# Commonwealth Grants Commission 2020 Methodology Review

Tasmanian Government Submission -

*Response to Urban Transport Consultancy Stage 2 (Jacobs Report)*

12 December 2018

Tasmania
Explore the possibilities

# CONTENTS

# Summary

Tasmania welcomes the comprehensive approach taken by the Consultant to construct an alternative model to assess urban transport costs. The Consultant has undertaken a comprehensive analysis of the drivers of urban transport costs and identified a number of potential cost variables that potentially will result in a more accurate method for estimating urban transport costs.

However, it is difficult to assess whether the Consultant's preferred model is reasonable without access to the model, the data, or the modelled outcomes for each Significant Urban Areas (SUA).

Tasmania's main concern with the Consultant's preferred model is that it appears too focussed on the peak commuter transport demand and has not given sufficient consideration to other demand drivers such as meeting the transport needs of passengers from low socio-economic areas.

The model also does not appear to capture diseconomies of scale and diseconomies of density that arise from smaller populations and less integrated transport networks.

These cost factors are particularly prevalent in Tasmania and their omission from the model may understate Tasmania's urban transport costs.

# Introduction

Tasmania welcomes the opportunity to provide comments on the Stage 2 - Final Report on Urban Transport prepared by Jacobs Australia (the Consultant) for the Commonwealth Grants Commission (CGC) as part of the 2020 Methodology Review.

Tasmania has previously expressed concern that the current assessment of urban transport net expenses does not fully capture all material cost drivers. The current assessment is simply a function of urban population size (Australian Bureau of Statistics SUAs of 20 000 and above). There are likely to be other cost drivers, for example, socio-demographic factors, that affect the demand for public transport services, and the topography of the urban area that is serviced.

The alternative model developed by the Consultant aims to capture more of the primary demand and supply drivers of urban transport which in turn drive per capita State government expenditures in this service area. The Consultant's report identifies that expenditure on public transport provision will vary across cities based on the transport task to be undertaken, the characteristics of the transport system and the specific characteristics of the city itself.

The Consultant also formed the view that a single recurrent expenditure model was sufficient, as capital expenditure was found to be highly correlated to recurrent expenditure, and did not require specific separate modelling.

Because the Consultant's report does not provide modelled outcomes for each State or the supporting State data, it is difficult to assess whether the preferred model produces reasonable estimates of net expenses for each State. The report enables States to comment only on the conceptual case, analytical framework and econometric analysis. Tasmania's comments are therefore limited to these matters.

# Conceptual basis

***Current method - recurrent expenses***

The conceptual case for the current assessment of urban transport expenses developed by the CGC (based on its consultant's advice to the 2010 Review[1]) is that urban transport per capita net expenses (total expenses less revenue) increases as the population of urban centres increases.

Possibly because of time constraints with the 2015 Methodology Review, the CGC did not undertake a comprehensive review of its model and decided to retain the 2010 model with some modifications to include public non-financial corporation urban transport expenses and cities with populations over 20 000.

The current method assumes urban population is a good proxy for the urban transport task (as measured by passenger-kilometres per capita). This is because, as urban areas get larger:

- there is more congestion and thus more public transport usage per capita; and
- travel journeys increase as urban areas become larger and travel distances grow.

The CGC compared passenger-kilometre data and per capita net expenses which showed that both data increased with population size. The CGC concluded that these results were consistent with its conceptual case.

The current CGC model is a simple linear-log regression function based on net expenses (dependant variable) and SUAs with populations greater than 20 000 (independent variable).

$Ei = a + b * Ln\ (population)$

$Ei$ = net expenses per capita

$a$ = intercept

$b$ = coefficient

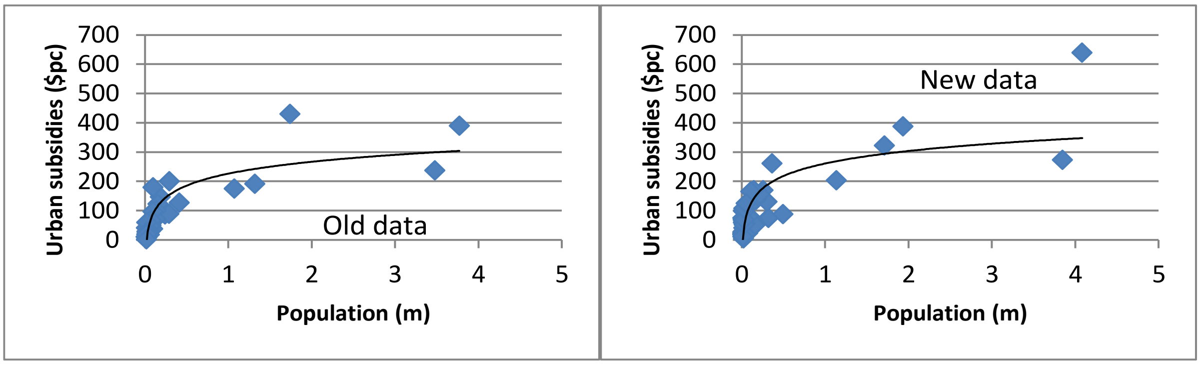
The population data is transformed by logarithm because expenses are observed to be non-linear with urban population due to factors such as economies of scale. That is, as urban populations increase in size there are operating efficiencies in the cost of delivery of transport services.

During the 2015 Methodology Review, some States raised concerns that, as there were only passenger-kilometre data for the eight capital cities, the conceptual case was based on a very limited data set. Also, as there was an uneven spread of city sizes, with a large number of smaller cities and few very large cities, the relationship derived from the data could be unduly influenced by the larger cities. The data point for Sydney could even be considered to be an outlier. This is shown in figure 1[2] where there is 'bunching' of data among the smaller cities and much fewer data points for the larger cities.

[1] *2010 Review of State Government Subsidised Urban Public Transport Services: Consultant Advice*, Institute for Sustainable Systems and Technologies, University of South Australia, April 2009.

[2] *2015 Review Proposed Assessments Staff Discussion Paper CGC 2013 - 07S,* October 2013.

Figure 1 - Relationship between public transport operating subsidy ($pc) and urban population (m), average of 2008-09 to 2011-12 and average of 2004-05 to 2006-07 - 2010 Review and 2015 Review with updated data.




***Consultant's approach***

For the 2020 Review, the CGC decided to undertake a review of its model and engaged a consultant to develop an alternative urban transport model. The aim is to better capture the recurrent and infrastructure costs in delivering urban transport. The Consultant investigated a broad range of potential variables in arriving at the preferred alternative model.

The Consultant noted that an ideal model is one where urban transport costs are a function of volume (the transport task) and factor prices. However, this structure only works if there is a single mode of transport, with similar population distributions, economic activity and geography between areas. As noted earlier, volume data (passenger kilometres per capita) is only available for the eight capital cities, so there are not enough data points to reliably undertake regression analysis using urban transport volume data. Therefore, proxies are needed for the volume variable.

The Consultant's recommended model framework uses expenditure on public transport as a function of the transport task, characteristics of the transport system and specific characteristics of the city itself. That is:

Ei = F(Di, Si, Ci)

Ei = net expenditure (dependent variable)

Di = demand related cost variables (proxies for volume)

Si = supply or network related cost variables (also proxies for volume, cost of provision and factor price effects)

Ci = city specific cost differences between SUAs

Data for the dependant variable (recurrent net transport expenditure) was provided by the States. However, the Consultant notes in the report that not all State data was derived in the same way, and so consistency was one of its major challenges (in addition to the need to use proxies for volume data) in developing the model. (page 7).

***Capital expenses***

The CGC separately assesses urban transport infrastructure costs within the investment assessment. This assessment is also based on population growth and city size as the key drivers of relative State costs.

The Consultant formed the view that a single expenditure model was sufficient as it found from the research evidence that capital expenditure was highly correlated to recurrent expenditure, and this obviated the need for a separate capital expenditure model.

Tasmania supports in principle the general approach taken by the Consultant as an improvement on the current assessment although we have some concerns with the form of the alternative model. These concerns are discussed in the following sections.

Without access to the data or modelled outcomes for each State, Tasmania reserves its support for the preferred model until further information can be provided.

QUESTION:

1. If a single expenditure model is adopted, how will the CGC treat urban transport infrastructure costs in the current investment assessment?

***Economies of scale***

The Consultant notes in research cited in the report that there is evidence that there are economies of scale as passenger numbers grow.

The Consultant argues that a linear-log relationship is used in the preferred model for the passenger numbers by mode variable to capture the increase in per capita expenses as the network becomes more complex, but that the rate of increase slows as passenger numbers increase.

Tasmania agrees with this observation. However, there is no specific discussion of the reverse of this situation, where small cities face diseconomies of scale. The linear-log form implies that as passenger numbers increase, the marginal per capita cost increase decreases. The converse of this argument is that as passenger numbers decline, the marginal per capita cost saving increases.

For small States such as Tasmania it is argued that servicing a smaller population base and dispersed urban form is costly on a per capita basis. That is, outer areas with low density and lower cost recovery from concession passengers requires more, rather than less, net spending per capita. This situation does not appear to be captured in the preferred model. Rather than a simple linear regression, the Consultant has not indicated in the report whether other functional forms were tested. These could have included a quadratic function or even a cubic function where there are initially diseconomies of scale when the population and demand is low, then transitioning to economies of scale as the population and demand increases.

The model includes population density as an independent demand variable although it is not transformed into a linear-logarithmic form, although other independent variables, such as passenger numbers, have been transformed. Non-transformation of the population density variable into logarithmic form assumes that economies of scale from increasing population density is being captured by the logarithm of other variables.

However, increasing population density could also contribute to economies of scale as greater urbanisation is likely to lead to economies of density[3] of transport systems. It is also recognised that increasing density will at some point lead to diseconomies of scale as systems become complex, congested and unable to meet demand.

[3] Economies of Density are generally characterised by a situation where unit costs are lower in relation to population density. The higher the population density, the lower the likely costs of infrastructure required to provide a service. One example would be the costs associated with providing electricity networks to urban compared to rural areas.

Economies/diseconomies of scale may also vary according to transport mode. For example, in Tasmania buses are the only form of public transport and for this transport mode economies of scale are harder to achieve compared to rail. This is because buses need to compete with other road traffic during peak periods as the population grows. This distinction does not appear to be modelled.

QUESTIONS:

1. The current CGC model uses the log of population which, in effect, accounts for population driven economies of scale. Is it a reasonable assumption by the Consultant to use population density as a demand variable that is not transformed into a logarithmic form?
2. Does the linear-log form correctly capture cost for smaller populations that face diseconomies of scale?

***Policy Neutrality***

Over many years successive State Governments have, at times, made policy choices to 'overinvest' in certain transport modes over other alternatives. For example, in Victoria some train services run well beyond distances where a bus connection would be a cheaper alternative to transfer lighter passenger loads to other destinations rather than to continue the journey by train. The introduction of light rail over cheaper alternatives is another example.

Section 2.1 of the Consultant's report addresses the need to develop a model that recognises the influence that policy factors have on expenditure levels and to subsequently limit the effect of these factors on GST funding shares. The Consultant states in the report:

> *'The principle of the CGC's advice on GST revenue distribution among states and territories (States) is horizontal fiscal equality (HFE). Therefore, the recurrent expenditure model must be independent of the policy of individual governments (policy neutral) and reflect what States do on average.*
>
> *In our opinion, policy neutrality and a reliable model can only be ensured following a two-step modelling process:*
>
> *1) Estimate a model that includes variables accounting for both policy-related and policy neutral cost drivers.*
>
> *2) Use this model to adjust the expenditure observations to policy neutral levels by removing the effect of policy variables on expenditure. Funds can then be allocated based on the relationships of these standardised expenditure levels.'*

In Table 2.1 the Consultant provides an illustrative example in relation to light rail to highlight the impact of a policy choice to introduce a higher cost alternative. This example suggests that the removal of policy driven expenditure in relation to light rail can be identified and removed.

Further, in Table 5.1, the Consultant identifies that policy related variables have been included in the modelling to remove the effect of policy related factors on subsequent funding shares. Tasmania supports this approach.

QUESTION:

1. It is not clear in the report what variables have been included to remove the effect of policy related factors on funding shares. For example, does the preferred model differentiate between types of rail (light rail and trams), and include ferries, and how have any policy related expenditure impacts been removed?

***Geography***

The Consultant has used the same spatial units as the current model used by the CGC except that there is no population cut-off for SUAs above 20 000. The geographic areas are therefore all SUAs as determined by the ABS[4] comprising one or more contiguous Statistical Areas level 2 (SA2s) containing one or more related Urban Centres joined using the following criteria:

- they are in the same labour market;
- they contain related Urban Centres where the edges of the Urban Centres are less than 5 km apart defined by road distance; and
- they have an aggregate urban population exceeding 10 000 persons.

It is noted that the SUA structure does not aggregate to State or Territories, as SUAs can cross State or Territory boundaries. An issue for the CGC is how cross boundary urban transport assessments are addressed. For example the Gold Coast - Tweed Heads SUA is in both Queensland and NSW and the Albury-Wodonga SUA is in NSW and Victoria. It is understood that in these cases the Consultant proposes that the estimates be apportioned to the States based on population or travel shares. Tasmania supports either of these approaches.

It is acknowledged that in order to develop a relatively manageable national model, SUAs are used as the primary spatial unit. Under the current model, this results in only four SUAs being included for Tasmania (Hobart, Launceston, Devonport and Burnie-Wynyard). However, with the removal of the 20 000 cut off, the number of SUAs only increases from four to five with the inclusion of Ulverstone.

Because expense data provided by Tasmania for Ulverstone and Devonport SUAs were estimated using population, this data has been excluded by the Consultant. This is because population is also a key variable in the regression. Similar exclusions for other States results in the total number of SUAs falling from 101 to 70. The Consultant considers that a complete dataset is not required and that a sufficiently diverse sample of SUAs enables the model to predict the missing values (for Ulverstone and Devonport). As all 70 data points represent 96 percent of total urban population, excluded SUAs represent only a small proportion of the total population. However, Tasmania would argue that while Ulverstone and Devonport have only small populations, they are unique and not representative of the national dataset due to the dispersed nature of the urban centres and the small populations served, and that they comprise largely non-commuter passengers.

Because there are only three SUAs for Tasmania out of a total of 70 SUAs, Tasmania's influence on the national model will be very small. By contrast, NSW accounts for the largest number of SUAs (31), significantly above its per capita share.

QUESTION:

1. Does past urban transport expenditure by NSW governments unduly influence the supply cost coefficients in the model? That is, is the model policy neutral?

[4] ABS cat 1270.0.55.004 - Australian Statistical Geography Standard (ASGS): Volume 4 - Significant Urban Areas, Urban Centres and Localities, Section of State.

***Treatment of satellite cities***

A key question raised by the Consultant is whether any SUAs should be combined. That is, whether there are any SUAs that should be considered as having a sufficiently integrated labour market with the neighbouring capital city to be, in effect, considered part of the larger SUA. These SUAs are considered as labour market integrated satellites.

The Consultant, using various geographic tests, (ie labour market is significantly integrated with the capital city) found that Yanchep in Western Australia was the only satellite city. Yanchep is a significant distance from Perth and is on the edge of the Perth SUA which includes Alkimos and Eglington. However, as there is no economic base at Yanchep it is primarily an outer suburb. Therefore, Tasmania agrees that Yanchep SUA should be combined with the Perth SUA.

# Variables considered

As noted, the Consultant developed an analytical framework that recurrent urban transport costs are a function of passenger demand related costs, network supply costs, and city specific costs.

The following comments are provided in relation to the variables that were considered by the Consultant to be suitable candidates for evaluation in their modelling.

## Passenger Demand Driven Cost Variables

***Low socio economic status***

The Consultant identified socio-economic status as a potential key determinant of public transport use and suggested that the ABS Socio-Economic Indexes for Areas (SEIFA) was likely to be a robust proxy to be used in its regression analysis.

Tasmania agrees with the Consultant that low socio-economic populations would be more reliant on public transport. This is because low socio-economic populations either do not have a car, or if they do, they cannot afford to use it regularly. Low income populations also tend to live further away from services where housing costs are lower. These journeys by low socio-economic populations are generally not commuter journeys but are to enable travel to appointments, shopping, and other non-work related purposes.

Tasmania's urban areas are characterised by small dispersed populations, and low-density development. The journey to work market is only really prevalent in the Hobart SUA, with little commuting on public transport in the rest of Tasmania[5].

The SEIFA variable was considered in Model 5, but it was rejected by the Consultant as the regression analysis test statistics were not superior to Models 1, 2 and 4. (Model 3 was rejected).

The key statistics from the Consultant's report are shown in the following table.

| Statistic (log linear form) | Model 1 | Model 2 | Model 4 | Model 5 |
|---|---|---|---|---|
| Adjusted $R^2$ | 0.77 | 0.73 | 0.76 | 0.69 |
| F Statistic | 48.44 | 47.12 | 32.0 | 40.17 |

However, in the case of Model 5 there is no technical proof or hypothesis test results provided as to why SEIFA was considered to not improve the fit of the model compared to the other models. It is therefore not possible to assess how significant the SEIFA variable is.

Non-linear relationships between the dependant variable and other variables have been attributed to scale effects and addressed by taking the natural logarithm of independent variables[6].

In the case of Model 1, the logarithm of passenger numbers was used and for Models 2, 3 and 4 logarithms of employment were used. However, in Model 5 there appears to be no effective scale effect adjustment as the logarithm has been applied to the slope variable. It is not clear from the Consultant's report that the slope variable would exhibit a diminishing rate of costs and, in addition, it is not related to transport demand. It is

[5] The Tasmanian Department of State Growth estimates that about 20 per cent or urban travel in the Hobart SUA is for journey to work. This is significantly lower in the North (about 10 per cent) and the North West (about 5 per cent).

[6] 'Economies of scale can be modelled with specific proxy variables, such as congestion or mode shares, which describe expense patterns. Alternatively, it can be accommodated through the choice of functional form by using logarithms. We have applied both approaches in our analysis'. Jacobs Urban Consultancy Stage 2 Report, page 42.

noted that the regression statistics for Model 5 are very similar with, or without, the logarithmic form of slope, which suggests that non-linearity may not be present.

If this is the case, then Model 5 does not include a variable that captures scale effects. It would be enlightening to know if it would have produced a better $R^2$ had it been in logarithmic form.

Tasmania is also concerned that, by excluding a variable to capture low socio-economic status, the model places greater emphasis on the commuter market as the key demand driver of urban transport costs. There is only limited consideration of non-work related trips. The journey to work demand is of significance only in the Hobart SUA with little commuting on public transport in the rest of Tasmania.

Tasmania is concerned that the Consultant has disregarded socio-economic factors such as SEIFA too readily and appears to have favoured the peak commuter as the key driver of passenger demand in the preferred model.

QUESTION:

1. Why have socio-demographic variables been omitted from the model, despite the report finding that socio-economic status is likely to be a key determinant of public transport use?

***Employment***

The inclusion of employment as a potential variable could be considered a proxy for the costs driven by journey to work passengers (commuters). Higher rates of employment would create a greater demand on commuter travel and thus transport costs. As there are economies of scale, the Consultant has applied this variable in logarithmic form.

While this would seem a reasonable proposition it does not necessarily hold for smaller SUAs that do not have the high levels of congestion of larger SUAs. As noted in Tasmania's case, a significant proportion of total passengers use urban transport for non-work related travel.

However, employment was not included as an independent variable in the preferred model because of multi-collinearity issues between population density and employment. Tasmania would agree with this decision.

***Population weighted density***

Tasmania accepts that population density is a key driver of urban transport costs. Clearly, as urban regions grow, and population density increases, there will be greater demand for, and associated costs of, providing public transport. The current model uses population rather than population density as its only variable. However, it is transformed into logarithmic form to capture scale effects. It is not clear why the Consultant did not consider applying a logarithmic form to population density instead of other variables in its modelling.

As noted earlier, population density could also be a driver of economies of scale and thus a log-linear transformation may also be appropriate. The Consultant may have used the logarithm of passenger counts to make the interpretation of the estimated coefficients easier and more directly comparable between SUAs. Population density per SUA can already be readily compared.

The Consultant has included population density in the preferred model. Tasmania agrees in principle that this is an appropriate variable.

QUESTION:

1. Why did the Consultant not consider applying a logarithmic form to population density instead of other variables?

***Student travel (School enrolments)***

It is noted that the CGC is considering including school transport in the urban transport assessment for the 2020 Methodology Review. Tasmania has previously expressed concerns with this proposal in its response to the CGC staff Draft Assessment Paper, particularly if it was applied to the current transport model.

School enrolments were considered by the Consultant, although excluded on the basis that school enrolments may be potentially too narrow in being representative of the broader population's transport needs and it did not appear to add explanatory power on top of that already offered by employment, especially if it exhibits high correlation with employment.

QUESTIONS:

1. Why was school enrolments considered as a possible driver of urban transport when the CGC already makes an assessment of school transport within the Schools assessment?
2. How are other student journeys treated? (e.g. urban students travelling on general public services).
3. The CGC is considering including school transport in the urban transport assessment. How does this accord with the Consultant's preferred model which does not include student travel?

***Network supply cost variables***

It is understood that the variables considered in this category relate to urban transport supply or availability differences between SUAs. For example, some SUAs will have a particular transport mode composition (rail, light rail, trams, buses and ferries) and each mode will have differing supply costs.

The Consultant also considered distance to work in this category of variables, because the average distance to work affects the cost of supply which will be dependent on rail and bus route length.

***Passenger numbers by mode (train and rail)***

The Consultant considered passenger numbers by mode and distance to work as proxies for passenger kilometres and congestion. This variable is designed to capture the higher cost of transport modes where higher demand requires more complex infrastructure such as trains compared to buses.

In Appendix D, the Consultant undertakes multi-collinearity tests of all considered variables to assess where independent variables may be correlated with each other, thus potentially undermining the statistical significance of an independent variable[7]. An assessment is made between population density and train and bus mode use/level and the correlation coefficients are both 0.79. This is above the threshold of 0.70 that the Consultant considers would be likely to significantly hamper the precision of the estimates.

However, in the Consultant's preferred model, both population density and passenger numbers for bus and train transport modes are included as independent variables which suggests that multicollinearity may be present. It is not clear why passenger numbers have been included when employment was rejected because the level of multi-collinearity is above its threshold.

***Distance (journey to work)***

This variable aims to capture the higher costs associated with longer distances required to journey to work.

The Consultant notes that this relationship is not clear cut, as greater commuter distances may make private car use more feasible so there is reduced demand for transport services the further the population lives from

[7] Severe multicollinearity is a problem because it can increase the variance of the coefficient estimates and make the estimates very sensitive to minor changes in the model. The result is that the coefficient estimates are unstable and difficult to interpret.

the city. This could be due to the fact that being further away from city centres means that urban transport service routes and frequency are less likely to be convenient compared to living closer to the city centre.

Tasmania would also argue that persons of low socio-economic status tend to live further from city centres and often there is little option but to use public transport even though it may be inconvenient or poorly serviced. These include many non-work related journeys.

Despite these conceptual issues, the Consultant has nevertheless included distance to work as an independent variable.

## City Specific Cost Variables

***Slope variable***

Tasmania would agree that topography is a key determinant of costs, particularly for the Hobart SUA. Recognition of slope would capture the higher maintenance and other operating costs on Tasmania's bus fleet due to the hilly terrain, in Hobart in particular. The Consultant considered a range of slope indicators such as the proportion of the area of land within Greater Capital City Statistical Areas (GCCSA) with zero slope area, land slope mean, and land slope Standard Deviation.

Ultimately, land slope mean was used in all the models considered.

It is not clear to what extent the slope indicator captures the areas that are specific transport routes. Presumably transport routes would avoid steeper sections but this may be at the cost of longer routes to avoid steep terrain. Nevertheless, land slope mean as a topographical variable is a positive inclusion in the Consultant's model, with hilly terrain a factor contributing to higher urban transport operating costs.

There does not appear to be a topographic factor to account for urban transport constraints such as river ways. For example, the Hobart SUA is bisected by the Derwent River and this constrains east-west traffic to a limited number of bridge crossings and, combined with the constraint of Mount Wellington, creates bottlenecks at lower levels of traffic than other similar sized cities.

QUESTIONS:

1. Does the model differentiate between an SUA that has high mean slope but the steeper areas within the SUA are not where public transport services occur and another SUA that also has high slope but where public transport routes occur?
2. Can other topographic features that create bottlenecks such as waterways and mountains be captured in the modelling?

## Econometric Modelling

The Consultant considered a range of demand, supply and cost variables, and after modelling a range of variables, found the following to be the best explanatory model, with an $R^2$ of 0.79 (linear-log model). The equivalent $R^2$ for the current CGC model is 0.64.

$$Exp_i = B_o + B_1\ dense_i + B_2\ Dist_i + B_3\ slope_i + B_4\ Ln\ pax_i\ train + B_5\ Ln\ pax_i\ bus + E_i$$

Dense = urban population weighted density - urban demand proxy for volume

Dist = distance to work (proxy for network complexity)

Slope = topography related cost

Ln pax bus/train = supply or network related variables by bus and train modes as proxies for congestion and volume

E = error term

The linear-log form of the model can be interpreted as indicative of scale effects in the wider sense as it suggests that the impact on per capita expenses from growth from additional passenger numbers increases at a slower rate as total volume increases.

The overall methodologies and econometric modelling used by the Consultant appears to be of sound quality. In most peer reviewed econometric research papers, the data and model codes are made available in order for the results to be replicated. However, no raw data, specific model outcomes or modelling codes have been provided, and so a detailed assessment and additional model testing is not fully possible.

Some minor issues that were found in the report are outlined below.

Throughout the report a high $R^2$ is often cited as being a primary desirable property for model selection. While this is largely true, a model with a high $R^2$ in particular is not always indicative of necessarily being a better fit than a model with a lower $R^2$. A high $R^2$ could be the result of a spurious regression, where the dependent and independent variables share some stochastic process that may indicate a false level of explanatory power that is driven by the specific sample selection.

Furthermore, a high $R^2$ could be driven by a model that is over fitted, since the inclusion of more independent variables will increase $R^2$, although these may not inhabit any real explanatory power. If there are too many terms compared to the number of observations, the estimated coefficients would mostly represent 'noise' and no longer provide an accurate basis for forecasting costs. However, this issue is partly recognised, through the use of the adjusted $R^2$ which accounts for additional explanatory variables that do not improve the model fit.

In terms of accurate variable selection for the final model, an additional method other than relying on the value of $R^2$ and adjusted $R^2$ for each model is to conduct restricted and unrestricted F-tests. This would accurately determine which independent variables significantly impact the dependent variable and should therefore be included in the final model.

Currently, other than analysing the value of the model $R^2$, the report often refers to the model F-statistic as a further important measure of model performance. While this is again mostly an accurate measure, the overall F-statistic only indicates the statistical significance of all independent variables jointly (specifically, testing whether all joint model coefficients are statistically significantly different from zero). If, for example, only one independent variable is statistically significant while the others are not, the model F-statistic will most likely be statistically insignificant and the model will be rejected. It is therefore important to take into account the individual significance of the independent variables measured through the individual coefficient t-statistics.

Most other potential concerns, such as multicollinearity arising when both population density and employment are included in a regression, have been recognised in the report. However, as noted earlier there appears to be a multicollinearity with the inclusion of the passenger count and population density variables.

Tasmania's queries with regard to the Consultant's modelling are briefly summarised as follows:

QUESTIONS:

1. Is urban transport demand captured twice in the Consultant's model by including both population density and passenger numbers? As noted, this may be a statistical problem as the two variables may be correlated - refer to earlier comments on the passenger numbers variable and multi-collinearity.
2. Why has a logarithmic form been used for the slope variable in Model 5? The Consultant argues the logarithmic form is used to address economies of scale. It is not clear why slope would exhibit economies of scale.
3. Why is R2 the primary basis for choosing the preferred model?
4. The data in the model 'clusters' at the lower end of SUA size with few data points at the higher population SUAs. Does this create an issue with determining a line of best fit? Should Melbourne and/or Sydney be treated as outliers?
5. What are the units of measurement and how has the data been used? For example, are the number of bus and train passengers based on the number of people who used either form of transport on Census night, irrespective of frequency or distance? The 2016 Census questionnaire only asks about the method used to travel to work, and there are no questions on other reasons for travel.
6. How is distance and slope derived?
7. As mentioned, it is difficult to assess the Consultant's model without access to model and the data. Will the CGC be able to provide the States with a copy of the model and the data to enable a more detailed evaluation?