25th Australasian Transport Research Forum
Incorporating the BTRE Transport Policy Colloquium
Canberra 2-4 October 2002

Features of seasonality and its effects on road freight transport task

Paul Sutcliffe
Australian Bureau of Statistics

Abstract
In validating the quality of reported data in the recently completed survey of Freight Movements (FMS 2000), conducted by the Australian Bureau of Statistics (ABS) for the 12 month period ended 31 March 2001, investigations identified some under-reporting. This prompted investigations into adjustment methods and the need for more detailed analysis to justify confidence in the data. What impact can the day of the week, the month of the year or other times of the year have on the road freight transport task?

This paper will focus on the effects of seasonality (including weekly and daily effects) on the road freight transport task. Key features identified from micro level data analysis undertaken by the ABS will be discussed, as will the techniques used by the ABS to conduct these analysis.

Contact author
Paul Sutcliffe
Director,
Statistical Services Branch
Australian Bureau of Statistics
Locked Bag 10
Belconnen ACT 2616
Ph: (02) 6252 6759

P.sutcliffe@abs.gov.au

GO BACK TO START SCREEN GO BACK TO INDEX

Introduction

Earlier this year the ABS released the 2000 Freight Movements Survey which provided statistics about the size and characteristics of the transport task, including freight flows between geographic areas. The statistics relating to freight moved by articulated vehicles were the first results using a new survey methodology. Previous Freight Movements Surveys were 'business-based' whereas the new approach is vehicle-based and has provided a new opportunity to understand trip characteristics.

This paper will discuss the main features of the FMS 2000 methodology and the strategies for ensuring that the new survey would be robust. During the data validation stage of the survey processing cycle there was some detailed analysis of a trip level dataset into the possibility of under-reporting of trips by survey respondents. The time series analysis techniques used provided an opportunity to understand aspects of seasonality of the road freight transport task.

Background

The 2000 Freight Movements Survey provides estimates of freight moved by road, rail, sea and air for the period 1 April 2000 to 31 March 2001. The statistics for the road component of the collection were based on a sample survey of articulated vehicles that were registered with a motor vehicle registry for road use at some stage during the 12 months ended 31 March 2001. Rigid trucks, other commercial vehicles and vehicles belonging to the defence services were excluded from the scope of the survey.

A sample of approximately 14,000 articulated vehicles was selected to report over 26 fortnightly periods within the reference year. This equated to a sample of approximately 500 selections in each fortnight. The sample size was chosen to give a suitable level of precision for total distance travelled, tonnes carried and tonne-kilometres travelled at the State/Territory of registration level. The selected articulated vehicles were asked to report on freight movements over a two week period by recording details of each trip (loaded or empty) made in a trip record book.

Selections were made from a population frame of 63,000 articulated vehicles compiled on 31 October 1999. To help correct for under coverage on the population frame, new articulated vehicles registered after 31 October 1999 were progressively added to the population frame during the enumeration period and supplementary samples drawn. Adjustments were also made to the estimation process to account for new and re-registered articulated vehicles not given a chance of selection in the survey.

Non- Sampling Error

The new survey collection methodology was tested using observational studies prior to the collection phase and the ABS made every effort to reduce non-sampling error to a minimum by careful design and testing of the questionnaire. In order to improve response rates and data quality, respondents were contacted 1 month prior to the start of the reporting period to encourage record keeping, provided with an information booklet and given a Freecall number to call if they required assistance.

While every effort was made, there were a number of issues which required resolution relating to the difficulty of collecting trip data for some vehicles selected in the survey. Standard processing and editing practices were followed to ensure the appropriate treatment for particular cases such as changed ownership of the vehicle during the collection period, reduction of respondent burden when multiple trips which had the same characteristics were performed, and the classification of unusual trips.

In retrospect under reporting of trips is not easily evident at the truck level. For example, there were many trucks that were operating only one week out of the two. During processing these were checked to determine if this was correct. Where a form had trips reported for both weeks it was very difficult to have an editing rule at truck level.

Data Validation Study

During the data validation phase of the collection cycle comparisons were made to determine if data reported in the first week of the reporting period was different to that reported in the second week. A dataset was created which had the estimates for five key variables: laden distance travelled; total distance travelled; weight carried; number of laden trips made; and total number of trips made for each day for each collection cycle. While the standard errors for each daily estimate were likely to be much larger than those used to publish an annual estimate it was expected to provide some insight into any reporting problems that might have existed.

Simple graphical analysis which plotted laden distance for each day of a cycle for all 26 reporting cycles provided possible evidence that respondents may have under-reported trips made during the second week. To check this the average laden distance between week one and week two was compared. In most cycles the average laden distances travelled were lower in the second week. After reviewing this simple analysis two main theories emerged. Firstly, there was a possible week 1 - week 2 effect, and secondly there was a possiblity that there

was less data reported in the first and last day. The questions then asked were: were these effects statistically significant? If so, what action should be taken? Since any adjustment to the published data would need to meet strict criteria and be approved by senior ABS staff, it was decided that, in order to determine if a reporting bias existed in a statistically significant sense, a more sophisticated analysis was required.

The main weakness of the simple analysis was that it was based on raw data which didn't take into account any effects such as public holidays. To estimate the week1/week2 effect, taking into account public holidays, a model based estimation approach TRAMO (Time Series Regression with ARIMA Noise, Missing Observations, and Outliers) was adopted. Generally, ARIMA modelling using seasonal series (e.g. 12 months, or 4 quarters) is straightforward but the TRAMO package had the flexibility to allow the 7 day period required for a more detailed analysis. The model using generalised regression allowing for stochastic and daily activity patterns is:

$F_t = b\ Y_t + O_t$

where the FMS daily data (F_t : t = 1,...,364) can be modelled by a vector (b) of regression parameters which we need to estimate, a regression matrix (Y_t) where we can fix components based on our known theories, such as the week1/week2 effect and first-last day effect, plus other parts assumed to follow an ARIMA model (O_t) which includes changing trend and daily activity patterns.

Table 1 : TRAMO Results

	Regression Parameters Week 1 to Week 2	Regression Parametrs Day 1 and Day 14	Under Reporting Effect	Standard Error	T-Tests for each regression parameter
Laden Distance	0.053	-0.099	-10.2%	1.6%	6.91, -4.33
Total Distance	0.042	-0.045	-8.0%	1.2%	6.81, -2.47
Laden Trips	0.024	na	-4.6%	1.4%	3.41
Total Trips	0.027	na	-5.2%	2.1%	2.55
Commodity Weight	0.054	na	-10.2%	1.8%	6.04

The results (table 1 above) indicated that the under-reporting effects were not the same for different variables. For example, the impact of the loss of trips is only half that of the distance travelled. This suggested some interaction between some of the variables. The important feature from our point of view was that the t-tests

scores were quite high and the standard errors reasonably low indicating that there was a statistical difference between data reported in week 1 and data reported in week 2.

A number of methods to adjust the FMS data were considered. Imputing for the 'missingness' explicitly by adding trips to the base dataset for estimation was the method chosen. The TRAMO analysis provided the imputation targets (half of the percentages shown in table 1) but the main difficulty was imputing appropriately for all of the key variables, not just one. In order to impute trips in an objective way based on the 'missingness' of each variable we set up appropriate imputation classes (A_k) and minimised the sum of squares of the adjustment at the imputation class level, thus meeting the constraints in table 1. In the end a number of different imputation class set-ups were investigated as well as using different variables in the least squares analysis. The best results achievable within the strict time constraints was the use of three imputation classes based on laden distance, while using both laden distance and tonnage in the least squares analysis. The final imputation class and allocations are presented in table 2 below.

Table 2 : Final Imputation Allocations

Imputation Classes (Laden Distance km)	Imputation Fractions	Number of Trips in Week 1	Number of Trips Imputed
Less Than 900km	0.103004	91501	9425
900km - 1999km	0.037344	2892	108
2000Km +	0.299213	381	114

The imputation fraction for differed for the final imputation classes for each trip which were based on the distance. The high imputation fraction for the longer trip imputation class reflects the interaction between the number of trips and distance travelled with respect to under reporting. The imputation was done by donor imputation within each imputation class and resulted in the adjustments at the total level shown in table 3. The time series analysis suggested adjustments are bracketed in bold.

Table 3 : Final Imputation Outcomes (Percentage Difference

	Laden Distance	Total Distance	Commodity Weight	Number of Laden Trips	Total Number of Trips
Outcome	5.15%	3.76%	5.09%	5.11%	2.90%
Target	5.05%	4.00%	5.10%	2.30%	2.60%

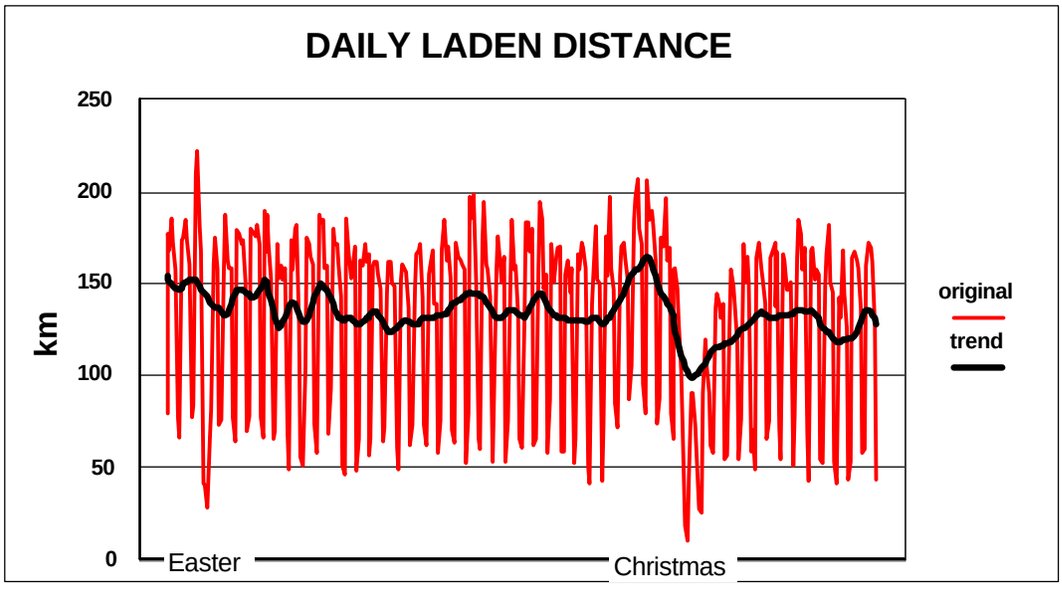
The final outcomes were a compromise between the ability to make the appropriate adjustment for each of the five key variables.

Road Transport Trend

Although the main outcome of the data validation study was to confirm and assure the quality of the published estimates the creation of the dataset for the validation exercise had some unexpected benefits in that it provided data relating to the road freight transport task which had been analysed in some depth.

The daily data is very volatile and the amount of road transport activity is influenced by many known events such as weekends, public holiday periods, and other particular irregular events. The standard time series technique to produce the trend is to remove seasonal and irregular components.

Graph 1 : Road Transport Daily Trend 1 April 2000 to 31 March 2001



The main feature of the daily trend is that the daily pattern seems to be changing throughout the year. There are large effects of public holidays on the data especially for Easter and Christmas periods. The surge prior to Christmas is most notable in the above graph as is the drop to half the normal activity throughout the Christmas period. There is no significant effect from the Sydney Olympics (September 2000).

Another notable feature of this data is that there appears to be a downward trend over the 12 month period. In actual fact there doesn't seems to be a recovery to pre Easter levels of the previous year and the daily distance is nearly 25,000 km

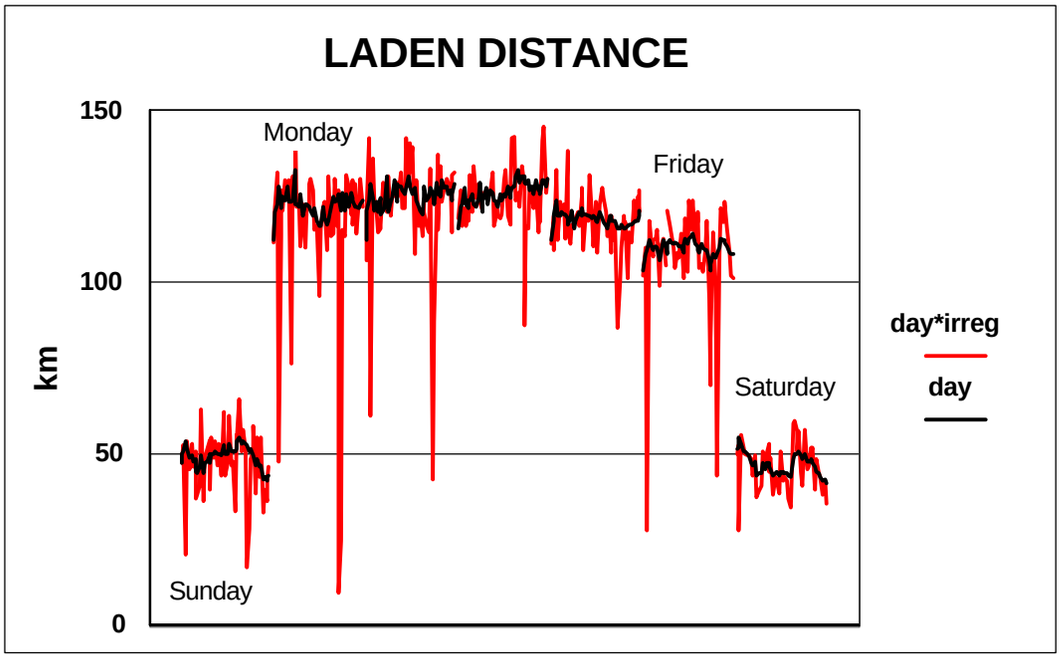
lower. At present there is no data available to be able to determine the trend for the following year.

Road Transport Daily Patterns

The extreme volatility in graph 1 indicated by the red lines shows the weekly pattern (cycle) for road transport Analysis of the daily patterns over the year show Saturday and Sunday to be more stable days because they do not contain public holidays (Graph 2 below) they have considerably less activity than week days.

While this could indicate that the road transport task is not a 24 hr 7 days a week process, there is also the possibility of a reporting day effect (i.e. trips are not recorded as starting or finishing on a weekend). This should is to be investigated in future development work for the survey.

Graph 2: Daily Patterns for the Year



Conclusion

The identification and subsequent adjustment for under-reporting of trips in the second week highlight the difficulty of collecting reliable and accurate trip level information. The value of the time series analysis conducted is that there is scope to use it during the live process (quality assurance and editing) to aid in the detection of unusual features of the data, such as under or mis-reporting or outliers.

The interesting feature was how the analysis of the daily data led to an understanding of the road freight task. While it is accepted that the data at daily level has high sampling errors the time series analysis has provided some conclusive evidence of the patterns in the daily data. Firstly, the changing tend throughout the year and the possibility that there might not have been a complete recovery from Christmas to Easter. The daily pattern analysis indicates that the road freight transport task may not be a 24 hrs seven day a week task with weekends appearing to be a relatively a quiet time for the industry.

Acknowledgments

This work was undertaken by the methodological staff in the ABS. Thanks go to Andrew Sutcliffe for time series analysis and Brett Fraser for the imputation and analysis of the data. Also, thanks to Donna Nicholson for providing editorial comment.

References

ABS Publication 9220.0, Freight Movements (2002)

Gomez, V and Maravall, A (1996) Programs TRAMO and SEATS Instructions for the User.