ASSESSMENT OF EXPOSURE PROXIES FOR MACROSCOPIC ROAD SAFETY PREDICTION

Constantinos Antoniou and George Yannis
National Technical University of Athens
antoniou@central.ntua.gr, geyannis@central.ntua.gr

Scan this QR code to access this poster online

ABSTRACT

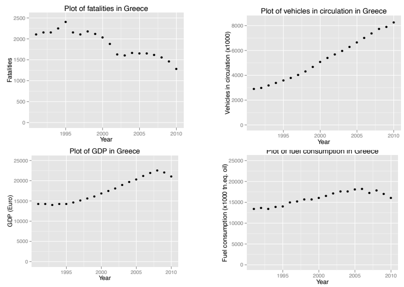
Road safety is a major global health problem and no effort should be spared in trying to limit its impacts. Modeling road safety is a complex task, which needs to consider both the quantifiable impact of specific parameters, as well as the underlying trends that cannot always be measured or observed. Macroscopic data are often not available, or not in the form that they are desired. Therefore, it is often required to attempt to consider alternative sources of data, which may be correlated with the modeled phenomenon.

The objective of this research is to investigate the suitability of alternative proxy variables for macroscopic road safety modeling, using three suitable exposure proxies: (i) number of vehicles in circulation, (ii) GDP and (iii) fuel consumption. Several structural time-series models have been developed for each proxy for two Mediterranean countries with many similar socio-economic characteristics: Greece and Cyprus.

Based on the findings of this analysis, a number of observations can be drawn. Proxy variables can provide reasonable results, when exposure data are not available. Furthermore, even in two countries with many similarities the selected proxy measure differs. This suggests that the underlying conditions that make a variable a suitable proxy for exposure is complex and needs further investigation.

Data

The fatalities in Greece show two distinct trends: an increasing one until 1995, following by a decreasing one thereafter. The number of fatalities depends strongly on a measure reflecting the amount of traffic. In Greece and Cyprus there are no traffic volume data available, so to forecast the fatalities, indirect measures such as the number of vehicles in circulation, the GDP or the fuel consumption may be used. (Data for Greece are shown - Figures of the data for Cyprus are available in the paper).



Methodology

Two structural time series models are considered in this paper: (i) the local linear trend (LLT) model and the (ii) latent risk time-series (LRT) model. Furthermore, a structured decision tree for the selection of the applicable model for each situation is outlined. Seemingly Unrelated Time-Series Equations (SUTSE), a third class of models, are also used in this approach as a preliminary step in establishing whether the two time-series may be correlated.

Decision logic

The following steps are considered:

1. Investigate exposure: Do the available exposure data make sense? Can any sudden changes in the level or slope be explained from some real events?

2. Develop a SUTSE (Seemingly Unrelated Time-Series Model) **model**: Establish whether the two series are statistically related. To achieve this, a SUTSE model is developed and based on the diagnostics, the modeler needs to decide whether the two time-series are correlated.

3. Depending on the output of the SUTSE model **determine whether an LLT or an LRT model should be pursued** (and develop it): If one or more of the null-hypotheses regarding the correlation of the disturbances (assuming the null hypotheses state that the correlations are equal to zero) is rejected, the time-series may be related and therefore an LRT can be estimated. In that case, of course, further analysis is needed, to investigate whether some of the level or slope components for the exposure and fatalities may be fixed. If, on the other hand, none of the hypotheses can be rejected, then there is no evidence that the two time-series are correlated and therefore an LLT model would be more appropriate.

Summary statistics (excerpt) of estimated SUTSE models (Greece)

The table (right) shows the SUTSE models estimated for Greece. Three models are first estimated, one for each proxy for the exposure: (i) vehicles in circulation, (ii) GPD, and (iii) fuel consumption. The beta coefficient indicates that none of these models suggest that the fatalities data and the exposure proxies are correlated for the considered time period (1991-2010). However, when one considers the trend of the fatalities time-series, two different trends appear: an increasing one until 1995 and a decreasing one thereafter.

	Veh 1991	GDP 1991	Fuel 1991	Veh 1995	GDP 1995	Fuel 1995
log likelihood	237.42	76.72	66.5714	60.9876	55.8204	39.2884
AIC	-474.53	-152.64	-132.343	-120.975	-110.641	-77.5767
Model Quality						
Box-Ljung test 1 Exposure	4.10*	0.599	0.808	0.366	3.224	0.593
Box-Ljung test 2 Exposure	4.44	0.614	1.090	0.852	3.230	1.372
Box-Ljung test 3 Exposure	4.62	0.619	1.128	0.957	3.260	1.401
Box-Ljung test 1 Fatalities	3.44	2.668	2.329	6.255*	4.671*	3.096
Box-Ljung test 2 Fatalities	4.55	3.233	2.646	11.835**	8.548*	6.410*
Box-Ljung test 3 Fatalities	6.67	3.281	2.689	13.737**	8.659*	6.501
Heteroscedasticity Test Exposure Proxy	0.238**	1.416	1.802	1.418	24.113**	1.138
Heteroscedasticity Test Fatalities	0.816	0.474	0.351	0.326	0.170	0.175
Slope (stratum 2)	7.42E-05	4.89E-04	1.12E-04	8.03E-13	1.86E-03	1.75E-03
Model H-matrix tests						
GDP Greece	5.17E-06	1.13E-09	2.42E-04	3.62E-06	1.02E-09	1.75E-09
GDP Greece	9.00E-05	1.62E-09	6.50E-05	2.69E-05	2.57E-08	2.60E-09
Correlation between fatalities and exposure						
Beta test	0.448	0.635	0.455	0.902	0.982	2.042
Significance	0.338	0.324	0.609	0.391	0.239	0.139

Therefore, three more SUTSE models were estimated, this time considering only the data from 1995 until 2010, i.e. the fatality data with the downward trend only.

Structural time-series models (Greece)

Based on the results of the SUTSE models analysis, three models are considered for Greece (for the period 1995-2010):

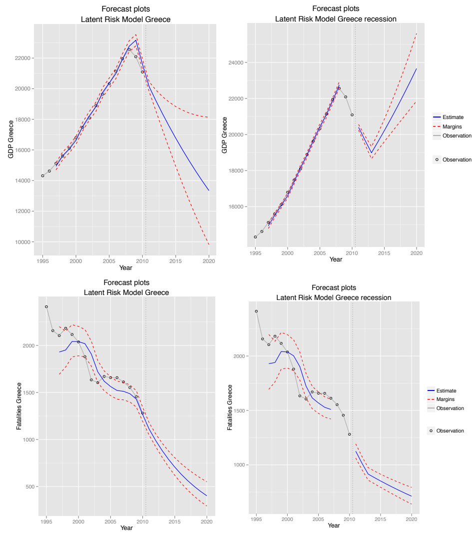
- An LLT model in which the fatalities are not assumed to be correlated with the available exposure measures
- LRT models in which the fatalities are considered to be correlated with the respective proxy to the exposure, i.e. GDP and fuel consumption.

Based on the diagnostics and predictive performance, the selected model is the restricted LRT using the GDP data.

Index	LLT	LRT - GDP Full	LRT - GDP Restricted	LRT - Fuel Full	LRT - Fuel Restricted
log likelihood	85.66	56.78	53.07	39.47	31.94
AIC	-171.20	-112.45	-105.65	-77.82	-63.39
Model Quality					
Box-Ljung test 1 GDP		3.32	3.15	1.75	0.648
Box-Ljung test 2 GDP		3.33	3.18	3.80	1.38
Box-Ljung test 3 GDP		3.35	3.20	3.97	3.19
Box-Ljung test 1 Fatalities	2.73	7.24**	7.44**	4.73*	10.17**
Box-Ljung test 2 Fatalities	3.63	14.66***	9.24**	9.91**	10.17**
Box-Ljung test 3 Fatalities	5.82	16.87***	15.83**	11.75**	11.68**
Heteroscedasticity Test GDP		22.34*	20.95*	1.36	1.28
Heteroscedasticity Test Fatalities	0.785	0.253	1.30	0.308	7.19
Normality Test standard Residuals GDP		10.77**	11.11**	0.707	0.777
Normality Test standard Residuals Fatalities	0.798	0.573	3.53	0.560	0.700
Normality Test output Aux Res GDP		1.02	0.374	0.360	0.448
Normality Test output Aux Res Fatalities	1.27	1.39	1.710	0.821	1.28
Normality Test State Aux Res Level exposure		0.13	0.169	1.18	0.802
Normality Test State Aux Res Slope exposure		8.06*	7.80*	0.233	0.420
Normality Test State Aux Res Level risk	1.61	0.929	0.699	0.795	0.621
Normality Test State Aux Res Slope risk	0.047	0.142	0.000	0.541	0.008
Model Q-matrix tests					
Level exposure		3.32E-06	-	3.59E-05	-
Level risk	3.91E-03 *	1.85E-03	-	1.91E-03	-
Slope exposure		2.4E-04*	2.4E-04*	1.1E-04*	1.5E-04*
Slope risk	1.25E-04 *	3.24E-06	-	1.34E-04	-
Transition Correlations					
Level exposure with Level risk		0.97		0.99	
Slope exposure with Slope risk		-1		-1	
Model H-matrix tests					
GDP Greece/Fuel consumption	1.00E-09	1.01E-09	2.14E-06	6.97E-09	9.62E-06
Fatalities Greece		4.75E-09	2.4E-03*	6.60E-08	4.7E-03*
Validation of predictive performance					
ME Fatalities 10	-900	-805	-263	-805	-210
MSE Fatalities 10	956930	757465	76753	757459	50784
ME Fatalities 7	-693	147	110	148	359
MSE Fatalities 7	551770	25790	15793	26023	137757
ME Fatalities 4	-131	-139	46	-139	208
MSE Fatalities 4	28162	31137	7061	30823	45604

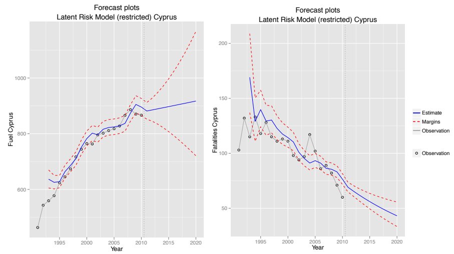
Model results and prediction validation for considered models (Greece)

The left subfigures show the results of the restricted LRT-GDP model. The projection of the GDP for Greece appears to follow a downward trend all the way to 2020. The reason for this trend is that the model detects the drop in the GDP in the last couple of years (due to the recession) but has no way to tell whether this trend will be reversed at some point. One way to overcome this would be to add an additional intervention variable to the model, which would indicate that the last few observations are part of a temporary recession phenomenon. This variable could then be used to indicate when the recession is expected to be over. Another way to indicate the same point (i.e. that these points are an intermediate disruption of an otherwise constant trend) would be to fix the slope of the exposure. However, the latter option would imply that the recovery would start from the first predicted point (i.e. 2011), which is clearly not the case. Therefore, the approach of an intervention recession variable has been selected, using 2013 as the last recession year. The (more reasonable) results of this model are showed in the right subfigures.



Forecasting results (Cyprus)

For Cyprus the best model is a restricted LRT model with fuel consumption as the proxy variable. Again, while occasionally a test is violated, there are no significant differences in the diagnostics tests across models, so the optimal model is selected mostly based on in-sample predictions. One interesting observation is that both the proxy variables considered as potentially correlated with fatalities (resulting from the SUTSE model tests), and the finally selected (from the LRT model results) are not the same for the two considered countries. Considering that the two countries have several similarities, this is an interesting finding, suggesting that the selection of proxy variables may be a very volatile process, dependent on many variables.



DISCUSSION AND CONCLUSION

Developing credible road safety forecasting models is a key prerequisite to assessing and improving future road safety. One of the key requirements (and often the weakest link) in this process is reliable and up-to-date exposure data. While some countries may have the appropriate data, e.g. vehicle-kilometers as the suitable variable for exposure, many countries and regions face limitations. One practical way to overcome this issue is to identify and use appropriate proxy variables that could be used instead of the actual exposure variables. In this research, three alternative (and in general widely available) variables are considered as suitable exposure proxies: (i) number of vehicles in circulation, (ii) GDP and (iii) fuel consumption. A number of different structural time-series models have been developed for each proxy for two Mediterranean countries with many similar socio-economic characteristics: Greece and Cyprus.

Based on the findings of this analysis, a number of observations can be drawn:

- Proxy variables can provide reasonable results, when exposure data are not available;
- Even in two countries with many similarities, such as Greece and Cyprus examined in this research, the selected proxy measure differs. This suggests that the underlying conditions that make a variable a suitable proxy for exposure is complex and needs further investigation.

The findings of this research also suggest a number of directions for future research. Beyond the obvious need for investigation of more proxy variables, as well as application in more countries and regions, a useful test would use data from a country or region that does have exposure data to compare the predictive results of models using the proxy measures versus those obtained with models directly using exposure. As the available data sample is rather small for such complicated models, it is expected that longer time-series would lead to better models. The investigation of the impact of other parameters (such as the size of the region) is also an interesting endeavor, as e.g. in smaller regions (such as Cyprus and Greece) the annual number of accidents can fluctuate a lot, compared to larger regions such as Germany or the US.

ACKNOWLEDGEMENTS

The authors are grateful to Dr. Sylvain Lassarre, Dr. Frits Bijleveld and Prof. Jacques Commandeur for their guidance and assistance throughout this research. The authors would also like to thank all the partners of the "DACOTA" project working group on time series analysis and forecasting, led by Dr. Heike Martensen and Dr. Emmanuelle Dupont, for their constructive comments and suggestions. The contribution of the road safety National Experts group of the European Commission in the data and information collection is also acknowledged.