Jemena Gas Networks (NSW) Ltd

2010-15 Access Arrangement Information

Appendix 7.6

UAG – Frontier Economics report

Public

30 June 2014

Page intentionally blank

UAG coefficients statistical methodology

A REPORT PREPARED FOR JEMENA

May 2014

Page intentionally blank

UAG coefficients statistical methodology

UAG coefficients statistical methodology

Figures

1 Background

The cost of replenishing gas that is lost, or unaccounted for, during delivery through the distribution network is a material component of Jemena Gas Network's (Jemena's) operating costs.

The AER's approach in the 2010 Access Arrangement (AA) for determining Jemena's allowed 'unaccounted for gas' (UAG) costs was the product of an 'efficient' forecast of UAG volumes (2.38% of forecast network receipts) and forecast unit rates to buy replenishment gas each year.

Jemena proposes to:

- justify the operating efficiency of UAG levels using the 5 year historical average of UAG levels for 2009 through 2013; and
- express the forecast rate of UAG as two different rates – one applied to withdrawals of daily metered customers and the other to non-daily metered customers. Note this change will also apply the rate of UAG to withdrawals rather than to receipts.

To assist in achieving the above goals, Jemena has engaged Frontier Economics (Frontier) to advise on:

- the strength of the statistical correlation between UAG and withdrawals for the two sets of customer markets;
- alternative valid statistical methodologies to express UAG in terms of withdrawals for the two markets and for determining the corresponding UAG rates/coefficients for the two markets from available data; and
- calculating those rates and the statistical relationship for the purposes of inclusion in Jemena's AA proposal.

This report presents the results of our investigations. We have estimated two regression models that fit the data well, using monthly and annual data, respectively, and we present the two sets of estimates for the required UAG coefficients corresponding to these two models. Both models have merit, but also some shortcomings due to the nature of the data. In particular, the results are somewhat sensitive to the sample period used in the estimation.

2 Data provided for analysis

2.1 Description of data series

Jemena has provided us with an Excel file that contains data relevant to the estimation of UAG rates. The data have been provided on a monthly basis.

Below we list the data series provided to us with a brief description of each series. All series have been provided for the period FY2002 to FY2013 and are measured in TJ.[1]

Receipts (R) = gas volumes received by Jemena from suppliers

Daily metered withdrawals (I) = gas volumes delivered to customers metered on a daily basis. These are mainly large industrial customers, and hence we indicate this quantity by I.

TMR (T) = the tariff market residual, calculated as the residual between receipts and daily metered withdrawals. This consists of the volumes delivered to the market segment of customers who are not metered on a daily basis, plus UAG.

UAG (U^e) = estimated annual unaccounted for gas. We have attached a superscript to this quantity because it is not based on direct measurement.

Billed volumes (B) = volumes billed each month to the market segment of customers who are not daily metered. This market segment consists mainly of residential and small business customers. We refer to this as the mass market.

We reserve the term U, without a superscript, for the true (unobserved) volume of unaccounted for gas.

We also use the term M for the actual (unobserved) withdrawals by the mass market.

Note that

$$T = R - I = M + U. \qquad (1)$$

U^e is a derived quantity obtained by using an estimate of M, say M^e, from billing data. By re-writing equation (1) we obtain:

---

1 FY2002 is the financial year ending on 30 June 2002, and similarly for the other financial years. For some series there are also a few months of data before and/or after this period. However, we restrict our analysis and comments to the data for the full financial years.

$$U^e = T - M^e = R - I - M^e. \quad (2)$$

2.2 Estimating UAG

UAG is not directly observed and needs to be estimated using equation (2). The important unknown quantity is the actual volume of withdrawals by the mass market (M). These volumes are metered at the customers' premises using accumulation meters, which are usually read on a three-monthly cycle for billing purposes. This means that the volumes are measured relatively infrequently and not across a common time period.

For the mass market, there is a discrepancy between the timing of meter reads and the time periods commonly used in data analysis, such as months, quarters or years. But this discrepancy becomes smaller the longer the time period used in the analysis. Thus, other things being equal, estimates of M based on annual data would be more accurate than estimates bases on quarterly data, and estimates based on monthly data would have the poorest accuracy.

Estimates based on five or ten years of data would be even more accurate than estimates based on annual data. However, the longer the period of data used to develop estimates of the UAG rate, the more likely it is that there has been a change in the underlying UAG rate due to changes in the configuration or state of the network.

At the same time, for the purpose of estimating statistical models, the number of observations used in the analysis takes on an important role – the fewer the observations that are used in a statistical estimation exercise, the less accurate the estimates of the coefficients in the model. As there are only 12 years of annual data available for analysis, and far fewer observations if we take averages across longer periods, statistical models estimated using such data are likely to be unreliable.

Therefore, selecting the appropriate periodicity for the data used to estimate UAG rates requires a trade-off to be made between minimising the discrepancy between meter reads and consumption within the selected calendar periods on the one hand, and the availability of a sufficient number of observations on the other hand.

The accuracy of the estimates is also affected by the methodology used to allocate billed volumes for the mass market to specific months, quarters or years. Utilities use a variety of procedures for this task, ranging from a pro-rata allocation of billed volumes to the months, quarters, or years based on the number of days of each billing period, to quite sophisticated adjustments that account for seasonal variations in consumption patterns. We have been advised that Jemena uses such an approach to estimate monthly withdrawals by the mass market segment consistent with physical flows. The next section discusses our understanding of Jemena's approach in more detail.

3 Jemena's proposed approach

3.1 Description of approach

The Excel file provided to us contains details of the data analysis undertaken by Jemena to obtain estimates of UAG rates, both in aggregate and separately for large industrial customers and the tariff market residual (TMR). There are two main parts to the analysis:

1. calculating the average historical UAG rate over the last five years
2. estimating the UAG rates that can be applied to daily metered withdrawals I and to the tariff market residual T (say α_I and α_T, respectively), so that UAG can split into the two components contributed by these two market segments; algebraically $U = \alpha_I I + \alpha_T T$.

Estimating the overall UAG rate

Jemena estimates the overall UAG rate using the last five years of annual data. The UAG rate in each year is calculated as $\alpha_{R_t} = \frac{U_t^e}{R_t}$ where t refers to the year, and U^e and R are defined in section 2.1. The estimated overall UAG rate is obtained by taking the average of these ratio values for the years 2009 to 2013 inclusive, i.e:

$$\bar{\alpha}_R = \frac{1}{5} \sum_{t=2009}^{2013} \alpha_{R_t} \qquad (3)$$

The value obtained is 2.239%.

In our view, this is a reasonable approach. As noted in section 2.2, estimates of UAG based on annual data are likely to be fairly accurate.

However, from a statistical point of view, a better approach would be to take the ratio between the total volumes of UAG and the total receipts over the five year period. This approach can be represented algebraically as follows:

$$\hat{\alpha}_R = \sum_{t=2009}^{2013} U_t^e \Big/ \sum_{t=2009}^{2013} R_t \qquad (4)$$

This calculation leads to a value of 2.237%. In this case, the difference between the two approaches is not material, but given other years of data the difference could be larger.

Estimating the UAG rates for the two market segments

In order to estimate the two rates for the two market segments, Jemena undertakes regression analysis using monthly data.

There is considerable merit in using monthly data. In order to obtain reliable separate estimates of the UAG rates for the large industrial and TMR segments using regression, one requires the relative size of the market segments to show reasonable variation in the estimation sample. Most of the variation in the relative size of the market segments is due to seasonal factors. The volume of TMR as a proportion of receipts using annual data ranges between 33% and 42% between 2002 and 2013. Using monthly data over the same period, there is a much wider range, from 24% to 53%. On the other hand, as indicated in section 2.2, the monthly estimates of UAG TJs are likely to be less precise than the annual estimates.

The form of the regression equation used is:

$$U_t^e = \alpha_I I_t + \alpha_T T_t + \varepsilon_t \quad (5)$$

where ε_t is the regression disturbance term, and the t subscript now refers to the month.

We have estimated this model with monthly data from July 2001 to June 2013 using least squares and the Stata statistical package. The estimation results are shown in Figure 1. The estimated coefficient for large industrials customers is $\hat{\alpha}_I = 0.356\%$ and for the TMR segment $\hat{\alpha}_T = 5.63\%$.[2] The fit of the model is very good, with an adjusted R^2 of 97.1%; the coefficient for TMR volumes is highly significant, and the coefficient for the industrial segment is almost significant at the 5% level.

Figure 1: Estimation results for Jemena's model specification

Source	SS	df	MS
Model	5127573.49	2	2563786.75
Residual	150576.36	130	1158.27969
Total	5278149.85	132	39985.9837

Number of obs	132
F(2, 130)	2213.44
Prob > F	0.0000
R-squared	0.9715
Adj R-squared	0.9710
Root MSE	34.034

uag	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
daily_tj	.0035557	.0018165	1.96	0.052	-.0000381	.0071495
tmr_tj_derived	.0562993	.0029994	18.77	0.000	.0503654	.0622332

Source: Frontier analysis

2 The symbol ^ signifies that these are estimates of the true parameters.

Note that the α_I and α_T are not UAG rates in the conventional sense, since they are defined as a proportion of withdrawals by the two market segments, not the receipts due to these withdrawals. However, since receipts by market segment are not observed, this is a more convenient representation. The corresponding conventional UAG rates can be derived as $\hat{\gamma}_I = \hat{\alpha}_I(1 - \hat{\alpha}_I) = 0.354\%$ and $\hat{\gamma}_T = \hat{\alpha}_T(1 - \hat{\alpha}_T) = 5.31\%$.

Consistency with overall five-year average UAG rate

The estimates obtained from the regression model are not guaranteed to be consistent with the five-year average UAG rate calculated from annual data according to equation (3). To obtain UAG rates that are consistent with this five-year average, Jemena undertakes the following calculations.

The first step is to take the UAG rate for the industrial market segment estimate estimated from the regression, i.e. $\hat{\alpha}_I = 0.356\%$ and apply it to metered withdrawals I_t in each of the last five years to obtain estimates of the UAG (TJ) losses in each year due for the industrial segment. The next step is to subtract this amount from the total UAG losses in TJ in each of the five years to obtain estimates of the UAG losses in TJ for the TMR market segment in each year. The third step is to divide the TMR market segment's UAG losses for each year by the total volume in TJ for the TMR segment for that year to obtain an estimate of the UAG rate α_{T_t} for the TMR segment for each year. The five yearly estimates are then averaged to obtain an estimate of α_T that is consistent with the five-year average overall UAG rate:

$$\bar{\alpha}_T = \frac{1}{5}\sum_{t=2009}^{2013} \alpha_{T_t} \qquad = 5.27\%$$

Note that this estimate is somewhat lower than the estimate obtained from the regression of 5.63%.

3.2 Review of the proposed approach

Endogeneity

From a theoretical econometric point of view, there is a serious issue with the model shown in equation (5). The TMR volume T is defined in equation (1) as:

$$T = M + U.$$

However, U is also the dependent variable in the regression model in equation (5). Under these circumstances, least squares estimation leads to statistically inconsistent estimates. This means that even in very large samples the estimated coefficients are biased.

However, in this case there is an easy solution to this problem. If we rewrite equation (5) as:

$$U_t^e = \alpha_I I_t + \alpha_T (M_t^e + U_t^e) + \varepsilon_t$$

This expression can be rewritten as:

$$U_t^e = \beta_I I_t + \beta_M M_t^e + \eta_t \qquad (6)$$

where $\beta_I = {}^{\alpha_I}/_{(1-\alpha_T)}$, $\beta_M = {}^{\alpha_T}/_{(1-\alpha_T)}$ and $\eta_t = {}^{1}/_{(1-\alpha_T)}\,\epsilon_t$.

Estimation of equation (6) will provide consistent estimates of β_I and β_M[3]. Consistent estimates of the original parameters, α_I and α_T, can be obtained using the equations $\alpha_I = {}^{\beta_I}/_{(1+\beta_M)}$ and $\alpha_T = {}^{\beta_M}/_{(1+\beta_M)}$. (7)

We have estimated the model in equation (6) and the results are shown in Figure 2.

Figure 2: Estimation results for equation (6) using monthly data

Source	SS	df	MS
Model	5109293.15	2	2554646.57
Residual	168856.706	130	1298.89774
Total	5278149.85	132	39985.9837

Number of obs	132
F(2, 130)	1966.78
Prob > F	0.0000
R-squared	0.9680
Adj R-squared	0.9675
Root MSE	36.04

uag	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
daily_tj	.004518	.0019111	2.36	0.020	.000737	.0082989
vol_tj	.0582683	.0033635	17.32	0.000	.051614	.0649225

Source: Frontier analysis

The fit of the model is also very good, with an adjusted R^2 of 96.7%[4]; the coefficient for the mass market is highly significant and the coefficient for the industrial sector is also very significant, with a p-value of 2%.

The coefficient for large industrials customers is $\beta_I = 0.452\%$ and for the mass market segment is $\hat{\beta}_M = 5.83\%$. The corresponding estimates for the original

---

3 We acknowledge that M_t^e may be measured with some error, and that this could also cause the estimates to be inconsistent. However, the impact of this is likely to be much smaller than the inconsistency caused by the endogeneity.

4 The adjusted R^2 value for this model is marginally worse than for the model in equation (5). This is to be expected, since in equation (5) the U term, implicitly included in the T term, contributes to explaining itself, namely, the U term on the left hand side of the regression model.

coefficients for large industrial customers and for the TMR segment are $\hat{\alpha}_I = 0.427\%$ and $\hat{\alpha}_T = 5.51\%$, respectively.

The corresponding conventional UAG rates[5] are $\hat{\gamma}_I = \hat{\alpha}_I(1 - \hat{\alpha}_I) = 0.425\%$ and $\hat{\gamma}_T = \hat{\alpha}_T(1 - \hat{\alpha}_T) = 5.20\%$.

Consistency with overall five-year average UAG rate

We have applied Jemena's method for ensuring that the estimated UAG rates are consistent with the five-year average overall UAG rate discussed above. Starting with the estimated value for the industrial segment UAG rate of $\hat{\alpha}_I = 0.427\%$, we have followed the steps in Jemena's calculations and obtained an estimate for the TMR market segment of:

$$\bar{\alpha}_T = \frac{1}{5} \sum_{t=2009}^{2013} \alpha_{T_t} \qquad = 5.16\%.$$

As a check on Jemena's procedure, we estimated a regression model that imposes a constraint on the estimated segment-level UAG rates consistent with the five-year average overall rate as part of the estimation procedure. The estimates we obtained using this approach were $\hat{\alpha}_I = 0.370\%$ and for the TMR segment $\hat{\alpha}_T = 5.260\%$.

These estimates are not very different from those obtained using Jemena's approach. Since Jemena's approach is much simpler to implement than the constrained regression estimation we have undertaken, we believe that Jemena's approach represents a reasonable compromise for regulatory purposes.

Stability of the results

To test the stability of the results, we have estimated both the monthly and the annual models using shorter time periods. The results of these investigations show that the coefficient estimates are fairly sensitive to the period chosen for the estimation. This is not an ideal situation. However, it does not appear that there are any other data sources that could be used to overcome this problem. Under the circumstances, we recommend that as long a period as possible be used to minimise the impact of this sensitivity. The results we report above are estimated on that basis.

We have also investigated the use of billed volumes for the mass market, with up to three lead terms, instead of using Jemena's imputed values for the monthly volumes for this market segment. The advantage of using billed volumes is that these are measured accurately. The disadvantage is that, temporally, they are not aligned well to the other variables in the model.

5 See the end of section 3.1 for a discussion on these different UAG rates.

Using this approach, the estimates of the UAG rate for large industrial customers are fairly similar to the corresponding monthly model estimates; but the coefficients for the billed volumes and its lead terms are harder to interpret. Consequently, we see no advantage in exploring this approach further. To some extent, these results provide validation of Jemena's procedure for imputing monthly flows for the mass market.

4 Conclusions

In this report we have examined the procedure developed by Jemena to obtain separate estimates for the UAG rates for daily metered customers and the tariff market residual segment of the market.

Jemena's proposal is to use:

- the equation $U_t = \alpha_I I_t + \alpha_T T_t$ with estimated coefficients of $\alpha_I = 0.427\%$ and $\alpha_T = 5.16\%$ when actual annual quantities of I_t and T_t are known; and
- the equivalent equation of $U_t = \beta_I I_t + \beta_M M_t$ with estimated coefficients of $\beta_I = 0.450\%$ and $\beta_M = 5.44\%$ when applying forecast annual values of I_t and M_t.[6]

In our opinion, Jemena's procedure is on the whole sound and appropriate for regulatory purposes. Our main conclusions regarding Jemena's approach are as follows.

Endogeneity

The regression model estimated by Jemena suffers from a technical problem known as endogeneity. Endogeneity leads to biased estimates of UAG rates, even in very large samples. We have suggested a relatively simple approach for overcoming the endogeneity problem and have applied it to obtain alternative estimates of the required UAG rates.

Consistency with overall five-year average UAG

Jemena has developed an approach that ensures that the estimated segment-level UAG rates are consistent with the five-year average overall UAG rate. As a check on the procedure, we estimated a regression model that imposes a constraint on the estimated segment-level UAG rates consistent with the five-year average overall rate as part of the estimation procedure. The estimates we obtained for the segment-level UAG rates using this approach are not very different from those obtained using Jemena's approach. Further, our approach is considerably more difficult to implement than Jemena's. Hence we believe that the procedure that Jemena has developed is reasonable for regulatory purposes.

Stability

The estimated UAG rates by segment vary somewhat depending on the sample that is used in the estimation. While this is not an ideal situation, there do not seem to be suitable alternative datasets available to overcome this issue. To

6 See equation (7) for the conversion equations linking the α coefficients and the β coefficients.

minimise the impact of sample size on the estimated UAG rates, we have recommended that as large a sample as possible be used in the estimation, which in this case is the sample of monthly observations from July 2002 to June 3013.

Page intentionally blank

FRONTIER ECONOMICS | MELBOURNE | SYDNEY
Frontier Economics Pty Ltd 395 Collins Street Melbourne Victoria 3000
Tel: +61 (0)3 9620 4488 Fax: +61 (0)3 9620 4499 www.frontier-economics.com
ACN: 087 553 124 ABN: 13 087 553 124