Resolving Some Methodological Issues in Road Safety Evaluations: The Victorian TAC Campaigns Revisited

Richard Tay (Presenter), CARRS-Q, QUT

Richard Tay (BSc, Texas Tech; MSc, Stanford; PhD, Purdue; Post-doc, MIT) is the Associate Professor in Road Safety at CARRS-Q. His current research interests include transport economics and modelling, public policy and applied econometrics, with special emphasis on issues relating to road safety. He has published extensively in major transportation and economics journals and is currently editing two special issues on road safety for *Logistics and Transportation Review* and *Applied Health Economics and Health Policy*. He has also served as a consultant for many multinational corporations and government agencies in Singapore, Hong Kong, New Zealand, Australia and the United States.

Abstract

Since their implementation, the TAC campaigns have been extensively evaluated and the mixed results obtained by different researchers have generated much public debate. The true effects of most countermeasures are unknown and may never be known. In estimating their impacts on road safety, analysts have to make a variety of assumptions regarding the relationships between variables and depending on their beliefs, the outcomes may differ considerably. The purpose of this paper is to review conceptually some of the differing assumptions adopted by the different researchers and empirically testing several of these assumptions and model specifications using the same data from previous studies (Cameron et al, 1993; Newstead et al, 1995; White et al, 2000), with the aim to improve our conceptual and methodological understanding of road safety evaluations.

Introduction

In an effort to reduce the social cost of road crashes, transport authorities in Australia, and around the world, have invested a considerable amount of scarce resources to design and implement a variety of countermeasures. Economic principles stipulate that all major policies and programs should be properly evaluated to ensure that the scarce resources are utilised efficiently. In practice, however, few road safety countermeasures have been adequately evaluated. One of the few exceptions is the advertising and enforcement campaign implemented by the Transport Accident Commission (TAC), which have been extensively evaluated since its implementation and the mixed results obtained have generated much public debate (Cameron et al 1993; Newstead et al 1995; White et al 2000, Cameron & Newstead 2000).

In the most widely cited evaluations by Monash University Accident Research Centre (MUARC), Cameron et al (1993) and Newstead et al (1995) estimated two log-linear models (double-log or multiplicative model) using monthly data from January 1983 to December 1992. The dependent variable used in the drink driving model was the number of serious crashes (fatal plus hospitalization crashes) during high alcohol hours and the independent variables used were a linear trend term, eleven dichotomous variables for monthly fluctuations, unemployment rate, alcohol sales, number of breath tests performed, and anti drink driving adstock which is the main measure of the amount of advertisements. For the speeding model, the dependent variable used was the number of serious crashes during low-alcohol hours and the independent variables used were a linear trend term, eleven dichotomous variables for monthly fluctuations, unemployment rate, number of speeding violations issued and the adstock for speeding. These evaluations concluded that most of the reduction in serious crashes in Victoria during the post-intervention period could be attributed to the TAC campaign.

The results obtained by MUARC were first publicly challenged by White et al (2000) who found that the models used by MUARC were not robust and several 'minor' modifications resulted in contrary outcomes. First, White et al (2000) replaced the unemployment rate with the 12-month centred moving average of the leading index of economic indicator and found that the estimated effect of the campaigns were significantly reduced, albeit still statistically significant, in the speed model, and more drastically, were insignificant in the alcohol model. Second, White et al (2000) replaced the quantitative measures for the
advertising and enforcement campaigns with qualitative measures and found that the qualitative measures performed just as well. Last, the starting month of the campaign was amended by a month, an issue that will not be discussed in this paper because it has no theoretical or methodological implications for other road safety evaluations.

The mixed results obtained by the different analysts are fairly common in the econometric evaluation of most public policies and programs. It is important to note that the true effects of most road safety countermeasures are unknown and may never be known, and that definitive evaluation does not exist. In estimating the impacts of countermeasures on road safety, analysts have to make a variety of assumptions regarding the relationships between variables and depending on their beliefs, the outcomes may differ considerably. Therefore, it is essential that all major policies and programs be evaluated and re-evaluated by different analysts using different methods to establish a broader frame from which a more reliable conclusion about their effectiveness can be drawn. More importantly, whenever possible, differences in the results obtained should be further examined to provide greater insight to improve our conceptual and methodological understanding of road safety evaluations.

The TAC campaign thus serves as a very interesting case study to highlight some of the recurring methodological issues in road safety evaluations. This paper will re-examine some of the issues raised in previous evaluations by utilising appropriate theories from economics, statistics, criminology and psychology, and empirically testing several of the differing assumptions made by previous researchers. Additionally, it will also test several other conventional wisdom and common model specifications used in the road safety literature.

**Efficacy of Different Economic Indicators**

While the role of economic activities has been well discussed (Hakim et al. 1991; Harry 1997; Tay 2003a), the relative efficacies of different economic indicators have received little attention in the literature (Tay 2003a). Over the last twenty years, nonetheless, a variety of economic indicators have been used and reported in the literature including employment rate (Parryka 1984), leading index (White et al 2000), retail index (Tay 2003b), new car sales (Tay 2001a), and unemployment rate (Newstead et al 1995) which is the most widely used among all the economic measures (Hakim et al 1991). Overall, the results seem to indicate that there is an inverse relationship between the level of economic activities and road safety regardless of which measure of economic activity has been used.

Although most of these economic indicators are very similar and are expected to be highly correlated, there are some differences among them which may render one measure as a better predictor of certain types of road crashes. In one of the few papers that explicitly examined the relative efficacies of different economic indicators in predicting different types of road crashes, Tay (2003a) argued that the specific measures of speed-related and alcohol-related crashes commonly used in Australia and their relationship to the two different economic indicators used in the previous evaluations of the TAC campaigns were the key to deciding which measure were better predictors in these models.

First, the dependent variable used in the alcohol model was the number of serious crashes during high alcohol hours, which corresponded roughly to crashes during the evenings and on weekends. It should be noted that much of the exposure during these time periods is likely to be discretionary travel and highly sensitive to current income and economic activities as measured by current unemployment rate and less sensitive to future income as measured by the leading index. Second, the consumption of alcohol, particularly the type and manner in which it is consumed (eg drinking at restaurants, pubs and hotels instead of drinking at home), is also likely to be more sensitive to current income than future income. Therefore, the current unemployment rate may be a better indicator of economic activities for this particular model.

Next, speed-related crashes, defined as crashes during low alcohol hours, coincide substantially with crashes during day times on weekdays. Trips taken during these hours are less likely to be discretionary trips but are more likely to be work-related and certain types of household travel, such as educational trips (to and from school) and grocery shopping, which are expected to be less income elastic, especially very short-term income fluctuations, but may be more elastic with respect to changes in medium term income. Also, crashes during these times are less likely to be caused by driver impairment such as drink driving or driving while fatigued but are more likely to be due to factors such as driving intensity and the vehicle and road conditions (age and maintenance of roads and vehicles), which are expected to be more sensitive to medium term income fluctuations. Lastly, fluctuations in medium term performance of the economy are also more likely than short-term fluctuations to affect the more mature driving population who are less likely to drive at night.
compared to younger drivers. Thus, conceptually, the moving average of the leading index may be a better predictor of crashes during low-alcohol-hours.

The empirical choice between the two economic indicators can easily be tested using the two approaches suggested by Gujarati (1995). The discriminating approach uses common goodness-of-fit measures such as the R-square, adjusted R-square, Amemiya's Prediction Criteria and Akaike's Information Criteria to select between competing models. Tay (2003a) showed that the model using unemployment rate fitted the data for alcohol-related crashes better whereas the model using the leading index fitted the data for speed-related crashes relatively better, confirming the conceptual arguments presented above.

In addition to the discriminating approach, Gujarati (1995) also recommended the discerning approach and suggested two common tests: the non-nested F-test and the Davidson-Mackinnon J-test. To conduct the non-nested F-test, a combined model using both economic variables was estimated. Two standard F-tests were then conducted to compare the explanatory power of the individual models with that of the combined model. Tay (2003a) conducted the non-nested F-test and found that the model using the unemployment rate was rejected in favour of the model using the leading index in predicting speed related crashes and vice-versa.

For each model, the Davidson-MacKinnon J-test required an additional model to be estimated (Gujarati, 1995). First, the model using the unemployment rate as the economic variable was re-estimated by including, as an additional independent variable, the predicted values from the model using the leading index. Next, the model using the leading index as the economic variable was re-estimated by including, as an additional independent variable, the predicted values from the model using unemployment rate. Last, two standard t-tests were conducted on the additional variables. Again, Tay (2003a) conducted the Davidson-McKinnon J-tests and found that the model using the leading index was rejected in favour of the model using the unemployment rate in predicting alcohol related crashes and vice versa.

**Structural Relationship between Countermeasures and Crashes**

The next issue raised by White et al (2000) was the use of dichotomous versus quantitative variables to capture the enforcement and advertising activities. Theoretically, both the enforcement and publicity campaigns are expected to produce deterrent effects on the amount of speeding and drink driving. In Victoria, the intense enforcement activities were designed to maximise specific deterrence by increasing the likelihood of apprehension significantly. On the other hand, the advertising campaigns were designed to supplement the enforcement activities by providing the general deterrence effect, in addition to changing drivers' attitude and behaviour through its fear appeals. Although there is a presumption in both theories that the relationship is monotonically increasing, at least at a very macro scale, both the deterrence theory and fear appeals model, however, are rather vague on the exact relationship or functional form. Moreover, economic theory posits that most short run production functions could exhibit increasing, constant or decreasing returns depending on the current input levels. Thus, this issue has to be resolved empirically if possible.

Both Newstead et al (1995) and White et al (2000) had assumed that the level of enforcement and advertising prior to the launch of the TAC campaign were negligible. With the advent of these activities, there were discrete and substantial increases in the values associated with these variables. Thus, it is clear that the amount of advertising and enforcement activities, especially in their logarithmic form, can be approximated very well by two step-functions or indicator variables. The correlations between the quantitative (in their logarithmic forms) and qualitative variables were extremely high, with estimated coefficients of 0.997 and 0.993 for enforcement and advertising respectively. It should be noted that high correlations among independent variables by themselves may not pose a serious estimation problem. For example, the correlation between the drink driving enforcement and advertising activities was also very high, with an estimated coefficient of 0.945 for the quantitative measures but the model was able to discriminate between their relative contributions. However, if the correlation is high enough to create an estimation problem, then care must be exercised in interpreting the results even though the ordinary least squares (OLS) estimator is theoretically still the best linear unbiased estimator.

Tay (forthcoming) empirically tested the choice of the two structural forms for the alcohol related crashes using both the discriminating and discerning approaches described above and could not find any evidence to reject either model in favour of the other. Therefore, both conceptually and empirically, there is no evidence to support the choice of one model over the other. This inability to differentiate between the two structural forms is due partly to the data used and more research should be conducted using data over a
wider range of values for the advertising and enforcement variables. One straightforward approach would be to include the actual data for the pre-campaign period instead of arbitrarily setting them to zero thereby creating the artificial step in the function.

**The Supportive Role of Advertising**

Despite extant research showing otherwise (Koenig and Wu, 1994; Tay 1999, 2001a, 2002b, 2003a,b; Tay & Watson, 2002; Ebel et al, 2003), there is a popular belief among many road safety professionals that road safety publicity campaigns do not work, or at best, have only a supplementary role to traffic law enforcement (Elliot, 1992). However, little research has been conducted to investigate the relationship between publicity and enforcement campaigns. The difficulty in testing this presumption is that most road safety campaigns in Australia and New Zealand utilize advertising as a support for enforcement and advertising always follows after enforcement and not vice-versa.

In two of the studies that explicitly examined this issue, Tay (2002a, forthcoming) included an interaction term in the models and found that the estimated coefficients were positive, albeit statistically insignificant for the alcohol model but significant for the speed model, while the estimated coefficients for the advertising term were negative and significant in both models. These results implied that the advertising campaigns by themselves had a significant effect in reducing serious crashes and more importantly, their impact was not reinforced by the level of enforcement as popularly believed.

It is interesting to note that a direct test of the alternate hypothesis that enforcement by itself does not work, or at best have only a supplementary role to advertising could be conducted using the TAC data. In the models using dichotomous variables for the campaigns, the estimated coefficients for the enforcement campaign simply captured the effect of the enforcement alone, whereas the estimated coefficients for the advertising campaigns captured the additional effects due to advertising, which were undertaken while the enforcement campaigns were in place. Tay (2002a, forthcoming) showed that the TAC enforcement campaigns by themselves were effective in reducing the number of serious crashes for both alcohol-related and speed-related crashes. Moreover, the introduction of the advertising campaigns was found to have an additional effect above the enforcement campaigns. Overall, the evidence from the TAC campaigns suggested that road safety advertising campaigns were effective and did not serve a supplementary role as popularly believed.

**Other Methodological Issues**

Obviously, the number of serious crashes per time period depends on numerous factors, many of which cannot be adequately captured in any empirical model or experiment. As a result, most models are developed with specific focuses. To focus the analysis on the impact of enforcement and publicity campaigns on the number of serious crashes, only these and a few major factors were included in the models estimated. The omission of other important variables may result in some specification bias. Consequently, the Regression Specification Test or the RESET test was conducted to check for specification errors arising from the omission of significant independent variables and non-linearity in the functional forms (Gujarati 1995). The RESET test did not find any evidence of a significant specification error in the models (Tay 2003a, forthcoming).

Newstead et al (1995) and White et al (2000) estimated a log-linear model, which is the most widely used functional form in road safety evaluations (Hakim et al, 1991). However, there is little theoretical foundation behind the choice of this specific functional form. Many other functional forms are equally plausible and a few studies that had explicitly tested the choice of functional forms did not support its choice (Zlatopher 1988; Tay 1999). Therefore, three other widely used functional forms, the linear (no log on both dependent and independent variables), semi-log (log of dependent but not independent variables) and linear-log (log of independent but not dependent variables) were also estimated as a comparison by Tay (forthcoming). Using the discriminating and discerning approaches outlined in the previous section, Tay (forthcoming) found that the semi-log model was rejected in favor of the log-linear model and the linear model was rejected in favor of the linear-log model but there was insufficient evidence to reject either the linear-log or the log-linear model using the Davidson-McKinnon J-test.

Although the OLS regression model is widely used in the road safety literature (Hakim et al, 1991), some researchers have questioned its suitability. One concern in using the standard regression model is that the number of fatal crashes may not have a Normal distribution but instead follow a Poisson distribution.
The Jarque-Bera Test of Normality (Gujarati 1995) was thus conducted by Tay (2002a, forthcoming) and the assumption of Normality was rejected in all the models estimated. Consequently, several Poisson regression models were also estimated for comparison. Since the Poisson regression model could only utilize count data, the dependent variable in its raw form was used but two models were estimated using the linear and logarithmic forms for the relevant independent variables. It should be noted that the results obtained were very similar to those obtained in the OLS models. For example, for the alcohol related crashes, the marginal effect of advertising in the linear model was estimated at -0.0058 compared to -0.0053 obtained in the OLS regression and the marginal effect of advertising in the linear-log model was estimated at -2.612 compared to -2.708 obtained in the OLS regression. These results are not surprising since OLS is still the best linear unbiased estimator even if the error terms do not have a Normal distribution. The Poisson regression model, however, is expected to produce more efficient and reliable estimates of the variance, which is crucial for hypothesis testing.

A common concern of using the Poisson regression model to analyze crash data is the possibility of over-dispersion in the data, which is checked using the t-tests proposed by Cameron and Trivedi (1990). The t-tests reject the null hypothesis of no over-dispersion for the linear model predicting alcohol-related crashes but not for the linear-log model predicting alcohol-related crashes or the linear model predicting speed-related crashes. As a result, the linear model for alcohol related crashes was re-estimated using the negative binomial model. Again, as expected, the results obtained were very close in both the Poisson and negative binomial models. For example, the marginal effects for the advertising and enforcement activities were -0.005822 and -0.0005622 respectively for the Poisson model whereas the corresponding estimates in the negative binomial model were -0.005796 and -0.0005593. These estimates were also very similar to the corresponding OLS estimates of -0.005322 and -0.0006339, attesting to the robustness and unbiased property of the OLS estimates.

Conclusion

The importance of conducting appropriate evaluations on any road safety countermeasures cannot be over-emphasised. Nevertheless, policy-makers and end-users should also be realistic in their demand on the analysts. In particular, it is important to understand that there is no perfect evaluation. In modelling the relationship between the policy variables and the performance measure, the analysts have to make many assumptions and depending on their choices, the outcome may differ considerably. It is therefore crucial that all road safety programs be evaluated and re-evaluated by different analysts imposing different assumptions and using different methodologies. It is also important for analysts to provide better conceptual explanations for their choices and explicitly test the model assumptions whenever possible. The robustness of the results can then be reviewed to provide a more complete understanding of the underlying relationship. More importantly, whenever possible, differences in the results obtained should be further examined to provide greater insight to improve our conceptual and methodological understanding of road safety evaluations.

Overall, using the data provided thus far, the results obtained for the TAC campaigns appeared to be fairly robust for models that were not explicitly rejected by the statistical tests performed. The advertising and enforcement campaigns were found to be effective in reducing the number of serious crashes in Victoria. Also, this study found no evidence that the advertising campaign played only a supportive role as popularly believed but had an independent effect in improving road safety.

References


Tay R (2002b). Exploring the effects of a road safety advertising campaign on the perceptions and intentions of the target and non-target audience to drink and drive. *Traffic Injury Prevention*


**Acknowledgement**

Support from the Motor Accident Insurance Commission is gratefully acknowledged but the views expressed by the author do not necessarily reflect those of the commission. The author also thanks Michael White for providing the data for the analysis.

**Keywords**

Evaluation Models, Advertising, Enforcement, Drink-Driving, Speeding.