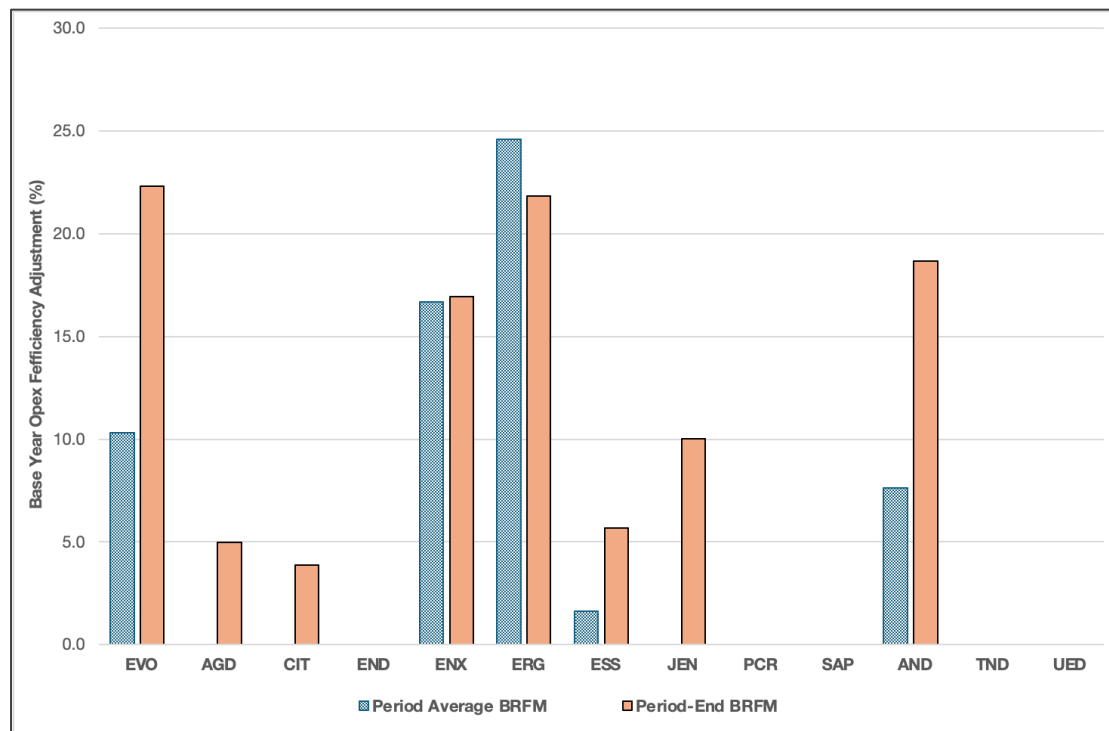
Note: * Averages of CD and TLG models.

Figure 6.6: Base-year efficiency adjustment by DNSP:* LSE-ADTT-GTC-QTT



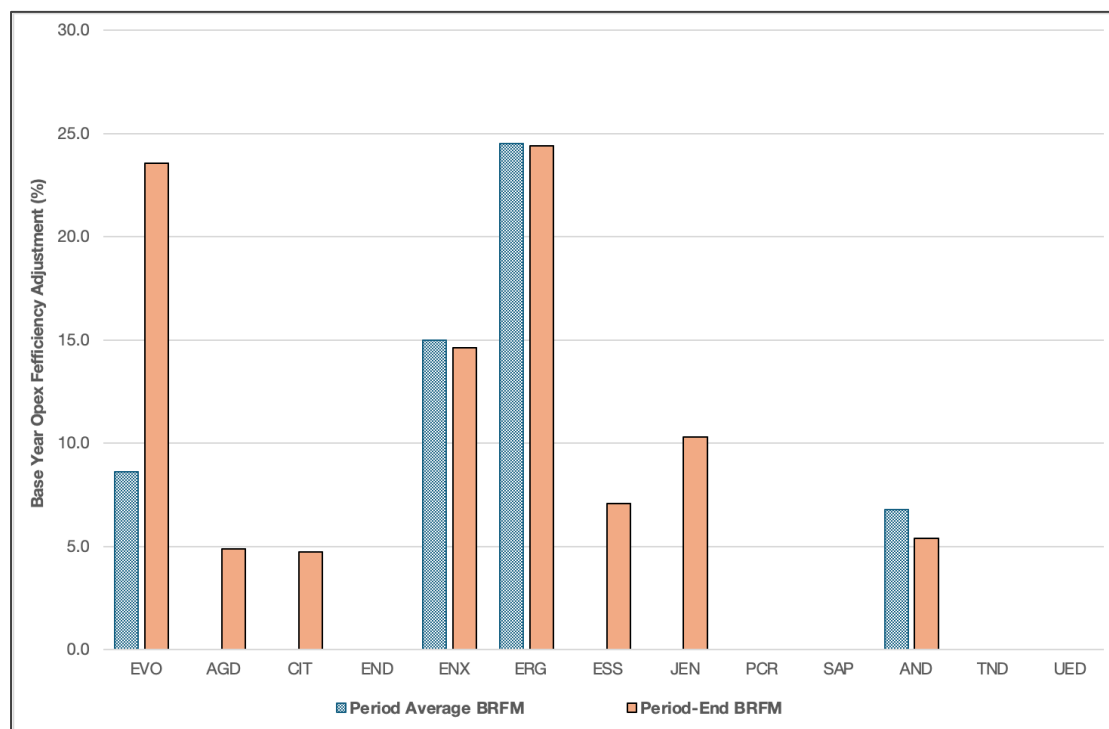
Note: * Averages of CD and TLG models.

Figure 6.7: Base-year efficiency adjustment by DNSP:* LSE-ADTT-GTC



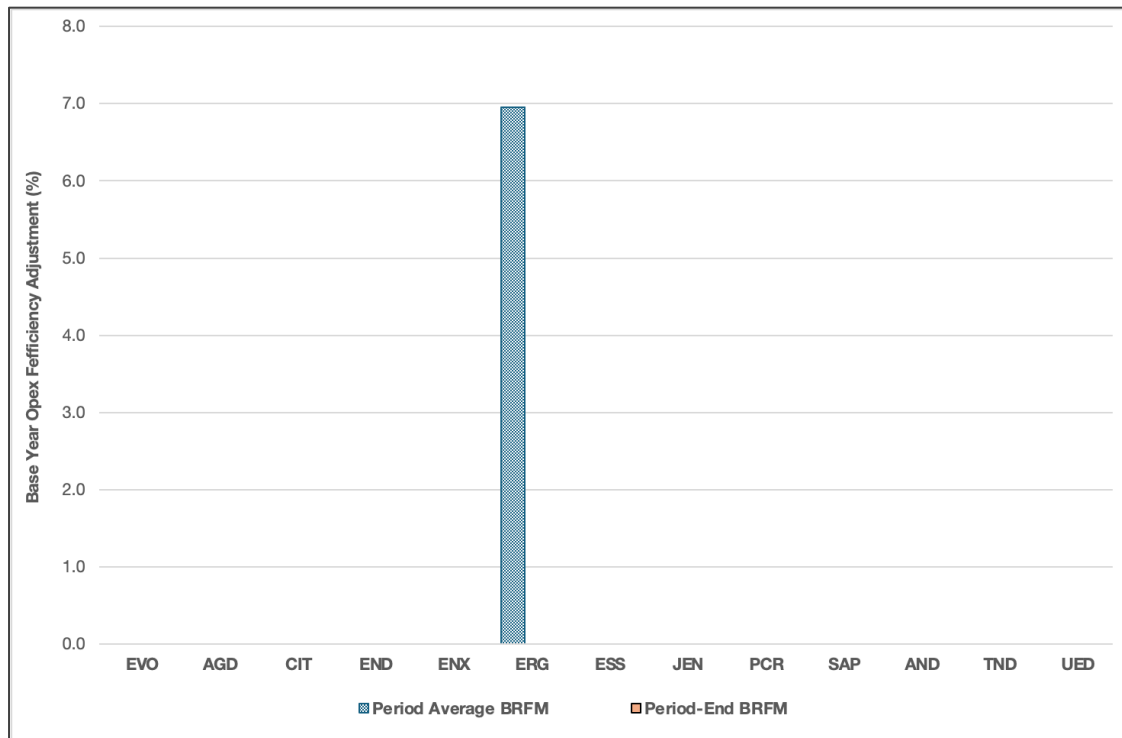
Note: * Averages of CD and TLG models.

Figure 6.8: Base-year efficiency adjustment by DNSP:* LSE-AJTT-QTT



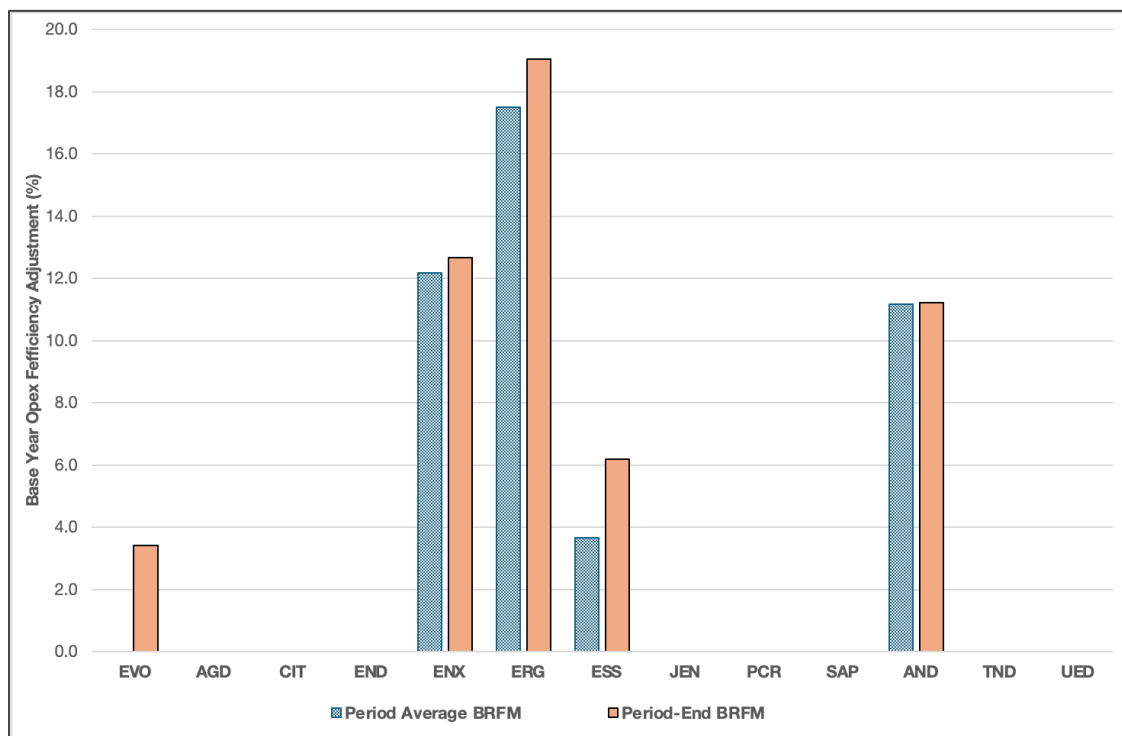
Note: * Averages of CD and TLG models.

Figure 6.9: Base-year efficiency adjustment by DNSP:* Kumb90-JTT-HN



Note: * Averages of CD and TLG models.

Figure 6.10: Base-year efficiency adjustment by DNSP:* SKumb90-JTT-HN-QTT



Note: * Averages of CD and TLG models.

Table 6.9 summarises the variability in estimated base-year efficiency adjustments across models and BRFM approach for each DNSP. For each DNSP, eight efficiency adjustments are considered (one for each model, using the average of the CD and TLG results). The Coefficient of Variation (CV) is the sample standard deviation of these eight estimates divided by the mean of the same nine estimates.²³

The period-end method generally produces higher base-year efficiency adjustments than the period-average method. The exceptions are ERG and SAP, as well for the DNSPs which the adjustment is zero under both approaches (END, PCR, TND and UED). For ENX, ERG and SAP, the difference between the two methods is less than one percentage point. The CV results indicate that the degree of variation across models also differs materially between DNSPs and methods. In general, the period-end method exhibits greater model variability than the period-average method, with EVO, ENX, ESS and SAP being the exceptions. For SAP, however, the high CV under the period-average method reflects its very small mean adjustment and therefore still corresponds to highly consistent results in absolute terms. The results also show that findings of no material inefficiency are robust across models. For DNSPs with zero efficiency adjustments where the CV is reported as NA, the adjustment is zero across all averaged models.

Table 6.9: BRFM Results by Method & DNSP

	<i>Period-Average</i>		<i>Period-End</i>		<i>Average of Methods</i>	
	<i>Mean (%)</i>	<i>CV</i>	<i>Mean (%)</i>	<i>CV</i>	<i>Mean (%)</i>	<i>CV</i>
EVO	4.2	116.9%	10.3	103.6%	7.1	110.1%
AGD	0.0	NA	1.8	138.0%	0.0	NA
AND	8.7	50.0%	9.6	58.6%	9.1	49.7%
CIT	0.0	NA	1.6	138.9%	0.4	165.0%
END	0.0	NA	0.0	NA	0.0	NA
ENX	12.0	46.7%	12.4	44.9%	12.2	45.7%
ERG	18.7	33.6%	18.0	42.7%	18.0	42.6%
ESS	2.2	110.3%	5.5	50.4%	3.8	65.0%
JEN	0.0	NA	3.8	138.1%	1.4	143.5%
PCR	0.0	NA	0.0	NA	0.0	NA
SAP	0.1	282.8%	0.0	NA	0.0	NA
TND	0.0	NA	0.0	NA	0.0	NA
UED	0.0	NA	0.0	NA	0.0	NA
<i>Average</i>	<i>3.5</i>	<i>51.9%</i>	<i>4.8</i>	<i>67.0%</i>	<i>4.0</i>	<i>59.1%</i>

²³ Note that the CV is sensitive to small means. Therefore, for DNSPs with small base-year adjustments, a high CV may still reflect highly consistent results, with little or no efficiency adjustment across models.

7 Network Utilisation Metrics

This section addresses the Justice and Equity Centre submission that the AER’s performance benchmarking should adopt the network utilisation metrics proposed by Institute for Sustainable Futures (ISF, 2024); particularly, Total Energy Throughput Utilisation (TETU) and Two-way Power Flow Utilisation (TPFU).

Section 7.1 discusses:

- How network utilisation is measured and reported in the AER’s current performance reporting framework for electricity networks and the way in which network utilisation is reflected in the benchmarking framework.
- Describes the limitations that the ISF sees in traditional network utilisation measures, primarily in terms of their continued accuracy when substantial amounts of energy are supplied from Customer Energy Resources (CER).
- Describes the proposed capacity utilisation measures and the proposed secondary asset-level metrics.

Section 7.2 discusses more generally how observations and proposed metrics of the ISF may be relevant to better reflecting network hosting services in the benchmarking framework. This discussion is brief and inconclusive because CER-related impacts on benchmarking are outside the scope of the current project and will be considered as part of the AER’s broader benchmarking development workstream scheduled for 2027.

The Justice and Equity Centre submission also suggested that the AER should standardise the calculation of network capacity and capacity utilisation, and identify efficient curtailment thresholds for CER. As explained in section 7.3, these topics are largely outside the scope of this paper.

7.1 Network Utilisation Measures

7.1.1 How network utilisation is currently captured

Network utilisation is reported in the AER (2026) *Network performance report 2026*. It is calculated by dividing a DNSP’s non-coincident maximum demand by the total capacity of its zone substation (ZSS) transformers. That is:

$$\frac{MD^{non-coincident}}{Capacity} = \frac{\sum_{i=1}^N MD_i^{non-coincident}}{\sum_{i=1}^N Capacity_i} \quad (7.1)$$

where:

- $MD_i^{non-coincident}$ is the non-coincident maximum demand supplied by the i th ZSS within a regulatory year.

- N stands for the total number of zone substations on the network, and
- $Capacity_i$ is the rated capacity of the i th ZSS.

Network utilisation is not used in the benchmarking framework, but there are measures related to $MD^{non-coincident}$ and $Capacity$. One of the outputs is ratcheted maximum demand (RMD) and one of the capital inputs is transformer capacity. However, there are differences of definition from the above measure. For a given year, RMD is defined as the highest $MD^{non-coincident}$ from 2006 up to that year. Transformer capacity is measured in megavolt amperes (MVA), and excludes the first stage ZSS transformer capacity of those systems that have two stage transformation from the higher voltages. Most relevant is the Multilateral Törnqvist total factor productivity (MTFP) index analysis, which includes both these measures, and therefore implicitly incorporates a measure of network utilisation.²⁴

7.1.2 Limitations of current utilisation measure and proposed new measures

The criticisms of the current network utilisation measure made by the Justice and Equity Centre and the ISF mostly relate to its lack of accounting for the hosting capacity of the network to accommodate customer energy exports, and of two-way energy flows. It is also argued that there are other network constraints that should be recognised, such as ‘minimum demand’ and voltage constraints, since they increasingly drive network investment. Remedying these gaps will require metrics that draw on data below the ZSS level, at “more granular spatial scales or asset-levels” (Institute for Sustainable Futures 2024, 5).

The proposed new metrics are:

- (a) *Total Energy Throughput Utilisation* (TETU), which is based on total electricity use, whether imported from the grid, exported to the grid or self-consumed. This is based on the idea that such a measure represents the total energy “facilitated by a grid connection”. However, a shortcoming of this proposal is the acknowledgement that the AER does not regard self-consumption as a ‘network output’ (Institute for Sustainable Futures 2024, 4).²⁵
- (b) *Two-way Power Flow Utilisation*, which is based on critical time-of-day and seasonal variations in two-way grid usage. This metric relates to capacity risk.

The ISF study also recommends three secondary asset-level metrics:

- voltage compliance
- reliability (SAIDI and SAIFI), and
- accumulated asset risk.

²⁴ Only RMD is included in the econometric opex cost function analysis.

²⁵ It is also noted that measuring self-consumption would require access to customer devices, and it would be difficult to measure accurately (Institute for Sustainable Futures 2024, 5).

The study is primarily focussed on performance reporting to aid planning and investment, which influences its preferences for a number of measures that together shed light on network conditions. The Justice and Equity Centre states that its proposed network utilisation metrics “would help assess and guide efficient investment and regulatory decisions, reduce stranded asset risks and inefficient expenditure decisions, and better support the efficient integration of consumer energy resources (CER)” (Justice and Equity Centre 2026, 4).

The metrics best suited for this purpose are not necessarily those best suited to economic benchmarking applications.

7.2 Incorporating Hosting Capacity in Benchmarking

As previously noted, CER-related impacts on benchmarking will be considered by the AER in 2027 as part of its benchmarking development program. Therefore, only a few comments are offered here, reflecting briefly on some of the issues raised by the Institute for Sustainable Futures.

Among the claimed limitations of the ‘traditional’ network utilisation measure, those that are possibly most relevant to benchmarking are:

- (i) lack of accounting for two-way flows on the network, or alternatively, the quantity of CER exports accommodated by the network; and
- (ii) non-measurement of network constraints relating to CER.

There is likely to be a trade-off between the amount of CER accommodated on the network and the reliability of supply or voltage compliance. This trade-off relates to the adequacy of the provision of network hosting capacity relative to the amounts of CER exports. That is, inadequate hosting capacity relative to the quantity of hosting services provided may lead to voltage noncompliance or reduced reliability.

With regard to the current benchmarking outputs and inputs, there is a limitation in measuring MD and Capacity at the ZSS, which does not include the CER exports in the system peak periods and the network’s capacity to accommodate CER in those periods. However, remedying this would only be part of the task of better reflecting CER into the outputs and inputs used.

Given the trade-off mentioned, measuring export services in outputs may require both quantity and quality metrics such as:

- A measure of the quantity of hosting services provided such as the proportion of total energy delivered that comes from CER exports, and
- The maintenance of voltage compliance.

The ISF states that power quality metrics typically “require high resolution and granularity data to model the low voltage (LV) networks to be able to run unbalanced power flow to

understand power quality”. However, “there may be some less complex ways of incorporating voltage (as perhaps the most pertinent CER power quality issue) into or alongside utilisation metrics” (Institute for Sustainable Futures 2024, 33). Voltage issues most relevant to solar generation are measured by the curtailment of exports caused by voltage tripping. An explicit metric would be “to document voltage ‘excursions’, by capturing the frequency that voltage levels reach or exceed recommended maximum or minimum operating bounds. Such indicators could include those required by the Victorian Essential Services Commission:

- Number/percentage of assets above the maximum voltage limit.
- Number/percentage of assets below the minimum voltage limit” (Institute for Sustainable Futures 2024, 33).

An impediment to developing such metrics is that, outside Victoria, there is a lack of availability of voltage data from smart meters. Hence, it is likely that it cannot at present be regularly reported by all DNSPs.

Other challenges associated with incorporating CER-related variables include: (a) formulating output value weights for CER-related outputs; and (b) ensuring that the capital inputs associated with providing hosting capacity are adequately captured by the input specification.

7.3 Other Justice and Equity Centre Submissions

7.3.1 Standardised calculation of network capacity

The Justice and Equity Centre argued that the AER should standardise the calculation of network capacity and its utilisation in order to improve the incentives under the regulatory framework and primarily to improve network planning. We have seen in the previous section that the TETU measure is proposed. However, there are potential limitations to this particular measure, firstly, in regard to whether household electricity production for self-consumption can properly be viewed as a service provided by the network, and secondly because measuring self-consumption would require access to customer devices within households, which may be impractical and costly to implement.

It is also noted above that the metrics used for benchmarking that are related to network utilisation, namely RMD and transformer capacity, do not include CER exports in the relevant peak hourly periods. There would be some challenges in remedying this:

- Since the MD measure is a non-coincident peak, there would be a question as to which periods should be considered the network peak periods. This may require granular data on CER exports by location, which can be assigned to or aligned with corresponding zone substations.
- Measuring peak period CER exports at granular levels would require smart meter data which is not universally available across the NEM.

Since the contribution of CER exports at the system peak will add to both demand and supply capacity, it may not have a major impact on the network utilisation measure unless a large share of peak period power comes from CER.

7.3.2 Framework for Efficient Curtailment

The Justice and Equity Centre argued that the AER should develop a framework for identifying efficient curtailment thresholds for customer energy resources (CER) at each stage of the energy transition. This relates to the planning question of what is the optimal balance between curtailing CER and investing in network capacity to accommodate it.

It is outside our scope to comment on this argument. However, the AER's website states that it is responsible for developing the customer export curtailment value (CECV) methodology and for calculating CECVs each year. CECVs measure the social detriment when CER exports are curtailed, and help guide efficient levels of investment by DNSPs to accommodate export services. The AER last updated the CECVs on 30 June 2025. The Justice and Equity Centre does not adequately clarify the relationship between its proposal and the current CECV.

Appendix A Review of Frontier Economics’ calculation of BC95 efficiency scores

This appendix reviews Frontier Economics’ alternative programming commands to calculate efficiency scores. Using these commands, Frontier Economics reports that it is able to calculate efficiency scores where the *sfpanel postestimation* command failed to return estimates. Hence, models where this failure arises, typically BC95 models with AJTT efficiency trends, need not be rejected as Quantonomics did. Frontier Economics’ method does not change the estimated model; it only estimates the efficiency score. By doing so more successfully, fewer of the BC95 models need to be rejected.

A.1 Background

The time-varying inefficiency SFA models, including BC95 and Kumb90 variants, that are presented in both the Quantonomics Phase 2 report and the Frontier Economics response, are estimated using the *sfpanel* community-contributed Stata command (Belotti et al. 2012). The reported efficiency scores and associated confidence intervals are produced by the *sfpanel postestimation* command: `predict <varname>, bc ci`. This command calculates the Battese and Coelli (1988) (“bc”) efficiency score based on the conditional expectation:

$$E[\exp(-u_{it}) | \epsilon_{it}] \quad (\text{A.1})$$

where $E[.]$ is the expectation operator, u_{it} is the inefficiency term at observation (i,t) , and ϵ_{it} is the composed residual. As previously stated, for some of the models tested, this post-estimation command returned missing efficiency scores.

A.2 Frontier Economics’ observations and alternative method

The relevant postestimation command is contained in the ado file *sfpanel_p.ado*. The calculations for the bc efficiency scores and confidence intervals are found in lines 602-640. The efficiency formulas differ for the BC95 and Kumb90 models, and confidence intervals are not available for the Kumb90 model.

Frontier Economics calculates the alternative efficiency scores in its program: ‘2026-04-09 BC95_models.do’, at lines 251-295. They only apply these efficiency scores to the BC95 models because this is where missing efficiency scores appear to mainly occur. They do not calculate confidence intervals.

The approach in *sfpanel_p.ado* is:

$$\exp(-\mu_1 + 0.5 \cdot \sigma_1^2) \cdot \text{normal}(-\sigma_1 + \mu_1/\sigma_1) / \text{normal}(\mu_1/\sigma_1)$$

Frontier Economics’ approach is:

$$\exp(-\mu_1 + 0.5 \cdot \sigma_1^2) \cdot \exp(\ln \text{normal}(-\sigma_1 + \mu_1/\sigma_1) - \ln \text{normal}(\mu_1/\sigma_1))$$

This second formulation is equivalent to the first, but uses *exp(.)* and *lnnormal(.)* in place of *normal(.)*.

Frontier Economics attributes the inferior performance of the first formulation to a numerical issue in the evaluation of very small normal probabilities. In particular, when the relevant normal cumulative distribution function is evaluated at a very large negative value, Stata may store the value as zero. This can then lead to division by zero or undefined values in the efficiency-score calculation.

The questions to be considered are:

- (a) Should Frontier Economics' calculation using *exp(lnnormal(.))* in place of *normal(.)* be accepted as a useful method for calculating efficiency scores where *sfpanel*'s 'bc' post-estimation command produces missing values?
- (b) If so, how can it be implemented in the context of Quantonomics' Stata programs?

A.3 Testing Frontier Economics' Method

This section replicates Frontier Economics' calculations. In the diagnostic checks, 'bc_stata' refers to the efficiency score produced by the *sfpanel postestimation* command 'predict ... bc'; and 'bc_frontier' refers to the calculated score using Frontier's program for BC95 models.

The tests examine whether:

1. bc_frontier matches bc_stata where the latter produces a non-missing value
2. bc_frontier fills missing values of bc_stata
3. bc_frontier itself produces missing values for efficiency scores
4. bc_frontier remains within the theoretical efficiency-score range of 0 to 1, and
5. any observed issues arise from the bc_frontier efficiency score calculation.

The following long-period BC95 files were tested:

- BC95-AJTT-HN, SFACD and SFATLG
- BC95-AJTT-HN-GTC, SFACD and SFATLG
- BC95-JTT-HN, SFACD and SFATLG
- BC95-JTT-HN-GTC, SFACD and SFATLG

Table A.1 presents the summary of the results.

Table A.1: Comparison of original spanel and Frontier-style efficiency-score calculations

<i>Model file</i>	<i>Specification</i>	<i>Missing Stata bc</i>	<i>Missing filled by Frontier</i>	<i>Missing Frontier score</i>	<i>Frontier scores outside [0,1]</i>	<i>Max absolute difference where both non-missing</i>	<i>Assessment</i>
BC95-AJTT-HN	CD	37	37	0	0	2.98e-08	Clean support
	TLG	19	19	0	0	2.98e-08	Clean support
BC95-AJTT-HN-GTC	CD	36	36	0	0	2.98e-08	Clean support
	TLG	54	54	0	0	2.97e-08	Clean support
BC95-JTT-HN	CD	0	0	0	0	2.98e-08	Clean equivalence test
	TLG	0	0	0	0	2.98e-08	Clean equivalence test
BC95-JTT-HN-GTC	CD	0	0	0	0	2.98e-08	Clean equivalence test
	TLG	554	389	0	165*	3.15e-23	Not clean; generated values outside theoretical range

Notes: * Values exceed 1 and should not be treated as valid without further investigation.

The results indicate that Frontier Economics' calculation method ('bc_frontier'):

- Is mathematically and numerically sound as a postestimation calculation.
- In several models, it is able to produce efficiency score estimates where the *sfp*anel postestimation command *predict ... bc* ('bc_stata') fails to produce efficiency score estimates.
- For non-missing efficiency scores, the bc_frontier method reproduces the bc_stata results to a high degree of precision.
- However, in some models the bc_frontier does not successfully produce efficiency scores, and hence does not improve on the bc_stata method in these cases.

Clear examples of improvement are: the BC95-AJTT-HN models (both CD and TLG), and the BC95-AJTT-HN-GTC Translog model, where in each case the bc_frontier estimates filled all of the missing efficiency scores from the bc_stata estimates.

Overall, the results indicate that Frontier Economics' method improves the availability of efficiency scores in some cases and does not produce inferior results to the sfpanel post-estimation command in any of the tested models.

A.4 Conclusions

Frontier Economics' alternative method of calculating efficiency scores can be accepted as a valid way of calculating BC95 efficiency scores. This acceptance is limited to the efficiency calculation method. The basis for this recommendation is that:

1. the method reproduces Stata's standard predict bc, bc ci central efficiency scores to numerical precision where Stata produces non-missing scores;
2. in the clean BC95-AJTT-HN cases, it fills Stata's missing central efficiency scores;
3. the filled scores in the clean cases remain within the theoretical 0 to 1 range; and
4. the method does not alter the estimated frontier model, but only recalculates the post-estimation central efficiency score.

Appendix B Kumb90 Efficiency Time Trend Specification

We tested a final-year time-trend normalisation for the Kumb90 models, under which the final year of the estimation period equals zero. This was intended to align the treatment of time with the BC95 and LSE models.

However, the final-year normalisation produced unusually low efficiency scores in the Kumb90 specifications. This appendix explains why changing the zero point affects the estimated inefficiency levels and why the centred time trend used in Phase 2 was retained.

B.1 Centred and final-year time-trend specifications

The centred time trend is defined as:

$$t^{ct} = yr - \bar{yr} \quad (B.1)$$

The final-year normalisation time trend is defined as:

$$t^{fy} = yr - \text{endyr} \quad (B.2)$$

The difference between these time trend specifications can be seen as:

$$\begin{aligned} yr &= t^{ct} + \bar{yr} \\ t^{fy} &= (t^{ct} + \bar{yr}) - \text{endyr} \\ t^{fy} &= t^{ct} - k \end{aligned} \quad (B.3)$$

For a 2006 to 2023 sample, $k = 8.5$.

B.2 Kumb90 Efficiency Specification

In the Kumb90 model, inefficiency for DNSP i in year t is defined as:

$$u_{it} = g_{it}u_i \quad (B.4)$$

where u_i is the underlying DNSP-specific inefficiency and g_{it} is the time-varying inefficiency scaling function, defined as:

$$g_{it} = \frac{1}{1 + \exp(\eta_{it})} \quad (B.5)$$

where η_{it} is the deterministic term governing how inefficiency evolves over time. Therefore:

$$u_{it} = u_i[1 + \exp(\eta_{it})]^{-1} \quad (B.6)$$

The efficiency score is defined as:

$$EFF_{it} = \exp(-u_{it}) \quad (B.7)$$

The relationship between the components is:

$$\eta_{it} \uparrow \Rightarrow g_{it} \downarrow \Rightarrow u_{it} \downarrow \Rightarrow EFF_{it} \uparrow \quad (B.8)$$

Thus, a lower value of η_{it} results in greater estimated inefficiency and a lower efficiency score. For example, in the Kumb90-JTT-HN specification, η_{it} is parameterised with a jurisdiction-specific time trend (JTT):

$$\eta_{it} = \delta_0 + \delta_{aus}t_{aus} + \delta_{nz}t_{nz} + \delta_{ont}t_{ont} \quad (B.9)$$

Here, each jurisdiction-specific time variable equals the time trend for observations in that jurisdiction and zero otherwise. δ_0 captures the common intercept to all firms, while δ_{aus} , δ_{nz} , δ_{ont} capture deviations from that baseline specific to each jurisdiction.

B.3 Implications for the choice of normalisation

When the efficiency time trend is normalised to equal zero in a given year, all jurisdiction-specific time variables in the JTT models, or DNSP-specific time variables in the AJTT models, are also zero in that year. The time-varying inefficiency scaler therefore becomes:²⁶

$$g_{it} = \frac{1}{1 + \exp(\delta_0)} \quad (B.10)$$

In that year, the scaler depends directly only on the estimated common intercept and is therefore the same for all DNSPs.²⁷ Under a centred time trend, the scaler takes this value at the midpoint of the estimation period. Under final-year normalisation, it takes this value in the final year.

Changing from a centred time trend to a final-year time trend does more than change the year at which the time trend equals zero. As shown in Equation B.3, the final-year-normalised trend is obtained by subtracting k from the centred trend: $t^{fy} = t^{ct} - k$. Using the JTT model for Australia as an example, the η_{it} under final-year normalisation is:

$$\eta_{it}^{fy} = \delta_0 + \delta_{aus}t_{aus}^{fy} \quad (B.11)$$

Substituting ($t_{aus}^{fy} = t_{aus}^{ct} - k$) gives:

²⁶ If the δ_0 is constrained to zero, then g_{it} would take value of 0.5.

²⁷ This **does not** imply that all DNSPs have the same efficiency, because efficiency also depends on the DNSP-specific time-invariant inefficiency term, u_i .

$$\eta_{it}^{fy} = \delta_0 + \delta_{aus}(t_{aus}^{ct} - k) \quad (B.12)$$

which can be rewritten as:

$$\eta_{it}^{fy} = \delta_0 + \delta_{aus}t_{aus}^{ct} - k\delta_{aus} \quad (B.13)$$

Holding δ_0 constant:

$$\eta_{it}^{fy} = \eta_{it}^{ct} - k\delta_{aus} \quad (B.14)$$

Thus, holding the coefficients fixed, changing the normalisation introduces an Australia-specific level shift equal to $-k\delta_{aus}$ in the inefficiency driver, η_{it} . As seen in equation (B.8), lower η_{it}^{fy} implies lower efficiency scores.²⁸

This mechanism explains why a final-year normalisation produces unusually low efficiency scores in the Kumb90 models. Recentring changes the year in which the jurisdictional indices and scalers are constrained to coincide, so the two forms of time trend normalisation produce different restricted specifications for the inefficiency patterns.

Frontier Economics recommends final-year normalisation for the BC95 models because it places the convergence point at the end of the estimation period and reduces the influence of historical efficiency trajectories on final-year comparisons. However, this reasoning is specific to the BC95 inefficiency structure and does not transfer directly to the Kumb90 model.

The centred time trend was therefore retained for the Kumb90 specifications because it avoids imposing large jurisdiction-specific level shifts, preserves the Phase 2 parameterisation and produces efficiency levels that are more consistent with the results from the other models. Frontier Economics also accepted the centred Kumb90 formulation in its evaluation.

Accordingly, final-year normalisation was retained for the BC95 models, while the centred time trend was retained for the Kumb90 models. The appropriate normalisation depends on the structure of the inefficiency specification and should not be imposed uniformly across models.

²⁸ The empirical results for the Kumb90 models show that estimates of δ_{AUS} are generally positive. If δ_{AUS} is negative, the direction of the effect would be reversed.

Appendix C Benchmarking Roll-forward Method

C.1 BRFM method using Period-Average Efficiency

The AER uses the benchmarking roll-forward model (BRFM) to estimate efficient opex from benchmarking results for the base year of a regulatory price review. In the established BRFM method:

- (a) Actual nominal opex is deflated by the opex input price index, and the sample period average is calculated (C_{PA}). Also, the actual (or year-end estimate) base-year opex is deflated by the opex input price index (C_{BY}). Subscripts PA and BY represent the sample period average and base year, respectively.
- (b) The period-average efficient opex (or ‘target opex’), C_{PA}^e , is calculated as:

$$C_{PA}^e = C_{PA} \times \frac{\theta_{PA}}{ACP} \quad (C.1)$$

where θ_{PA} is the period-average estimated efficiency score; and ACP is the ‘adjusted comparison point’ defined as:

$$ACP = \frac{BCP}{1 + OEF\%} \quad (C.2)$$

where OEF% is an ex-post adjustment for the combined effect of OEFs on efficient opex, and BCP is the *benchmark comparison point*, 0.75, which is a discretionary adjustment by the regulator to account for factors such as modelling imprecision and the combined effect of any unaccounted for OEFs.

For time-varying inefficiency models, θ_{PA} could either be a simple average of the efficiency scores estimated for each year of the sample period, or a weighted average (with weights equal to the estimated efficient real opex in each year of the sample period). These two methods will generally give different results.

- (c) Then the roll-forward of efficient opex to the base year is performed as follows:

$$C_{BY}^e = C_{PA}^e \times \exp(\dot{C}) \quad \text{where: } \dot{C} = \dot{Y} - \dot{P} \quad (C.3)$$

where 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 efficient real opex; 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:

$$\dot{Y} = \frac{1}{e_Y} \sum_{m=1}^M \varepsilon_m \cdot \dot{y}_m \quad (\text{C.4})$$

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:

$$\dot{P} = (1 - e_Y)\dot{Y} - \varphi\dot{z} - \lambda T \quad (\text{C.5})$$

where $\dot{z}$ is the rate of change in the OEF, φ is the elasticity of real variable cost with respect to the included OEF (underground cable share), λ is the parameter on the time-trend variable used to measure technical change, and T is the cumulative number of years between the sample average and the base year. For models that do not use a linear trend for technology change (such as the GTC models), λT will refer to the total cumulative effect of technical change on efficient opex. If the set of dummy variables used to specify GTC use the last period as the excluded base period, then λT may equal zero.

(d) The base year percentage opex inefficiency (δ_{BY}) is then given by:

$$\delta_{BY} = \max\left\{1 - \frac{C_{BY}^e}{C_{BY}}, 0\right\} \quad (\text{C.6})$$

The efficient base-year nominal opex can be obtained by adjusting base-year actual nominal opex by δ_{BY} .

C.2 BRFM method using Period-End Efficiency

If the BRFM is applied between the end of the sample period and the base year, then the preceding formulation needs to be modified as follows:

- In equations (C.1), (C.2) and (C.3), the period-average, denoted by subscript PA, will be replaced by the period-end, denoted PE. Two cases can be considered for the period-end: (i) the last year of the sample period; or (ii) an average of the last three years of the sample period.
- In equation (C.2), OEF% should relate to the ‘period-end’, rather than the period-average. See section 6.2.3.
- Also, in equation (C.2), the BCP relevant to the period-end need not be the same as that used for the period-average. This is related to AusNet’s point noted in section 6.1.2, and is discussed in section 6.2.4.

- In equations (C.3), (C.4) and (C.5), the dot over a variable will represent the *cumulative change in that variable from the period-end to the base year*.
- In equation (C.4), the elasticities of real opex with respect to each output are all evaluated at the sample midpoint (averaged over all three jurisdictions in the sample). A question that might arise is whether they should be evaluated at the period-end timeframe (and, possibly, whether they should represent only the Australian average).
- In equation (C.5), T will represent the cumulative number of years between the period-end and the base year. If the period-end is the last year of the sample, T would typically equal 1. If the period-end was an average of the last three years of the sample, T would typically equal 2. If the GTC specification of technical change is used, then λT will refer to the effect of technical change on efficient opex only over the short period between the period-end and the base year.

Appendix D Sensitivity Analysis for Alternative Sample Periods Results

Table D.1 Sensitivity Analysis

		2006-2019		2006-2020		2006-2021		2006-2022		2006-2024	
		CD	TLG	CD	TLG	CD	TLG	CD	TLG	CD	TLG
BC95-JTT-HN	Comp. Issues*	! (a)	!	!	!	!	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		2.7%		2.6%		0.5%		6.3%		8.9%
	MVs % - Total		13.1%		13.2%		12.8%		14.2%		15.0%
	# DNSPs MV		0		0		0		1		1
	Pseudo-R ²	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.999	0.998
	BIC	-264.1	-311.9	-287.6	-342.1	-305.0	-365.7	-322.6	-387.0	-367.7	-450.1
BC95-JTT-HN-GTC	Comp. Issues*	✓	!	✓	X(c)	✓	✓	✓	✓	✓	X(c)
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare coeff	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		0.0%		0.0%		2.4%		1.4%		0.0%
	MVs % - Total		12.6%		12.8%		13.5%		13.3%		12.9%
	# DNSPs MV		0		0		0		0		0
	Pseudo-R ²	0.998	0.997	0.998	0.997	0.998	0.998	0.998	0.998	0.998	0.997
	BIC	-250.0	-357.3	-276.5	-409.6	-299.7	-362.1	-327.5	-395.1	-367.7	-583.3
BC95-AGECJTT-HN	Comp. Issues*	✓	!	✓	!	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		2.7%		3.6%		0.0%		4.1%		7.3%
	MVs % - Total		13.2%		13.6%		13.1%		14.0%		15.7%
	# DNSPs MV		0		0		0		0		1
	Pseudo-R ²	0.998	0.999	0.998	0.999	0.998	0.999	0.998	0.998	0.998	0.998
	BIC	-258.7	-308.7	-281.4	-338.4	-293.3	-357.0	-312.0	-379.6	-355.5	-444.0

Table D.1 (cont.)

		2006-2019		2006-2020		2006-2021		2006-2022		2006-2024	
		CD	TLG	CD	TLG	CD	TLG	CD	TLG	CD	TLG
BC95-AGECJTT -HN-GTC	Comp. Issues*	✓	✓	✓	✓	✓	✓	✓	✓	X(c)	!
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		1.6%		2.6%		2.4%		1.4%		0.0%
	MVs % - Total		13.2%		13.6%		13.4%		13.5%		12.6%
	# DNSPs MV		0		0		0		0		0
	Pseudo-R ²	0.998	0.998	0.998	0.998	0.998	0.999	0.998	0.999	0.994	0.996
	BIC	-240.6	-288.7	-265.4	-321.2	-280.9	-344.9	-308.3	-377.3	-462.1	-574.4
BC95-JTT -HN-QTT	Comp. Issues*	!	!	✓	!	✓	✓	✓	✓	✓	X(c)
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		0.0%		0.0%		0.0%		0.9%		0.0%
	MVs % - Total		13.3%		13.6%		13.4%		13.5%		13.6%
	# DNSPs MV		0		0		0		0		0
	Pseudo-R ²	0.998	0.997	0.998	0.997	0.998	0.998	0.998	0.998	0.998	0.996
	BIC	-287.4	-343.4	-310.7	-373.3	-324.4	-391.8	-341.6	-411.4	-393.6	-557.2
Kumb90- JTT-HN	Comp. Issues*	✓	✓	✓	✓	✓	✓	✓	✓	!	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		55.5%		51.8%		34.1%		5.9%		0.0%
	MVs % - Total		27.2%		23.6%		11.1%		1.3%		0.0%
	# DNSPs MV		8		7		4		0		0
	Pseudo-R ²	0.994	0.995	0.994	0.994	0.994	0.994	0.994	0.994	0.994	0.994
	BIC	-1140.1	-1126.2	-1215.0	-1196.8	-1313.3	-1289.1	-1411.7	-1383.2	-1578.8	-1558.6

Benchmarking Opex Model Development 2026

Table D.1 (cont.)

		2006-2019		2006-2020		2006-2021		2006-2022		2006-2024	
		CD	TLG	CD	TLG	CD	TLG	CD	TLG	CD	TLG
Kumb90-JTT-HN-GTC	Comp. Issues*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	!	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✗	✓	✗	✓	✗	✓	✓	✓	✓
	MVs % - Aust.		65.9%		56.9%		43.8%		9.0%		0.0%
	MVs % - Total		40.3%		31.1%		14.0%		2.8%		0.0%
	# DNSPs MV		9		7		7		1		0
	Pseudo-R ²	0.994	0.995	0.994	0.994	0.994	0.994	0.994	0.994	0.994	0.994
	BIC	-1102.1	-1089.9	-1173.5	-1157.7	-1266.4	-1244.7	-1362.2	-1335.9	-1518.6	-1502.0
Kumb90-AJTT-HN	Comp. Issues*	!	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✗	✓	✗	✓	✗	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		16.5%		65.6%		73.6%		64.3%		47.4%
	MVs % - Total		9.5%		48.2%		50.5%		38.2%		21.2%
	# DNSPs MV		2		9		9		8		6
	Pseudo-R ²	0.995	0.995	0.994	0.995	0.994	0.995	0.994	0.995	0.994	0.994
	BIC	-1102.8	-1077.1	-1163.1	-1157.7	-1263.7	-1261.3	-1371.7	-1359.3	-1557.5	-1545.8
Kumb90-AJTT-HN-GTC	Comp. Issues*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✗	✓	✗	✓	✗	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗
	MVs % - Aust.		31.3%		71.3%		72.6%		66.5%		49.8%
	MVs % - Total		22.5%		55.4%		52.6%		43.2%		32.8%
	# DNSPs MV		4		9		10		8		6
	Pseudo-R ²	0.995	0.995	0.994	0.995	0.994	0.995	0.994	0.994	0.994	0.994
	BIC	-1062.8	-1041.0	-1122.5	-1125.0	-1219.1	-1218.3	-1324.4	-1313.5	-1501.8	-1493.6

Table D.1 (cont.)

		2006-2019		2006-2020		2006-2021		2006-2022		2006-2024	
		CD	TLG	CD	TLG	CD	TLG	CD	TLG	CD	TLG
Kumb90- AGECJTT-HN	Comp. Issues*	!	!	!	!	!	!	!	✓	!	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		47.3%		39.0%		26.0%		10.9%		0.0%
	MVs % - Total		25.6%		20.1%		13.3%		3.3%		0.0%
	# DNSPs MV		6		5		4		2		0
	Pseudo-R ²	0.995	0.995	0.994	0.994	0.994	0.994	0.994	0.994	0.994	0.994
	BIC	-1130.8	-1123.7	-1199.2	-1182.2	-1270.5	-1246.4	-1345.3	-1314.5	-1519.2	-1491.6
Kumb90- AGECJTT -HN-GTC	Comp. Issues*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	X	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	X
	MVs % - Aust.		57.1%		51.3%		46.2%		21.7%		0.0%
	MVs % - Total		32.6%		29.6%		21.4%		6.3%		0.0%
	# DNSPs MV		8		7		6		2		0
	Pseudo-R ²	0.994	0.995	0.994	0.994	0.994	0.994	0.994	0.994	0.994	0.994
	BIC	-1096.2	-1089.6	-1168.5	-1152.5	-1249.7	-1227.3	-1336.8	-1307.7	-1496.1	-1472.9
Kumb90-JTT -HN-QTT	Comp. Issues*	!	!	✓	✓	!	✓	!	✓	!	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	X	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		36.3%		50.8%		42.8%		10.4%		0.0%
	MVs % - Total		25.9%		24.6%		15.2%		2.2%		0.0%
	# DNSPs MV		5		6		6		2		0
	Pseudo-R ²	0.995	0.995	0.994	0.995	0.994	0.994	0.994	0.994	0.994	0.994
	BIC	-1141.0	-1134.2	-1212.8	-1198.3	-1305.3	-1283.8	-1401.0	-1374.5	-1567.2	-1550.1

Table D.1 (cont.)

		2006-2019		2006-2020		2006-2021		2006-2022		2006-2024	
		CD	TLG	CD	TLG	CD	TLG	CD	TLG	CD	TLG
Kumb90-AJTT -HN-QTT	Comp. Issues*	!	✓	!	✓	!	✓	!	!	!	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	X	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		12.6%		12.8%		74.0%		75.6%		7.7%
	MVs % - Total		13.3%		9.8%		42.3%		34.1%		8.7%
	# DNSPs MV		2		2		9		10		1
	Pseudo-R ²	0.995	0.995	0.995	0.995	0.995	0.995	0.995	0.995	0.994	0.995
	BIC	-1081.4	-1063.6	-1150.4	-1134.9	-1254.7	-1236.6	-1363.3	-1339.1	-1556.5	-1521.6
LSE-ADTT	Comp. Issues*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		4.4%		1.5%		8.2%		18.6%		3.2%
	MVs % - Total		13.1%		14.3%		14.3%		16.4%		16.1%
	# DNSPs MV		1		0		1		2		0
	Pseudo-R ²	0.981	0.984	0.981	0.983	0.980	0.983	0.980	0.983	0.980	0.982
LSE-ADTT -GTC	Comp. Issues*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		0.0%		0.0%		0.0%		0.0%		0.4%
	MVs % - Total		13.0%		13.2%		12.9%		12.9%		15.8%
	# DNSPs MV		0		0		0		0		0
	Pseudo-R ²	0.981	0.984	0.981	0.983	0.981	0.983	0.980	0.983	0.980	0.982

Table D.1 (cont.)

		2006-2019		2006-2020		2006-2021		2006-2022		2006-2024	
		CD	TLG	CD	TLG	CD	TLG	CD	TLG	CD	TLG
LSE-AJTT	Comp. Issues*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		7.7%		7.7%		5.8%		10.0%		7.7%
	MVs % - Total		18.7%		18.5%		18.4%		19.3%		19.1%
	# DNSPs MV		1		1		1		1		1
	Pseudo-R ²	0.981	0.984	0.981	0.984	0.981	0.984	0.981	0.984	0.981	0.984
LSE-AJTT -GTC	Comp. Issues*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		7.7%		7.7%		7.7%		9.5%		7.7%
	MVs % - Total		19.8%		19.8%		19.9%		20.1%		21.4%
	# DNSPs MV		1		1		1		1		1
	Pseudo-R ²	0.981	0.984	0.981	0.984	0.981	0.984	0.981	0.984	0.981	0.984
LSE-AJT Tx -GTC	Comp. Issues*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		7.7%		7.7%		7.7%		9.5%		8.1%
	MVs % - Total		19.0%		19.5%		19.4%		20.1%		21.1%
	# DNSPs MV		1		1		1		1		1
	Pseudo-R ²	0.982	0.984	0.981	0.984	0.981	0.984	0.981	0.984	0.982	0.984

Table D.1 (cont.)

		2006-2019		2006-2020		2006-2021		2006-2022		2006-2024	
		CD	TLG	CD	TLG	CD	TLG	CD	TLG	CD	TLG
LSE-ADTT -QTT	Comp. Issues*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		7.7%		4.1%		8.7%		11.3%		7.3%
	MVs % - Total		14.9%		15.5%		15.0%		14.8%		17.3%
	# DNSPs MV		1		1		1		1		1
	Pseudo-R ²	0.981	0.984	0.981	0.983	0.981	0.983	0.980	0.983	0.980	0.982
LSE-ADTT- GTC-QTT	Comp. Issues*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		0.5%		0.0%		0.0%		0.0%		1.6%
	MVs % - Total		13.6%		13.7%		13.1%		12.8%		15.4%
	# DNSPs MV		0		0		0		0		0
	Pseudo-R ²	0.981	0.984	0.981	0.983	0.981	0.983	0.980	0.983	0.980	0.982
LSE-AJTT -QTT	Comp. Issues*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		8.8%		7.7%		7.7%		10.9%		7.7%
	MVs % - Total		18.5%		18.8%		18.9%		19.5%		19.4%
	# DNSPs MV		1		1		1		1		1
	Pseudo-R ²	0.982	0.984	0.981	0.984	0.981	0.984	0.981	0.984	0.981	0.984

Table D.1 (cont.)

		2006-2019		2006-2020		2006-2021		2006-2022		2006-2024	
		CD	TLG	CD	TLG	CD	TLG	CD	TLG	CD	TLG
LSE-AJTT- GTC-QTT	Comp. Issues*	X (d)	X (d)	X (d)	X (d)	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		7.7%		7.7%		7.7%		10.4%		8.1%
	MVs % - Total		19.9%		19.9%		19.6%		20.3%		21.8%
	# DNSPs MV		1		1		1		1		1
	Pseudo-R ²	0.981	0.984	0.981	0.984	0.981	0.984	0.982	0.984	0.982	0.984
LSE-AJTtx -QTT	Comp. Issues*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		8.2%		7.7%		6.2%		9.5%		7.7%
	MVs % - Total		18.6%		19.0%		18.5%		19.2%		19.2%
	# DNSPs MV		1		1		1		1		1
	Pseudo-R ²	0.982	0.984	0.982	0.984	0.982	0.984	0.982	0.984	0.982	0.984
LSE-AJTtx -GTC-QTT	Comp. Issues*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		8.8%		7.7%		7.7%		10.9%		8.5%
	MVs % - Total		19.1%		19.3%		19.4%		20.3%		21.1%
	# DNSPs MV		1		1		1		1		1
	Pseudo-R ²	0.982	0.984	0.981	0.984	0.982	0.984	0.982	0.984	0.982	0.984

Table D.1 (cont.)

		2006-2019		2006-2020		2006-2021		2006-2022		2006-2024	
		CD	TLG	CD	TLG	CD	TLG	CD	TLG	CD	TLG
LSE-ADTT	Comp. Issues*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		7.7%		7.7%		7.7%		7.7%		7.7%
	MVs % - Total		17.1%		18.0%		16.1%		15.1%		19.2%
	# DNSPs MV		1		1		1		1		1
	Pseudo-R ²	0.982	0.984	0.982	0.984	0.981	0.984	0.981	0.984	0.980	0.983
LSE-ADTT -PW-2015-2022 ⁽¹⁾	Comp. Issues*									✓	✓
	Output coefs - signs									✓	✓
	- significance									✓	✓
	UGshare sign									✓	✓
	MVs % - Aust.										7.7%
	MVs % - Total										20.3%
	# DNSPs MV										1
	Pseudo-R ²									0.980	0.983
LSE-AJTT -PW-2015-2022 ⁽¹⁾	Comp. Issues*									✓	✓
	Output coefs - signs									✓	✓
	- significance									✓	✓
	UGshare sign									✓	✓
	MVs % - Aust.										8.5%
	MVs % - Total										21.6%
	# DNSPs MV										1
	Pseudo-R ²									0.982	0.985

Table D.1 (cont.)

		2006-2019		2006-2020		2006-2021		2006-2022		2006-2024	
		CD	TLG	CD	TLG	CD	TLG	CD	TLG	CD	TLG
LSE-AJTT -PW-2015	Comp. Issues*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Output coefs - signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	- significance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	UGshare sign	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MVs % - Aust.		14.8%		12.3%		15.4%		15.4%		8.5%
	MVs % - Total		21.8%		21.9%		21.7%		21.6%		21.1%
	# DNSPs MV		2		2		2		2		1
Pseudo-R ²		0.982	0.984	0.982	0.984	0.982	0.984	0.982	0.984	0.982	0.984

Note: The computational-performance classifications are defined as follows: ✓ indicates that no convergence or computational issues were identified. ! indicates that convergence or flat-region warning was recorded, but estimates were produced. X(b) indicates that the estimation failed or did not produce usable results. X(c) indicates that estimates were produced, but one or more coefficients, standard errors or z-statistics were missing. X(d) indicates that one or more variables were omitted because of collinearity. (1) The piecewise models with a second knot in 2022 collapse to the models with a single knot for samples ending before 2023.

References

- Ausgrid. 2026. 'Response AER's Benchmarking Model Review Stage 2'. March 27.
- Australian Energy Regulator (AER). 2023. *Incentivising and Measuring Export Service Performance*. Final Report.
- Australian Energy Regulator (AER). 2025. *Annual Benchmarking Report; Electricity Distribution Network Service Providers*.
- Australian Energy Regulator (AER). 2026. *Network Performance Report 2026*.
- Battese, G. E., and T. J. Coelli. 1995. 'A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data'. *Empirical Economics* 20 (2): 325–32. <https://doi.org/10.1007/BF01205442>.
- Battese, George, and Tim Coelli. 1988. 'Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Function and Panel Data'. *Journal of Econometrics* 38: 387–99.
- Belotti, Federico, Silvio Daidone, Giuseppe Ilardi, and Vincenzo Atella. 2012. 'Stochastic Frontier Analysis Using Stata'. *Stata Journal*, 1–39.
- Cameron, A. Colin, and Pravin K. Trivedi. 2010. *Microeconometrics Using Stata*. Revised. Stata Press.
- Cornwell, Christopher, Peter Schmidt, and Robin C. Sickles. 1990. 'Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels'. *Journal of Econometrics* 46 (1–2): 185–200. [https://doi.org/10.1016/0304-4076\(90\)90054-W](https://doi.org/10.1016/0304-4076(90)90054-W).
- Economic Insights. 2012. *Econometric Estimates of the Victorian Gas Distribution Businesses' Efficiency and Future Productivity Growth*. Report Prepared for SP AusNet.
- Economic Insights. 2015. *Domestic Transmission Capacity Services Benchmarking Model: Testing Further Specifications*. Report Prepared for ACCC by John Fallon, Michael Cunningham and Tim Coelli.
- Evoenergy. 2025. *Re: 2025 DNSP Annual Benchmarking Report – Draft Report and Results*.
- Frontier Economics. 2026. *Response to Phase 2 Consultation on Econometric Benchmarking Models*. Prepared for Evoenergy.
- Greene, William H. 2012. *Econometric Analysis*. 7th edn. Pearson.
- Institute for Sustainable Futures. 2024. *Reimagining Network Utilisation in the Era of Consumer Energy Resources*. University of Technology Sydney.
- Justice and Equity Centre. 2026. 'Annual Benchmarking Reports 2025 – Quantanomics Phase 2 Report'. March 26.
- Lawrence, Denis. 2007. *Victorian Gas Distribution Business Opex Rate of Change*. Report by Meyrick and Associates for Envestra, Multinet and SP AusNet. Canberra.
- Lawrence, Denis, and Erwin Diewert. 2006. 'Regulating Electricity Networks: The ABC of Setting X in New Zealand'. In *Performance Measurement and Regulation of Network Utilities*, edited by Tim Coelli and Denis Lawrence. Edward Elgar.
- Pacific Economics Group. 2004. *Predicting Growth in SPI's O&M Expenses*. Report Prepared for SPI. Madison, Wisconsin.

- Peyrache, Antonio. 2024. *Review of the AER's Estimated Non-Reliability Output Weights Used in the TFP and MTFP Benchmarking Models*. Centre for Efficiency and Productivity Analysis (CEPA).
- Quantonomics. 2025a. *Economic Benchmarking Results for the Australian Energy Regulator's 2025 DNSP Annual Benchmarking Report*. Report Prepared for Australian Energy Regulator by Michael Cunningham, Joseph Hirschberg and Alice Giovani.
- Quantonomics. 2025b. *Electricity Distribution Benchmarking Opex Model Development*. Report Prepared for Australian Energy Regulator by Michael Cunningham, Valentin Zelenyuk, Alice Giovani and Joe Hirschberg.
- Quantonomics. 2025c. *Nonreliability Output Index Weights ABR25*. Memorandum Prepared for the AER by Michael Cunningham, Joe Hirschberg and Alice Giovani.
- Quantonomics. 2025d. *Output Index Weights - Supplementary Analysis*. Memorandum. <https://www.aer.gov.au/system/files/2025-11/Quantonomics%20-%20Memorandum%20-%20Output%20Index%20Weights%20-%20Supplementary%20Analysis%20-11%20November%202025.pdf>.
- Wang, Zhichao, and Valentin Zelenyuk. 2024. 'Random Versus Explained Inefficiency in Stochastic Frontier Analysis: The Case of Queensland Hospitals'. In *Advances in Econometrics*, edited by Christopher F. Parmeter, Mike G. Tsionas, and Hung-Jen Wang. Emerald Publishing Limited. <https://doi.org/10.1108/S0731-905320240000046013>.