www.pwc.com.au

Data Retention Grant Allocation

Methodology

Attorney-General's Department

07 July 2016

pwc

Disclaimer

This Report has been prepared by PricewaterhouseCoopers (Australia) Pty Limited (PwC) in our capacity as advisors to the Attorney-General's Department in accordance with Official Order 10008303 dated 27th January 2016.

The information, statements, statistics, material and commentary (together the "Information") used in this Report have been prepared by PwC from information provided by the Attorney-General's Department and from discussions held with a range of client stakeholders. PwC has relied upon the accuracy, currency and completeness of the Information provided to it and takes no responsibility for the accuracy, currency, reliability or correctness of the Information and acknowledges that changes in circumstances after the time of publication may impact upon the accuracy of the Information. The Information may change without notice and PwC is not in any way liable for the accuracy of any information used or relied upon by a third party.

Furthermore PwC has not independently validated or verified the Information provided to it for the purpose of the Report and the content of this Report does not in any way constitute an audit or assurance of any of the Information contained herein.

PwC has provided this advice solely for the benefit of the Attorney-General's Department and disclaims all liability and responsibility (including arising from its negligence) to any other parties for any loss, damage, cost or expense incurred or arising out of any person using or relying upon the Information.

Executive summary

The Data Retention Industry Grants Programme (DRIGP) exists to assist eligible telecommunications service providers meet their data retention obligations under the Telecommunications (Interception and Access) Amendment (Data Retention) Act 2015. A grant of up to $128.4 million has been made available for this purpose, and Guidelines published on 7 January 2016 defined principles under which these funds would be allocated. One of these Guidelines[1] provides for an independent consultant to be appointed to assist the Department in the analysis of the grant applications, and the application of the methodology set out later in the document.[2] PwC was appointed to this role, and this report summarises the approach that we took to producing a recommended grant distribution that is not only compliant with the specific Guidelines set out by the Department, but also consistent with principles of equity and transparency.

The purpose of this document is to summarise the overall process undertaken by PwC, from the application for funds to the allocation recommendation as specified in the Guidelines. After an extensive screening process by the Department, 15 providers withdrew their applications, while 15 applicants were found ineligible for compensation under the Guidelines.[3] 180 Providers in total, therefore, were deemed to be "Eligible Applicants".

The grant amount for each Provider deemed to be eligible under the Guidelines is determined by a total score of up to 100 points[4]. Comprising this total score:

- a maximum of 25 points are determined by the Enterprise Scale, which allows for differences in cost due to the relative size of each Provider and

- a maximum of 75 points are determined by an analysis of the relationship between each Provider's Estimated Costs, and

    i. the number of Eligible Service(s);
    ii. the type(s) of Eligible Service(s) offered;
    iii. the number of subscribers;
    iv. gross annual revenue and
    v. the anticipated data storage required to meet the data retention obligations.

This estimated relationship is called the "Typical Implementation Impact".

We used a technique called multivariate regression to estimate the Typical Implementation Impact. This technique also measured the relative importance of each of the five factors listed in the second bullet point above in determining Estimated Cost.

Of the five factors listed, our analysis showed that all except for the Eligible Service type were statistically significant contributors to Estimated Cost. The 75 points available to be allocated for the Typical Implementation Impact were split across the four remaining factors, according to how important the factor was in determining implementation cost. Individual provider scores were allocated based on a linear scale for each factor, with reference to each Provider's declared factor

[1] DRIGP Programme Guidelines, Jan 2016, Sections 29-30

[2] Ibid. Sections 59 – 67

[3] Ibid. Section 38

[4] Ibid. Section 61 – 62

values. The total score out of 100 was calculated by adding the Enterprise Scale score and the Typical Implementation Impact score.

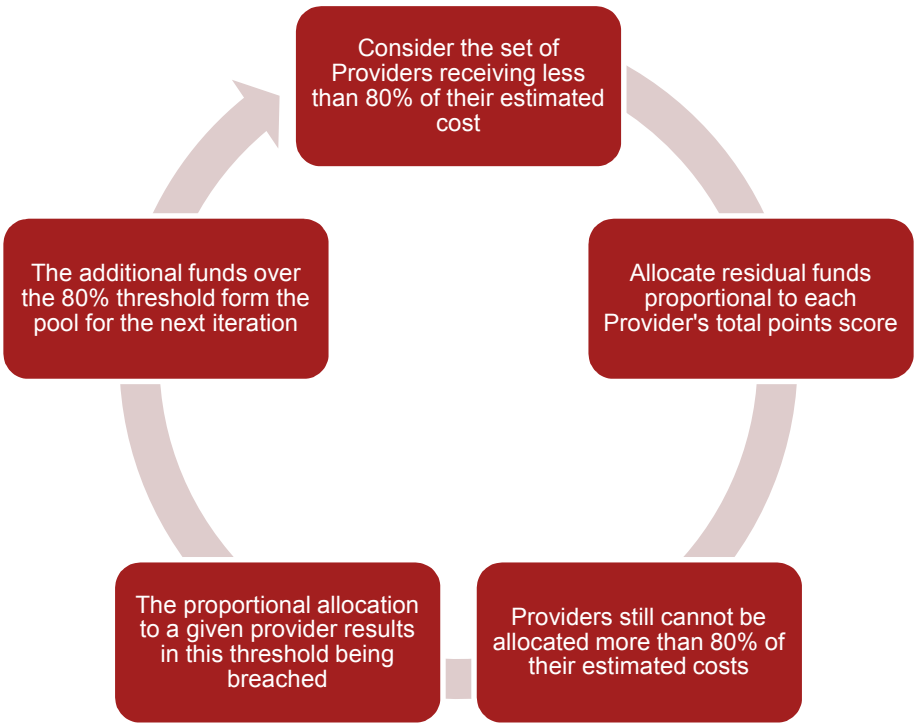
Using our recommended methodology, the total of the scores allocated across all Eligible Applicants was 7,006, an average (mean) score per provider of 38.92 out of 100. If the total available funding grant were divided linearly and proportionately across these scores, each point would be worth $18,321. If this allocation were applied unadjusted, a large number of providers (88 per cent) would be reimbursed with grant amounts higher than their estimated costs. Moreover, 74 per cent of providers would be reimbursed with amounts over double their estimated costs. The Guidelines note that that funding allocations will not reimburse full costs, but are a financial contribution to the typical up-front costs of compliance. Therefore, the Programme Delegate is given the discretion to determine maximum and minimum constraints for the allocation.[5]

In consultation with the Programme Delegate, funds were allocated based on each Provider's modelled cost, but applying a maximum threshold of 80 per cent of the Provider's estimated cost. This allowed the grant to be based on modelled cost and remain contributory, whilst tolerating variation between providers below this threshold.

Once the initial allocation has been constrained to an upper limit of 80 per cent of estimated cost, there were Eligible Applicants at the bottom end of the distribution whose allocations given this maximum threshold would be very low in dollar terms. Therefore, the Programme Delegate determined that a minimum grant size of $10,000 should be set so that the very smallest providers are not financially penalised for operating on a small scale.

Constraining the allocation in the manner described above, there remained $90 million out of the total Programme Grant unallocated. These remaining funds were allocated iteratively as described in the diagram below:

Iterative allocation process



[5] Ibid. Sections 24c, 65

The iterative process continued until the entire Programme grant has been allocated. This produces our recommended allocation, which is provided in detail in Appendix E of this report. A variation on the above approach allocates the residual funds proportional to each provider's Typical Implementation Impact score (out of 75) rather than the total score (out of 100). This alternative allocation is provided in Appendix F.

Our recommended allocation considers both the equity and transparency objectives contained within the Guidelines, in addition to the narrow focus of a methodology consistent with points and modelled cost. We consider the constraints that we have applied to the allocation of funds as the minimum necessary so that Providers were not compensated with amounts significantly greater or less than their estimated costs. In the recommended allocation, only 2 Eligible Applicants were compensated at less than the constrained maximum of 80 per cent. The minimum reimbursement rate was 47 per cent, and this was in the case of an applicant whose estimated costs far exceeded those modelled based on their independent variable characteristics.

By contrast, in the unconstrained allocation, over 89 per cent of providers would be allocated more than 80 per cent of their estimated costs, and 4 providers allocated less than 20 per cent.

In summary, the recommended allocation:

- ✓ Allocates the entire Programme Grant
- ✓ Allocates no more than 80 per cent of estimated cost to any Eligible Applicant
- ✓ Allocates no less than 47 per cent of estimated cost to any Eligible Applicant
- ✓ Allocates no less than $10,000 to any Eligible Applicant
- ✓ Addresses the overall Programme Objectives as specified in the DRIGP Guidelines

Contents

1 Background

The Data Retention Industry Grants Programme (DRIGP) exists to assist eligible telecommunications service providers meet their data retention obligations under the *Telecommunications (Interception and Access) Amendment (Data Retention) Act 2015*.

A pool of grant funding of up to $128.4 million has been made available for this purpose, and Guidelines published on 7 January 2016 by the Attorney General's Department defined principles under which these funds would be allocated. One of these Guidelines[6] provides for an independent consultant to be appointed to assist the Department in the analysis of the grant applications, and the application of the methodology set out later in the document.[7] PwC was appointed to this role, and this report summarises the approach that we took to producing a recommended grant distribution that is not only compliant with the specific Guidelines set out by the Department, but also consistent overall with principles of equity and transparency.

The purpose of this document is to summarise the overall process undertaken by PwC, from the application for funds to the allocation recommendation as specified in the Guidelines. The text will occasionally indicate references to Appendices, which will generally present technical analysis supporting our findings and recommended allocation.

1.1 The application process

Companies operating telecommunications infrastructure in Australia that:

- offer at least one service deemed 'eligible' by the department[8] and
- were impacted financially as a result of complying with the new data retention requirements[9]

were given the opportunity to apply for a DRIGP grant. Compliance costs expended prior to 30 October 2014 were deemed ineligible for compensation under the DRIGP.[10] The application process closed on 23 February 2016, and at close of business on this date, 210 providers had submitted applications for a grant. After an extensive screening process by the Department, 15 providers withdrew their applications, while 15 applicants were found ineligible for compensation under the Guidelines. The balance of this report describes the methodology used to determine the size of the grant allocation for the remaining 180 Providers ("Eligible Applicants"). The total cost of implementation estimated by the 180 Eligible Applicants was $198.5 million.

1.2 Funding model

The grant amount for each Eligible Applicant is determined by a total score of up to 100 points.[11] Comprising this total score:

- a maximum of 25 points are determined by the Enterprise Scale

---

6 DRIGP Programme Guidelines, Jan 2016, Sections 29-30

7 Ibid. Sections 59-67

8 Ibid. Sections 43 – 45. See also Appendix A for a list of Service Types.

9 DRIGP Programme Guidelines, Jan 2016, Section 38. To be eligible, this cost must be either forecast by a Data Retention Implementation Plan (DRIP) already submitted separately to the department, or a Statement of Work attached to the application in the case of already compliant companies.

10 Ibid.

11 Ibid. Sections 61-62

- a maximum of 75 points are determined by an analysis of the cost data and other variables submitted in the provider's application (Typical Implementation Impact).

The requirement of the Guidelines for points to be assigned to Eligible Applicants in this way was based on the contention that smaller providers might be expected to have higher unit compliance costs, and therefore, funds were explicitly set aside for the purpose of compensating them. Additionally, in the Department's view, the Guidelines direct that the funding model has to be agnostic between points accruing from the Enterprise Scale and the Typical Implementation Impact.[12] As will be described in Section 3, one of the consequences of a pure allocation on this basis is that smaller providers would be on average overcompensated relative to their estimated costs. In the final analysis, we developed legitimate adjustments to the pure allocation consistent with a strict interpretation of the Guidelines. Looking at the broader intent of the DRIGP as a whole however, there are many other possible approaches to allocating the funds, but we feel that the approach outlined in Section 3 below is optimal, subject to the objectives and constraints outlined in the Guidelines.

1.2.1 Enterprise Scale Score

One of the aims of the DRIGP is to address smaller providers' greater expected need for financial assistance to comply with their data retention obligations.[13] Since there are some relatively fixed costs of compliance, it is anticipated that some of the financial impact on a smaller provider will be proportionately greater than for a larger provider.

An Enterprise Scale, based on the size of the Applicant (as determined by gross annual revenue for the most recent full financial year) was developed to lend support to smaller providers. Eligible Applicants with annual gross revenue of up to $3 million were allocated a score of 1 to 25 points (with the smallest businesses receiving the most points). Eligible Applicants with revenues above $3 million received 0 points out of the 25 available. This enterprise scale is provided in Appendix B.

1.2.2 Typical Implementation Impact

The Typical Implementation Impact is derived by calculating the typical cost of achieving compliance based on the information provided by Eligible Applicants on their costs of compliance. The Guidelines indicate that only the following variables may be taken into account when assessing the typical cost to a provider:

i Number of Eligible Service(s)
ii Type(s) of Eligible Service(s)[14]
iii Number of subscribers
iv Gross annual revenue (turnover) for the most recent full financial year
v Anticipated data storage required to meet the data retention obligations, as at 13 April 2017.

[12] Based on DRIGP Programme Guidelines, Jan 2016, Sections 60-62; Email from AGD Data Retention Implementation Team, 10th March 2016: *"The $128.4m will not be divided into separate proportionate pools of money to determine the allocation of the Enterprise Scale and Typical Implementation Impact. The grant amount the applicants receive will be based on the ordering of the total points score."*

[13] DRIGP Programme Guidelines, Jan 2016, Section 18

[14] This includes natting capability, for which we developed a separate dummy variable

2 Our regression-based approach

2.1 Applications dataset

We were provided with a dataset from the Department comprising the following fields collected from Provider applications:

1. **Provider information** (*Application Number, DRIP Reference Number*);
2. **Service provision** (*Number of Eligible Services, List of Eligible Service Types, NAT Employed*);
3. **Financial Information** (*DRIP agreed, Date DRIP Submitted to CAC, DRIP not yet agreed, Statement of Work Submitted, Financial Year Start, Sales Revenue Turnover*);
4. **Enterprise Score Scale**;
5. **Number of Subscribers**;
6. **Anticipated Data Storage Volume**;
7. **Total Estimated Project Cost**

PwC was not supplied with the Provider names or any other identifying information as part of this dataset. This was to ensure that the modelling process and the consequent allocation of funds were done without reference to, or knowledge of, the identity of individual providers.

These data were used to develop a dataset for analysis containing those variables listed in section 1.2.2. Altogether, the dataset comprised information from 180 Eligible Applicants.

2.2 Best fit regression model form

The first task was to establish the extent to which each independent variable was correlated to the cost of compliance for the Eligible Applicants. The independent variables to be tested are explicitly listed in the Guidelines:[15]

> '[The] Typical Implementation Impact … will be derived from calculating the typical cost of achieving compliance based on the analysis of information provided by all Eligible Applicants on the cost of compliance and subsequent weighting of the following variables:
>
> i Number of Eligible Service(s);
> ii Type(s) of Eligible Service(s);
> iii Number of subscribers;
> iv Gross annual revenue (turnover) for the most recent full financial year;
> v Anticipated data storage required to meet the data retention obligations, as at 13 April 2017.'

A technique called multivariate regression was applied across the whole final dataset of Eligible Applicants to make this determination. A variety of different linear and non-linear models were tested to ensure that the best fitting relationship was recommended for use. In addition to identifying the form of the best fit regression model, the number of independent variables to be included was also tested. To this end, dummy variables[16] were developed to

[15] DRIGP Guidelines Section 61b

[16] A dummy variable consists of a binary value (1 or 0) against each Provider depending on whether a particular characteristic is present or not. For example, the natting dummy variable we developed assigns a value 1 to those providers with natting capability

isolate the types of eligible services provided, and the statistical significance of these dummy variables tested as part of the regression model.

The best statistical fit was found to be a mixture between a LOG-LOG and LIN-LOG model; where the implementation cost is expressed in logarithmic form, and the independent variables are expressed either in linear form or in logarithmic form, depending on their statistical fit. The interpretation of this type of model is that there exists a relationship between a unit or percentage change in the independent variables and a percentage change in implementation costs. As an example of a LIN-LOG relationship in our best fit model, an increase of 100,000 subscribers is estimated to result in an increase of 3 per cent in implementation costs for the average eligible provider. Conversely, as an example of a LOG-LOG relationship, a 10 per cent increase in revenue is estimated by our model to result in a 1.6 per cent increase in costs. The technical justifications for selecting this model form are described more fully in Appendix C.

The analysis found that, as independent variables to predict implementation costs, there was statistical significance at over the 99 per cent level for:

- the number of services,
- the number of subscribers,
- gross annual revenue and
- the anticipated data storage required.

As the statistical analysis in Appendix C demonstrates, there was consistently low overall statistical significance for any given cohort of dummy variables developed to isolate the types of services provided, or for natting capability. The implication of this finding is that - although the number of services provided is a significant factor in determining the implementation cost - there is little evidence from the final applicant dataset of significant cost variation in response to different service types. Therefore, the best fit model does not include any service-type dummy variables or the natting capability dummy variable.

Equation 1 and Table 1 summarise the best-fit regression model used to estimate average implementation costs.

Equation 1: Best fit regression model

$$\ln(Implementation\ cost) = \beta_1 + \beta_2(\#services) + \beta_3(\#subscribers) + \beta_4 \ln(Revenue) + \beta_5 \ln(Storage)$$

R^2 = 54.1%, F-test significance = 100%

Table 1: Best fit Regression model statistics

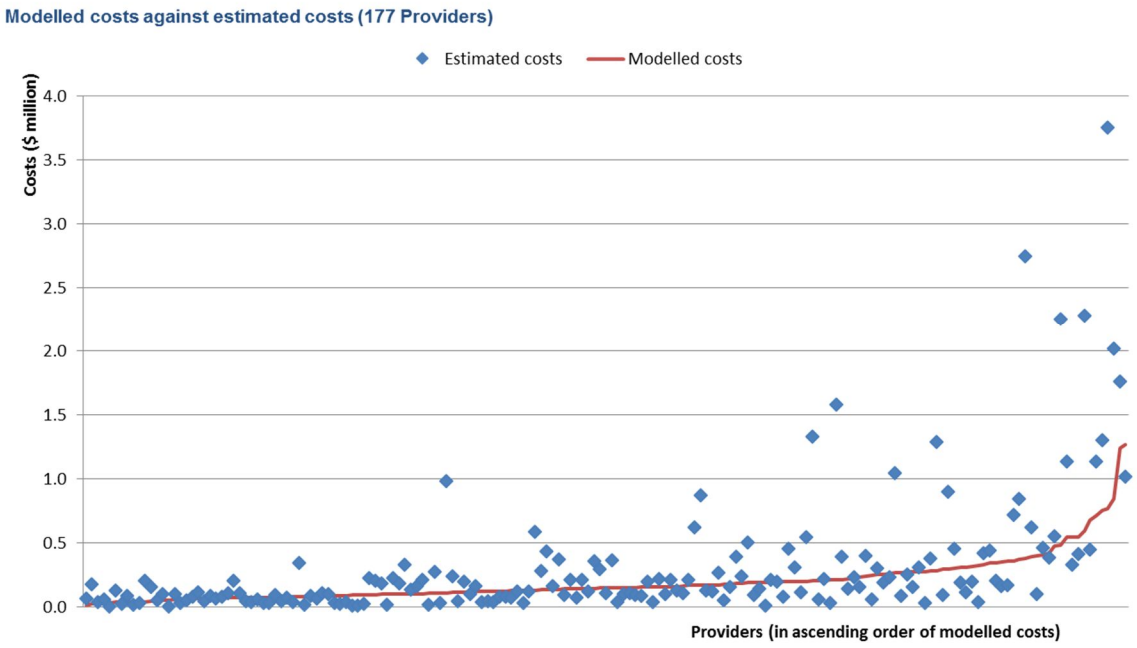
Variable	Co-efficient (β_i)	t-statistic	Significance level
Intercept (Constant)	8.20	19.19	>99%
# of services	0.04	3.73	>99%
# of subscribers)	2.75×10^{-7}	3.40	>99%
Ln (Revenue)	0.14	5.22	>99%
Ln (Storage)	0.16	4.87	>99%

and a 0 to those without. If a dummy variable is statistically significant, it implies that the factor being tested by the variable forms part of the model to explain the dependent variable – in this case, the implementation cost.

2.3 Modelled Implementation Costs

Figure 1 below presents a plot of the estimated and modelled implementation costs for each Eligible Applicant. In the interests of graphical clarity and for the protection of potentially commercial-in-confidence material, the three providers with the highest estimated costs have been removed from Figure 1. The significant curvature at the top end of the graph means that with these three providers included, the variation at the lower end of the scale cannot be as readily observed. Although the top three providers have been redacted in Figure 1, the same model and methodology has been applied consistently across all Eligible Applicants to calculate the final grant allocations, in full adherence to the DRIGP guidelines.

Figure 1: Modelled costs from regression analysis against estimated costs



The red curve is produced by plotting the modelled implementation cost for each eligible provider in ascending order from left to right across the chart (according to the y-axis scale). The blue diamonds on the chart represent each eligible provider's estimated implementation cost as per the dataset. If one traces a vertical line on the chart to the red line from any blue diamond, one establishes the difference between the estimated cost and modelled cost for an individual provider, again using the y-axis scale. This difference could be positive or negative depending on whether the provider has estimated costs greater than or less than our model predicts given the magnitude of their independent variables discussed above. Since the providers are in ascending order of modelled costs from left to right across the chart, the blue dots situated towards the left hand side of the graph represent the estimated costs of providers with very low modelled costs. The blue diamonds situated towards the right hand side of the graph represent the estimated costs of those providers with relatively high modelled costs, and as would be expected at this end of the distribution, we see a much larger variance in estimated costs compared with the modelled costs (illustrated by looser clustering around the modelled costs line).

2.4 Determination of points

As discussed above, the best fit model identified four (non-dummy and non-constant) independent variables whose impact on implementation cost was statistically significant. The DRIGP guidelines[17] call for 75 points to be allocated to each provider to reflect the Typical Implementation Impact, determined by weighting the independent variables listed in section 2.2 above.

A simple approach to allocating the 75 points for the Typical Implementation Impact would be to split them equally across these four variables, allowing a maximum of 18.75 points (75 ÷ 4) to be allocated to any given eligible provider for each of the four independent variables included in the best fit model. This would imply an equal weighting across all of the independent variables. Underpinning this approach is the observation that the statistical significance level of each variable indicates the likelihood of a relationship with estimated cost, and hence the weighting that should be applied. As shown in Table 1 above, the variables have very similar and high significance levels.

An alternative approach is to calculate the relative contribution of a change in each independent variable to the variation of the implementation cost. In order to determine these relative contributions, each co-efficient presented in Table 1 above must be normalised to account for the large differences in scale across the suite of independent variables. This calculation is shown in Equation 2 below:

Equation 2: Normalising the best fit regression co-efficients

$$Normalised\ \beta_i = \beta_i * \frac{S.D.\ Independent\ variable_i}{S.D.\ Estimated\ cost}$$

S.D. = Standard deviation

Table 2: Best fit regression model standardised co-efficients

Variable $_i$	Co-efficient (β_i)	S.D (variable $_i$)	S.D (estimated cost)	Normalised Co-efficient	Weighting
# of services	0.04	7.65	1.40	0.22	21.9%
# of subscribers	2.75×10^{-7}	1,052,218	1.40	0.21	20.7%
Ln (Revenue)	0.14	2.95	1.40	0.30	29.7%
Ln (Storage)	0.16	2.45	1.40	0.28	27.7%

The normalised co-efficients indicate the relative sensitivity of implementation cost to changes in each of the independent variables. The relatively higher normalised co-efficients in Table 2 for revenue and storage indicate that cost is more sensitive to percentage changes in these variables than to changes in the number of subscribers and services. These co-efficients can be used to weight the 75 points as shown in Table 2 above. Table 3 summarises the two approaches to the allocation of points outlined above, and provides the range from zero of each of the independent variables.

[17] DRIGP Guidelines, Jan 2016, Section 61b

Table 3: Allocation of points across independent variables

Variable	Variable range (zero to maximum)	Equal weighting		Weighting by co-efficients	
		Maximum points	Range per point	Maximum points	Range per point
# of services	45.00	18.75	2.40	16.45	2.81
# of subscribers	16.16	18.75	557,697	15.50	653,551
Ln (Revenue)	10,456,810	18.75	1.28	22.29	1.09
Ln (Storage)	15.91	18.75	0.85	20.76	0.76
	Total	75		75	

The ranges per point for each independent variable were calculated by dividing the variable range from zero to the maximum by the maximum point totals. For example, the maximum declared number of services across all Eligible Applicants was 45. In the equal weighting case, a linear scale for 18.75 points was calculated from zero to 45 whereby an additional 2.4 services (45 ÷ 18.75) would yield an additional point. Hence the provider declaring the maximum 45 services would receive the full 18.75 points, but a provider declaring 24 services would receive 24 ÷ 2.4 = 10 points.

The same approach would be used where the maximum points for each independent variable were determined using the standardised co-efficients derived in Table 2 above. In this case, for number of services, a linear scale for 16.45 points was calculated from zero to 45 whereby an additional 2.81 services (45 ÷ 16.45) would yield an additional point. Hence the provider declaring the maximum 45 services would receive the full 16.45 points, but a provider declaring 24 services would receive 24 ÷ 2.81 = 8.54 points.

As stated in the Guidelines, the maximum number of points for Typical Implementation Impact that could potentially be allocated to any single eligible provider is 75.

Each Eligible Applicant would therefore be allocated points using the thresholds above for each variable according to the methodology selected. Our recommendation to the Department is to use the 'weighting by co-efficients' approach to determining the maximum points thresholds for the Typical Implementation Impact as we feel that of the two methodologies presented, this alternative best reflects the intent of Section 61b of the DRIGP guidelines. However, we do present an allocation based on the equal weighting approach as an alternative in Appendix F (Table 8).

The total score out of 100 is derived by adding the Enterprise Scale Score and the Typical Implementation Impact Score.[18]

[18] Ibid. Section 62

3 Allocation of Funds

The total of the scores allocated across all Eligible Applicants was 7,006, an average (mean) score per provider of 38.92 out of 100. If the total available funding grant were divided linearly and proportionately across these scores, each point would be worth $18,321. This produces the unconstrained allocation described in Appendix D, which carries a number of shortcomings with respect to the Guidelines:

- A large number of providers (88 per cent) would be reimbursed with grant amounts higher than their estimated costs. 74 per cent of providers would be reimbursed with amounts over double their estimated costs. The Guidelines note that that funding allocations will not reimburse full costs but are a financial contribution to the typical up-front costs of compliance.[19]

- The linear allocation of funds to points does not take into account that the implementation cost for larger providers may be materially different from smaller providers due to, for example, system complexity. Since the model takes the average cost across all providers, this will tend to disadvantage those businesses at the top end of the distribution. Indeed, the top 3 providers are reimbursed less than 10 per cent of their estimated costs under this distribution. This contention is supported by the graph in Figure 1, which shows rapidly increasing costs at the top end of the distribution and a much wider estimated cost variation.

For these reasons, this unconstrained allocation is not recommended by PwC for use by the Department. The sections below indicate how the above issues were addressed within the provisions of the Guidelines in order to reach a distribution that is consistent with all of the aims of the DRIGP.[20]

3.1 Minimum and maximum allocation limits

The Guidelines allow for and anticipate the imposition of a maximum and minimum grant size.[21] As seen above, one of the key drawbacks of the unconstrained allocation is that a large proportion of providers would receive considerably more than both their estimated cost and the implementation cost modelled using the regression analysis.

According to the Guidelines,[22] the intention of the grant is to be a contribution to the costs of compliance, not a full reimbursement. This maximum threshold is determined and recommended to the Minister at the discretion of the Programme Delegate. In light of the Programme objectives, the Programme Delegate determined that allocating funds based on each Provider's modelled cost, but applying a maximum threshold of 80 per cent of each Eligible Applicant's estimated cost allows the grant to remain contributory, whilst allowing variation between providers according to the regression model.

Once the initial allocation has been constrained to an upper limit of 80 per cent of estimated cost, there are Eligible Applicants at the bottom end of the distribution whose allocations given this maximum threshold would be very low in absolute terms. Therefore, in accordance

[19] Ibid. Section 65

[20] Ibid. Sections 16-18; particularly allowing for the fact that the typical cost of compliance may be higher for larger, more established Providers.

[21] Ibid. Section 65

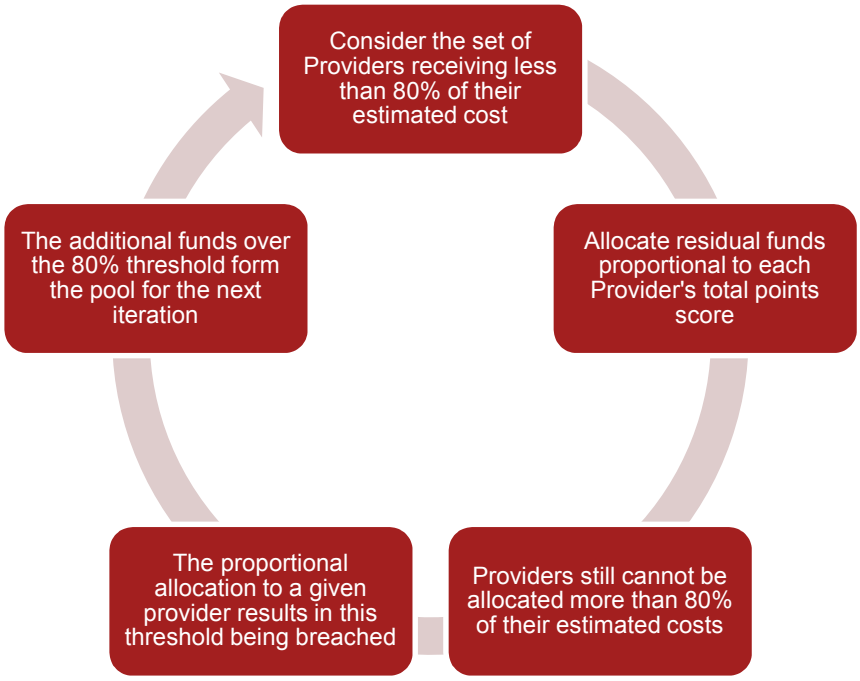
[22] Ibid.

with the Guidelines,[23] the Programme Delegate determined that a minimum grant size of $10,000 should be set to ensure that the very smallest providers are not financially penalised for operating on a small scale. This minimum threshold affects only the 3 Eligible Applicants at the bottom of the distribution and is designed to be consistent with the objective of the DRIGP to address smaller providers' greater expected need for financial assistance to comply with their data retention obligations. As will be clear from the allocations presented in Appendices E-F, these 3 Providers at the bottom of the distribution are overcompensated relative to the maximum threshold of 80 per cent of Estimated Costs.

3.2 Allocating the remaining funds

Constraining the allocation in the manner described in section 3.1 leaves around $90 million of the Programme Grant unallocated. A process for allocating these remaining funds was developed, based on each eligible provider's total score. This process is described in Figure 2 below:

Figure 2: Iterative allocation process



The iterative process continues until the entire Programme grant has been allocated. This produces our recommended allocation, which is provided in detail in Appendix E of this report. A variation on the above approach allocates the residual funds proportional to each provider's Typical Implementation Impact score (out of 75) rather than the total score (out of 100). The main difference is that in the latter approach, the minimum reimbursement rate is 30 per cent, compared with 47 per cent for our recommended approach. This alternative allocation is provided in Appendix F, and its high level characteristics are shown in Table 4.

As Table 4 shows, our recommended allocation (shaded) considers both the equity and transparency objectives contained within the Guidelines, in addition to the narrow focus of a methodology consistent with points and modelled cost. We consider the constraints applied

[23] Ibid. Sections 24c, 60

to the allocation of funds as the minimum necessary to ensure that Providers were not compensated with amounts significantly greater or less than their estimated costs.

To illustrate: in the recommended allocation, only 2 Eligible Applicants were compensated at less than the constrained maximum of 80 per cent of estimated costs. The minimum reimbursement rate was 47 per cent, and this was in the case of an applicant whose estimated costs far exceeded those modelled based on their independent variable characteristics. By contrast, in the unconstrained allocation, over 89 per cent of providers would be allocated more than 80 per cent of their estimated costs, and 4 providers allocated less than 20 per cent.

Table 4: Comparison of funding allocation alternatives

Objective	Initial unconstrained allocation	Constrained: Reallocation based on total points scored	Constrained: Reallocation based on Implementation Impact points scored
Allocation of $ based on points	✓	✓	✓
Allocation methodology incorporates Enterprise Scale	✓	✓	✓
No provider allocated more than 80% of estimated cost[24]	✗	✓	✓
No provider allocated less than 47% of estimated cost	✗	✓	✗
No provider allocated less than $10,000	✗	✓	✓
Allocation of $ reflects regression model	✓	✓	✓
Allocation uses the entire Programme Grant	✓	✓	✓
Allocation weighted in favour of total points	✗	✓	✗

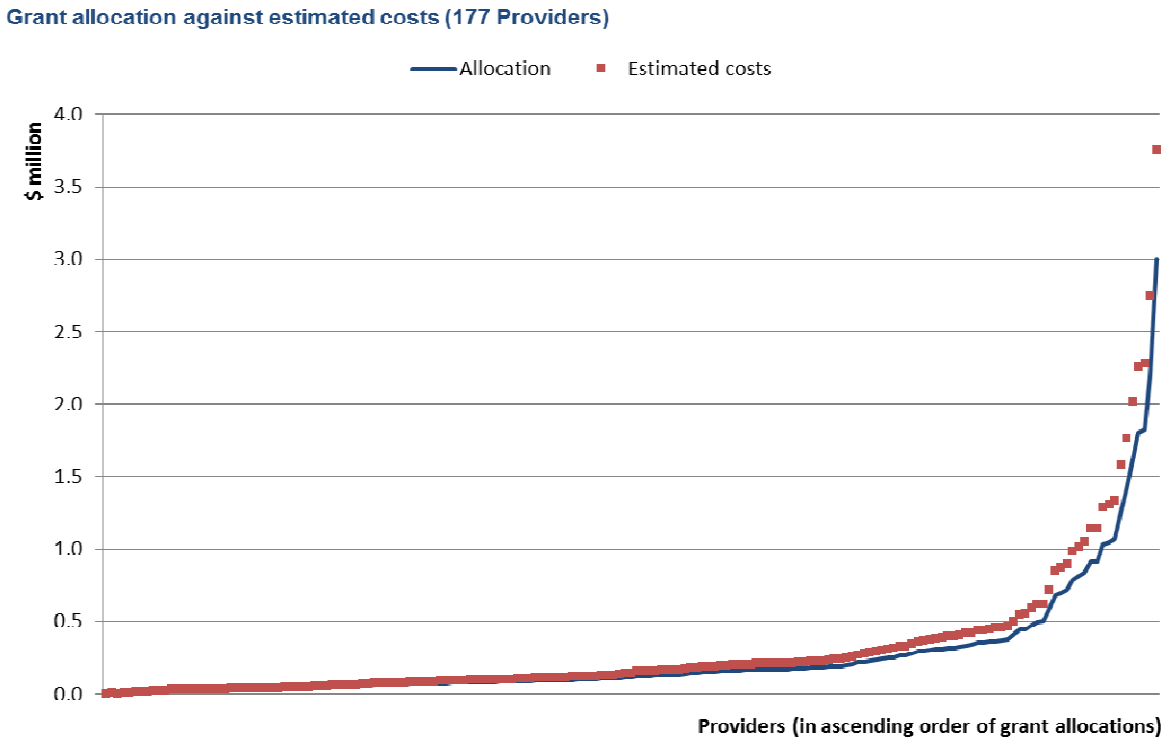
Figure 3 illustrates the reimbursement rates for our recommended allocation. The red diamonds plot each Eligible Applicant's estimated cost in ascending order from left to right. The blue line plots the grant allocations to each provider. As with Figure 1, if a vertical line on the chart is traced to the blue line from any red diamond, the difference between the estimated cost and grant allocation for an individual provided can be established using the y-axis scale. Since in our recommended allocation the vast majority of Eligible Applicants are allocated 80 per cent of their estimated costs, the blue line only appears to diverge from the

[24] Where this estimated cost is greater than $10,000

red diamonds at the top end of the distribution, where the gap in absolute dollar terms will be higher.

As with Figure 1, in the interests of graphical clarity, and the protection of commercial-in-confidence material, the three providers with the highest estimated costs have been removed from Figure 3.

Figure 3: Comparison of grant allocations with estimated costs



Appendix A
List of service types

Internet Access

DSL/ADSL Broadband-Operate
NBN Broadband-Operate
Satellite-Operate
Dial-up-Operate
Cable-Operate
Mobile Data (3G/4G/LTE)-Operate
Wifi-Operate
Transit-Operate
Pairing-Operate
Other Internet Access-Operate
Hybrid Fibre Coaxial (HFC)-Operate

Data Link

Hybrid Fibre Coaxial (HFC)-Operate
Dark Fibre-Operate
Ethernet-Operate
Virtual Private Network (VPN)-Operate
Multiprotocol Label Switching (MPLS)-Operate
Mobile (3G/4G/LTE)-Operate
Wireless-Operate
Other Radio Frequency (RF)-Operate
Wide Area Network (WAN)-Operate
Radio-Operate
Layer 2 or Layer 3 Service-Operate
Other Data Link-Operate

Managed Service

Cloud-based services-Operate
Storage-Operate
Other hosted services-Operate
Web Hosting-Operate
Telephony-Operate

Telephony

Fixed Line/PSTN (Public Switched Telephone Network)-Operate
Integrated Services Digital Network (ISDN)-Operate
Multiline-Operate
Fax-Operate
Pager-Operate
Mobile-Operate
In-bound services (13, 1300 etc)-Operate
Satellite Telephony-Operate
Other Telephony-Operate

Messaging

Email Hosting-Resell
Email Relay and Transmission (SMTP, POP, IMAP)-Resell
Short Messaging Service (SMS)-Resell
Internet Chat and Messaging-Resell
Other Messaging-Resell

IP-based Communications

Voice over Internet Protocol (VoIP)-Resell
Video Conferencing-Resell
Other Over-the-top (OTT) Services-Resell

Appendix B
Enterprise scale

Revenue ($ per annum)	Points awarded
0 to 119,999	25
120,000 to 239,999	24
240,000 to 359,999	23
360,000 to 479,999	22
480,000 to 599,999	21
600,000 to 719,999	20
720,000 to 839,999	19
840,000 to 959,999	18
960,000 to 1,079,999	17
1,080,000 to 1,199,999	16
1,200,000 to 1,319,999	15
1,320,000 to 1,439,999	14
1,440,000 to 1,559,999	13
1,560,000 to 1,679,999	12
1,680,000 to 1,799,999	11
1,800,000 to 1,919,999	10
1,920,000 to 2,039,999	9
2,040,000 to 2,159,999	8
2,160,000 to 2,279,999	7
2,280,000 to 2,399,999	6
2,400,000 to 2,519,999	5
2,520,000 to 2,639,999	4
2,640,000 to 2,759,999	3
2,760,000 to 2,879,999	2
2,880,000 to 2,999,999	1
More than 3,000,0000	0

Appendix C
Selecting the model form

A variety of different linear and non-linear (logarithmic) models were tested to ensure that the best fitting relationship was recommended for use. In addition to identifying the form of the best fit regression model, the number of independent variables to be included was also tested. To this end, dummy variables[25] were developed to isolate the types of eligible services provided, and the statistical significance of these dummy variables tested as part of the regression model. This Appendix provides further statistical information about the model that was recommended, and some of the alternatives that were tested.

Recommended model

Input Variables	Co-efficient	Std. Error	t-Statistic	P-Value
Intercept	8.197	0.427	19.19	$6.375 * 10^{-45}$
# Services	0.040	0.012	3.73	$2.579 * 10^{-4}$
# Subscribers	$2.75 * 10^{-7}$	$8.074 * 10^{-8}$	3.40	$8.218 * 10^{-4}$
Log(Revenue)	0.141	0.027	5.22	$5.015 * 10^{-7}$
Log(Storage Volume)	0.158	0.032	4.87	$2.442 * 10^{-6}$

Output variable	Log (Cost)
R^2	0.54
Adjusted R^2	0.53

Our recommended model is a combination of linear and non-linear variables. The output variable (implementation cost), revenue and storage volume are expressed in natural logarithms, reflecting a non-linear relationship between them. The t-statistic tests the probability that the co-efficient were in fact to be zero, given the correlation calculated. The P-value expresses this probability as a number between zero and one. Note that the P-values for each of these independent variables are very close to zero, indicating very strong statistical significance. The R^2 values indicate that over half of the variation in estimated costs is attributable to the four variables in our recommended model. The DRIGP Guidelines allow for the inclusion of variables to test the relationship between service type and cost as part of the model, and we have tested these alternative models explicitly below.

We acknowledge that there may be additional drivers of cost that are not covered by the DRIGP Guidelines and therefore were not considered in our analysis. It is possible that the inclusion of these variables in this regression could yield a higher R^2 value (and therefore a better statistical fit), however it is not clear that the impact on the final grant allocation would be anything other than negligible.

[25] A dummy variable consists of a binary value (1 or 0) against each Provider depending on whether a particular characteristic is present or not. For example, the natting dummy variable we developed assigns a value 1 to those providers with natting capability and a 0 to those without. If a dummy variable is statistically significant, it implies that the factor being tested by the variable forms part of the model to explain the dependent variable – in this case, the implementation cost.

Alternative 1 – LIN-LIN Model

This model assumes a linear relationship between cost and the independent variables. None of the variables are transformed prior to the regression being estimated.

Input Variables	Co-efficient	Std. Error	t-Statistic	P-Value
Intercept	728,129	366388	1.99	0.048
# Services	-71,961	34391	-2.09	0.039
# Subscribers	4.05	1.09	3.72	0.00027
Revenue	0.000243	0.000279	0.87	0.39
Storage Volume	1.52	0.84	1.82	0.07

Output variable	Cost
R^2	0.78
Adjusted R^2	0.77

Although this model appears to explain a higher percentage of the cost variation, indicated by the higher R^2 value, there are a number of reasons to prefer our recommended model presented above. First, the negative sign on the co-efficient on number of services implies that each additional service offered reduces overall cost. This is an example of spurious correlation, whereby a relationship is found that cannot be defended logically. Another indication that this model is less robust than our recommended model above is the very high P-value on the revenue variable, which would indicate a 39 per cent chance that the revenue co-efficient is, in fact, equal to zero; that is, that revenue is not correlated with cost at all. Generally, the P-values are higher in this model than for our recommended model, and on balance therefore, we rejected this alternative.

Alternative 2 – LIN-LOG Model

This model assumes that a linear change in the independent variables leads to a fixed percentage change in costs. The results are presented below.

Input Variables	Co-efficient	Std. Error	t-Statistic	P-Value
Intercept	11.39	0.128	89.3	$6.16 * 10^{-148}$
# Services	0.064	0.012	5.22	$5.06 * 10^{-07}$
# Subscribers	$3.51 * 10^{-07}$	$3.79 * 10^{-07}$	0.93	0.36
Revenue	$-6.07 * 10^{-11}$	$9.72 * 10^{-11}$	-0.62	0.53
Storage Volume	$3.42 * 10^{-07}$	$2.91 * 10^{-07}$	1.178	0.24

Output variable	Log(Cost)
R^2	0.39
Adjusted R^2	0.37

This model exhibits high P-values for number of subscribers, revenue and storage volume. It also explains only around 40% of the cost variation.

Alternative 3 – LOG-LIN Model

This model assumes that a percentage change in the independent variables lead to a linear change in costs. The results are presented below:

Input Variables	Co-efficient	Std. Error	t-Statistic	P-Value
Intercept	-12,501,721	2,505,721	-4.99	$1.45 * 10^{-06}$
Log(# Services)	315,385	475,242	0.66	0.51
Log(# Subscribers)	367,890	184,757	1.99	0.05
Log(Revenue)	363,298	190,847	1.90	0.06
Log(Storage Volume)	611,157	200,705	3.05	0.003

Output variable	Cost
R^2	0.22
Adjusted R^2	0.20

This model has an unacceptably high P-value for the Log(#services) variable. It is the model with the least explanatory power, with an R^2 value of only around 20 per cent.

Alternative 4 – LOG-LOG Model

This model assumes that a percentage change in the independent variables lead to a percentage change in costs. In this sort of model, the co-efficients are interpreted as 'elasticities'. The results are presented below:

Input Variables	Co-efficient	Std. Error	t-Statistic	P-Value
Intercept	7.34	0.42	17.40	$5.32 * 10^{-40}$
Log(# Services)	0.30	0.08	3.72	0.000268
Log(# Subscribers)	0.05	0.03	1.73	0.086
Log(Revenue)	0.15	0.03	4.80	$3.318 * 10^{-06}$
Log(Storage Volume)	0.18	0.033	5.28	$3.77 * 10^{-07}$

Output variable	Log(Cost)
R^2	0.50
Adjusted R^2	0.49

This model is a good statistical fit with all co-efficients in the logical (positive) direction, low P-values, and reasonable explanatory power. Our recommended model is a variant of this model form, the only difference being that we use a linear services and subscriber number variable instead of expressing them in Log form. This improves the overall explanatory power and statistical fit.

Alternative 5 – Model including Natting and High Level Service Type

This alternative uses the recommended model form to examine whether the explanatory power can be improved by adding 'service type' dummy variables. The results are presented below:

Input Variables	Co-efficient	Std. Error	t-Statistic	P-Value
Intercept	8.33	0.48	17.37	$2.35 * 10^{-39}$
# Services	0.07	0.02	3.29	0.00121
# Subscribers	$2.24 * 10^{-07}$	$8.61 * 10^{-08}$	2.61	0.01
Log(Revenue)	0.12	0.028	4.45	$1.58 * 10^{-05}$
Log(Storage Volume)	0.16	0.033	4.77	$4.04 * 10^{-06}$
NATTING	-0.16	0.150	-1.09	0.28
Internet Access	0.41	0.216	1.89	0.06
Datalink	0.11	0.174	0.65	0.51
Managed Service	-0.50	0.182	-2.77	0.0063
Telephony	-0.21	0.183	-1.17	0.24
Messaging	0.05	0.172	0.30	0.77
IP Communications	-0.17	0.184	-0.89	0.38

Output variable	Log(Cost)
R^2	0.58
Adjusted R^2	0.55

Looking at the P-Values of the new variables, the only one of significance to cost is 'Internet Access'; and even here there is a 6 per cent chance of the co-efficient being zero. The R^2 values are not significantly different to those of our recommended model. Based on this evidence, we therefore reject the inclusion of a 'service-type' variable as part of the regression model.

Alternative 6 – Model including Natting and All Service Types

This alternative uses the recommended model form to examine whether the explanatory power can be improved by adding dummy variables to encompass the full suite of eligible services as enumerated in Appendix A. The results are presented below:

Input Variables	Co-efficient	Std. Error	t-Statistic	P-Value
Intercept	8.865	0.587	15.103	0
# Services	9.201	5.207	1.767	0.081
# Subscribers	0	0	0.929	0.355
Log(Revenue)	0.085	0.037	2.29	0.024
Log(Storage Volume)	0.178	0.045	3.984	0
NATTING	-0.203	0.195	-1.038	0.302
DSL/ADSL Broadband-Operate	-9.202	5.203	-1.769	0.08
NBN Broadband-Operate	-9.119	5.155	-1.769	0.08

Input Variables	Co-efficient	Std. Error	t-Statistic	P-Value
Satellite-Operate	-9.35	5.523	-1.693	0.094
Dial-up-Operate	-8.901	5.092	-1.748	0.084
Cable-Operate	-8.698	5.211	-1.669	0.099
Mobile Data (3G/4G/LTE)-Operate	-7.904	5.287	-1.495	0.138
Wifi-Operate	-8.932	5.213	-1.713	0.09
Transit-Operate	-8.872	5.246	-1.691	0.094
Pairing-Operate	-9.88	5.247	-1.883	0.063
Other Internet Access-Operate	-8.977	5.238	-1.714	0.09
Hybrid Fibre Coaxial (HFC)-Operate	-9.718	5.406	-1.798	0.076
Dark Fibre-Operate	-9.027	5.221	-1.729	0.087
Ethernet-Operate	-8.624	5.206	-1.657	0.101
Virtual Private Network (VPN)-Operate	-9.199	5.268	-1.746	0.084
Multiprotocol Label Switching (MPLS)-Operate	-9.179	5.156	-1.78	0.078
Mobile (3G/4G/LTE)-Operate	-10.825	5.163	-2.097	0.039
Wireless-Operate	-9.2	5.202	-1.768	0.08
Other Radio Frequency (RF)-Operate	-8.741	5.13	-1.704	0.092
Wide Area Network (WAN)-Operate	-9.102	5.221	-1.743	0.085
Radio-Operate	-10.09	5.208	-1.937	0.056
Layer 2 or Layer 3 Service-Operate	-8.969	5.236	-1.713	0.09
Other Data Link-Operate	-8.483	5.027	-1.688	0.095
Cloud-base services-Operate	-9.675	5.246	-1.844	0.068
Storage-Operate	-8.745	5.263	-1.662	0.1
Other host eservices-Operate	-9.415	5.147	-1.829	0.071
Web Hosting-Operate	-9.657	5.244	-1.842	0.069
Telephony-Operate	-9.186	5.25	-1.75	0.084
Fixed Line/PSTN (Public Switch eTelephone Network)-Operate	-9.209	5.18	-1.778	0.079
Integrated Services Digital Network (ISDN)-Operate	-8.815	5.458	-1.615	0.11
Multiline-Operate	-8.914	5.367	-1.661	0.1
Fax-Operate	-9.036	5.113	-1.767	0.081
Pager-Operate	0	0	NA	NA
Mobile-Operate	-18.93	11.647	-1.625	0.108

Input Variables	Co-efficient	Std. Error	t-Statistic	P-Value
In-bound services (13, 1300 etc)-Operate	-9.731	5.197	-1.872	0.064
Satellite telephony-Operate	-10.35	4.33	-2.39	0.019
Other Telephony-Operate	-9.142	5.301	-1.725	0.088
Email Hosting-Operate	-9.573	5.205	-1.839	0.069
Email Relay and Transmission (SMTP, POP, IMAP)-Operate	-8.916	5.172	-1.724	0.088
Short Messaging Service (SMS)-Operate	-7.091	5.392	-1.315	0.192
Internet Chat and Messaging-Operate	-8.524	5.162	-1.651	0.102
Other Messaging-Operate	-8.987	5.364	-1.676	0.097
Voice over Internet Protocol (VoIP)-Operate	-9.26	5.248	-1.764	0.081
Video Conferencing-Operate	-8.844	5.23	-1.691	0.094
Other Over-the-top (OTT) Services-Operate	-9.631	5.244	-1.837	0.07
DSL/ADSL Broadband-Resell	-9.024	5.2	-1.735	0.086
NBN Broadband-Resell	-9.139	5.15	-1.774	0.079
Satellite-Resell	-8.841	5.033	-1.756	0.082
Dial-up-Resell	-9.151	5.186	-1.765	0.081
Cable-Resell	-8.828	5.156	-1.712	0.09
Mobile Data (3G/4G/LTE)-Resell	-9.087	5.384	-1.688	0.095
Wifi-Resell	-8.493	5.29	-1.606	0.112
Transit-Resell	-8.604	5.187	-1.659	0.101
Other Internet Access-Resell	-9.202	5.167	-1.781	0.078
Dark Fibre-Resell	-9.839	5.429	-1.812	0.073
Ethernet-Resell	-9.27	5.177	-1.79	0.077
Virtual Private Network (VPN)-Resell	-9.412	4.968	-1.895	0.061
Multiprotocol Label Switching (MPLS)-Resell	-8.93	5.396	-1.655	0.101
Mobile (3G/4G/LTE)-Resell	-9.467	5.299	-1.786	0.077
Wireless-Resell	-9.07	5.129	-1.769	0.08
Other Radio Frequency (RF)-Resell	-9.922	5.724	-1.733	0.086
Wide Area Network (WAN)-Resell	-8.851	5.735	-1.543	0.126
Layer 2 or Layer 3 Service-Resell	-8.217	4.968	-1.654	0.102
Other Data Link-Resell	-8.683	5.283	-1.643	0.104

Input Variables	Co-efficient	Std. Error	t-Statistic	P-Value
Cloud-base services-Resell	-8.524	5.645	-1.51	0.134
Storage-Resell	-10.763	7.66	-1.405	0.163
Other host e services-Resell	-10.894	5.343	-2.039	0.044
Web Hosting-Resell	-11.303	5.651	-2	0.048
Telephony-Resell	-8.488	5.431	-1.563	0.122
Fixed Line/PSTN (Public Switch Telephone Network)-Resell	-8.945	5.345	-1.674	0.098
Integrate Services Digital Network (ISDN)-Resell	-9.224	5.209	-1.771	0.08
Multiline-Resell	-9.171	5.285	-1.735	0.086
Fax-Resell	-8.769	5.048	-1.737	0.086
Mobile-Resell	-9.301	5.235	-1.777	0.079
In-bound services (13, 1300 etc)-Resell	-9.199	5.201	-1.769	0.08
Satellite telephony-Resell	-8.937	4.858	-1.84	0.069
Other Telephony-Resell	-9.688	5.236	-1.85	0.067
Email Hosting-Resell	-8.95	5.058	-1.769	0.08
Email Relay and Transmission (SMTP, POP, IMAP)-Resell	-8.714	5.578	-1.562	0.122
Short Messaging Service (SMS)-Resell	-9.417	5.346	-1.762	0.082
Internet Chat and Messaging-Resell	-6.265	4.776	-1.312	0.193
Other Messaging-Resell	-8.944	5.273	-1.696	0.093
Voice over Internet Protocol (VoIP)-Resell	-9.46	5.13	-1.844	0.068
Video Conferencing-Resell	-8.658	5.25	-1.649	0.103
Other Over-the-top (OTT) Services-Resell	-8.404	5.465	-1.538	0.128

Output variable	Log(Cost)
R^2	0.79
Adjusted R^2	0.58

In addition to the large number of insignificant service type dummy variables, the difference between the R^2 and adjusted R^2 value will be noted. This indicates that any correlation found is likely to be spurious, and that the variation truly explained by this model is 58 per cent, a value not far removed from that of our recommended model above. For these reasons, we reject this alternative in favour of our recommended approach described at the start of this section.

www.pwc.com.au