Telstra's Ordinary Access Undertaking for ULLS

The Impact of Distribution Area Design on Customer Access Network Investment Costs

Iain Little

Zoran Peroski

John Zeitsch

9 March 2009

Contents

A Executive Summary

1. The Customer Access Network (CAN) is the copper telecommunications network from Telstra's local exchange buildings to customers' premises. The CAN is (simplistically) comprised of five major physical components, which includes equipment within the telephone exchange building, conduit containing main network cables, pillars, conduit containing distribution network cables, and lead-ins.

2. Equipment in the telephone exchange is the main point of aggregation at which all of the cabling serving an exchange area must terminate. The area served by one telephone exchange is known as an exchange service area (ESA). Each ESA is divided into a number of distribution areas (DAs). Each DA is served by a pillar which acts as another point of aggregation in an ESA.

3. The TEA model calculates the forward-looking efficient costs of providing the CAN within Band 2 (urban) ESAs in Australia. A commonly held engineering view that relates to the TEA model and other engineering models of a CAN, is that DA design (the number of and dimensions of distribution areas and the location of pillars in DAs) collectively, has minimal impact on CAN cost.

4. To test this hypothesis we need to look at all factors affecting investment cost as generated in the TEA model. We split factors into three parts affecting Band 2 ESAs.

- How the size of the CAN in Band 2 affects CAN investment cost;
- How characteristics of the ESAs in Band 2 affects CAN investment cost; and,
- How changes to DA design in Band 2 impact on the total investment costs for the CAN.

5. We examine these effects using regression analysis which is a commonly used statistical technique which in this case is applied to data from version 1.2 of the TEA model. The analysis finds that:

- The regression results indicate that investment in the CAN, as generated by the TEA model, exhibit slightly increasing returns to scale. A 1 per cent increase in inputs used to construct the CAN could support a 1.04 expansion in the CAN.
- the characteristics of the network in an ESA has a small positive impact on investment costs of the average ESA — a 1% change in the average perimeter of an ESA or the density and dispersion of an ESA will impact the average ESA Band 2 investment costs by 0.044%, 0.007% and 0.012% respectively;
- DA design has no practical impact on the investment cost of the CAN:
  - While the regression analysis shows that ESAs with 1% longer average distance from the pillar to the copper centre will have 0.027% lower investment costs, in practice, pillars are typically located on or near the boundary of the DA; and,

- Increasing the average DA size by 25% would reduce the predicted investment costs in Band 2 by 0.020% (or less than 1c per service per month).

6. Overall we find that the size of the CAN in band 2 largely explains investment cost as predicted using the TEA model. Variations in the characteristics of the network served in Band 2 explain a small portion of CAN investment costs. Given the immaterial size of the saving achieved from an impractical restructure of DA design in Band 2, we conclude that DA design does not materially impact the investment cost of the CAN.

B Introduction

7. The TEA model calculates the forward-looking efficient costs of providing the CAN – the copper telecommunications lines from Telstra's local exchange buildings to customers' premises (the unbundled local loop) – within Band 2 areas of Australia.[1] The CAN is (simplistically) comprised of five major physical components. Sequentially from the exchange to the customers premises these include:

- The customer side block on the Main Distribution Frame (MDF), intra-exchange cabling and associated facilities in the exchange building;
- Conduit containing main network cables;
- Pillars;[2]
- Conduit containing distribution network cables; and,
- Lead-ins.

8. The MDF in the telephone exchange is the main point of aggregation at which all of the CAN cabling with an ESA must terminate. Each ESA is divided into a number of DAs. Each DA is served by a pillar which acts as another point of aggregation in an ESA.

9. Distribution network cables connect individual customers in the DA to a pillar and main network cables join the pillars within an ESA to the exchange building. The distribution and main network cables are housed in PVC conduits that are installed underground. Lead-ins are copper cables that connect customers' premises to the distribution cables in the street.

10. The TEA model takes as given the several real-world characteristics of Telstra's CAN infrastructure, including:

- Number and location of customer premises;
- The number of and the dimensions of DAs and the location of pillars;
- The number of and dimensions of ESAs and the location of exchange buildings in the ESAs; and,
- The subset of the conduit routes that take the minimum distance between premises, pillars and exchanges.

11. In reviewing the TEA model the ACCC has come to the conclusion that assuming the location of pillars in DAs are fixed accords with best practice engineering rules and practices. The ACCC noted:

[1] For greater detail of the TEA model V1.2 see the Telstra Efficient Access (TEA) Model Overview.
[2] Within the TEA model V1.2 cabinets are either eliminated or converted to pillars in the analysis. This does not impact on this regression analysis as cabinets can serve the role of a pillar.

In recognition of actual circumstances, the ACCC has generally accepted the following simplifications to the fully forward-looking TSLRIC approach: certain key features of an existing network such as exchanges and pillars are kept constant. This is often referred to as the scorched node approach where the locations of particular nodes are assumed to be fixed. [3]

The ACCC considers that given the starting point of scorched node and the need to model a copper network, the TEA model is broadly based on best practice engineering rules and practices. [4]

12. Within any ESA there may be a large degree of variation in the design of the DAs that comprise it. This is because real world obstacles exist such as water courses, private property, roads, recreational areas etc., which govern the actual routes and locations where real-world CAN infrastructure can exist. This means that from ESA to ESA:

- The sizes of DAs varies, both in terms of number of addresses and geographic area served;
- The number of DAs is different for each ESA; and,
- Pillars, pits, manholes and joints are situated at different relative locations within each DA.

13. This analysis quantitatively measures, using a statistical regression model and economic theory of a firm's cost minimising incentive, how changes to DA design impact on the total investment costs for the CAN as calculated by version 1.2 of the TEA model. In particular, the analysis examines:

- The variation in DA design between different band 2 ESAs;
- Whether such variations in DA design are the cause of differences in the total investment costs for different ESAs; and,
- If DA design characteristics are found to impact on total investment costs, the extent of the impact that DA design has upon total investment costs.

14. The remainder of this report is structured as follows. Section C outlines what regression analysis is and why it is the most appropriate technique in this context. Section D addresses the issue of functional form for the regression analysis. The regression equation, data, results and associated analysis are set out in section E and Section F concludes the analysis.

C Regression modelling (why and how)

15. There are two approaches that could be used to establish the impact of DA design on the cost of constructing the CAN as calculated by the TEA model. These include:

---

[3] ACCC (2008), *Assessment of Telstra's Unconditioned Local Loop Service Band 2 monthly charge undertaking: Draft Decision*, November 2008, at page 36.

[4] Ibid, pp 72.

- Simulation analysis using the TEA model. This would require altering the DA design in the TEA model and then re-calculating the investment cost of the CAN; and,

- Regression modelling to establish the stable mathematical relationship between DA design variables and investment costs using all DAs in Band 2 ESAs as the data set.

16. Simulation analysis would be a very time consuming task. For example, the effect of pillar location on CAN construction costs could be analysed by altering the location of the pillar in each DA, and then calculating construction cost given the new pillar locations. This would require engineers to determine whether the alternative pillar location is feasible (that is, that there are no obstacles to installation) and, if so, to redesign the cable routing to the alternative pillar location to ensure that the resulting network would be functional. Such assessments/calculations would need to be undertaken for each of the approximately 50,000 DAs in the TEA model.

17. In contrast, regression modelling can be used to test whether the variability in the characteristics of the CAN across ESAs can explain the variability in CAN investment costs across ESAs as calculated by version 1.2 of the TEA model. Regression modelling is extremely robust and has been widely accepted among statisticians for many years. It is most commonly used in applied fields such as the physical, health, social and life sciences. For example, the Journal of Econometrics[5], amongst others, is a global monthly publication that exists primarily to advance the science of regression analysis in applied economics and business problems.

18. Regression analysis determines the relationships between variables that 'best fits' the data. For example, assume that there is a linear relationship between CAN investment cost and the number of premises connected to the CAN. The graphs below illustrate the relationships that a regression analysis might produce (represented by the red lines) given a set of hypothetical data (given by the blue dots). Diagram 1 illustrates what the relationship might look like if the relationship between investment cost and number of premises connected is positive (Case 1) and negative (Case 2). The red line in each of the cases is 'best fit' because it most closely approximates the relationship between the two variables, that is, it minimises the sum of the squared vertical distances between the observed values for the data and the line of best fit (u).

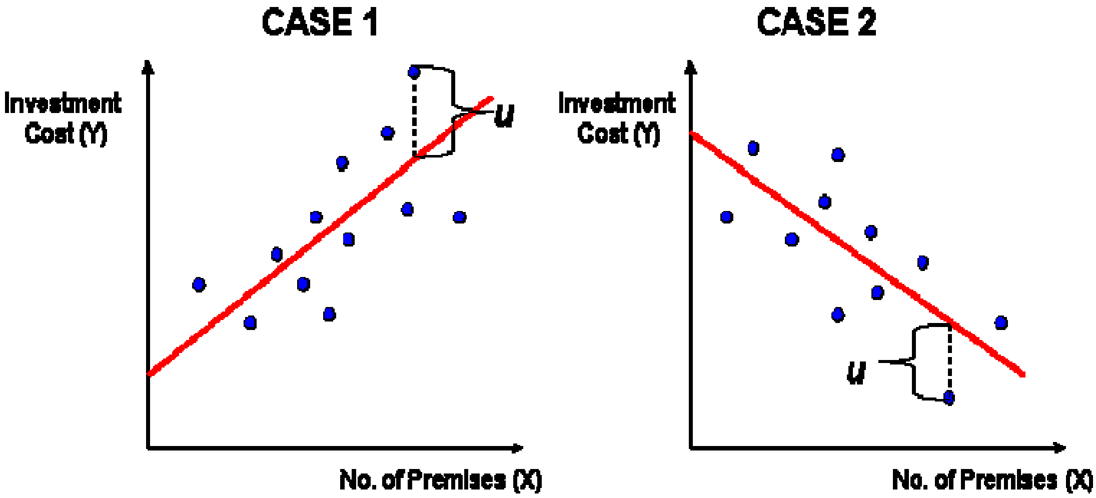
19. The 'vertical distance' depicted in both cases in figure 1 and defined as u in paragraph 18 is known as a random 'disturbance'. It is the amount the best fit line and the actual data for which we are estimating a stable relationship are 'disturbed' from their stable relationship as estimated by the regression.

20. The disturbance from the line of best fit arises for several reasons, primarily because for every possible variable that influences total CAN investment cost data may not be available or measurable. Therefore the disturbances or vertical distances u attempt to capture how much information in total CAN investment cost is unobserved, does not exists or is unable to be

[5] http://www.elsevier.com/wps/find/journaldescription.cws_home/505575/description?navopenmenu=-2

explained by the data that already exists. Therefore, the smaller the vertical distances u the less information that influences total CAN investment cost is unobserved and the stronger the regression relationship.

Figure 1: Regression of linear functional form



21. The strength of the relationship derived from the regression analysis is captured by a metric know as 'R-squared' which makes use of the sum of the vertical squared distance[6] between the data points and the line of best fit. The R-squared metric is equal to 1 if, in our example, the line of best fit perfectly describes the relationship between CAN investment cost and the number of premises connected to the CAN, that is, if all of the blue dots in diagram 1 are positioned on the red line and the sum of the vertical squared distance is equal to zero. Conversely if the relationship is found to be weak the R-squared measure will be closer to 0 and the sum of the vertical squared distance will be large.

22. The R-squared statistic therefore represents the proportion of the variation in y that is explained by the regression equation. For example if R-squared is 0.80, then 80% of the variation in y is explained by the regression equation

23. The econometrician William H Greene has noted that:

In terms of the values one encounters in cross-sections[7], an R^2(R-squared) of 0.5 is relatively high.[8]

D Functional Form

24. The standard (linear) functional form for a regression relationship of the type hypothesised in paragraphs 18 to 20 can be expressed algebraically as follows.

---

[6] For a complete analysis of how the random disturbance is used to calculate the R-squared metric see Greene, W., H..*Econometric Analysis,* Sixth Edition, Pretence Hall, 2008, pp 32-38.
[7] A cross-section is a data set that is taken at one particular point in time across different individuals, or firms etc.
[8] Greene, op. cit, pp. 38.

$$Y = b_0 + b_1 x_1 + u \text{ ;}$$

Where:

Y = investment cost (the dependent variable);

X_1 = the number of premises (the explanatory variable);

b_0 = the value of investment if the number of premises is zero (the constant coefficient);

b_1 = the extent to which investment cost changes as the number of premises changes (the coefficient of the explanatory variable); and,

u = the difference between the value of investment cost expected by the regression analysis and the value actually observed (the random disturbance).

25. The regression model described in paragraph 24 and Figure 1 has a linear functional form. It is referred to as a linear[9] functional form because the regression line that it exhibits is a straight line. The interpretation of the regression, in words, is that 'X has a linear impact on y'. Each time one of the x variables changes by 1 unit (meters, centimetres, kilograms, premises etc) investment costs (y) change by b1 units, and this is true no matter what the values of x and y are.

26. The drawback of the linear functional form for a regression is that it allows for only a linear relationship between variables. In our example, the linear functional form implies that each time the number of premises goes up by one unit investment cost goes up by b1 dollars. This relationship would not fit the data well if the cost of adding each additional premise decreases (or increases) with the total number of premises already served. For example, the cost of building the CAN network to the first premise may be more or less expensive than the millionth premise. Thus the linear functional form model imposes severe restrictions on the relationship between CAN construction costs and CAN characteristics.

27. A less restrictive functional form is the logarithmic functional form. This has the following standard form:

$$\ln(Y) = b_0 + b_1 \ln(x_1) + u \text{ .}$$

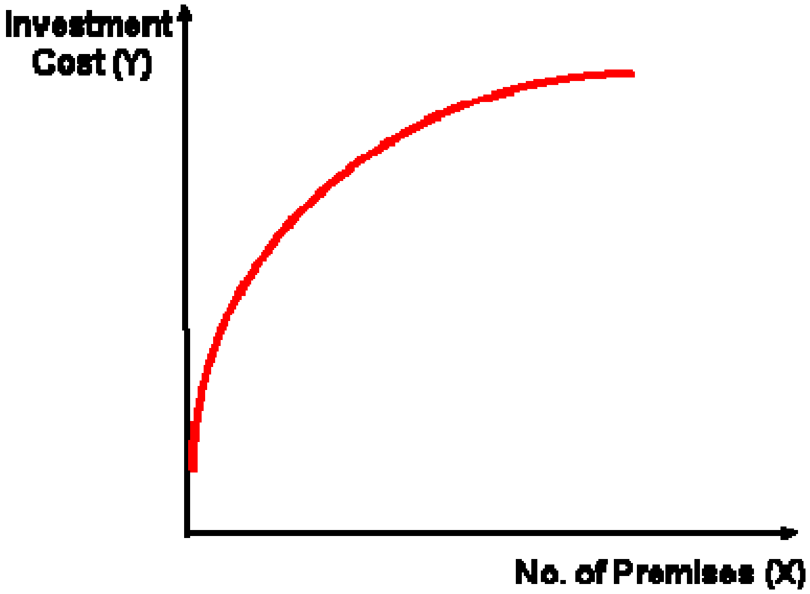
28. This functional form has the same basic components as the standard linear functional form (and the x and y variables are the same). However, in the logarithmic model the x1 and y variables are transformed by taking their natural logarithm (ln) prior to the regression analysis and prior to obtaining a value for b1.

29. Taking the natural logarithm of the x1 and y variables, as far as regression modelling is concerned, changes the interpretation to be placed upon the estimated value of b1. That is, b1 is now interpreted as the percentage change in Y given a 1 percent change in X1.

[9] Linear within regression analysis technically refers to linearity in the parameter estimates b0, b1 etc. Nonlinearity of the data such as quadratic terms do not make the model non-linear.

30. Diagram 2 illustrates the logarithmic functional form describing the relationship between investment cost and the number of premises connected to the CAN. This functional form allows the effect of the number of premises on investment cost to vary depending on the number of premises served.

Figure 2: Regression of logarithmic functional form



31. One difficulty with the models so far outlined is that they include only one explanatory variable and also do not allow for the possibility that an explanatory variable could have an indirect effect on the dependent variable via its effects on other explanatory variables. For example, as the number of premises increase it would be reasonable to assume that the length of distribution and mains conduit would also change, however not necessarily at a constant rate. This indirect effect is known as an interaction effect.

32. Interaction effects and non-constant rates of change in variables can be accommodated in regression analysis by way of a third functional form for a regression model, know as a transcendental logarithmic (translog) functional form. The translog functional form is an extended version of the basic logarithmic functional form outlined in paragraph 27.

33. The translog functional form, for the case of 1 explanatory variable (x1) can be expressed as follows.

$$\ln(Y) = b_0 + b_1 \ln(x_1) + b_{11}(1/2)\ln(x_1)\ln(x_1) + u$$

where:

Y	= investment cost;
x_1	= the number of premises;
b_0, b_1, b_{11}	= coefficients estimated by regression analysis
u	= the random disturbance.

34. The term ln(x1)ln(x1) in the equation above makes it possible to not only establish what percentage change in y1 results from a percentage change in x1 (which is what the standard logarithmic functional form allows), but also allows for the relationship between the x and y percentage changes to be different for different values of x. For example, the percentage change in investment cost given a 1% increase in premises from 1000 premises can be different with the translog functional form to the percentage change in investment cost given a 1% increase in premises from 1 million premises.

35. Therefore the more variables a functional form can posses the more desirable and 'flexible' it is. As the well known empirical economist Larry Lau has noted:

> *Flexibility of functional form is desirable because it allows the data the opportunity to provide information about critical parameters. An inflexible functional form often prescribes the values, or at least the range of values, of critical parameters (which should ideally be) free to attain any set of theoretically consistent values.*[10]

36. The translog functional form is what Lau classes as a 'flexible functional form'[11] as the translog functional form imposes minimal restrictions on the relationship between[12] the response variable (CAN investment costs) and the explanatory variables (inputs into the construction of the CAN and its design characteristics).

37. The translog function can also be used to concurrently accommodate the environment in which the CAN must exist, such as its geographic location, all within one regression equation. For example, it is possible within one regression model to test and account for differences in investment costs between the states and territories of Australia, and test whether moving the placement of pillars has an impact on investment costs.

38. An example of a three variable translog functional form regression, which includes two explanatory variables and one environmental variable, can be written as below:[13]

$$\ln(Y) = b_0 + b_1 \ln(x_1) + b_2 \ln(x_2) + b_{11}(1/2)\ln(x_1)\ln(x_1) + b_{22}(1/2)\ln(x_2)\ln(x_2)$$
$$+ b_{12}\ln(x_1)\ln(x_2) + \sum_{i=1}^{7} DState + u$$

where:

Y = investment cost;

x1, x2 = the direct explanatory variables of investment cost, here assumed to be two;

State = an environmental variable that captures for example if an ESA is in a particular state or territory of Australia;

[10] Lau, L. (1986). "Functional Forms of Econometric Model Building." In Griliches, Z. And Intriligator, M.D. eds., *Handbook of Econometrics,*V.3, pp.1513-1566.

[11] Technically a flexible functional form is one in which the functional form can approximate an arbitrary twice continuously differentiable function to the second order.

[12] This is not an exhaustive list of the properties of flexible functional forms, however the remaining properties are extremely technical in nature and will not add to the understanding of why such forms are appropriate.

[13] This form applies when symmetry is assumed. That is for example ln(a)ln(b) = ln(b)ln(a), therefore there is no need to include the additional symmetric variables and estimate additional parameters in the regression (nor mechanically possible to estimate the regression).

b_0, b_1, b_2, b_{11}, b_{22}, b_{12} and D are coefficients estimated by regression analysis;

u = the random disturbance.

39. In the following section the, data used in the modelled translog functional form regression is summarised and the regression model results are presented.

E The regression equation, data, regression results and interpretation

E.1 Regression equation

40. The TEA model provides information in relation to investment costs for each ESA in band 2 areas. The following are likely possible drivers of the cost in each ESA:

- Main and distribution conduit and cable;
- Joints used to connect lengths of cable;
- Pillars, pits, and manholes;
- The number of premises connected to the CAN;
- The length of lead-in cable (cable from the customer premises to the distribution cable in the street);
- The relative location of pillars;
- The average size of a DA;
- The perimeter length of an ESA
- The density of each ESA;
- The average dispersion of premises in an ESA; and,
- The state or territory of Australia each ESA is located.

41. The major components of the CAN such as cable and conduit lengths, joints, pits, pillars, and manholes etc will exhibit a degree of correlation. For example, correlation between conduit and cable lengths are expected since cable is laid inside conduit, and generally a number of joints are required for each given length of cable.

42. A high degree of correlation between explanatory variables in a regression is known as multicollinearity. Multicollinearity can cause statistical problems for the results of a regression if it is not addressed by aggregating, removing or using proxy variables for those variables that are highly multicollinear.[14] To avoid any potential problems associated with

[14] See Griffiths, W. E., Hill C., R., and Judge, G., G., (1993), *Learning and Practicing Econometrics* .John Wiley & Sons, pp. 435.

multicollinearity of the explanatory variables of CAN investment cost, the major components of the network (conduit, cable, joints, pits, pillars, and manholes) were aggregated into a single quantity index of the CAN. A quantity index allows for a simple interpretation of all of the components of the CAN without loss of any explanatory power, Appendix 1 outlines the quantity index.

43. Environmental variables were included in the cost function to control for factors relating to each ESA — density, dispersion and perimeter length. In addition interaction terms were constructed from the environmental variables. Interaction terms capture the effect on CAN investment costs from the natural interaction of two environmental variables. It may be the case that variables such as density and dispersion of an ESA or dispersion and perimeter measure of an ESA are inter-related. For example, if two ESAs are of identical area, but one ESA servers a larger number of customers, the ESA with more customers will have a higher density (calculated as customers divided by area). Further if one of the ESAs consist mainly of free-standing buildings spread out evenly (greater dispersion) throughout the ESA and the other consists primarily of apartment blocks with multiple units within each block situated in a small area of the ESA, the former will have greater dispersion. The interaction of density with dispersion will have distinctly different impacts within each ESA. Including such interaction terms allows for the possible impact of these relationships on CAN investment costs for each individual ESA.

44. Variables relating to DA design were added to the list of investment cost drivers to test the effect that these variables have on investment costs. These variables were – the relative location of pillars, the average size of DAs, a squared DA Size variable and an interaction term to examine if pillar location and DA size are related. Including a squared term of the 'DA_Size' variable allows for the impact of 'DA_Size' on investment cost to vary with changes in 'DA_Size'.

45. There may also be factors that influence construction costs that are particular to certain states or territories in Australia. These could include different street layouts, council restrictions and suburb design between states/territories etc. To account for differences between states a series of binary qualitative variables known as dummy variables are included, one for each state or territory excluding the chosen state/territory as the basis of comparison which is implicitly represented by the constant term in the model. The variable takes the value of 1 if the ESA is in the specified state/territory and 0 otherwise and is interpreted relative to the chosen base state/territory of comparison

46. The translog functional form cost model that captures these investment cost drivers and DA design variables is written as follows:

$$\ln TC_i = \beta + \alpha_1 \ell n(CAN\_INDEX_I) + \alpha_2 0.5(\ell n(CAN\_INDEX_I))^2$$

$$+ D_1 NSW_I + D_2 QLD_I + D_3 NT_I + D_4 WA_I + D_5 SA_I + D_6 VIC_I + D_7 TAS_I$$

$$+ \varsigma_1 \ell n(LOCATION_I) + \varsigma_2 \ell n(DA\_SIZE_I) + \varsigma_3 (\ell n(DA\_SIZE_I))^2$$

$$+ \varsigma_4 \ell n(LOCATION_I) \ell n(DA\_SIZE_I)$$

$$+ \varsigma_5 \ell n(DENSITY_I) + \varsigma_6 \ell n(PERIMETER_I) + \varsigma_7 \ell n(DISPERSION_I) +$$

$$+ \lambda_1 \ell n(DENSITY_I) \ell n(PERIMETER_I) + \lambda_2 \ell n(DENSITY_I) \ell n(DISPERSION_I)$$

$$+ \lambda_3 \ell n(PERIMETER_I) \ell n(DISPERSION_I) + \mu_i$$

Where:

TC_i = Total cost in the *i*th ESA in Band 2;

β = Constant;

CAN_INDEX_i = Quantity index measure of the components of the CAN in the *i*th ESA;

NSW_i = Equals 1 if ESA is in New South Wales, zero otherwise;

QLD_i = Equals 1 if ESA is in Queensland, zero otherwise;

NT_i = Equals 1 if ESA is in the Northern Territory, zero otherwise;

WA_i = Equals 1 if ESA is in Western Australia, zero otherwise;

SA_i = Equals 1 if ESA is in South Australia, zero otherwise;

VIC_i = Equals 1 if ESA is in Victoria, zero otherwise;

TAS_i = Equals 1 if ESA is in Tasmania, zero otherwise;

$LOCATION_i$ = The average relative position of the pillars in DAs in the *i*th ESA;[15]

DA_SIZE_i = The average size of a DA in the *i*th ESA;

$DENSITY_i$ = The density of the *i*th ESA;

$PERIMETER_i$ = The perimeter length of the *i*th ESA;

[15] The relative position is calculated as the linear distance between the pillar location and the average geographic location of premises in the DA connected to that pillar.

$DISPERSION_i$ = The average dispersion of a DA in the *i*th ESA;

μ_i = random disturbance term;

ℓn = Natural logarithm operator; and,

α , D , ς & λ = are parameters to be estimated via the least squares method.

47. We note that a typical cost function would also include input price variables, such as the prices of capital, labour and materials used in the construction of the CAN.[16] In the TEA model, the prices of capital, labour and materials in the underlying vendor prices for plant and equipment do not vary across ESAs. Because input prices are constant across ESAs, an input price variable would be indistinguishable from the constant term (β) in the regression model.

E.2 Data

48. Data for this study are derived from version 1.2 of the TEA model and the Cable Plant Records 2 (CPR2) system. Table 1 gives the definitions of each variable used in the regression modelling.

[16] See for example Bloch. H., Madden. G. and Coble-Neal. G., (2001). *The cost structure of Australian telecommunications.* The Economic Record vol. 239, Issue 77. pp. 338-350 and Kiss, F., and Lefebvre, B.J. (1987), *"Econometric Models of Telecommunications Firms: A Survey"*, Economique, Paris, Vol. 38, No. 2, pp. 307-374.

Table 1: TEA model data used

Variable	Description	No. Obs.	Description	Source
TC	Total Cost	584	Sum of CAN investment costs for each ESA calculated by the TEA model v1.2. All costs incurred in building the Band 2 network are included.	TEA model V1.2
CAN_Index	Output quantity index.	584	Main and distribution conduit length, pits, pillars, joints plus length of cable and lead-ins formed into a quantity index	Calculated from TEA model V1.2
DNSW**	Dummy New South Wales	173 =1 & 411 = 0	Dummy variable which equals 1 if the ESA is within the state and 0 otherwise.	Calculated from TEA model V1.2
DQLD	Dummy Queensland	118 = 1 & 466 = 0	Dummy variable which equals 1 if the ESA is within the state and 0 otherwise.	Calculated from TEA model V1.2
DNT	Dummy Northern Territory	3 = 1 & 581 = 0	Dummy variable which equals 1 if the ESA is within the state and 0 otherwise.	Calculated from TEA model V1.2
DSA	Dummy South Australia	34 = 1 & 550 = 0	Dummy variable which equals 1 if the ESA is within the state and 0 otherwise.	Calculated from TEA model V1.2
DWA	Dummy Western Australia	63 = 1 & 521 = 0	Dummy variable which equals 1 if the ESA is within the state and 0 otherwise.	Calculated from TEA model V1.2
DSA	Dummy Australian Capital Territory	13 = 1 & 571 = 0	Dummy variable which equals 1 if the ESA is within the state and 0 otherwise.	Calculated from TEA model V1.2
DVIC	Dummy Victoria	167 = 1 & 417 = 0	Dummy variable which equals 1 if the ESA is within the state and 0 otherwise.	Calculated from TEA model V1.2
DTAS	Dummy Tasmania	13 = 1 & 571 = 0	Dummy variable which equals 1 if the ESA is within the state and 0 otherwise.	Calculated from TEA model V1.2
Location	Pillar Location	584	Actual pillar location in a DA relative to the copper centre[17] of that DA.	Calculated from CPR2
DA Size	Average size of a DA	584	The average size of a DA in each ESA.	Constructed from TEA model V1.2 data as the ratio of area in an ESA to the number of DAs.

[17] The copper centre is the geographically-weighted average location of customer addresses in each DA.

Density	Density of ESA	584	The density of each ESA.	Constructed from TEA model V1.2 data as the ratio of premises in an ESA to area.
Perimeter	Perimeter length of the ESA	584	The perimeter length of each ESA.	Calculated from CPR2
Dispersion	Average customer distance from the copper centre in a DA	584	The average distance of customers from the copper centre of each DA within the ESA	Calculated from CPR2

** The Australian Capital Territory is the base dummy variable of comparison.

Table 2 provides summary statistics and units of measurement of the variables used in the statistical analysis and obtained from the TEA model V1.2 and CPR2 databases respectively.

Table 2: Descriptive statistics

Abbreviation	Units of measurement	Mean	S.D.
TC	Dollars ($)	$34,811,866	$19,496,063
CAN_Index	Percentage (%)	0.635	0.353
DNSW	N/A 0 or 1	0.296	0.457
DQLD		0.005	0.071
DNT		0.202	0.402
DWA		0.108	0.310
DSA		0.058	0.234
DVIC		0.286	0.452
DTAS		0.022	0.145
Location	Metres (m)	192.523	57.994
DA Size	Square kilometres (km^2)	0.362	0.317
Density	Premises per square kilometre ($p.km^2$)	616.415	421.221
Perimeter	Metres (m)	196.589	105.544
Dispersion	Metres (m)	151.424	54.452

E.3 Regression results

49. The initial regression results of the estimated model of CAN investment costs, including all variables listed above are given in Table 3.

Table 3: Initial regression results for the CAN investment cost model

Variable	Coefficient	Standard Error	P-Value
β	17.523*	0.477	0.000
$\ln(CAN\_INDEX_I)$	0.965*	0.013	0.000
$(\ln(CAN\_INDEX_I))^2$	-0.002	0.004	0.630
$D_1 NSW$	0.045*	0.011	0.000
$D_2 QLD$	0.056*	0.012	0.000
$D_3 NT$	0.103*	0.024	0.000
$D_4 WA$	0.061*	0.012	0.000
$D_5 SA$	0.059*	0.014	0.000
$D_6 VIC$	0.057*	0.012	0.000
$D_7 TAS$	0.063*	0.016	0.000
$\ln(LOCATION_I)$	-0.054*	0.013	0.000
$\ln(DA\_SIZE_I)$	0.065	0.056	0.247
$(\ln(DA\_SIZE_I))^2$	0.012*	0.004	0.008
$\ln(LOCATION_I)\ln(DA\_SIZE_I)$	-0.018	0.010	0.057
$\ln(DENSITY_I)$	0.112*	0.057	0.048
$\ln(PERIMETER_I)$	0.055	0.065	0.396
$\ln(DISPERSION_I)$	-0.061	0.084	0.470
$\ln(DENSITY_I)\ln(PERIMETER_I)$	-0.015*	0.005	0.002
$\ln(DENSITY_I)\ln(DISPERSION_I)$	-0.007	0.010	0.478
$\ln(PERIMETER_I)\ln(DISPERSION_I)$	0.017*	0.009	0.042
R-Squared^	0.9971		

* Individually statistically significant at the 5% level of significance.
^ Adjusted R-squared statistic reported.

50. At the bottom of Table 3 is a row titled R-squared[18] which, as explained in paragraph 21, is a measure of how well the regression line fits the underlying data. As the R-squared for this regression model is 0.9971 we

[18] The reported statistic is the adjusted R-squared.

can conclude that the model provides a very good approximation of the factors driving CAN investment costs in the TEA model V1.2.

51. The column in Table 3 headed Coefficient shows the impact of the variable given in the first column on the natural logarithm of investment cost, holding all other variables in the model including interaction variables constant. In general a negative coefficient implies the variable has a negative relationship with investment cost and a positive coefficient implies a positive relationship. However, where a direct driver of investment cost appears more than once (for example, 'DA_Size' and the square of 'DA_Size') in the translog regression, the sign of an individual coefficient has no direct interpretation. It is the net impact of all of the coefficients of the direct driver in combination that determine the variables overall impact.

52. The column headed P-value can be used to represent the proportion of times we could expect to reject the null hypothesis (that is, that the value of a coefficient is equal to zero) when in fact the coefficient was equal to zero.[19] In this analysis we have chosen to use a 5 per cent significance level. Thus coefficients with a p-value at or below 0.05 are "statistically significant" and are marked with a single asterisk in Table 3 above.

53. Several statistical tests were undertaken on these results to determine whether any refinement to the hypothesised model of CAN investment costs were necessary. In particular, given that the variables of the translog cost function enter the regression a number of times, an F-test[20] can be performed to assess whether those variables have a 'joint' statistical effect on investment costs. The results of the F-test are presented in Table 4 below. All jointly statistically significant coefficients are marked with an asterisk.

[19] Technically speaking the P-value is defined as the probability of rejecting the null hypothesis that a variable is statistically equal to zero, when in fact the null hypothesis is true i.e. the variable is equal to zero.
[20] The F-test is a statistical test that assesses whether the coefficients of a group of variables are jointly statistically significant.

Table 4: F-test for the joint exclusion of variables

Null Hypothesis	Test Statistic	P-Value	Result of test
H_0: That all **CAN_Index** coefficients are jointly equal to 0.	2994.460 *	0.00	All **CAN_Index** coefficients are jointly **statistically different to 0.**
H_0: That all **Location** coefficients are jointly equal to 0.	9.244*	0.002	All **Location** coefficients are **jointly statistically different to 0.**
H_0: That all **DA_Size** variable coefficients are jointly equal to 0.	24.119*	0.000	All **DA_Size** coefficients are **jointly statistically different to 0.**
H_0: That all **Density** coefficients are jointly equal to 0.	3.564*	0.014	All **Density** coefficients are **jointly statistically different to 0.**
H_0: That all **Perimeter** coefficients are jointly equal to 0.	14.622*	0.000	All **Perimeter** coefficients are **jointly statistically different to 0.**
H_0: That all **Dispersion** coefficients are jointly equal to 0.	4.797*	0.002	All **Dispersion** coefficients are **jointly statistically different to 0.**

* Jointly statistically significant at the 5% level of significance.

54. The results of the F-tests indicate that jointly all variables in the translog cost function are statistically significant and aide in explaining CAN investment costs in some combination

55. The results in Tables 3 and 4 demonstrate that the model in Table 4 is statistically justified.

E.4 Interpretation of results

56. The overall effect of a statistically significant variable on investment cost can be summarised by an elasticity. Elasticities provide a convenient way of summarising the interpretation of the total impact of a variable of interest when interaction effects are included in a regression model (for example, ‘a 1 percent change in x brings about a b percent change in y’).

57. We can calculate the elasticity of investment cost with respect to the size of the network put in place. The average value of the output cost elasticity is 0.966[21] which is interpreted as a 1% increase in outputs brings about a 0.966% increase in costs. The inverse of this elasticity,[22] is a measure of the

[21] T-tests of the null hypothesis of lqn + lqn2 = 1 show that the output cost elasticity is statistically different to 1 even at the 1% level of significance, with a p-value of 0.0023.
[23] See Diewert, W. E. (1974), “Applications of Duality Theory”, pp106-171 in Frontiers of Quantitative Economics, Volume 2, M. D. Intriligator and D. A. Kendrick (eds.), Amsterdam: North-Holland.

returns to scale of the CAN investment function. Thus the calculated scale elasticity is 1.04 which suggests the existence of slightly increasing returns to scale for the investment in the CAN.[23]

58. We note that the implied scale elasticity of 1.04 is in line with previously estimated scale elasticities for telecommunications networks. Bloch, Madden and Coble-Neal (2001),[24] for example, estimate scale elasticities for Australian telecommunications networks in the order of 0.94 to 1.14. Kiss and Lefebvre (1987)[25] survey 36 telecommunications cost studies from the 1950's to the mid-1980's, in which scale elasticities are calculated. The reported range of scale elasticities is from 0.94 to 1.75.

59. The results of the estimated cost function also indicate that the characteristics of the network also impact on investment cost of the CAN. The impact (all else equal) of each, relative to the average ESA in Band 2, is summarised:

- ESAs with 1% higher densities cost 0.007% less;
- ESAs with 1% longer perimeters cost 0.044% more; and,
- ESAs with 1% greater average DA dispersion cost 0.012% less.

60. In relation to the final point above, it follows that, if two ESAs have identical lengths of conduit, copper, trenching etc and the same number of pits, pillars, manholes and premises etc, but one ESA has a larger dispersion, then the ESA with the larger dispersion will cost less. This is because in practice (all else being equal) ESAs with larger dispersion contain less concrete and asphalt. Thus, it follows that (all else equal) ESAs with greater dispersion have lower breakout and reinstatement costs than ESAs with lower dispersion, and hence have lower costs.

61. In addition, we find that DA design has a statistically significant but very small theoretical impact on investment cost. For example, the elasticity of the 'Location' variable is calculated as -0.027 implying that ESAs with 1% longer average distance from the pillar to the copper centre will have 0.027% lower investment costs. Thus, it follows that placing pillars close to the copper centre of a DA will result in higher investment cost. In practice, pillars are typically located on main roads on or near the boundary of the DA (as this is where the main network is typically placed). To move a pillar off the main road toward the copper centre of the DA would require additional main network costs being incurred.

62. The calculated elasticity of the DA-Size variable is -0.048. To increase the average size of a DA holding the size of an ESA constant requires reducing the number of DA's in an ESA. Reducing the number of DAs in each ESA by 16 (and those ESAs with 16 DAs or less, to 1) results in the average size of DAs increasing by approximately 25%. This hypothetical change would result in the investment cost reducing by 0.020% (or less than 1 cent per month).

---

[24] Bloch et al Opt. Cit., (2001)
[25] Kiss et al Op. Cit. , (1987)

63. Given the size of this saving and the extent of the restructure that would be required to achieve it, we conclude that DA design has no practical impact on the investment cost of the CAN.

F Conclusion

64. Overall, the analysis presented shows that the regression analysis (presented in Table 3) explains investment costs very well. In particular, the explanatory variables chosen for the analysis explain greater than 99 percent of the variation in the log of CAN investment costs between ESAs. The regression also displays a measure of RTS in line with previous research, which adds further validity to the regression model.

65. In terms of the characteristics of an ESA, the results of the estimated cost function also indicate that the characteristics of the network impact on investment cost of the CAN in Band 2.

66. In terms of DA design, the analysis indicates that DA design variables have an extremely small impact on Band 2 investment costs. In particular keeping all other variables constant:

- The location of pillars has no material impact on investment costs; and,
- The average size of DAs has no material impact on investment costs.

67. These results show that changing the DA design in the TEA model would not materially impact investment costs.

APPENDIX A: The Conduit Index

68. A measure of the aggregate of:

- Mains and distribution conduit;
- Mains and distribution cable;
- Lead-in cable;
- Pits;
- Pillars; and,
- Manholes,

was derived.

69. As the cost of (for example) a kilometre of mains conduit differs from the cost of a kilometre of distribution conduit or the costs of manholes of different sizes differs the aggregate of quantities can not be found simply by adding the different quantities of the CAN in each respective ESA. A measure of the quantities across ESAs can be derived via the calculation of an index.

70. While there are a variety of index number formulations that could be used to derive a measure of the quantities of the CAN in each respective ESA a Tornqvist quantity index formulation was implemented in this study as this index formulation provides an exact aggregation of the quantities of the CAN in each respective ESA where the underlying function is a translog function as used in this study.

71. The Tornqvist log-change quantity index is written as follows.

$$\ln Q_{st}^{T} = \sum_{i=1}^{N}\left(\frac{w_{i1} + w_{i2} + \ldots + w_{in}}{n}\right)\left(\ln q_{i1} - \ln q_{i2} - \ldots - \ln q_{in}\right)$$

Where:

Q_{St}^{T} = is the quantity index;

w_{i1} = is the cost share of the first output of the CAN in the ith ESA;

w_{i2} = is the cost share of the second output of the CAN in the ith ESA;

w_{in} = is the cost share of the nth output of the CAN in the ith ESA;

q_{it} = is the quantity of the first output of the CAN in the ith ESA;

q_{is} = is the quantity of the second output of the CAN in the ith ESA;

q_{is} = is the quantity of the nth output of the CAN in the ith ESA;

$\ln$ = the natural logarithm operator.

72. When the quantity of outputs across ESAs is expressed as an index, one ESA has to be chosen as the base ESA. In the base ESA the index has a value of one. It is possible

that the value of the quantity index for ESAs other than the base ESA could be affected by the ESA chosen to be the base ESA. To avoid this possibility a so called transitive Tornqvist quantity index was used in this study. Coelli et al provide a discussion of why transitive index numbers are required when calculating aggregates across different entities such as different ESAs[26].

[26] Tim Coelli, D.S Prasada Rao and George Battese 2002, *An Introduction to Efficiency and Productivity Analysis*, Kluwer Academic Publishers, pp. 84-87.