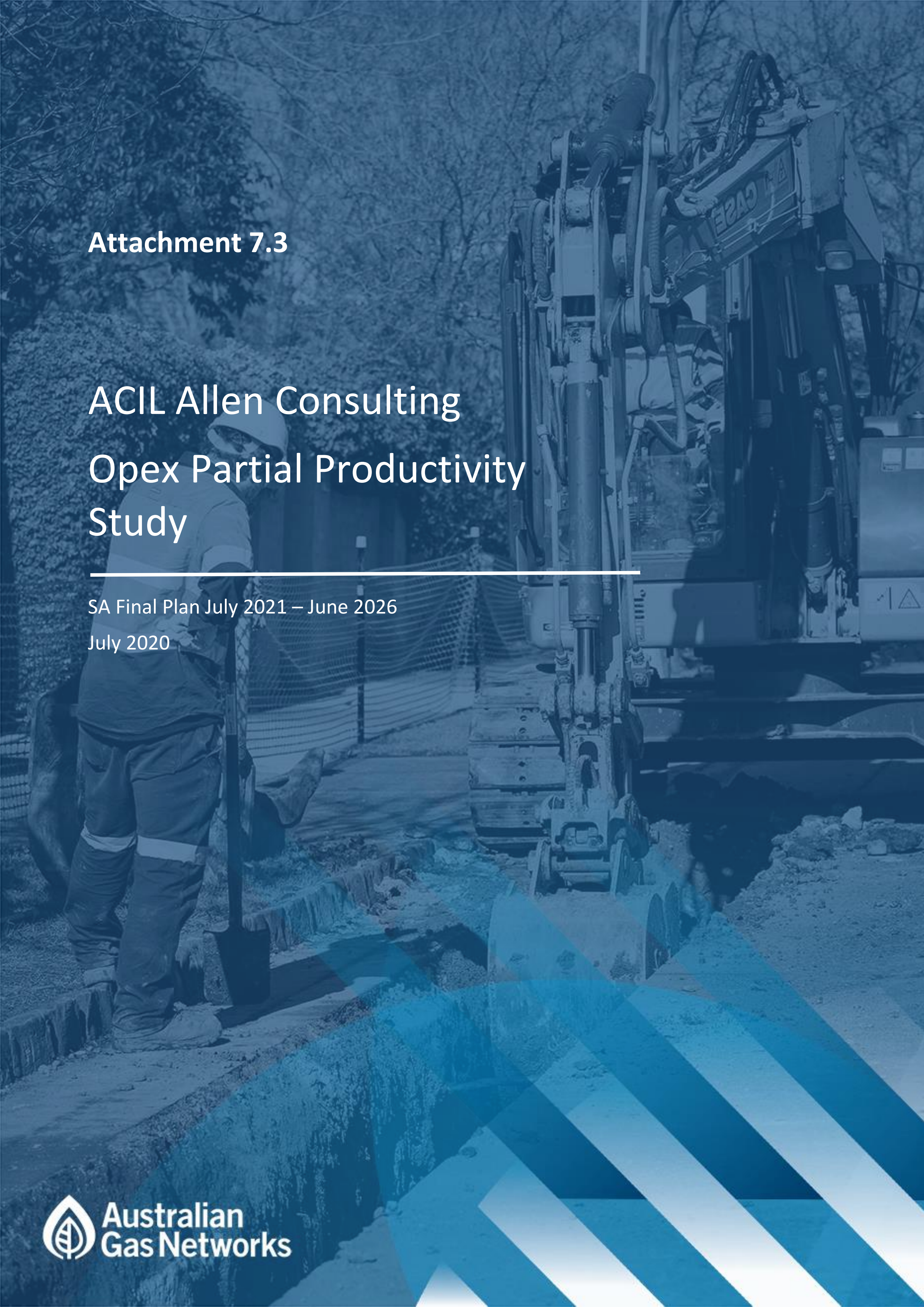

Attachment 7.3

ACIL Allen Consulting

Opex Partial Productivity Study

SA Final Plan July 2021 – June 2026

July 2020

Australian Gas Networks

REPORT TO
AUSTRALIAN GAS NETWORKS LIMITED
21 MAY 2020

OPEX PARTIAL PRODUCTIVITY STUDY

AUSTRALIAN GAS NETWORKS LIMITED

ACIL ALLEN CONSULTING PTY LTD
ABN 68 102 652 148

LEVEL NINE
60 COLLINS STREET
MELBOURNE VIC 3000
AUSTRALIA
T+61 3 8650 6000

LEVEL NINE
50 PITT STREET
SYDNEY NSW 2000
AUSTRALIA
T+61 2 8272 5100

LEVEL FIFTEEN
127 CREEK STREET
BRISBANE QLD 4000
AUSTRALIA
T+61 7 3009 8700

LEVEL SIX
54 MARCUS CLARKE STREET
CANBERRA ACT 2601
AUSTRALIA
T+61 2 6103 8200

LEVEL TWELVE
28 THE ESPLANADE
PERTH WA 6000
AUSTRALIA
T+61 8 9449 9600

167 FLINDERS STREET
ADELAIDE SA 5000
AUSTRALIA
T +61 8 8122 4965

ACILALLEN.COM.AU

1

INTRODUCTION 2
1.1 Scope of work 2
1.2 Report structure 3

2

COST FUNCTION ANALYSIS 4
2.1 Overview of cost function analysis 4
2.2 Model functional form 4
2.3 Econometric approach 6
2.4 Estimation techniques 8

3

PRODUCTIVITY STUDY DATA 11
3.1 Sample of gas distribution businesses 11
3.2 Data limitations and issues 12
3.3 Data definitions 13

4

ESTIMATED COST FUNCTIONS 15
4.1 Cost functions estimated 15
4.2 Model validation and testing 15
4.3 Two output specification: Energy and Customer numbers 17
4.4 Two output specification: Customer numbers and mains length 18

5

OPEX PARTIAL PRODUCTIVITY FORECASTS 20
5.1 Inputs to calculating opex partial productivity 20
5.2 Calculating opex partial productivity 22
5.3 Opex partial productivity forecasts 22

TABLES

TABLE 4.1 ESTIMATED COBB-DOUGLAS FUNCTION: TWO OUTPUTS: ENERGY THROUGHPUT AND CUSTOMERS 17
TABLE 4.2 ESTIMATED COBB-DOUGLAS FUNCTION: TWO OUTPUTS: CUSTOMER NUMBERS AND MAINS LENGTH 18
TABLE 5.1 TWO OUTPUT OPEX COST FUNCTION REGRESSION ESTIMATES WITH ENERGY AND CUSTOMER NUMBERS AS OUTPUTS 20
TABLE 5.2 TWO OUTPUT OPEX COST FUNCTION REGRESSION ESTIMATES WITH CUSTOMER NUMBERS AND MAINS LENGTH AS OUTPUTS 21
TABLE 5.3 FORECAST CHANGES IN GROWTH DRIVERS: CUSTOMERS AND ENERGY THROUGHPUT MODEL 21
TABLE 5.4 FORECAST CHANGES IN GROWTH DRIVERS: CUSTOMERS AND MAINS LENGTH MODEL 21
TABLE 5.5 ANNUAL OPEX PARTIAL PRODUCTIVITY FORECASTS, CUSTOMERS AND ENERGY OUTPUTS, FGLS MODEL 23
TABLE 5.6 ANNUAL OPEX PARTIAL PRODUCTIVITY FORECASTS, CUSTOMERS AND ENERGY OUTPUTS, SFA MODEL 23
TABLE 5.7 ANNUAL OPEX PARTIAL PRODUCTIVITY FORECASTS, CUSTOMERS AND MAINS LENGTH OUTPUTS, FGLS MODEL 23
TABLE 5.8 ANNUAL OPEX PARTIAL PRODUCTIVITY FORECASTS, CUSTOMERS AND MAINS LENGTH OUTPUTS, SFA MODEL 24

1 INTRODUCTION

ACIL Allen Consulting (ACIL Allen) has been engaged by Australian Gas Networks Limited (AGN) to provide productivity analysis in support of the preparation of AGN's Access Arrangement (AA) proposal for South Australia for the period from 1 July 2021 to 30 June 2026.

This expert report has been prepared to assist AGN to develop expenditure forecasts to be included in the AA proposal. Under the Terms of Reference for the study, ACIL Allen has been asked to provide a forecast of the operating expenditure (opex) partial productivity growth rate that applies to AGN's South Australian networks for the period 2021 to 2026.

This involves estimating an opex cost function which is used to calculate an opex partial productivity growth rate forecast split into three components: technology, returns to scale and operating environment.

In conducting the analysis ACIL Allen has had regard to:

— historical and forecast cost, input and output data provided by AGN
— publicly available information from other gas distribution businesses, such as regulatory submissions, regulators' final decisions, and annual reports

In this study we estimate the rate of technical change as the average across four separate econometric models, utilising two separate estimation techniques, Feasible GLS and Stochastic Frontier Analysis (SFA) as well as two sets of outputs, one with customer numbers and energy throughput as outputs and another with customer numbers and mains length as outputs.

The average estimate of the opex productivity growth factor is estimated in this study to be **0.17** per cent per annum.

1.1 Scope of work

The scope of work undertaken can be summarised as follows:

1. Update the database used to estimate the cost functions
   - Data was obtained from AGIG for its gas distribution businesses, AGN SA, AGN VIC and Multinet
   - Data was also be obtained from public sources such as the AER website, individual distribution businesses' websites and annual reports
2. Estimate the cost functions with the updated data, using FGLS and SFA estimation methodologies
3. Calculation of forecast opex partial productivity and output growth over the forecasting horizon with the preferred calibrated models (using AGIG's forecasts of the inputs for the next regulatory period)
4. Preparation and deliver a final report outlining our findings.

1.2 Report structure

The report is structured as follows:

— section 2 provides an overview of cost function analysis including a description of the possible functional forms and estimation techniques used to estimate cost functions
— section 3 describes the data used in the partial productivity analysis including their limitations and deficiencies
— section 4 presents the set of estimated opex cost functions and describes the model selection and validation process
— section 5 calculates the opex partial productivity forecasts based on the most appropriate estimated opex cost functions

2 COST FUNCTION ANALYSIS

2.1 Overview of cost function analysis

In production economics, econometric cost functions provide a useful tool for estimating the least cost means of production and can be used to explore efficient costs for energy networks. A cost function is a function that measures the minimum cost of producing a given set of outputs in a given production environment in a given time period.

By modelling the output quantities, the input prices, and the operating conditions in which the business operates, a minimum-cost function (a theoretical concept), yields an estimate of the periodic costs incurred by an efficient business to deliver those services in that environment.

Let $x = (x_1, \dots \dots, x_M)'$, $w = (w_1, \dots \dots, w_M)'$, $q = (q_1, \dots \dots, q_N)'$ and $z = (z_1, \dots \dots, z_J)'$ denote vectors of input quantities, input prices, output quantities and environmental variables respectively. Mathematically, the cost function is defined as:

$$c(w, q, z, t) = \min_{x \geq 0} \{w'x : x\ can\ produce\ q\ in\ an\ environment\ characterised\ by\ z\ in\ period\ t\}$$

2.2 Model functional form

To facilitate the estimation of econometric cost functions it is necessary to assume a functional or algebraic form that can approximate the unknown, theoretical cost function. The functional form imposes certain assumptions, which may be more or less strict, about the relationships between model variables (outputs, input prices and costs) including relating to economies of scale and elasticities of substitution and hence the shape of the underlying cost function.

The desirable features of a functional form are as follows:

- it captures the underlying technology of an industry adequately
- it is non-decreasing in prices
  - i.e. if prices increase then costs increase
- it is non-decreasing in outputs
  - i.e. if outputs increase then costs increase
- homogeneity in prices
  - i.e. if you double prices, you double costs
- it has a smooth function
- linearity in its parameters.

The functional forms applied most commonly in the econometric cost benchmarking literature are:

- Cobb-Douglas: a linear in logs functional form that makes relatively stricter assumptions about the functional form
- Translog: a flexible functional form that allows for linear, quadratic and interaction terms in the logarithms of the output quantity and input price variables.

In general, increased flexibility in the functional form may be desirable in terms of more closely reflecting reality and allowing for a greater range of possible estimated outcomes. However, the more flexible forms such as a Translog cost function require estimation of a large number of parameters which may introduce econometric problems (e.g. multicollinearity).

A range of practical criteria are typically used to determine the functional form used including reducing estimation problems (including multicollinearity and loss of degrees of freedom when sample size is small), ease of interpretation (some functional forms have an intrinsic and intuitive economic interpretation and in which the functional structure is clear) and computational ease.[1]

The choice of an appropriate functional form is of vital importance in the estimation of an econometric cost function. If the functional form applied is not appropriate, then the estimated cost function will be mis-specified and any associated opex productivity forecasts will be unreliable.

Cobb-Douglas function

The Cobb-Douglas function assumes a log-linear functional form where the natural logarithm of opex is linear in the logarithm of the output quantities and the input price.

For a Cobb-Douglas function with:

- two output variables:
  - energy throughput (E)
  - customer numbers (C)
- two input variables:
  - capital services proxied by the constant price RAB (R)
  - opex price (P)
- a single operating environment variable, customer density (CD)
- a time trend capturing technological changes

the function takes the form:

$$\ln(Opex) = a + b_1 Time + b_2 \ln(E) + b_3 \ln(C) + b_4 \ln(P) + b_5 \ln(R) + b_6 \ln(CD)$$

To ensure homogeneity in prices, the coefficient on the opex price variable (P), b_4 is restricted to equal 1. This is dealt with in the estimation process by subtracting ln(P) from both sides of the equation so that the dependent variable in the regression becomes ln(Opex) minus ln(P) and the price variable disappears from the right hand side of the equation.

The Cobb-Douglas function imposes a constant elasticity of opex to each of the outputs regardless of the scale of the business. From the above specification, this implies that a 1 per cent increase in customer numbers (C) will result in a b_3 per cent increase in opex, regardless of whether the firm is large or small.

The sum of the coefficients of the output variables gives an indication of the type of returns to scale present in the sample. If the coefficients b_2 and b_3 sum to less than 1, operating costs increase at a slower rate than the outputs, implying increasing returns to scale. We would expect this to be the case for gas distribution businesses. This is because gas distribution businesses benefit from increasing efficiency resulting from economies of scale as they move from smaller to larger scale production.

The Cobb-Douglas functional form is useful to the extent that it reflects the underlying production technology of gas distribution businesses which are subject to increasing returns to scale. This functional form has been applied in a number of previous studies of gas distribution businesses.

[1] Fuss, M, McFadden D. and Mundlak, Y,"A Survey of Functional Forms in the Economic Analysis of Production" in Fuss, M and McFadden D. (Eds) (1978), *Production Economics: A Dual Approach to Theory and Applications*.

Translog functional form

The Translog functional form allows for linear, quadratic and interaction terms between the output and input variables.

The Translog is an example of a flexible functional form which is considerably less restrictive than the Cobb-Douglas. It allows for linear, quadratic and interaction terms in the natural logarithms of output and input prices. Its main advantage over the Cobb-Douglas is that it allows the degree of returns to scale to vary with firm size, something that the Cobb-Douglas does not allow.

Extending the two output, two input and single environmental variable Cobb-Douglas function above to the more flexible Translog function, we get:

$$\ln(Opex) = a + b_1 Time + b_2\ln(E) + b_3\ln(C) + b_4\ln(P) + b_5\ln(R) + b_6\ln(CD) + 0.5b_{22}\ln(E)^2 + b_{23}\ln(E)\ln(C) + 0.5b_{33}\ln(C)^2 + 0.5b_{44}\ln(P)^2 + 0.5b_{55}\ln(R)^2 + 0.5b_{66}\ln(CD)^2 + b_{24}\ln(E)\ln(P) + b_{25}\ln(E)\ln(R) + b_{26}\ln(E)\ln(CD) + b_{34}\ln(C)\ln(P) + b_{35}\ln(C)\ln(R) + b_{36}\ln(C)\ln(CD) + b_{45}\ln(P)\ln(R) + b_{46}\ln(P)\ln(CD) + b_{56}\ln(R)\ln(CD)$$

In this instance the Translog function requires twenty-one explanatory variables, compared to six for the Cobb-Douglas.

Under certain restrictions, the Translog function reduces to a Cobb Douglas function. This will be the case when the coefficients on all the quadratic and interaction terms are zero.

While the Translog is more suitable due to its flexible form, it becomes unsuitable in the face of a limited sample size. Small samples will have insufficient degrees of freedom to reliably estimate the parameters of the model. Furthermore, there is likely to be strong multicollinearity between the explanatory variables due to numerous terms involving transformations of the same variables and interaction among variables. This problem is exacerbated when the sample size is small.

In the absence of a limited sample size, the translog function's main advantage is its flexibility. By being less restrictive than the Cobb-Douglas function, its use reduces the risk of model mis-specification that could arise from imposing restrictions on the functional form that are not valid.

Choice of functional form

As part of this study we initially considered both the Cobb Douglas and Translog functional forms as potential options.

While the Translog offers extra flexibility, this is achieved at additional cost. One of the requirements of the Translog specification is a sample size that is large enough to provide sufficient degrees of freedom to reliably estimate the cost function and also to overcome the problem of multicollinearity between the explanatory variables.

For the sample of 123 observations available for this study, it is our view that the sample size is insufficiently large to reliably estimate a translog cost function. For this reason, we opt for the simpler but more restrictive Cobb-Douglas form, where the parameter estimates can be readily interpreted and are reasonably robust to changes in the estimation technique applied.

As mentioned previously, the main disadvantage of the Cobb-Douglas functional form is the degree of its restrictiveness, for example by imposing the restriction that the percentage response of opex to a given 1% change in an output, is the same whether the firm is large or small. If the restrictions imposed by the Cobb-Douglas are not valid, then the estimated model is mis-specified and the opex productivity forecasts obtained from the model are subject to error.

2.3 Econometric approach

Different cost function estimation approaches may be applied. This study adopts an econometric approach estimating the opex cost functions. The econometric approach is a parametric approach that aims to establish a statistical relationship between operating costs and the individual cost drivers.

The main advantages of the econometric approach are that it allows for:

— statistical testing to choose between competing models

- differences in operating environment such as scale and density to be controlled for across firms, something which is not possible within many non-parametric methods.

The main disadvantages are:

- the conventional econometric method does not separate statistical noise from inefficiency
  - this is where Stochastic Frontier Analysis (SFA) deviates from the conventional econometric approach by attempting to split one from the other through the introduction of a composite error term
- the econometric method is reliant on the functional form of the model to be chosen so as to reflect the appropriate production technology of the firms in question
- it is subject to a number of data limitations and statistical problems which may bias the results.

A large range of possible estimation techniques is possible. In this study we consider pooled OLS, Feasible GLS (FGLS) and SFA. These are discussed in greater detail in section 2.4.

2.3.1 Choice of variables

The possible output, input and operating environment variables that may be specified from the data available are shown below.

Possible outputs

- energy throughput (TJ)
- mains length
- customer numbers

In 2013 Economic Insights[2] was commissioned by the AER to provide advice on economic benchmarking of electricity Network Service Providers (NSPs), including the appropriate choice of outputs, inputs and operating environment variables.

The report recommended that any chosen output should reflect services provided to customers and that the output itself should be significant. The report makes a distinction between billed outputs, which are used as the basis for billing customers, such as energy throughput and customers, who pay fixed charges and functional outputs such as system capacity, system reliability and system security.

The report also recommends the use of customer numbers as an output, given that customers pay fixed charges on their bills and that these charges reflect activities that the NSP must undertake regardless of the amount of energy delivered. Economic Insights also recommend the use of energy delivered as an output as it is the service that is directly consumed by the customers. They do note however, that the inclusion of energy delivered is 'more arguable' because provided there is sufficient capacity, changes in energy throughput will have only a marginal impact on operating costs. Economic Insights also recommended the use of system capacity as an output variable in the cost function.

In a more recent study for Jemena Gas Networks, published in 2015, Economic Insights estimated an econometric variable cost function for Australian and New Zealand using customer numbers and gas throughput as their output variables. More recent work published in 2016 conducted by Economic Insights for Multinet Gas concluded that gas throughput was no longer a statistically significant determinant of real opex, and instead used customer numbers and network length as their choice of output variables. In their 2019 work for Jemena Gas Networks[3], Economic Insights opted for a model with customer numbers and mains length as the two outputs. In our analysis we opt for two sets of models, using customer numbers and gas throughput as well as customer numbers and mains length as the outputs.

Inputs

- capital services (constant price Regulatory Asset Base (RAB))

[2] Economic Insights (2013), Economic Benchmarking of Electricity Network Service Providers, Report prepared for the AER, 25 June 2013.
[3] Economic Insights (2019), 'Relative Efficiency and Forecast Productivity Growth of Jemena Gas Networks (NSW)', 24 April 2019.

— opex price index (weighted price index described below)

While Economic Insights recommended that use of physical measure of input capital services, constant price RAB is used as a proxy for capital instead of mains length mainly to avoid significant multicollinearity issues that arise from the presence of mains length in the denominator of the network density variable.

The opex price index is the index recommended by the AER for network service providers.[4] This is a weighted opex price index formed using the following Australian Bureau of Statistics (ABS) indexes and weights:

— electricity, gas, water and waste services (EGWWS) wage price index (WPI) —62 per cent
— intermediate inputs: domestic producer price index (PPI) —19.5 per cent
— data processing, web hosting and electronic information storage PPI —8.2 per cent
— other administrative services PPI —6.3 per cent
— legal and accounting PPI —3 per cent
— market research and statistical services PPI —1 per cent.

ACIL Allen sourced these indexes from the ABS and calculated the weighted index.

Environmental variables

— network density (customers per km of network length)

According to the AER[5], the main criteria for selecting an environmental variable should be that it has a material impact, it is exogenous and outside the control of the NSP and that it is a primary driver of costs. Based on these criteria, Economic Insights (2013) states that density variables are likely to be the most important operating environment factors that will affect efficiency comparisons between NSPs. In their 2015 work for Jemena Gas Networks[6], Economic Insights found that customer density was an important explanatory variable of real opex. In our view customer density is an appropriate choice for an environmental variable that is supported by considerable statistical evidence.

Since the data set used has only one environmental control variable, the likelihood of correct model specification is limited. However, while this does not invalidate the results, it suggests that the results may not be robust enough to rely on deterministically, are subject to a degree of error and need to be interpreted cautiously.

2.4 Estimation techniques

The study tested the following cost function estimation techniques:

— Pooled OLS
— Feasible Generalised Least Squares (FGLS)
— Stochastic Frontier Analysis (SFA)

Each of the techniques is discussed below.

2.4.1 Pooled OLS

The standard econometric technique to estimate a cost function is ordinary least squares (OLS). OLS fits a linear relationship between the dependent variable and a set of explanatory variables, in our case a set of outputs, inputs and operating environment variables. The line of best fit is chosen so as to minimise the sum of squared errors of the model.

The OLS estimator is considered BLUE (the best linear unbiased estimator) under a set of restrictive assumptions:

— the dependent variable is a linear function of a set of independent variables plus a disturbance term
— the expected value of the disturbance term is zero (unbiasedness)

[4] See AER, 2013, Better Regulation, Expenditure Forecast Assessment Guidelines for Electricity Distribution, p. 154-155.
[5] See AER, 2013, Better Regulation, Expenditure Forecast Assessment Guidelines for Electricity Distribution, p. 158
[6] Economic Insights (2015), Relative Opex Efficiency and Forecast Opex Productivity Growth of Jemena Gas Networks.

- disturbances have uniform variance and are uncorrelated (homoscedastic, no serial correlation)
- observations on the independent variables are fixed in repeated samples
- there are no exact linear relationships between independent variables (no perfect multicollinearity).

Pooled OLS treats the entire sample as if it is a single cross section. It does not recognise that the data has two dimensions, both across time and firms. This approach therefore does not recognise the panel structure of the data. This is not an issue if there is no heterogeneity across firms or if the heterogeneity can be captured entirely by existing explanatory variables in the model. However, in this and similar analyses, this is difficult due to the lack of environmental variables and the uncertainty about the comparability of the data. The inability to capture all environmental variables in the model will disadvantage the firm that has higher opex costs that arise due to environmental factors. Such are firm will appear to be inefficient relative to its peers in any productivity benchmarking exercise.

2.4.2 Feasible GLS (FGLS)

An alternative estimation method to that of OLS is Feasible Generalised Least Squares (FGLS).

OLS estimates are only efficient under the assumption of homoscedasticity and no serial correlation in the residuals. When these assumptions are violated, the OLS estimates are inefficient, although they remain unbiased. By inefficient we mean that the estimator no longer has the minimum variance among the class of linear unbiased estimators.

In this circumstance, the usual formula of the variance-covariance matrix is incorrect and the estimated variance-covariance matrix will be biased. In this context, interval estimation and hypothesis testing can no longer be trusted.

To address these problems, two solutions have been developed.

The first is a set of heteroscedasticity and autocorrelation consistent variance-covariance matrix estimators for the OLS estimator, which eliminate the bias in the variance-covariance matrix (albeit only asymptotically). These then allow OLS and other estimators to be employed with more confidence.

These heteroscedasticity and autocorrelation consistent variance-covariance estimators are sometimes referred to as robust variance estimates. Wherever possible in this study, we present t statistics based on heteroscedasticity and autocorrelation consistent variance-covariance estimators.

Alternatively, another estimator which explicitly recognises the heteroscedasticity and autocorrelation of the disturbances is FGLS, which can produce a linear unbiased estimator with smaller variances than OLS. This is done by using additional information such as that large disturbances are likely to be large because their variances are larger, or that large and positive error values in one period are likely to be followed by large and positive error values in the following period.

While OLS estimation minimises the sum of squared residuals, FGLS minimises an appropriately weighted sum of squared residuals, which gives lower weights to those residuals that are expected to be large because their variance is large or those residuals that are expected to be large because other residuals are large.

This approach results in a more efficient estimator than that obtained through OLS regression under heteroscedasticity or serial correlation. Under the classical OLS assumptions of spherical disturbances, the OLS estimator is the most efficient.

2.4.3 Stochastic Frontier Analysis (SFA)

The standard econometric approach is typically interpreted as all deviations from the predicted values of the model are due to inefficiency. This interpretation is an assumption, whereas in truth, the error term (i.e. ‘deviation’) is due to three causes: measurement error and other statistical noise, firm heterogeneity outside of management control, and managerial inefficiency.

Just like the standard econometric approach, SFA aims to model the relationship between operating costs, outputs and environmental variables. However, SFA separates the error term into two components:

— an inefficiency term, and
— a random error component.

This split attempts to remove the influence of random noise from the estimate of firm inefficiency. However, these two terms can only be interpreted as such if all firm heterogeneity outside of management control is accounted for within the model. If such factors are not taken into account within the model, then this firm heterogeneity will enter, most likely, both terms as well as affect estimates of the other parameters. This will result in errors in the estimated model parameters which will then feed through into the opex productivity growth forecasts.

SFA uses maximum likelihood estimation to model the relationship between opex and its drivers. The model takes the form:

$$\ln(C_{it}) = \beta_0 + \sum \beta_j \ln(x_{jit}) + v_{it} + u_{it}$$

where, in the ideal case:

— C_{it} is the opex for firm i at time t
— x_{jit} refers to all output and environmental drivers of opex j for firm i at time t
— v_{it} captures the effect of random factors such as unusual weather conditions for firm i at time t
— u_{it} captures the inefficiency for firm i at time t.

The statistical noise term is assumed to follow a normal distribution with mean zero and variance σ^2:

$$v_{it} \sim N(0, \sigma_v^2)$$

The inefficiency term is assumed to follow a one-sided non-negative truncated normal distribution with mean μ and variance equal to σ^2:

$$u_{it} = N^+(\mu, \sigma_u^2)$$

It is logical for the inefficiency term to remain positive because a business cannot reduce costs below the minimum possible level for a given set of outputs at a given set of input prices.

Just as in the standard econometric approach, SFA requires additional assumptions about the functional form of the cost function. If the underlying production technology of the industry is not reflected in the choice of cost function, there is a risk that this mis-specification could lead to biased estimates which will introduce errors into the productivity growth forecasts.

Because of the separation of the error term into two separate components, estimation of SFA cost models are more computationally demanding than conventional econometric methods. Moreover, separating the random and inefficiency components of the error term requires a large number of data points. This is a significant drawback in our case, where we have data on only nine firms and a limited number of observations.

3 PRODUCTIVITY STUDY DATA

This section describes the sample of Australian gas distribution firms used in the opex productivity analysis. Information is also provided on:

— the sources of the data used in the study
— limitations to data that are available publicly
— qualifications regarding the extent to which ACIL Allen has been able to verify the accuracy and comparability of the data.

3.1 Sample of gas distribution businesses

The cost function analysis presented in this study uses data from nine Australian gas distribution businesses serving urban populations and that are subject to economic regulation, namely:

— ATCO Gas Australia (WA)
— Australian Gas Networks South Australia (SA)
— Australian Gas Networks Victoria (VIC)
— Multinet Gas (VIC)
— AusNet Services (VIC)
— Jemena Gas Networks (NSW)
— Australian Gas Networks Queensland (QLD)
— Allgas Energy (QLD)
— Evoenergy (formerly ActewAGL)

ACIL Allen has compiled a dataset for the nine Australian gas distribution businesses. The data were largely sourced from AGN itself and public reports including:

— gas distribution business Access Arrangement Information statements
— regulatory determinations by the AER and jurisdictional regulators
— annual and other reports published by the businesses
— consultant reports prepared as part of access arrangement review processes
— Australian Gas Networks who provided an updated set of data for their Victorian, South Australian and Queensland networks

ACIL Allen has sourced data for the historical period from 2004-05 to 2018-19. A longer time series is best when estimating the underlying productivity trend, but if the historical period contains results which are not reflective of the forecast period then the estimated productivity growth may be distorted. In this case a shorter time series may be appropriate.

We consider that productivity trends should be measured over complete business cycles to reduce the risk that cyclical factors cause shifts in the long run trend. It is our view that utilising data from 2004-05 onwards approximately incorporates two distinct business cycles and is suitable for the purposes of this study. Moreover, data points preceding 2004-05 are likely to be out of date, containing historical conditions that are not likely to be reflective of the forecast period.

Over the historical period the study relies to the greatest extent possible on data from reported actual costs and outputs, rather than on forecasts.

The key data items used in the econometric analysis are:

— Customer numbers
— Energy throughput (TJ)
— Mains line length (km)
— Operating expenditures ($m)
— Regulatory Asset Base (RAB)

3.2 Data limitations and issues

3.2.1 Data comparability and suitability for the productivity analysis

To a large extent, this study relies on data that were reported publicly by the gas distribution businesses and, in most cases, verified by the relevant economic regulator. Where data has been provided to ACIL Allen directly, the data are reported on a basis that is consistent with the regulatory data in the Access Arrangement Information statements. In particular, the study uses the expenditure categories reported within the gas distribution businesses' Access Arrangements, including the operating expenditure and capital expenditure categories.

It is our opinion that the data used in the study is robust and appropriate for indicative productivity analysis, particularly as the majority of the data has been subject to scrutiny by the relevant economic regulator and in many cases also by expert consultants engaged by the economic regulators. There remains, however, some uncertainty about data comparability that cannot be resolved. Possible differences in the comparability of cost categories and other inevitable shortcomings in the analysis mean that the productivity forecasts produced should be treated as indicative, not exact.

By 'indicative' we mean that the results of the analysis should be used cautiously, giving careful consideration to the limitations of the dataset and modelling described further below in this section. These limitations suggest that the models parameter estimates will be subject to a degree of uncertainty and that model results should not be regarded as being particularly precise.

Other shortcomings that limit the ability of the models in this study to represent the gas distribution businesses' true cost functions include:

— the limited data available for this study e.g. a richer data set with a broader range of cost inputs, outputs and operating environment factors could be used to create model specifications that better account for the variation between the gas distribution businesses
— potential data errors that have not been identified
— the limitations of the modelling techniques in terms of their ability to accurately estimate the true opex cost parameters.

3.2.2 Small number of firms

A key limitation of this study is that the sample includes only nine firms. While other studies have tried to rectify this situation by significantly expanding the sample size to include firms from international jurisdictions, this is likely to exacerbate other problems such as the failure to account for operating differences between jurisdictions.

A larger sample size will help to improve the accuracy of the model parameter estimates. This will be beneficial if the additional businesses are subject to the same regulatory requirements and business conditions to those already in the sample. This is unlikely to be the case if the businesses are from different international jurisdictions.

3.2.3 Multicollinearity between explanatory variables

An issue arises in the specification of econometric models when there is a high degree of multicollinearity between the explanatory variables in a regression. Multicollinearity is a phenomenon in which the predictor variables in a regression are highly correlated with each other. When this happens, it becomes difficult to measure the impact of any specific variable in the model, despite the model performing reasonably well as a whole.

A model with collinear explanatory variables will tend to be characterised by:

— imprecise coefficient estimates leading to high standard errors and statistical insignificance

— erratic shifts in the coefficients in response to small changes in the model

— the presence of theoretically inconsistent coefficients.

The presence of multicollinearity is problematic because we are attempting to estimate separate elasticities for each variable within a cost function. If these variables do not exhibit sufficient independent variation then it will not be possible to reliably disentangle the separate effects of each variable.

The multicollinearity problems expands exponentially when estimating the Translog cost function, which contains quadratic and interaction terms for each output, input and operating environment term.

For this reason, it is our opinion that the high degree of multicollinearity between explanatory variables and the small sample size make it impossible to reliably estimate a Translog cost function in this instance. We have therefore limited ourselves to the Cobb-Douglas specification, despite its more restrictive functional form.

3.2.4 Different accounting treatment of opex

When modelling opex, different accounting practices for capitalising costs can potentially disadvantage those businesses that capitalise a smaller percentage of their expenditure. These businesses will show higher levels of opex compared to those businesses that capitalise a larger percentage of their expenditure onto their balance sheets.

3.2.5 Missing environmental variables and model mis-specification

Data limitations are such that we are only able to control for a small number of operating environment variables. Failure to control for important environmental or operational differences can potentially lead to biased results. The key operating environment variable specified in the cost functions is customer density. In previous benchmarking studies of gas distribution businesses this has been shown to be a significant explanator of differences in operating and capital costs.

Additional operating environment variables related to network age (proxied by the proportion of mains length not made of cast iron or unprotected steel) and service area dispersion (proxied by the number of city gates) are also possible candidates for inclusion into the cost function. Unfortunately, ACIL Allen do not have the data necessary to include these additional operating environment variables. The exclusion of these environmental variables, and potentially other significant operating environment variables could reduce the accuracy of the inefficiency measure that can be attributed to actions of the gas distribution businesses. However, this is not in itself a reason to discount the cost function analysis in this report but suggests that care be exercised in the degree of precision that is attached to the model estimates.

3.3 Data definitions

The following describes the data items used in the analysis.

Operating expenditure

The operating expenditure amounts used in this benchmarking study reflect the costs classified as operating expenditure within each businesses' Access Arrangement. This typically includes a range of operating costs (including network operations, regulatory costs and billing cost), maintenance costs (including for pipelines, meters and network control) and other management and administration costs.

As had been identified in previous benchmarking reports, unaccounted for gas (UAFG) is treated differently between the jurisdictions. As a result, it has been excluded from operating costs for this study. Debt raising costs have also been removed where included in reported operating expenditure. This has also been done to account for differences in the treatment of these costs over time and between the businesses. Other operating expenditure items removed to aid comparability and to remove costs that are outside the control of the gas distribution businesses are carbon costs and government levies.

The operating expenditure data sourced for the study were reported in a range of nominal and constant dollar values within the source documents. All dollar amounts have been placed on a common basis using the Australian Bureau of Statistics All Groups, Weighted average of eight capital cities, CPI (Series ID: A2325846C).

Regulatory asset base (RAB)

The measure of RAB is the closing value for each year.

Network length

The network length for the gas distribution businesses includes the mains that the businesses classify as low, medium and high pressure distribution mains and transmission pressure mains operated above 1,050kPa.

Customers

The customer number measure is the total number of customers including residential and non-residential volume customers and contract customers.

Gas delivered

The gas delivered measure is the total gas delivered to the above customers measured in Terajoules (TJ).

4

ESTIMATED COST FUNCTIONS

4.1 Cost functions estimated

Three separate estimation methods were applied to our Cobb-Douglas operating cost function. These are:

— pooled ordinary least squares (OLS)

— feasible GLS (FGLS) (with heteroscedastic panels)

— Stochastic Frontier Analysis (SFA) (with time invariant inefficiency).

The results of all five methods are presented in section 4.3 and 4.4. The results presented in section 4.3 are for a model specification including two outputs (energy throughput, which is measured as TJ of gas throughput and customers) and in section 4.4 for a model specification with energy throughput and mains line length as the outputs.

Before presenting the estimated regression models themselves, we describe the model validation and testing process undertaken to choose the most appropriate models for the purpose of forecasting productivity growth.

4.2 Model validation and testing

In order to assess the suitability of the various estimated opex cost functions a number of assessment criteria are applied. These are:

— theoretical coherence
  - estimated coefficient signs make economic sense

— statistical testing and performance
  - estimated coefficients are statistically significant and the model fit is good

— robustness to changes in estimation technique
  - estimated coefficients are stable across estimation techniques

The above steps can be considered part of a validation process where each step should be satisfied before a model specification is accepted. Each of the steps are described in more detail below.

4.2.1 Theoretical coherence

In assessing the suitability of any model, it is important that the selected model is consistent with economic theory. This means that the model should contain the theoretical drivers of operating costs as well as variables that control for operating conditions between firms.

The estimated coefficients on each of the explanatory variables must have a theoretically correct sign. That is, increases in drivers such as energy throughput or customers must lead to increases in predicted operating costs from the model. If they do not, then the model is not consistent with economic theory and can be considered suspect.

Similarly, the customer density of the gas network is expected to have a negative relationship to operating expenses. As the network grows denser, a given level of outputs is able to be produced at a lower cost compared to a network with lower customer density.

The magnitude of the coefficients should also be consistent with our expectations of the production technology of the industry. We would expect there to be economies of scale with regard to opex in the gas distribution business. This is both logical and supported by a significant number of empirical studies of both gas and electricity distribution businesses.

Increasing returns to scale with respect to opex implies that the sum of output coefficients is less than one.

4.2.2 Statistical testing

Statistical significance of estimated coefficients

An important step in assessing the suitability of any explanatory variable is its statistical significance. Not only should the variable be of a sign and magnitude that is consistent with economic theory, but standard hypothesis testing should indicate that the variable is statistically significant either at the 1% or 5% significance levels, though it is preferable to achieve statistical significance at the 1% significance level. The significance level can be interpreted as the probability of rejecting the null hypothesis that no relationship exists between the explanatory and dependent variable when it is actually true. This means that under the 1% level of statistical significance there is only a 1% probability of concluding that a relationship exists when it does not, while the probability of incorrectly concluding that a relationship exists is five times as large at the 5% level of statistical significance. The 1% level of statistical significance therefore provides a considerably higher degree of confidence that our conclusions are actually true. While the 1% level of statistical significance provides a higher level of confidence, both the 1% and 5% levels of statistical significance are used in common practice.

A statistically significant result is one which is unlikely to have occurred by chance. Each estimated coefficient in the regression models has an associated t-statistic or p value. If the estimated p value is less than 0.01 then that coefficient is statistically significant at the 1 per cent significance level. A p-value that is less than 0.05 is significant at the 5 per cent level of significance. The lower the observed p value on a coefficient the greater the probability that a statistically significant relationship exists between the dependent variable and the explanatory variable concerned.

In this study we consider the statistical significance of each variable at both the 1% and 5% levels of statistical significance as a guide to whether a particular variable should be included in a given model. However, it is important to note that in small samples, which we are limited by in this study, it may be difficult to achieve statistical significance for a variable even though the variable is economically significant. This means that some professional judgement needs to be applied where a variable is failing to achieve the desired level of statistical significance but the coefficient estimate is still consistent with economic theory and with other empirical studies.

In this instance we chose to retain the variable in the model even if it failed to satisfy statistical significance. While it is not surprising for the models to produce insignificant estimates given the small sample sizes, statistically insignificant estimates still provide useful information and an insignificant estimate is better than a zero estimate as long as the high standard error is noted.

Goodness of fit

The most commonly used measure of the goodness of fit of a linear regression model to the observed data is the coefficient of determination, also known as R^2. It represents the proportion of the variation in the dependent variable that is explained by variation in the independent variables. In assessing a model’s suitability and fitness for purpose, we would prefer the within sample fit to be strong.

However, in the model validation process, the R^2 is just one of a wide suite of tools available. While it is important to emphasize that goodness of fit is a desirable feature of any model, there are factors other than in sample fit that need to be taken into account. For example, a high R^2 is no guarantee that a model will have any predictive ability.

Tests for heterogeneity and serial correlation

Additional statistical tests are carried out to assess the presence of heteroscedasticity and autocorrelation in the model residuals. To test for heteroscedasticity, we apply a Modified Wald test for groupwise heteroscedasticity[7]. Serial correlation is tested for via the Wooldridge Lagrange Multiplier (LM)[8] test for autocorrelation in panel data.

4.2.3 Robustness to changes in estimation method

The estimated coefficients should be reasonably stable across a range of estimation techniques. Very large deviations in the coefficients across different models, particularly where the coefficients become theoretically implausible is a sign that the model may not be correctly specified or that the technique is not the most appropriate to use.

4.3 Two output specification: Energy and Customer numbers

Table 4.1 shows the results of the estimated Cobb-Douglas cost functions with two outputs, energy (TJ of gas throughput) and customer numbers. The table shows the estimated coefficients for each of the estimation techniques in column 2 to column 5. Column 1 shows the relevant variable that the coefficient applies to. The numbers in parentheses under each coefficient estimate are the standard errors that apply to each estimate.

A key characteristic of this model is that both the customer numbers and energy throughput variables have both a positive and statistically significant coefficient for both the OLS and FGLS specifications. This is consistent with economic theory. In the case of the SFA model, only the customer numbers coefficient was found to be statistically significant.

The coefficients on the customer numbers and energy throughput variables both sum to less than one, a result which is consistent with increasing returns to scale. The negative coefficients on the time trend in the FGLS and SFA models imply increasing productivity over time arising from technological change. The time trend was found to be statistically insignificant for all three estimation approaches. Despite this, we have opted to retain the time trend variable in the model on the basis that the variable still contains useful information.

Moreover, the negative coefficient on the customer density coefficients indicate that more dense networks have a cost advantage those that are less dense, a result which is consistent with economic theory.

Additional diagnostic testing indicated the presence of heteroscedasticity and serial correlation in the residuals, suggesting that the FGLS specification is preferable to the pooled OLS model. For this reason, we exclude the OLS model from the opex productivity calculations in the next chapter.

TABLE 4.1 ESTIMATED COBB-DOUGLAS FUNCTION: TWO OUTPUTS: ENERGY THROUGHPUT AND CUSTOMERS

Variables	OLS	FGLS	SFA
Time	0.00187	-0.000157	-0.00112
	(0.00373)	(0.00273)	(0.0035)
Customers	0.581***	0.556***	0.894***
	(0.0715)	(0.0661)	(0.138)
Energy	0.165***	0.162***	-0.00103

7 See Greene (2000), Econometric Analysis, Upper Saddle River, NJ: Prentice-Hall

8 Wooldridge, J.M (2002), Econometric Analysis of Cross Section and Panel Data, Cambridge, MA: MIT Press

Variables	OLS	FGLS	SFA
	(0.0239)	(0.0213)	(0.109)
RAB	0.147**	0.174***	-0.0408
	(0.0634)	(0.0596)	(0.151)
Customer density	-0.509***	-0.473***	-0.267
	(0.0827)	(0.0803)	(0.223)
Constant	-12.69*	-8.555	-8.407
	(7.493)	(5.446)	(7.014)
Observations	123	123	123
R-squared	0.955		

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

SOURCE: ACIL ALLEN

4.4 Two output specification: Customer numbers and mains length

Table 4.2 shows the results of the estimated Cobb-Douglas cost functions with customer numbers and mains length as the outputs. Both the outputs have theoretically correct positive signs, which together sum to less than one, a necessary condition for increasing returns to scale. To avoid a multicollinearity problem, the customer density variable is not included in this regression. This is because it is defined as the number of customers divided by the network length, both variables of which are included separately in the specified models.

TABLE 4.2 ESTIMATED COBB-DOUGLAS FUNCTION: TWO OUTPUTS: CUSTOMER NUMBERS AND MAINS LENGTH

Variables	OLS	FGLS	SFA
Time	-0.00245	-0.00389	-0.0011
	(0.00418)	(0.00338)	(0.00246)
Customers	0.178**	0.235***	0.627***
	(0.072)	(0.0696)	(0.172)
Mains length	0.538***	0.583***	0.266
	(0.0957)	(0.093)	(0.21)
RAB	0.190***	0.067	-0.0414
	(0.0694)	(0.0656)	(0.134)
Constant	-4.241	-1.595	-8.454*
	(8.386)	(6.758)	(4.882)
Observations	123	123	123
R-squared	0.944		

Note: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

SOURCE: ACIL ALLEN

Testing for group-wise heteroscedasticity rejected the null hypothesis of no heteroscedasticity at the 1 per cent significance level. The Wooldridge test of serial autocorrelation in panel data failed to reject the null hypothesis of no autocorrelation. These results provide evidence in support of group wise heteroscedasticity in the panel, but not of autocorrelation.

For this reason, we prefer the FGLS model which accounts for heteroscedasticity in the panel over the pooled OLS model which imposes an assumption of homoscedasticity on the disturbances.

In the subsequent section of this report, we do not use the results for the pooled OLS model to generate the opex partial productivity forecasts, instead focussing on the FGLS and SFA models.

5 OPEX PARTIAL PRODUCTIVITY FORECASTS

In this section we calculate the opex partial productivity growth forecasts based on the two output models estimated in the previous section.

This is done by taking the parameter estimates from the two output cost function models (i.e. the FGLS and SFA cost models) and combining them with AGN's forecasts of customer numbers, energy throughput, RAB, and pipeline length over the next regulatory control period to obtain forecasts of AGN's opex partial productivity. While we present the opex partial productivity growth rates from all the variables represented in the cost function, we focus on the rate of technical change represented by the time trend coefficient in the estimated cost functions, following the standard approach recommended by the AER.

5.1 Inputs to calculating opex partial productivity

The parameter estimates from the models are shown in **Table 5.1** and **Table 5.2** below.

Table 5.1 shows the regression estimates for the estimated models with energy and customer numbers as the outputs. The table also shows the output weights implied by the regression estimates. In the case of the FGLS model, customer numbers receive an output weight of 77.4% while energy throughput has a weight of 22.6%. In the case of the SFA model, customer numbers receive an output weight of 100%, with energy throughput receiving an output weight of zero.

TABLE 5.1 TWO OUTPUT OPEX COST FUNCTION REGRESSION ESTIMATES WITH ENERGY AND CUSTOMER NUMBERS AS OUTPUTS

Coefficients	FGLS	SFA
Time	-0.0001572	-0.0011239
Customers	0.5558284	0.8936792
Energy	0.161873	-0.0010312
RAB	0.1740685	-0.0407733
Customer density	-0.4728912	-0.2668525
Output weights		
Customers	77.4%	100%
Energy	22.6%	0%

SOURCE:ACIL ALLEN CONSULTING

Table 5.2 shows the regression estimates for the estimated models with customer numbers and network length as the outputs. In the case of the FGLS model, customer numbers receive an output weight of 28.7% while network length has a weight of 71.3%. In the case of the SFA model, customer numbers receive an output weight of 70.2%, with line length receiving an output weight of only 29.8%.

TABLE 5.2 TWO OUTPUT OPEX COST FUNCTION REGRESSION ESTIMATES WITH CUSTOMER NUMBERS AND MAINS LENGTH AS OUTPUTS

Coefficients	FGLS	SFA
Time	-0.0038949	-0.0011005
Customers	0.2347014	0.6268975
Mains length	0.5825121	0.2661463
RAB	0.0669515	-0.0414289
Output weights		
Customers	28.7%	70.2%
Mains length	71.3%	29.8%

SOURCE:ACIL ALLEN CONSULTING

Table 5.3 and **Table 5.4** show AGN's forecast growth drivers of opex over the period from 2021 to 2026. Average customer number growth is expected to be 1.33 per cent per annum, while energy throughput is expected to decline by 1.24% per annum. Applying the output weights from the energy and customer numbers regressions, the average weighted output growth from the FGLS model is expected to be 0.75% per annum, while the equivalent calculation from the SFA model produces an average weighted output growth of 1.34% per annum.

The RAB and customer density are projected to grow at an annual average rate of 2.09% and 0.70% per annum respectively.

TABLE 5.3 FORECAST CHANGES IN GROWTH DRIVERS: CUSTOMERS AND ENERGY THROUGHPUT MODEL

Year	Customers	Energy	Weighted average output: FGLS	Weighted average output: SFA	RAB	Customer density
2021	1.33%	-1.40%	0.71%	1.33%	4.33%	0.37%
2022	1.35%	-0.92%	0.84%	1.36%	1.47%	0.82%
2023	1.35%	-1.78%	0.64%	1.35%	1.63%	0.82%
2024	1.31%	-0.84%	0.82%	1.31%	0.93%	0.78%
2025	1.26%	-1.72%	0.59%	1.26%	0.84%	0.75%
2026	1.22%	-0.75%	0.78%	1.22%	0.32%	0.72%
Average	**1.33%**	**-1.24%**	**0.75%**	**1.34%**	**2.09%**	**0.70%**

SOURCE: AGN

Applying the output weights from the customer numbers and network length regressions, the average weighted output growth from the FGLS model is expected to be 0.84% per annum, while the equivalent calculation from the SFA model produces an average weighted output growth of 1.13% per annum. Mains length is forecast to grow at an average rate of 0.64% per annum over the next regulatory period.

TABLE 5.4 FORECAST CHANGES IN GROWTH DRIVERS: CUSTOMERS AND MAINS LENGTH MODEL

Year	Customers	Mains length	Weighted average output: FGLS	Weighted average output: SFA	RAB
2021	1.33%	0.96%	1.07%	1.22%	4.33%
2022	1.35%	0.53%	0.77%	1.11%	1.47%

Year	Customers	Mains length	Weighted average output: FGLS	Weighted average output: SFA	RAB
2023	1.35%	0.53%	0.76%	1.10%	1.63%
2024	1.31%	0.52%	0.75%	1.07%	0.93%
2025	1.26%	0.51%	0.73%	1.04%	0.84%
2026	1.22%	0.50%	0.71%	1.01%	0.32%
Average	**1.33%**	**0.64%**	**0.84%**	**1.13%**	**2.09%**

5.2 Calculating opex partial productivity

The opex partial productivity growth rate can be broken down into three components, namely:

— technical change
— returns to scale
— changes in operating environment.

Technical change is represented by the time trend in the regression. The productivity gains associated with technical change (A) is estimated as:

$$Technology\ (A) = -Time\ coeffcient$$

Returns to scale are productivity gains that arise as a result of increasing business size over time. The productivity gains from returns to scale (B) are calculated as:

$$Returns\ to\ scale\ (B) = \big(1 - (Customer\ elasticity + Energy\ elasticity)\big) \times (\%\Delta\ in\ weighted\ average\ output\ growth)$$

The above formula relates to the two-output specification with customer numbers and energy throughput as the outputs. In the model with mains length as an output instead of energy throughput, we would adjust the formula by replacing the energy elasticity, with the estimated elasticity on the mains length variable.

Operating environment partial productivity is calculated as the RAB and customer density coefficients multiplied by each of their respective changes in each year. The total operating environment contribution to opex partial productivity is the negative of the sum of the RAB and customer density contributions.

$$Operating\ environment\ factors\ (C) = (RAB\ elasticity \times \%\Delta\ in\ RAB) + (Customer\ density\ elasticity \times \%\Delta\ in\ Customer\ density)$$

The opex partial factor productivity growth rate is estimated from these three elements, using the formula:

$$Opex\ partial\ productivity\ growth\ rate = A\ (Technology) + B\ (Returns\ to\ scale) - C\ (Operating\ environment\ factors)$$

While in previous determinations the opex partial productivity growth rate was calculated as the sum of the three components shown above, we now follow the standard approach adopted by the AER and calculate the recommended opex productivity growth factor based only on the rate of technical change represented by the time trend in the estimated cost functions.

5.3 Opex partial productivity forecasts

In this section we calculate opex partial productivity forecasts using the estimated cost functions and AGN's projected growth drivers.

For the models using energy throughput and customer numbers as outputs, the FGLS model predicts an increase in productivity growth attributed to technical change of 0.02 per cent per annum over the period 2021 to 2026 (see **Table 5.5**). The SFA model predicts an increase in in the productivity growth attributed to technical change of 0.11 per cent per annum over the same period (see **Table 5.6**).

TABLE 5.5 ANNUAL OPEX PARTIAL PRODUCTIVITY FORECASTS, CUSTOMERS AND ENERGY OUTPUTS, FGLS MODEL

Year	Technology (A)	Returns to scale (B)	Operating environment factors (C)	PP Opex growth rate (A+B-C)
2021	0.02%	0.20%	0.58%	-0.36%
2022	0.02%	0.24%	-0.13%	0.39%
2023	0.02%	0.18%	-0.10%	0.30%
2024	0.02%	0.23%	-0.21%	0.46%
2025	0.02%	0.17%	-0.21%	0.39%
2026	0.02%	0.22%	-0.28%	0.52%
Average (2021 to 2026)	**0.02%**	**0.21%**	**-0.06%**	**0.28%**

SOURCE:ACIL ALLEN

TABLE 5.6 ANNUAL OPEX PARTIAL PRODUCTIVITY FORECASTS, CUSTOMERS AND ENERGY OUTPUTS, SFA MODEL

Year	Technology (A)	Returns to scale (B)	Operating environment factors (C)	PP Opex growth rate (A+B-C)
2021	0.11%	0.08%	-0.28%	0.46%
2022	0.11%	0.09%	-0.28%	0.48%
2023	0.11%	0.07%	-0.28%	0.47%
2024	0.11%	0.09%	-0.25%	0.45%
2025	0.11%	0.06%	-0.23%	0.41%
2026	0.11%	0.08%	-0.20%	0.40%
Average (2021 to 2026)	**0.11%**	**0.08%**	**-0.25%**	**0.44%**

SOURCE:ACIL ALLEN

For the models using customer numbers and network length as outputs, the FGLS model predicts an increase in the productivity growth from technical change of 0.39 per cent per annum over the period 2021 to 2026 (see **Table 5.7**). The SFA model predicts an increase in productivity growth from technical change of 0.11 per cent per annum over the same period (see **Table 5.8**).

TABLE 5.7 ANNUAL OPEX PARTIAL PRODUCTIVITY FORECASTS, CUSTOMERS AND MAINS LENGTH OUTPUTS, FGLS MODEL

Year	Technology (A)	Returns to scale (B)	Operating environment factors (C)	PP Opex growth rate (A+B-C)
2021	0.39%	0.19%	0.29%	0.29%
2022	0.39%	0.14%	0.10%	0.43%
2023	0.39%	0.14%	0.11%	0.42%
2024	0.39%	0.14%	0.06%	0.46%
2025	0.39%	0.13%	0.06%	0.47%
2026	0.39%	0.13%	0.02%	0.50%

Year	Technology (A)	Returns to scale (B)	Operating environment factors (C)	PP Opex growth rate (A+B-C)
Average (2021 to 2026)	**0.39%**	**0.15%**	**0.11%**	**0.43%**

SOURCE:ACIL ALLEN

TABLE 5.8 ANNUAL OPEX PARTIAL PRODUCTIVITY FORECASTS, CUSTOMERS AND MAINS LENGTH OUTPUTS, SFA MODEL

Year	Technology (A)	Returns to scale (B)	Operating environment factors (C)	PP Opex growth rate (A+B-C)
2021	0.11%	0.13%	-0.18%	0.42%
2022	0.11%	0.12%	-0.06%	0.29%
2023	0.11%	0.12%	-0.07%	0.30%
2024	0.11%	0.11%	-0.04%	0.26%
2025	0.11%	0.11%	-0.03%	0.26%
2026	0.11%	0.11%	-0.01%	0.23%
Average (2021 to 2026)	**0.11%**	**0.12%**	**-0.07%**	**0.29%**

SOURCE:ACIL ALLEN

Following the advice of Armstrong (2001), we combine the forecasts derived from the two separate specifications and methods to improve forecast accuracy[9].

Armstrong suggests equal weights as a starting point where there is no additional knowledge about which method is the most accurate. If we follow this advice, then a simple average of the four separate average partial productivity measures should be used.

Using equal weights, we take a simple average of the four separate average technical change measures. This results in an average forecast opex productivity growth factor of **0.17 per cent per annum.**

The FGLS time trend coefficients measure the residual productivity change across the all distribution businesses, meaning that the measured productivity change includes both the expansion of the industry frontier as well as catch up to the industry frontier. This suggests that the FGLS models overstate the productivity growth achieved by frontier businesses when there is significant catch-up across the whole industry. The SFA models, in contrast, theoretically capture only the change in the efficiency frontier. This suggests that the true opex productivity factor could be even lower than the average calculated over all four models.

Nevertheless we have chosen to be conservative and to use the average of both the FGLS and SFA models because any differences between the models could also be due to underlying statistical uncertainty in the estimated coefficients, rather than theoretical differences in the nature of the productivity gains that is captured by the models.

9 See Armstrong J. S (2001), Principles of forecasting: A Handbook for Researchers and Practitioners, Kluver Academic Publishing p. 417-439.