Review of the AER's estimated non-reliability output weights used in the TFP and MTFP benchmarking models.

Antonio Peyrache*

a.peyrache@uq.edu.au

*Director of the Centre for Efficiency and Productivity Analysis (CEPA), School of Economics, The University of Queensland, Australia.

The Centre for Efficiency and Productivity Analysis (CEPA)

CEPA (https://economics.uq.edu.au/cepa) was established at the University of Queensland in 2003 to provide a focal point for research, consultancy and training in efficiency and productivity analysis in Australia and the Asia/Pacific Region. The staff of the Centre are involved in all aspects of efficiency and productivity analysis, including firm-level, industry-level, and international comparisons, in both private and government sectors. Measurement methods used include: index numbers, data envelopment analysis (DEA), stochastic frontiers and econometric production models.

The principal aims and objectives of the Centre are to undertake theoretical and applied research, develop useful end-user-driven short courses and workshops and to develop computer software for use by private and public sector organisations.

Contents

1 Introduction

In this report we undertake a review of the estimated non-reliability output weights that are currently used by the Australian Energy Regulator (AER) to run a benchmarking exercise for energy service providers in Australia. The AER benchmarking exercise consists of a methodological procedure that can broadly be classified into three steps. *In step one*, a Leontief cost function is estimated via non-linear least squares. *In step two*, the estimated parameters of the Leontief cost function are used to calculate a set of weights for the non-reliability outputs. *In step three*, these non-reliability output weights are used to construct multilateral Tornqvist indexes of productivity change. The three steps are interconnected, in the sense that any computational mistake done in one of the steps will carry over and change the results of the final benchmarking analysis.

The AER found that there was a mistake in the computation of the non-reliability output weights used in the multilateral total factor productivity (MTFP) index number computation. The mistake was made in step one of the aforementioned procedure, during the estimation of the Leontief cost function. This meant that the associated non-reliability output weights (and the resulting benchmarking exercise) were affected by this numerical error. This error in the computation of the weights was corrected in the 2020 AER benchmarking exercise. Following this revision, the AER decided to conduct an independent review on how these output weights are determined and calculated. The AER hired us to provide this independent review, with a special focus on the following three requirements:

1. review current calculation of non-reliability output weights under the Leontief cost function method used in the AER's TFP and MTFP models to ensure they are currently correctly estimated. This applies to both the distribution network service providers (DNSP) and the transmission network service providers (TNSP) modelling under the index number approach adopted by the AER;

2. discuss the advantages and disadvantages of the Leontief cost function method for estimating these output weights;

3. list and discuss alternative methods to the current (Leontief) model for estimating these output weights and discuss the advantages/disadvantages of each.

In this brief report we are going to illustrate the results of our own independent calculations for the non-reliability output weights. We have also considered feedback received from stakeholder submissions on a draft of this report from Ausgrid, Evoenergy and Jemena. We do not feel that this feedback changes significantly the original recommendations in our draft report and have included (in section 7) an appendix to respond to this feedback.

We employ available data provided by the AER on which to base these calculations. We used two different datasets to implement our calculations: one for the transmission network service providers (TNSP) and the other one for the distribution network service providers (DNSP). These datasets are publicly available and can be downloaded at the following link:

- https://www.aer.gov.au/authors/economic-insights[1].

This is also the set of data that has been used by the AER to conduct its annual benchmarking exercise. It should be pointed out here that these two datasets are an updated version of the datasets used by the AER to conduct the revision of the output weights in 2020. The numerical differences in the datasets are small, but these small numerical differences carry over to the calculation of the non-reliability output weights. We decided to use the updated datasets in this report, since the numerical differences in our calculated non-reliability output weights (compared to the AER) fall within the normal range of variation of numerical optimization and are not significantly different from the outcome that one would expect from the non-updated datasets.

[1] The specific links are: https://www.aer.gov.au/documents/economic-insights-2020-benchmarking-data-files-distribution for the DNSP; and https://www.aer.gov.au/documents/economic-insights-2020-benchmarking-data-files-transmission for the TNSP.

The methodology for the estimation of the Leontief cost function and the subsequent computation of non-reliability output weights is described in two separate Economic Insights reports: Lawrence et al. (2020b) describe it for the transmission network; and Lawrence et al. (2020a) discuss it for the distribution network. These two reports are available online at the following links (respectively for the TNSP and DNSP):

- https://www.aer.gov.au/documents/economic-insights-benchmarking-results-aer-transmission-october-2020
- https://www.aer.gov.au/documents/economic-insights-benchmarking-results-aer-distribution-october-2020

TFP and MTFP techniques measure the relationship between multiple outputs and multiple inputs enabling a comparison of productivity across production units that share the same set of inputs and outputs. In this respect it is important to be clear and transparent about the choice of inputs and outputs used in the AER benchmarking exercise. The set of inputs and outputs used in the transmission and distribution cases are similar in the aforementioned reports.

On the input side we have the opex cost and three measures of capital infrastructure: the overhead lines, the underground lines and transformers and other assets. The dataset used by Economic Insights and provided to us by the AER also includes Annual User Costs (AUC) for each of the three capital stock measures: this is the total imputed cost of maintaining and enhancing the three capital infrastructure measures used in the benchmarking assessment. This means that the unit price for each capital stock input can be obtained and used in the estimation of the Leontief cost function. The opex cost quantity index is determined by deflating the opex cost dollar value by a price deflator (which adjusts for inflation pressures on the inputs accounted for by the opex measure). This means that there are four input quantities and four input prices to be used in the Leontief cost function model. This set of inputs is the same for the distribution and the transmission case.

On the output side, three outputs are common to both the TNSP and DNSP. These are: energy delivered, ratcheted maximum demand and circuit length. The fourth output is the number of customers, but its definition varies slightly between the two groups of energy service providers. For the TNSP dataset the number of distribution customer numbers is used and for the DNSP the number of customers is used. There are no available output prices (or output revenue shares) that can be used for building an index of output quantity change. This means that to obtain weights for these non-reliability outputs, one needs to compute a cost function and then apportion cost to the various outputs. For completeness it is important to mention here that, although not used in the Leontief cost function estimation, a reliability output is used in the AER index number exercise. This output is used to provide a quantification of the reliability of the TNSP and DNSP networks and it is obtained using a measure of the value of customer reliability (VCR). The VCR provides a measure of how consumers value energy not supplied.

The current report will focus on providing a full revision of the calculations for the non-reliability output weights based on these two datasets and the specified set of inputs and outputs. We will use one panel dataset with 13 time periods (2006-2018) for 5 TNSP, for a total of 65 observations. And one panel dataset with 13 time periods (2006-2018) for 13 DNSP, for a total of 169 observations.

This report will focus on the first two steps of the AER benchmarking exercise:

- Step 1: estimation of the Leontief cost function parameters via non-linear least squares.
- Step 2: computation of the non-reliability output weights based on the estimated parameters of the Leontief cost function.

Numerical mistakes can happen both in the first and the second step. We therefore conducted a computation of both steps to check the numerical correctness of the non-reliability output weights that are used in the two benchmarking reports on transmission and distribution.

In section 2 we describe the Leontief cost function model and we provide our own set of estimates based on the non-linear least squares facility of STATA. In section 3 the non-reliability output weights are computed and we compare them to the ones obtained by the AER. In section 4 some critical comments on the Leontief cost function method are provided. We point out how the Leontief cost function is very flexible in fitting the data, but may present some challenges in terms of the numerical stability of the parameter estimates obtained via non-linear least squares. In section 5 some potential improvements on the current method are provided. These improvements are intended to provide a potentially more numerically stable set of results, within the current Leontief cost function framework. In this section we also point out the possibility of exploring in the future the use of direct cost benchmarking. In section 7, we respond to feedback received from stakeholder submissions on this report.

2 Estimation of the Leontief Cost Function

Consider a panel dataset of service providers, where there are $k = 1, \ldots, K$ firms (service providers) and there are $t = 1, \ldots, T$ time periods for which the inputs and outputs of the firms are observed. For each firm in each time period, x_i^{kt} is the quantity of input $i = 1, \ldots, M$ used by firm k in time period t, w_i^{kt} the price of input i faced by the same firm, and y_j^{kt} the quantity of output $j = 1, \ldots, N$ produced by the firm. The cost function represents the minimum cost of production that can be achieved for a given level of output, at the prevailing input prices and in a given time period. The following is the definition of the cost function:

$$C(y, w, t) = \min_x \{wx : x \, \text{can produce} \, y \, \text{in time period} \, t\} \tag{1}$$

and the input demands for such a cost function are given by the first derivatvie of this function with respect to input prices (see Färe and Primont (1995) and O'Donnell (2018)). The cost function used by the AER in its benchmarking exercise takes the following specific functional form:

$$C(y^{kt}, w^{kt}, t) = \sum_i w_i^{kt} \sum_j (a_{ij}^k)^2 y_j^{kt}(1 + b_i^k t) \tag{2}$$

where the superscript index kt is identifying firm k in time period t. Notice that the cost function is different for different firms, since the parameters of the cost function (a_{ij}^k, b_i^k) are dependent on k. This is a Leontief cost function, in the sense that no allowance is made for substitution possibilities among inputs in the production of a given output. In other words the inputs are perfect complements. The parameters a_{ij}^k represent the optimal quantity of input i that is required to produce one unit of output j in firm k. The parameters b_i^k are time trend parameters and they account for the shift overtime of these input requirements. Taking the first derivative of the Leontief cost function with respect to prices returns the input demand functions for each firm in each time period:

$$x_i^{kt} = \sum_j (a_{ij}^k)^2 y_j^{kt}(1 + b_i^k t), \, \forall i, k, t \tag{3}$$

Since the quantity of input used on the left hand side and the quantity of output produced (right hand side) are observed, this equation can be estimated in order to obtain values for the parameters of interest (a_{ij}^k, b_i^k). This involves estimating one regression for each input and each firm using T observations. The purpose of taking the square of the coefficients a_{ij}^k is to impose non-negativity constraints on those coefficients. These input demand functions are estimated by Economic Insights using non-linear least squares (nls). It should be stressed that the Leontief coefficients a_{ij}^k depends on i, k, j: this means that these coefficients not only are different for different inputs and outputs, but they also vary across firms. The time trend parameter b_i^k is also different for different inputs and different firms.

In order to check the results of the Economic Insights reports, we have used the non-linear least squares facility in STATA to obtain estimates of the parameters of the Leontief cost function (via the input demand equation estimation). To obtain our estimates, we have set the initial parameter value estimates to the ones reported in the Economic Insights report[2]. We obtained results for these parameters similar to the ones reported in the Economic Insights report. A full set of results that includes the point estimates for each of the coefficients is reported in the excel file associated with this submission. We are therefore confident that the parameter estimates obtained by the Economic Insights report are close to the one we obtained, and well within the tolerance level of standard numerical optimization procedures. The Leontief cost functions can be then recovered by using the estimated values of these parameters.

The estimation of the Leontief cost function parameters represent the first step in the building of MTFP indexes used by the AER and highlighted in the Economic Insights report. From the parameter estimates of the Leontief cost function, it is possible to compute the cost share attributed to each output by the Leontief cost function, and then average these shares to obtain the average share for the whole dataset. We explain this procedure proposed by Economic Insights in the next section and then provide the results of our own computation for the non-reliability output weights.

3 Computation of Non-reliability output weights

The non-reliability output weights are derived from the estimated parameters of the Leontief cost functions. Using a slightly different notation from the one used in the Economic Insight reports, these weights can be computed in the following way:

$$r_j = \frac{\sum_t \sum_k C_j^{kt}}{\sum_t \sum_k \sum_j C_j^{kt}}, \forall j \tag{4}$$

where $C_j^{kt} = \sum_i w_i^{kt}(a_{ij}^k)^2 y_j^{kt}(1 + b_i^k t)$ is the predicted cost associated with output j in firm k and time period t (obtained as the sum over all input costs i). The parameters of the Leontief

[2]We took the numerical value of these estimates approximated to the third decimal place.

cost function are the point estimates obtained via non-linear least squares. These are the weights subsequently used in the index number benchmarking exercise by the AER. One can readily compute also these same weights for each observation in each time period as follows:

$$r_j^{kt} = \frac{C_j^{kt}}{\sum_j C_j^{kt}}, \forall k, t, j \tag{5}$$

The average weight is obtained as a weigthed average of this last set of weigths where the weighting system is the cost share of each observation:

$$\alpha_j^{kt} = \frac{C_j^{kt}}{\sum_t \sum_k \sum_j C_j^{kt}}, \forall k, t, j \tag{6}$$

It is easy to verify that $r_j = \sum_t \sum_k \alpha_j^{kt} r_j^{kt}$. We follow the Economic Insights report and only provide the results of our computations for the average non-reliability output weights r_j. In the excel file associated with this submission, we report the full set of non-reliability output weights r_j^{kt} for all observations in all time periods. Any numerical mistake will affect both the observation specific weights r_j^{kt} and their average r_j.

We computed both the shares in equation (5) and their average in equation (4) using the aforementioned weighting scheme. Our results in terms of the computation of the non-reliability output weights do not differ substantially from the ones found by Economic Insights. As we mentioned in the introduction, we are using a slightly different (updated) dataset compared to the Economic Insighnts computation. Moreover we are using a different non-linear least squares optimizer, which may return slighlty different values for the estimated parameters of the Leontief cost function. Given this numerical differences in the dataset and the estimated coefficients, the resulting differences in the calculated non-reliability output weights are of a negligible order and within the expected tolerance level of numerical optimization.

As discussed above the non-reliability output weights are obtained using the estimated parameters of the Leontief cost function. Since our set of estimates of the Leontief cost function will differ (even if by a small margin due to the numerical optimizer used) from the one used by the AER, it is important to keep in mind that this small differences will carry over to the computation of the non-reliability output weights[3].

3.1 Transmission

	Economic Insights (Original)	Economic Insights (Corrected)	Our Calculation
Energy throughput	23.11%	14.91%	14.95%
Ratcheted max. demand	19.44%	24.71%	24.90%
End-user customer numbers	19.90%	7.59%	7.53%
Circuit length	37.55%	52.79%	52.62%

Table 1: Non-reliability output weights for the TNSP

Table 1 reports the values of the non-reliability output weights for the TNSP. The first column reports the original incorrect value of these weights as computed by Economic Insights (pag. 4, table 1.1) and the second column reports the corrected weights as computed by Economic Insights. Our calculations are reported in the last column. The differences between these sets of numbers are in general not so large. In fact the differences between the second and third columns are small and well within the tolerance level bounds of standard numerical optimization methods. There is a difference of 0.04% in the energy throughput weight; a difference of 0.19% in the weight of the ratcheted maximum demand; a difference of 0.06% in the number of customers; and a difference of 0.17% in the circuit length. These are small differences, especially if contrasted with the reported uncorrected shares from the Economic Insights report which are here shown in the first column. The differences in this case are quite substantial. Energy throughput changes by 8.20%, ratcheted maximum demand by 5.27%, End-user customer numbers by 12.31% and circuit length by a staggering 15.24%.

[3] To be sure, the set of estimates associated with the Leontief cost function parameter estimates will differ from the ones used by the AER due to the choice of the non-linear optimization algorithm used and the various numerical tolerance levels and stopping criteria used by this algorithm.

From this exercise we conclude that the current method of calculation of the non-reliability output weights used by the AER is numerically correct and within the tolerance levels of the non-linear least squares numerical optimizer.

3.2 Distribution

	Economic Insights (Original)	Economic Insights (Corrected)	Our Calculation
Energy throughput	12.46%	8.58%	8.63%
Ratcheted max. demand	28.26%	33.76%	33.79%
End-user customer numbers	30.29%	18.52%	18.26%
Circuit length	28.99%	39.14%	39.33%

Table 2: Non-reliability output weights for the DNSP

Table 2 reports the values of the non-reliability output weights for the DNSP. The first column reports the original incorrect value of these weights as computed by Economic Insights (pag. 4, table 1.1) and the second column reports the corrected weights as computed by Economic Insights. Our calculations are reported in the last column. The differences between these sets of numbers are small and well within the tolerance level bounds of standard numerical optimization methods. There is a difference of 0.05% in the energy throughput weight, a difference of 0.03% in the weight of the ratcheted maximum demand; a difference of 0.26% in the number of customer; and a difference of 0.19% in the circuit length. Again, these are small differences, especially if contrasted with the reported uncorrected shares from the Economic Insights report. The differences in this case are quite substantial. Energy throughput changes by 3.88%, ratcheted maximum demand by 5.50%, End-user customer numbers by 11.77% and circuit length by a 10.15%.

From this exercise we conclude that the current method of calculation of the non-reliability output weights used by the AER is numerically correct and within the tolerance levels of the non-linear least squares numerical optimizer.

4 Some critical comments on the methodological approach used by the AER

In this section we will provide some critical comments on the main advantages and shortcomings of the Leontief cost function model used to determine non-reliability output weights. The major benefit of using the current specification of the Leontief cost function is that it represents a very flexible functional form. The Leontief function is usually regarded as a very inflexible functional form because it requires fixed coefficients of production, therefore not allowing for substitution possibilities among inputs. This would be regarded in practice as a feature that provides poor prediction performance in fitting the data. It should be noted however that the Leontief cost function as used by the AER has a large number of parameters, since the input demand functions are input and firm specific. This provides a level of flexibility that allows a very good fitting to the data, possibly better than the more commonly used translog functional specification. A proper account of the ability of the Leontief cost function to provide a good fit to the data is discussed in the next sub-section.

The main potential shortcoming of the current Leontief cost function method for the computation of non-reliability output weights, mainly comes from the fact that it is based on non-linear least squares (nls). Non-linear least squares requires to solve a non-linear optimization program in order to determine the parameter estimates. This raises two potential issues that do not affect the standard ordinary least square problem. *First*, there can be several local optima in terms of the value of the objective function (the sum of squares). Given any initial parameter value, the resulting optimum found by the solver is not necessarily a global optimum. This problem is normally solved by using a grid of values for the initial parameter estimates and then comparing all the found optimal solutions for each set of initial parameter values. In the case of the AER benchmarking exercise, the Leontief cost function has 5 parameters and therefore a grid search on a 5-dimensional space becomes computationally intractable. The *second* problem is also a numerical problem and is related to the fact that for any given optimal value of the objective function that the solver has found, there may exist several optimal solutions in terms of the parameter values. Therefore the parameter estimates of the coefficients of the Leontief cost function are not necessarily unique. This second problem may affect the computation of the non-reliability output weights if the non-linear least squares problem is solved using different solvers (say STATA, or Shazam, or R, or Matlab). Each solver will likely provide a different set of estimates and it is not clear how one can choose from them. While the potential for these shortcomings has been identified theoretically, it was beyond the requirements of this report to determine if this was the case in the AER benchmarking exercise. This is an area the AER may want to explore further in the future. We will discuss these potential benefits/shortcomings of the Leontief cost function approach in the following sub-sections and we will provide some partial solutions that could be adopted by the AER in section 5.

4.1 Flexibility of the Leontief cost function specification

The Leontief cost function is not estimated directly but, as explained earlier, its input demand functions are estimated one by one independently from each other. Notice that each one of these input demand equations is estimated separately for each input and each firm. There is a total number of 20 input demand equations for the TNSP and a total of 65 input demand equations for the DNSP. Each equation is estimated using the T time periods available for each input and each firm. Given that there is a total number of four outputs, this means that there are 5 coefficients to be estimated using T observations.

Although the Leontief cost function is normally considered an inflexible functional form since it does not allow for substitution possibilities among inputs, it should be noted here that there is a total of $5 \times M \times K = 20 \times K$ number of parameters to fit a cost function for the whole panel dataset. This is because there are 5 parameters (one parameter for each output plus the time trend parameter) for each input demand equation and $M \times K$ of such input demand equations to be estimated (where K is the number of service providers and M the number of inputs). In both the transmission and distribution case, this means that the cost function provides an extremely flexible functional form. To see this, re-write the cost function in the following way:

$$C(y^{kt}, w^{kt}, t) = \sum_j C_j(y_j^{kt}, w^{kt}, t) \tag{7}$$

with

$$C_j(y_j^{kt}, w^{kt}, t) = \sum_i w_i^{kt}(a_{ij}^k)^2 y_j^{kt}(1 + b_i^k t) \tag{8}$$

Notice that an implicit hypothesis is that the cost functions for the different outputs can be separated and therefore cost can be apportioned to each specific output separately. Now, notice that the cost function associated with output j has a total of $M \times K$ parameters for $K \times T$ observations. This means that as long as $M < T$ the model is identified. Moreover it should be noted that if we were to use a translog specification for the cost function associated with output j as a benchmark for flexibility of the functional form, for such a function there would be a total of $\frac{(M+1)(M+2)}{2}$ parameters to be estimated. Since in general $K > M$, the Leontief cost function so specified is more flexible than the translog cost function. In fact, in the transmission case the Leontief output specific cost function will have a total of 25 parameters and the translog function a total of 15 parameters: this means that the Leontief function achieves a higher level of flexibility compared to the translog function. In the distribution case the Leontief cost function will have a total of 65 parameters and the translog function would still have only 15 parameters: both functions are estimated using 169 observations, but the Leontief function will have the potential to achieve much higher flexibility in fitting the data, given the higher number of parameters. The reason why flexibility is important in this context is that the computation of the non-reliability output weights will depend on the ability of the cost function estimates to predict observed cost accurately. A more flexible functional form will provide predicted cost estimates that are closer to the observed ones than a less flexible functional form would, and in this respect the Leontief cost function adopted by the AER will provide a better fit to the data.

4.2 Numerical Stability of the Leontief cost function estimates

The input demand functions derived from the Leontief cost function are estimated using non-linear least squares (nls). This means that one has to solve a non-linear optimization program in order to compute the estimates of the parameters of interest and the predicted total cost of production. There are several numerical challenges in estimating a regression using nls, due to the non-linear nature of its underlying optimization program. The two most important ones are: 1) in non-linear optimization there is no guarantee that one obtains a global optimum[4]; 2) there may be multiple solutions in terms of the underlying parameter values that will support any given optimal solution, meaning that the solution in terms of the parameter estimates may not be unique.

The first problem should not be difficult to solve even in the current modelling of the Leontief cost function and it is highly recommended that the AER investigates in the future the possibility of considering minor modifications to the Leontief cost function specification that can be estimated via more reliable convex programming methods, which would ensure uniqueness and numerical stability of the optimal solution. This is not a point that should be underestimated and the next section will suggest some ways of implementing this.

As for the second point, there is no simple way forward, but a good method would rely on having criteria for choosing among alternative solutions in order to obtain a numerically stable value for the estimates that can be easily reproduced via convex programming. At the very least, the method used should allow to make an assessment of the spread of the solution values. This will be addressed in the next section as well.

[4]Any non-convex optimization program will in general have multiple solutions known as local optima. Among those solutions, one or more can have the optimal (maximal or minimal) value of the objective function to be optimized. These particular local optima are also known as global optima.

5 Alternatives to the current AER approach

Given the Leontief function is already quite a flexible approach that fits the data well, it is unlikely that generilizing this functional form (for example to a constant elasticity of subsitution (CES) function) would provide a significant better fit to the data, although it would be interesting to explore this in more detail in the future. In particular, it would be interesting to explore the possibility of extending this very same framework to what is known as the constant elasticity of subsitution (CES) production function and its associated cost function. The CES function will return both the Cobb-Douglas and the Leontief specifications as special cases, and it could represent a good way of introducing input substitution possibilities and testing them to check if the inputs are substitute or complement. Although appealing from an economic theory perspective, this avenue will share the same numerical problems encountered with the Leontief cost function (i.e. numerical instability and reliance on the non-unique parameter values for the computation of the output weights). Moreover, given the very good fit to the data shown by the Leontief function, it is not clear if the complications associated with this extensions would be justified in terms of returning a better representation of the cost structure of the Australian energy service providers.

However, in light of the issues raised above, some potential improvements to the current Leontief approach can be suggested. These includes:

- modifying the time trend specification to a simple linear time trend;
- modifying the non-linear least squares program to obtain a minimization of absolute deviations (that can be converted easily into a linear program);
- use the linear program associated with the minimization of least absolute deviations to assess the numerical stability of the Leontief parameter values estimates.

While outside the scope of this review, we will also make the additional point that the AER could consider alternatives to the productivity index number (PIN) methodology that do not require the calculation of output weights, but rather use direct computation of the cost functions to be used for benmchmarking.

The Leontief cost function for observation $k = 1, \ldots, K$ in time period t is the following:

$$C(y^{kt}, w^{kt}, t) = \sum_i w_i^{kt} \sum_j (a_{ij}^k)^2 y_j^{kt} (1 + b_i^k t) \tag{9}$$

where i is indexing the inputs and j is indexing the outputs. The input demands are:

$$x_i^{kt} = \sum_j (a_{ij}^k)^2 y_j^{kt} (1 + b_i^k t), \; \forall i, k, t \tag{10}$$

and these are used to estimate the regression models. For each regression model associated with one particular input demand function, the set of parameter estimates will be obtained by minimizing a least squares criteria, therefore by solving the following optimization program for each regression (for each firm k and each input i):

$$\min_{a_{ij}^k, b_i^k} \sum_t \left[x_i^{kt} - \sum_j (a_{ij}^k)^2 y_j^{kt} (1 + b_i^k t) \right]^2 \tag{11}$$

This is a non-linear optimization program. The non-linearity of this program is coming from: the squared parameters a_{ij}^k; from the fact that the time trend coefficient is multiplying these squared parameters. Currently the AER solves this least squares problem using non-linear least squares. An alternative to this method can be implemented by noting that for a given value of b_i^k the non-linearity of the program derives only from the squared Leontief parameters. By defining $z_j^{kt} = y_j^{kt}(1 + b_i^k)$, one can solve the following constrained quadratic program for a given selected value of b_i^k:

$$\begin{aligned} \min_{a_{ij}^k} \quad & \sum_t \left[x_i^{kt} - \sum_j a_{ij}^k z_j^{kt} \right]^2 \\ s.t. \quad & a_{ij}^k \geq 0, \forall j \end{aligned} \tag{12}$$

This program is a quadratic program, in particular a non negative least squares problem, therefore admitting a unique solution for a each given value of the parameter b_i^k. A grid search over the parameter b_i^k can be used to return the optimal solution. This provides a more suitable alternative to the use of non-linear least squares. Moreover, inspection of the optimal value of the objective function for various values of b_i^k will also inform on possible multiple solutions in terms of this parameter, thus representing a good robustness exercise. The question remain if the Leontief specification can be further simplified to allow for a simplified estimation strategy.

5.1 Simplification of the time trend specification

A first step towards the simplification of the optimization program linked to the least squares estimation of the Leontief cost function relates to the treatment of the time trend. It should be immediately noted that it is not clear if a time trend should be included at all. The effect of the time trend specification is in most cases to adapt the cost function with changes in time. A negative sign for this coefficient means that cost is decreasing and the input demand is also decreasing over time, therefore signalling technical progress (i.e. a more efficient use of resources overtime). A positive sign of the coefficient b_i^k means that cost is increasing in time and the input demand increasing as well, signalling that there is technical regress (a deterioration of the efficiency of use of resources overtime). Although one may want to make allowance for the possibility of technical progress, it is not clear why one should allow for technical regress. In fact, it is not clear what would cause a deterioration in the technology used to produce energy, apart from a more inefficient use of resources. Allowing the possibility of technical regress in the current specification may represent a form of cost plus allowance for those service providers that have increased their cost overtime. Alternatively, one may interpret a negative sign of this coefficient to reflect cost changes over time due to other factors that are not explicitly specified in the model or related to technical change. This may include changes in the regulatory environment (e.g. higher vegetation management responsibility) or changes associated with the penetration of distributed energy resources which may increase the cost of maintaining the network.

Apart from these interpretational issues, on a more technical note, the effect of time on the input demands is given by the derivative of the input demand function with respect to time:

$$TC^{kt} = b_i^k \sum_j (a_{ij}^k)^2 y_j^{kt} \tag{13}$$

This quantity is the effect of the time trend on the input quantity demanded: if it is positive, it provides the additional input quantity needed to sustain production; if it is negative, it provides the savings in the input quantity that would still sustain production. This quantity depends on two factors: the time-invariant coefficient b_i^k; the total output production $\sum_j (a_{ij}^k)^2 y_j^{kt}$. Since the Leontief coefficients a_{ij}^k are time-invariant, the time variation in the second quantity comes from the variation in time in the output volume y_j^{kt}. And since in the current dataset used in the AER benchmarking exercise these quantities do not vary in time substantially, the effect of the time trend on the input quantities is more or less constant in practice. And since TC is more or less constant, one can include it directly as a linear trend by writing the following input demand function:

$$x_i = \beta_i^k t + \sum_j (a_{ij}^k)^2 y_j \tag{14}$$

derived by its associated cost function:

$$C(y^{kt}, w^{kt}, t) = \sum_i w_i^{kt} \sum_j (a_{ij}^k)^2 y_j^{kt} + \sum_i w_i^{kt} \beta_i^k t \tag{15}$$

The derivative of the demand function with respect to time will return a constant time trend of β_i^k. Notice that $\beta_i^k \neq b_i^k$, but it will be approximately equal to TC^{kt} in the current dataset. There will be a small loss of fitting (due to the fact that the TC component will be assumed constant instead of being allowed to be time varying), but this is the more negligible the more stable the TC^{kt} is in its time variation (in the original specification). Moreover, there is a theoretical argument in favour of a simpler specification, in the sense that a simpler model should be preferred to a more complicated one (parsimony principle), unless the more complicated one provides a much better fitting to the data. Our analysis points to the fact that a simpler time trend specification is justified in the case of the AER benchmarking exercise.

Implementing this small change in the functional form specification of the time trend will substantially decrease the computational complexity of the model. The regression estimates of the parameters will now be the outcome of the following least squares program:

$$\min_{a_{ij}^k, \beta_i^k} \sum_t \left[x_i^{kt} - \beta_i^k t - \sum_j (a_{ij}^k)^2 y_j^{kt} \right]^2 \tag{16}$$

This is still a non-linear optimization program, but the non-linearity is only arising from the square of the Leontief coefficients a_{ij}^k. Since the purpose of taking the square of these coefficients is justified on the ground that the Leontief coefficients should not be negative, one can equivalently re-write the program as a constrained program:

$$\begin{aligned} \min_{\alpha_{ij}^k, \beta_i^k} \quad & \sum_t \left[x_i^{kt} - \beta_i^k t - \sum_j \alpha_{ij}^k y_j^{kt} \right]^2 \\ s.t. \quad & \alpha_{ij}^k \geq 0, \, \forall j \end{aligned} \tag{17}$$

These two simple modifications have now transformed the non-linear least squares problem into a quadratic program that can be easily solved using standard solvers. Since quadratic programming is a convex optimization program, this also means that this strategy will return a unique solution in terms of the objective function, therefore solving the problem intrinsic in non-linear optimization of not being able to rule out if the solution is a local or a global optimum. Moreover, due to the much more simple optimization, this also means that the program solution will always be found, contrary to non-linear algorithms that may fail convergence. Finally, the solution of this program will also be independent from any grid search or initial values for the parameter estimates. All in all, these minor modifications will solve the first computational problem highlighted in the previous section. Even if this problem is solved, there is still the second problem of having potentially many alternative optimal values in terms of the parameter estimates. This will be discussed in the next sub-section. It should also be emphasized here that since the program is a quadratic one, it is possible to include additional constraints to prevent technical regress from happening, if one wishes to do so.

5.2 Minimization of Mean Absolute Deviations

The previous sub-section has suggested a minimal change to the way in which time trend is included in the input demand function that permits to write the associated least squares problem as a quadratic program. Although this solves the problem of potentially having a local optima, by guaranteeing always convergence to a global optimum, it does not solve the problem of the multiplicity of solutions in terms of the parameters of the Leontief cost function a_{ij}^k. Although this problem cannot be eliminated in the current framework adopted by the AER for benchmarking, some additional modifications can be implemented to at least quantify the extent to which this multiplicity of solutions may affect the non-reliability output weights. In the conclusion section of this report, we will highlight how one could avoid this problem altogether by computing the cost benchmark using index numbers derived directly from a cost function.

In order to take a further step in the direction of simplification, one could change the minimization criteria from least squares to sum of absolute deviations. The absolute deviation is $|x_i^t - \beta_i t - \sum_j \alpha_{ij} y_j^t|$ and it corresponds to the absolute value of the deviation of the observed value of the dependent variable from the predicted one. The program associated with the minimization of absolute deviations will return the following:

$$\begin{aligned} \min_{\alpha_{ij}^k, \beta_i^k} \quad & \sum_t \left| x_i^{kt} - \beta_i^k t - \sum_j \alpha_{ij}^k y_j^{kt} \right| \\ s.t. \quad & \alpha_{ij}^k \geq 0, \, \forall j \end{aligned} \tag{18}$$

This program can be transformed and solved as a linear program and it will coincide with a quantile regression that is looking at the expected median instead of the expected mean. The associated linear program will take the following form (where we introduce the additional decision variable u_t):

$$\begin{aligned} \min_{\alpha_{ij}^k, \beta_i^k, u_t} \quad & \textstyle\sum_t u_t \\ s.t. \quad & \alpha_{ij}^k \geq 0, \, \forall j \\ & x_i^{kt} - \beta_i^k t - \textstyle\sum_j \alpha_{ij}^k y_j^{kt} \leq u_t, \, \forall t \\ & \beta_i^k t + \textstyle\sum_j \alpha_{ij}^k y_j^{kt} - x_i^{kt} \leq u_t, \, \forall t \end{aligned} \tag{19}$$

We are now in a position to assess the numerical stability of the coefficient estimates. Before discussing this in the next section, it is also interesting to notice that one can make another step and look at this same program as an efficiency program, by making deviations one sided and de-facto allowing for inefficiency in the use of inputs:

$$\begin{aligned} \min_{\alpha_{ij}^k, \beta_i^k} \quad & \sum_t \left(x_i^{kt} - \beta_i^k t - \sum_j \alpha_{ij}^k y_j^{kt} \right) \\ s.t. \quad & \alpha_{ij}^k \geq 0, \, \forall j \\ & x_i^{kt} - \beta_i^k t - \textstyle\sum_j \alpha_{ij}^k y_j^{kt} \geq 0, \, \forall t \end{aligned} \tag{20}$$

The difference between the observed values for the inputs and the predicted values from this last program, will return a measure of inefficiency in the use of inputs, therefore signalling how far the observation is from the Leontief vertex given by the Leontief coefficients.

5.3 Assessing the numerical stability of the coefficient estimates

The numerical stability of the estimated Leontief coefficients can be quantified in the following way. Take the objective function optimal solution of program (19) and call it U^*. We can now search among all the parameters that support this optimal solution, the smallest and the largest for each one of them. Consider one particular α_{ij}^k coefficient. To find its minimal possible value one solves the following linear program:

$$\begin{aligned} \min_{\alpha_{ij}^k, \beta_i^k, u_t} \quad & \alpha_{ij}^k \\ s.t. \quad & \alpha_{ij}^k \geq 0, \forall j \\ & x_i^{kt} - \beta_i^k t - \textstyle\sum_j \alpha_{ij}^k y_j^{kt} \leq u_t, \forall t \\ & \beta_i^k t + \textstyle\sum_j \alpha_{ij}^k y_j^{kt} - x_i^{kt} \leq u_t, \forall t \\ & \textstyle\sum_t u_t = U^* \end{aligned} \tag{21}$$

The last three constraints of this program guarantee that the solution will be in fact a solution that provides the same value for the objective function of program (19). In practice, we are constraining this program to look among alternative optimal values of the parameters, the one that is the minimal for α_{ij}^k. Call the solution to this program m_{ij}^k. Similarly, one could solve the following program to find out the maximal possible value that α_{ij}^k may take at the optimal:

$$\begin{aligned} \max_{\alpha_{ij}^k, \beta_i^k, u_t} \quad & \alpha_{ij}^k \\ s.t. \quad & \alpha_{ij}^k \geq 0, \forall j \\ & x_i^{kt} - \beta_i^k t - \textstyle\sum_j \alpha_{ij}^k y_j^{kt} \leq u_t, \forall t \\ & \beta_i^k t + \textstyle\sum_j \alpha_{ij}^k y_j^{kt} - x_i^{kt} \leq u_t, \forall t \\ & \textstyle\sum_t u_t = U^* \end{aligned} \tag{22}$$

If we call the solution of this program M_{ij}^k, then it must be that for each Leontief coefficient, its optimal value must be between these bounds: $m_{ij}^k \leq \alpha_{ij}^k \leq M_{ij}^k$. In practice if these numerical bounds are tight, then this will guarantee that the associated non-reliability output weights will be numerical stable and similar across alternative optimal solutions. If these bounds are large, then the associated non-reliability output weights may vary significantly.

5.4 Direct Cost Benchmarking

Although potential numerical problems may exist in the current Leontief cost function framework adopted by the AER benchmarking exercise, they could be overcome altogether by moving to what shall be defined as direct cost benchmarking. This means using the cost function to directly compute the major driver of costs and in so doing providing a guidance to set appropriate price caps.

While beyond the requirements of this project, which were to explore other feasible approaches to determining the non-reliability weights, this is something that the AER may want to consider at some point in the future when it is more broadly reviewing its benchmarking exercise methodology.

Direct cost benchmarking would involve computation of the cost function using mathematical programming methods such as data envelopment analysis (DEA). In such a case, one would estimate a cost function as for step 1 of the AER benchmarking approach, although this would be done with a different method that does not suffer from the potential numerical problems that may affect the current Leontief approach adopted by the AER and described above. But then, instead of using the cost function to estimate output weights in step 2 (and Tronqvist indexes in step 3), one would directly compute a productivity change index using the cost function. Since one is using the estimated cost function to directly compute the productivity index, we shall refer to this method as a direct cost benchmarking. There could be two main advantages of using this method: the first one would be that the resulting productivity index would not suffer from the potential numerical instability that has been identified as a feature of the current AER benchmarking approach; the second one would be that one has the ability of decomposing this productivity index into components related to the underlying causes of cost change. In particular one could identify two main effects: the effect on cost of an increase in input prices; and the effect on cost of an increase in the volume of outputs. Such an approach would help in further understanding the drivers of cost change for the energy service providers.

6 Conclusion

As explained in the previous sections the AER benchmarking approach makes use of what can be described as a three step procedure:

1. estimate the specified Leontief cost function (via non-linear least squares);
2. use the parameter estimates from the Leontief cost function to derive non-reliability output weights;
3. use the so obtained non-reliability output weights to build Tornqvist indexes of productivity change.

The current report has focused on the first two steps of this procedure by providing an independent review of all the numerical computations conducted by the AER at the first and second step. We found that our own computations of the non-reliability output weights are very close to the ones computed by the AER, where the differences are within the normal tolerance level that one would expect from non-linear optimization procedures. This points to the fact that the current AER method of computing non-reliability output weights is substantially correct.

In section 4 it has been explained why estimating the Lenontief cost function via non-linear least squares may lead to numerically unstable results, with multiple solutions for the estimates of the coefficients that may lead to alternative set of weights for the non-reliability outputs. In section 5 some possible improvements of the current methodology have been proposed that may improve the numerical stability of the results without changing the underlying 3-step approach to the computation of the Tornqvist productivity indexes and with minimal changes to the specification of the Leontief cost function. In these sections it was discussed how the Leontief cost function currently employed by the AER represents a very flexible functional form due to the relatively large number of parameters used in its definition. Finally, it has been discussed various possible modification of the current setting used by the AER. One of these possibilities is the use of what we have called direct cost benchmarking, a method that relies on mathematical programming to compute directly efficient cost and its main drivers.

7 APPENDIX: Responses to stakeholder submissions.

In this appendix we respond to some critical comments raised by some service providers on a draft of this report.

7.1 Jemena

Jemena made the following three major comments on the draft report provided by CEPA:

1. a suggestion to use more recent data for the computation of the output weights;
2. a comment on the impact of the revised capitalization approach on output weights and inconsistency between input and output weights due to different capitalization methods;
3. introduction of a fixed cost component in the cost function.

The first two points raised by Jemena were outside the scope of the work undertaken by CEPA, as agreed with the AER. CEPA was asked to conduct a review of the current methodology using the same data that were used to compute the last round of output weights. The main purpose of the analysis undertaken by CEPA was to check that the computation of this output weights was done correctly. Although outside the scope of the analysis undertaken by CEPA, our assessment of these first two points raised by Jemena is in general supportive, in the sense that the use of more recent data and a homogenization of the capitalization methods for both the inputs and outputs would be advisable.

The third point raises an important question about the presence of fixed costs in the production process. In the current specification of the Leontief cost function there is no allowance for such fixed costs. We note however that the proposal of including a fixed cost component is in the same direction as the CEPA report proposal of linearizing the time trend component in the Leontief cost function. In fact, this means adding an intercept to the regressions of the input demand functions defined in equation (14). The other two service providers have raised concerns over the introduction of this time trend specification and implicitly over the introduction of the type of fixed costs proposed by Jemena. One of the reasons for this concern raised by Ausgrid and Evoenergy is that the definition of the output weights needs to be revised if such an approach is adopted. In fact, by adding an intercept to the input demand regressions (or by linearizing time trends), one is violating the cost separability assumption that underlies the current Leontief framework, which allows to apportionate overall cost to the separate outputs. It should be noted however that in this alternative specification the cost function would take the following form:

$$C(y^{kt}, w^{kt}, t) = \sum_i w_i^{kt} \sum_j (a_{ij}^k)^2 y_j^{kt} + \sum_i w_i^{kt} \beta_i^k t + \sum_i w_i^{kt} \gamma_i^k \tag{23}$$

Even in this specification, output weights could be computed by using the separable part of the cost function $\sum_i w_i^{kt} \sum_j (a_{ij}^k)^2 y_j^{kt}$. Since the fixed costs are paid irrespective of the level of production, this provides a good rationale to do so. We are in general supportive of the proposal of including an intercept in the input demand equations.

7.2 Evoenergy and Ausgrid

Evoenergy and Ausgrid raised the following main critical points about CEPA's recommendations:

1. agreed on the fact that the linearization of the time trend would simplify computation, but at the cost of loosing the Leontief interpretation.

2. support for the proposal of introducing a quadratic programming specification approach to estimate the parameters of the Leontief cost function;

3. a concern about the potential use of the LAD approach, by claiming that its main use should be confined to situations with the presence of outliers in the data;

4. a general criticism of direct cost benchmarking based on the fact that if adopted direct cost benchmarking would deny service providers the possibility of tracking their own performance relative to other service providers.

With regard to the first point on the linearization of the time trend, we notice that a Leontief cost function requires that the production function is of the fixed coefficients type, which indeed implies that the associated dual cost function is linear in the input prices. Notice that there is (to the best of our knowledge) no common accepted strategy for the introduction of time trends in such a specification. The cost specification proposed by CEPA can still be interpreted as a Leontief cost function, since it is linear in prices and its first derivative provides input demand functions that depend on the level of output only (i.e. they are independent of input prices as required by the Leontief specification). In equation (15) we now report the cost function associated with such a specification.

The second point raised by Evoenergy is supportive of CEPA proposal of using quadratic programming as a way of avoiding numerical problems that may be encountered by non-linear least squares. For a given value of the time trend coefficient (in the current non-linear specification adopted by the AER), the optimization program becomes a quadratic program, or more specifically a non-negative least squares problem. This admits a unique solution in terms of the parameter estimates for the given value of the time trend parameter. One could then conduct a grid search over alternative values of the time trend parameter and choose the one that minimizes least squares, thus achieving a global optimum. Since the grid search is conducted over one parameter only (the time trend) and since non-negative least squares is computationally fast, this becomes a computationally tractable strategy. We notice, however, that there may still exists (in principle) alternative points on the grid that may deliver a minimum, therefore uniqueness of the solution cannot be guaranteed, although this would allow to identify these alternative solutions. This would still represent a valuable alternative to non-linear least squares, since multiplicity of solutions can be easily checked and the computational implementation would be faster and more accurate.

The third point raised by the service providers is in regard to the use of least absolute deviations. It is true, as pointed out correctly by the service providers, that a least absolute deviations problem is less prone to be influenced by the presence of outliers. The introduction of least absolute deviations in section 5.3 of the report was not done with the intention of dealing with outliers. It was done with the intention of illustrating another alternative way in which the analysis could be conducted. The subsequent analysis on the sensitivity of the solution in terms of the decision variables is somehow standard in linear programming and may help in establishing numerical bounds on such coefficients. These bounds can then be potentially used to check the sensitivity of the output weights to alternative optimal solution values for of the Leontief coefficients.

Finally, in the fourth point raised concerns over the use of direct cost benchmarking. We first acknowledge that this section of the report (5.4) was beyond the scope of this report. We notice however that a direct cost benchmarking approach allows for the same sort of analysis that is currently conducted using the PIN methodology. This means that each service provider would be provided with the same information as currently done in the AER benchmarking exercise.

Evoenergy and Ausgrid also raised the following of points that were not directly addressed in the CEPA report:

1. a suggestion to have a more frequent update of the output weights;
2. the possible presence of multicollinearity and its impact on the computation of output weights;
3. the potential non-linear Opex changes in time that are not allowed in the current AER specification, as well in the various specifications proposed by CEPA.

The first point raises a general question about the frequency of updates of the output weights. This was outside the scope of the report for which CEPA was engaged by the AER.

The second point raises concerns over the presence of multicollinearity. The main effect of multicollinearity in a regression model is to inflate the standard errors of the estimates. How this affects the computation of the output weights is unclear, in the sense that the standard errors of the cost function parameter estimates are not used to assess the sensitivity of the output weights with respect to these standard errors. We notice however that the AER is using the standard practice of computing the output weights using point estimates, which is the generally accepted practice in econometrics. Using standard errors to provide measures of uncertainty around the output weights is complicated and outside the scope of the current report. Moreover, the possible use of additional years of data in future rounds may help to alleviate the problem.

The third point about non-linearities in the evolution of the Opex change actually points to the fact that direct cost benchmarking would help mitigate this problem. The only alternative solution in the current framework is to include additinal components in the time trend specification. The Leontief cost function model in its current specification is already extremely flexible (as pointed out in the report) and the inclusion of additional parameters may cause overfitting.

References

Färe, R. and Primont, D. (1995). *Multi-OutputProduction and Duality: Theory and Applications.* Kluwer Academic Publishers.

Lawrence, D., Coelli, T., and Kain, J. (2020a). Economic benchmarking results for the australian energy regulator's 2020 dnsp annual benchmarking report. *Economic Insights Report.*

Lawrence, D., Coelli, T., and Kain, J. (2020b). Economic benchmarking results for the australian energy regulator's 2020 tnsp annual benchmarking report. *Economic Insights Report.*

O'Donnell, C. J. (2018). *Productivity and efficiency analysis.* Springer.