Olubusoye, Olusanya Elisa; Korter, Grace Oluwatoyin; Salisu, Afees Adebare

Article

MODELLING ROAD TRAFFIC CRASHES USING SPATIAL AUTOREGRESSION MODEL WITH ADDITIONAL ENDOGENOUS VARIABLE

Statistics in Transition New Series

Provided in Cooperation with:
Polish Statistical Association

Suggested Citation: Olubusoye, Olusanya Elisa; Korter, Grace Oluwatoyin; Salisu, Afees Adebare (2016) : MODELLING ROAD TRAFFIC CRASHES USING SPATIAL AUTOREGRESSIVE MODEL WITH ADDITIONAL ENDOGENOUS VARIABLE, Statistics in Transition New Series, ISSN 2450-0291, Exeley, New York, Vol. 17, Iss. 4, pp. 659-670, http://dx.doi.org/10.21307/stattrans-2016-045

This Version is available at:
http://hdl.handle.net/10419/207835

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes. You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

https://creativecommons.org/licenses/by/4.0/
MODELLING ROAD TRAFFIC CRASHES USING SPATIAL AUTOREGRESSIVE MODEL WITH ADDITIONAL ENDOGENOUS VARIABLE

Olusanya Elisa Olubusoye, Grace Oluwatoyin Korter, Afees Adebare Salisu

ABSTRACT

Road traffic crashes have become a global issue of concern because of the number of deaths and injuries. The model of interest is a linear cross sectional Spatial Autoregressive (SAR) model with additional endogenous variables, exogenous variables and SAR disturbances. The focus is on RTC in Oyo state, Nigeria. The number of RTC in each LGA of the state is the dependent variable. A 33×33 weights matrix; travel density; land area and major road length of each LGA were used as exogenous variables and population was the IV. The objective is to determine the hotspots and examine whether the number of RTC cases in a given LGA is affected by the number of RTC cases of neighbouring LGAs and an instrumental variable. The hotspots include Oluyole, Ido, Akinyele, Egbeda, Atiba, Oyo East, and Ogbomosho South LGAs. The study concludes that the number of RTC in a given LGA is affected by the number of RTC in contiguous LGAs. The policy implication is that road safety and security measures must be administered simultaneously to LGAs with high concentration of RTC and their neighbours to achieve significant remedial effect.

Key words: road traffic crashes, generalized spatial two-stage least squares estimator, instrumental-variable estimation, spillover effects.

1. Introduction

Spatial models that accommodate forms of cross-unit interactions are features of interest in social sciences, biostatistics and geographic sciences. A simple spatial model augments the linear regression model by including an additional
Right Hand Side (RHS) variable known as a spatial lag and considers spill over effects either in the dependent variable or in the disturbance term. When the focus is on the dependent variable, the spillover effect is modelled by including a RHS variable known as a spatial lag. Each observation of the spatial-lag variable is a weighted average of the values of the dependent variable observed for the other cross-sectional units. The matrix containing the weights is known as the spatial-weighting matrix. This model is frequently referred to as a Spatial-Autoregressive (SAR) model.

A generalized version of this model also allows for the disturbances to be generated by a SAR process and for the exogenous RHS variables to be spatial lags of exogenous variables. The combined SAR model with SAR disturbances is referred to as SARAR model. In modelling the outcome for each unit as dependent on a weighted average of the outcomes of other units, SARAR models determine outcomes simultaneously. This simultaneity implies that the ordinary least squares estimator will not be consistent. The RHS variables are a spatial lag of the dependent variable, exogenous variables and spatial lags of the exogenous variables.

The SARAR model allows for additional endogenous RHS variables. Thus, the model of interest is a linear cross sectional SAR model with additional endogenous variables, exogenous variables and SAR disturbances.

The focus is on RTC in Oyo state, Nigeria. The number of RTC in each LGA of the state is the dependent variable. A 33×33 weights matrix; travel density; land area and major road length of each LGA were used as exogenous variables and population was the IV.

Road accidents are a global scourge characteristic of our technological era with a list of victims that grows insidiously longer day by day. Our roads which are supposed to take us places often become sources of sorrow and venues of loss. Most families have found themselves mourning, surrounded by indifference that is all too common as if this was a price or an unavoidable tribute the societies have to pay for the right to travel. The frequencies of deaths, injuries, environmental degradation, material losses and the economic impact of this seemingly unpreventable phenomenon is a global issue of concern.

Worst still, road traffic injuries cost low and middle income countries between 1 and 2 percent of their gross national product which is more than the total development aid received by these countries (WHO, 2004). In Africa, RTC are adjudged to be the second leading cause of deaths between the ages of 15 and 44. In Nigeria, RTC death rate is 162 deaths per 100,000 populations (Ogbodo and Nduoma, 2011). This is against the world average of 22 deaths per 100,000 populations (Sukhai et al., 2011).

Thus, the Nigerian RTC death rate is disproportionately high when compared to the world average by over 636 percent. Young adults’ deaths in RTC will affect the workforce of the nation. This will in turn impact negatively on economic activities and indirectly cause a devaluation of the country’s gross domestic product.
Consequent upon similar occurrences around the globe, the United Nations General Assembly have designated 2011 to 2020 as a decade of action on road safety for dedicated intervention by governments to bring down the estimated rise in deaths from RTC by 50 percent (Ki-moon, 2012).

This paper, therefore, considers spatial autoregressive dependence in RTC cases, the error term and an additional endogenous RHS variable. In other words, the objective is to determine the hotspots and examine whether the number of RTC cases in a given LGA is affected by the number of RTC cases of neighbouring LGAs and an instrumental variable. The basic underlying assumptions include the existence of spillover effects across the study area and that RTC are caused by certain actions of man, which may not necessarily be aimed at causing accidents.

Section 2 describes past works. Section 3 describes the SARAR model with an additional endogenous variable and the estimation techniques. Section 4 discusses data and results. Section 5 discusses the summary and conclusion.

2. Literature review

Road traffic injuries are a major public health problem. Leveque et al. (2001) chose Years of Potential Life Lost (YPLL) to analyse the trends during the period 1974-1994 and the relative impact of the traffic injuries death on total mortality and on total avoidable mortality in Belgium. The paper analyzed the geographical trends over a 20-year period at the district level. The geographical analysis showed marked differences between districts. Even though a favourable trend was observed for the traffic injuries and deaths in Belgium it was necessary to highlight the important slowing down of this trend during the most recent years. The study concluded it was also necessary to underline the importance of geographical disparities in the distribution of YPLL rates within the entire population.

Cirera et al. (2001) described the characteristics of Motor-Vehicle (MV) injury cases admitted to Emergency Departments (ED), and assessed factors related to injury severity and hospital admission. The subjects were MV injury patients, aged 16 or over and admitted to four EDs in the city of Barcelona (Spain), from July 1995 to June 1996. Severity was assessed with the abbreviated injury scale and the injury severity score. Univariate and bivariate descriptive statistical analyses were performed, as well as multiple logistic regressions. For the 3791 MV-injury cases included in the study period, a larger contribution of cases was noted for males (63.1%), for cases younger than 30 years (55.3%) and for motorcycle or moped occupants (47.1%). After adjusting for age, sex and the presence of multiple injuries, pedestrians, followed by moped and motorcycle occupants were at a higher risk of a more severe injury. Correspondingly, these user groups also showed a higher likelihood of a hospital admission when
attended to in an ED. Injury cases attended to in the ED during night hours were also at a higher risk of a hospital admission.

Geurts et al. (2005) compared characteristics of accidents occurring in black zones to those scattered all over the road. Identifying dangerous accident locations and profiling them in terms of accident-related data and location/environmental characteristics provided new insights into the complexity and causes of road accidents. The case study was a Belgian peri urban region. A technique of frequent item sets (data mining) was applied for automatically identifying accident circumstance that frequently occurred together for accidents located in an outside black zone. Results showed that accidents occurring in black zones were characterized by left-turns at signalized intersections, collisions with pedestrians, losing control of the vehicle (run-off-roadway) and rainy weather conditions. Accidents occurring outside black zones (scattered in space) were characterized by left turns on intersections with traffic signs, head-on collisions and drunken road users. Furthermore, parallel collisions and accidents on highways or roads with separated lanes occurring at night or during the weekend were frequently occurring accident patterns for all accident locations.

Labinjo et al. (2009) explored the epidemiology of Road Traffic Injury (RTI) in Nigeria. The focus was on populations affected and the risk factors for RTI. The RTI rates for rural and urban respondents were not significantly different. Increased risk of injury was associated with male gender among those aged 18-44 years. Simple extrapolations from the survey suggested that over 4 million people were injured and as many as 200,000 potentially killed due to this menace annually.

Aderamo (2012), in a bid to proffer measures to reduce the scourge of road traffic casualties examined the spatial variation of RTC in Nigeria. The models developed relate total number of road accidents, population estimates, length of roads and number of registered vehicles for the country. The regression results revealed that MV deaths had a positive association with population estimate and length of roads. Also, population estimate and length of roads had a significant effect on MV injuries.

In order to enhance prioritization of the burden of RTC, Chandran et al. (2013) calculated Years of Life Lost (YLL) and reduction in life expectancy using population and crash data from Brazil’s ministries of health and transport. The potential for reduction in crash mortality was calculated for hypothetical scenarios reducing death rates to those of the best performing region and age category. For males and females at birth, RTC reduced the life expectancy by 0.8 and 0.2 years respectively. The study concluded that many YLL for men and women could be averted if all rates matched those of the lowest-risk region and age category.

To reduce RTC casualties and improve safety and security on roads, usually, highway improvement project selection requires screening thousands of road segments with respect to RTC for further analysis and final selection into improvement projects. Kelle et al. (2013) described a two-step procedure for selecting potential RTC locations for inclusion in highway improvement projects.
The first step of the proposed methodology used odds against observing a given crash count, injury count and run-off road count as measures of risk and a multi-criteria pre-selection technique with the objective of decreasing the number of prospective improvement locations. The second step was based on a composite efficiency measure of estimated cost, benefit and hazard assessment under budget constraint. To demonstrate the two-step methodology, the study analyzed 4 years of accident data at 23000 potential locations, from which final projects were selected.

With a special focus on truck drivers, Rancourt et al. (2013) developed different scheduling algorithms embedded within a tabu search heuristic. This was in attempt to improve safety amongst long haul carriers in vehicle routing and scheduling in the United States. The methods and results confirmed the benefits of using a sophisticated scheduling procedure for long haul transportation.

Metaheuristic algorithms, such as simulated annealing and tabu search are popular solution techniques for vehicle routing problems (VRPs). Harwood et al. (2013) focussed on single VRP similar to travelling salesman problem, and investigated the potential for using estimation methods on simple models with time-invariant costs, mimicking the effects of road congestion. Working with standard VRPs, where the costs of the arcs did not vary with advancing time, changes to the total cost were evaluated following a neighbourhood move simple process. In the cases where a time-varying aspect such as congestion was included in the costs, the calculations became estimations rather than exact values.

3. SARAR model with additional endogenous variables and estimation techniques

\[ y = Y\pi + X\beta + \lambda W_1y + u \]  
\[ u = \rho W_2u + \varepsilon \]  

where \( y \) is an \( n \times 1 \) vector of observations on the dependent variable; \( Y \) is an \( n \times p \) matrix of observations on \( p \) RHS endogenous variables; and \( \pi \) is the corresponding \( p \times 1 \) parameter vector; \( X \) is an \( n \times k \) matrix of observations on \( k \) RHS exogenous variables (where some of the variables may be spatial lags of exogenous variables), and \( \beta \) is the corresponding \( p \times 1 \) parameter vector; \( W_1 \) and \( W_2 \) are \( n \times n \) spatial weighting matrices (with 0 diagonal elements); \( W_1y \) and \( W_2u \) are \( n \times 1 \) vectors typically referred to as spatial lags, and \( \lambda \) and \( \rho \) are the corresponding scalar parameters typically referred to as SAR parameters and \( \varepsilon \) is an \( n \times 1 \) vector of innovations.

The model in equations 1 and 2 is a SARAR model with exogenous regressors and additional endogenous regressors. Spatial interactions are modelled through
spatial lags, and the model allows for spatial interactions in the dependent variable, the exogenous variables and the disturbances.

The model in equations 1 and 2 is a first-order SAR process with first order SAR disturbances, also referred to as a SARAR(1,1) model, which is a special case of the more general SARAR \((p, q)\) model. We refer to a SARAR \((1, 1)\) model as a SARAR model. Setting \(\rho = 0\) yields the SAR model \(y = Y\pi + X\beta + \lambda W_1 y + \epsilon\). Setting \(\lambda = 0\) yields the model \(y = Y\pi + X\beta + u\) with \(u = \rho W_2 u + \epsilon\), which is sometimes referred to as the SAR error model. Setting \(\rho = 0\) and \(\lambda = 0\) causes the model to reduce to a linear regression model with endogenous variables.

The spatial-weighting matrices \(W_1\) and \(W_2\) are taken to be known and nonstochastic. These matrices are part of the model definition, and in many applications, \(W_1 = W_2\). Let \(\bar{y} = W_1 y\), let \(\bar{y}_i\) and \(y_i\) denote the \(i\)th element of \(\bar{y}\) and \(y\), respectively, and let \(w_{ij}\) denote the \((i, j)\)th element of \(W_1\). Then,

\[
\bar{y}_i = \sum_{j=1}^{n} w_{ij} y_j.
\]

The weights \(w_{ij}\) will typically be modelled as inversely related to some measure of distance between the units. The SAR parameter \(\lambda\) measures the extent of these interactions.

The innovations \(\epsilon\) are assumed to be independent and identically distributed or independent but heteroskedastically distributed, Drukker et al. (2013).

The Generalized Spatial Two Stage Least Squares (GS2SLS) estimation technique was used.

### 4. Analysis and discussion of results

#### 4.1. Data description

In February 1988, the Federal Government created the Federal Road Safety Commission (FRSC) through Decree No. 45 of 1988 as amended by Decree 35 of 1992 referred to in the statute books as the FRSC Act cap 141 Laws of the Federation of Nigeria (LFN). Prior to this time, the Nigerian Police Force established in 1930 was saddled with the responsibility of keeping RTC records. In most instances, especially when the form of RTC was not serious, such events were not usually reported. Thus, it is possible that this data does not contain all the RTC that occurred within the study time and area. Nonetheless, data for this study are reliable and despite possible omissions the findings will not be negatively influenced.

Data on number of RTC and traffic volume was collected from the RS11.3 FRSC Oyo sector command. This research enjoyed the good cooperation of the
FRSC, which is a good potential to obtain quality results. The study focussed on area of land, total length of major roads, travel densities and the residential population for every LGA.

The approximate locations of the road crashes were estimated from the records obtained from the RS 11.3 Oyo FRSC command. The records provided an indicative description of the locality where the RTC occurred and in some cases the site was described using nearest landmarks such as filling stations, roundabouts, stores, markets, garages, institutions or road intersections. Therefore, using the locations and/or nearest landmarks provided in the records together with the Google Earth image, the existing digital road networks and knowledge of the area of study, it was easy to place points on the approximate locations where RTC occurred.

The Google Earth image was particularly helpful because it provided a photographic view of the area of study together with the associated landmarks and road networks. Through the instrumentality of Global Positioning System (GPS) and the information on locations of RTC for each of the unit commands, the coordinates of each RTC location were obtained on the geographic coordinates system of the world.

The coordinate locations generated were subsequently exported into ArcGIS and were plotted as point locations. The point locations, which represent RTC locations, were overlaid on the road network of the LGAs of Oyo State. This made it possible to clip the Oyo state geo-referenced map within the ArcGIS environment to ascertain where each RTC location falls. Therefore, the number of RTC points within each LGA was counted and recorded accordingly.

The Geographic Information System (GIS) was used to create polygon shapefile for the study area. The shapefile was used to create spatial contiguity weight matrix based on the queen criteria. The lengths of the roads and the area encompassing each LGA were calculated using the measuring tool in the software environment.

Population was defined as the population figures for each LGA as reported by the National Bureau of Statistics for the 2006 population census in Nigeria. Travel density was defined as traffic count for each FRSC unit command divided by total major road length in kilometres within each LGA. Major road length was the total length of roads in kilometres from each settlement to another within each LGA. Area was defined as the area per square kilometre encompassing each Local Government Authority. Accident was defined as the total cases of RTC recorded in each LGA whose location was identified by taking the longitudes and latitudes. The logarithms of all the variables were taken for the purpose of this study.

4.2. RESULTS

Figure 1 shows the distribution of RTC across LGAs, with darker colours (blue) representing higher values of the dependent variable. Spatial patterns of RTC are clearly visible. The hotspots for RTC are within Oluyole, Ido, Akinyele,
Egbeda, Atiba, Oyo East, and Ogbomosho South LGAs. This is followed by Ibadan North, Ibadan North West, Ibadan South East, Ibadan South West, Lagelu, Afijio, Oyo West, Ori-Ire and Ogbomosho North LGAs. Each of these categories of LGAs happens to be neighbours.

The content of the spatial-weighting matrix is summarized in Table 1. Some basic information about the normalized contiguity matrix, including the dimensions of the matrix and its storage is displayed. The number of neighbours found is reported as 156, with each LGA having 5 neighbours on average. Each LGA has a minimum of 2 neighbours and a maximum of 8 neighbours. This is an indication that no LGA is isolated, thus spatial dependency is bound to result based on neighbourhood characteristics. The frequency of RTC in a LGA is dependent on the frequency of RTC in the contiguous LGAs. Therefore, to achieve maximum remedial effect when administering road safety remedial measures to a LGA, it is significant to take cognizance of the neighbours of that particular LGA.

Figure 1. Number of Road Traffic Crashes cases for Local Government Areas in Oyo State, Nigeria
Table 1. Summary of Spatial Contiguity Weighting Matrix

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensions</td>
<td>33 x 33</td>
</tr>
<tr>
<td>Stored as Links</td>
<td>33 x 33</td>
</tr>
<tr>
<td>Total</td>
<td>156</td>
</tr>
<tr>
<td>Min</td>
<td>2</td>
</tr>
<tr>
<td>Mean</td>
<td>4.727273</td>
</tr>
<tr>
<td>Max</td>
<td>8</td>
</tr>
</tbody>
</table>

The GS2SLS parameter estimates for the spatial-autoregressive model with spatial-autoregressive disturbances (SARAR) with additional endogenous variable are as shown in Table 2.

Table 2. SARAR Model with Additional Endogenous Variable - GS2SLS

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>P &gt;</th>
<th>Z</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>-0.417</td>
<td>0.85</td>
<td>0.624</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major Road Lengths</td>
<td>-0.104</td>
<td>0.54</td>
<td>0.849</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Travel Density</td>
<td>0.009</td>
<td>0.18</td>
<td>0.962</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area</td>
<td>-0.175</td>
<td>0.34</td>
<td>0.606</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lambda( ( \lambda ) )</td>
<td>1.205</td>
<td>0.37</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rho ( ( \rho ) )</td>
<td>-1.179</td>
<td>0.89</td>
<td>0.185</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Given the normalization of the spatial-weighting matrix, the parameter space for \( \lambda \) and \( \rho \) is taken to be the interval \(-1\) and \(1\) (Kelejian and Prucha, 2010).

The estimate of \( \lambda \) is positive, large, and significant, indicating strong SAR dependence in RTC cases. In other words, the RTC frequencies for a given LGA is strongly affected by the RTC frequencies in the neighbouring LGAs. One possible explanation for this result may be the existence of freeways linking LGAs. Another may be the result of high travel density on major road networks of particular LGAs. Apparently, there is spillover effect of frequencies of RTC across the neighbourhood.

The estimated \( \rho \) is negative and moderate, indicating moderate spatial autocorrelation in the innovations. An exogenous shock to one LGA will cause moderate changes in RTC rate for neighbouring LGAs. For example, an addition of a new freeway linking a LGA will cause moderate changes in RTC rate for neighbouring LGAs. However, maximum road safety remedial effect will only be achieved through new freeways linking all contiguous LGAs simultaneously.

The estimated \( \beta \) vector does not have the same interpretation as in a simple linear model, because including a spatial lag of the dependent variable implies that the outcomes are determined simultaneously. The parameter estimates for population, major road lengths, travel densities and area are not significant. This means that each of the exogenous variables do not contribute significantly to the
incidence of RTC in Oyo state. The signs of the coefficients suggest the following:

Controlling for spatial lag and the LGAs, population is negatively related to the number of RTC occurring within the localities. All other things being equal, LGAs with larger residential populations tend to have fewer RTC. The coefficient indicates the expected number of RTC for every person living in the LGA. One percent increase in population will generate 0.42 percent decrease in the number of RTC cases within each LGA. The high number of people could be the result of high economic activities in the state, which may lead to high traffic density that will slow traffic and thus reduce the frequency of RTC.

Major road lengths characteristics produce a negative coefficient. This means the existence of a freeway link crossing a LGA inversely impact the incidence of RTC cases for the period. One percent increase in major road length will generate 0.10 percent decrease in the number of RTC cases within each LGA. Thus, an increase in major road lengths will generate a decrease in the number of RTC cases within each LGA. Invariably, the existence of more freeways linking LGAs will reduce travel densities, which may in turn reduce RTC. On the other hand, low travel densities could give rise to opportunities for increased speed by drivers which may result in a higher number of RTC.

Travel densities are positively related to the number of accidents. An increase in travel densities will lead to an increase in the number of RTC cases. This suggests that traffic generated tend to be associated with larger crashes. The high intensity of vehicles in urban areas as a result of economic activities and the existence of few major road networks in rural areas could be responsible for this result. One percent increase in travel densities will generate 0.01 percent increase in the number of accident cases.

The sign of the coefficient for area is negative. There are fewer RTC cases in larger LGAs. Every decrease in area per square kilometre encompassing a LGA will lead to an increase in the number of accident cases. This means an increase in the area of administration of LGAs will lead to less RTC cases in each LGA. Every one percent decrease in area per square kilometre encompassing a LGA will generate 0.18 percent increase in the number of RTC cases. This suggests that several factors other than location are responsible for RTC occurrences across the study area.

5. Summary and conclusion

The study concludes that the number of road traffic crashes in a given Local Government Area is affected by the number of road traffic crashes for neighbouring Local Government Areas. Thus, the policy implication of our result is that safety and security measures must be administered within Local Government Areas that have high concentration of road traffic crashes along with
their neighbouring Local Government Areas in order to achieve significant remedial effect.

Although this study examined a few possible exogenous variables for its investigations, improvements could be made by using more precisely measured variables. Also, the significance of more trip generators, such as land use activities, generators of economic activities and different employment activities could be the focus of future works. Studies revealing the causes of accidents such as speed violation, dangerous driving, driving under alcohol influence amongst several other factors will be useful for the implementation of safety and security measures. To enhance succinct investigation into road traffic crashes characteristics, data for more years should generally be used for accident modelling in the theoretical framework of spatial panels.

Additionally, future studies should examine post estimation issues such as how a change in one exogenous variable potentially changes the predicted values for all the observations of the dependent variable in Local Government Areas across the State. Finally, to reduce road traffic accidents, improve safety and security measures and achieve maximum remedial effect policy designs and decision-making on transportation and road networks construction/maintenance should incorporate spillover effects and take cognizance of exogenous variables across each Local Government Area in the state.

**REFERENCES**


