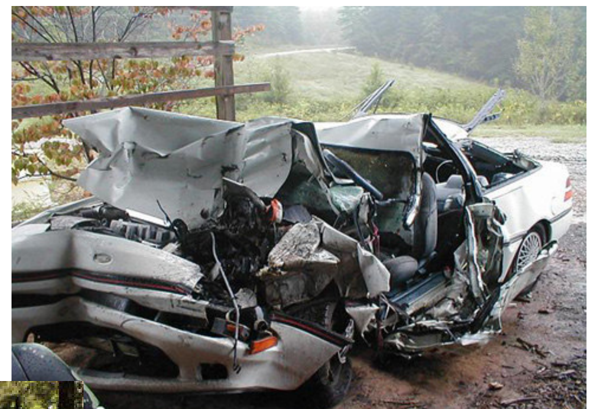
Consideration of socioeconomic and transportation related variables in fatal crash frequency analysis



Prepared by

Jonathan Aguero

For

Course ECON 490 Introductory Econometrics

Instructor: Li Wang

June 28, 2004

Introduction

According to a World Health Organization/World Bank report "*The Global Burden of Disease"*(Murray and Lopez, 1996), deaths from non-communicable diseases are expected to climb 77% from 1990 to 2020 (from 28.1 million to 49.7 million) and traffic accidents are the main cause of this rise. Road traffic injuries are expected to take third place in the rank order of disease burden by the year 2020.

Another study of the World Health Organization, the publication *"Injury: A Leading Cause of the Global Burden of Disease"* (1999), reports that the leading injury-related cause of death among people aged 15 to 44 years is traffic injuries. Of the 5.8 million people who died of injuries in 1998, 1,170,694 died as a direct result of injuries sustained in a motor vehicle accident.

It is clear that deaths due to car crashes is one of the biggest problems on public health around the world and developed countries are not an exception. Just in 2002, 42,815 persons died in 38,309 crashes in USA (FARS, 2004), 1614 of them in the state of Pennsylvania.

Although car accidents are by nature determined by individuals (e.g. the driver, the car, the intersection), it is practically impossible to study, at individual level, the influence of spatially defined factors like land use, demographic characteristics, and traffic volume, among others. In most of the car accident studies, the data are grouped in spatial units that range from intersection or road section level to zip code or county level.

The spatial units of study might be selected as the smallest physical unit that can represent the potential causes or risk factors of traffic accidents. In this case the unit selected was counties, because it was not feasible to get important socioeconomic and transportation related data, which have being used in the past as risk factors, at smaller spatial units (e.g. census tracks). The study area is the state of Pennsylvania.

In this study the risk factors were divided into two main categories: socioeconomic and transportation related factors. The former category attempts to describe the user characteristics while the later one seeks to explain the system characteristics. User characteristics involve drivers and other vehicle occupants, as well as pedestrians and other users of the transportation facilities. On the other hand, transportation related factors involve the characteristics of the transportation infrastructure.

Among socioeconomic factors, the following were included in this paper:

- *Percent of persons under 16.*
- *Percent of persons between 16 and 24.*
- *Percent of persons over 65.*
- *Percent of males.*
- *Percent of urban population.*
- *Percent of population under poverty.*
- *Driving under Influence (DUI) arrests.*

The transportation related factors studied are:
- *Daily Vehicle-Miles Traveled.*
- *Miles of roads by functional classification.*
- *Mean travel time.*

Different studies, including Doherty *et al.* (1998) and Kam (2003), have shown that young and old drivers have higher risk of car accidents, as well as higher risk for male drivers against females. Therefore, the variables percent of persons between 16 and 24, percent of persons over 65, and percent of males by county were included in the analysis.

Younger population groups are often associated with high risk of road accidents (Noland and Qudduss, 2004). The higher risk is associated with the lack of awareness about the danger of the roads and the condition of pedestrian.

The percent of urban population measures the urbanization of the county and therefore it is an indicator of land use intensity. Higher intensities are normally associated with higher accident risk. The results of Noland and Qudduss (2004) showed that land use and area deprivation are associated with car accidents. In this study the percent of persons living under poverty is used as an indicator of area deprivation.

Another important socioeconomic factor is the participation of drivers under the influence of alcohol or other drugs in car crashes. Gary et al (2003) found that in 1998 dry counties in Kentucky (those were alcohol sell is prohibited) had fewer alcohol related traffic crashes and fewer driving under influence (DUI) arrests per 1000 licensed drivers. For this paper the number of DUI arrests was used to related alcohol consumption and crashes frequency.

In transportation-related factors, vehicle-miles traveled is often used as exposure indicator (Miaou et al, 2003) along with the number of miles of different functional classes per county (Noland and Qudduss, 2004). In addition to these factors the mean travel time to work was tested. The hypothesis is that higher travel times to work results in higher exposure and therefore higher risk.

The purpose of this paper is to test if the socioeconomic and transportation related factors mentioned above are significant predictors of fatal crashes for the counties of the state of Pennsylvania in 2002. The paper is organized as follows: the next section describes the sources and nature of the data analyzed in this study, followed by the presentation of the regression results, and finally some conclusions.

Data Description

The variables employed in this work have already being mentioned in the introduction. The sample size is 67 and the study year is 2002, which is the most updated year of data for most of the variables. The notable exception is the data from the US. Census Bureau; for these data the year is 2000, which is the year of the last census. Table 1 shows the final set of variables used as well as their corresponding descriptions. Natural logarithm transformation was performed for the dependent variable in order to normalize it. Several independent variables were also transformed using the natural logarithm.

Table 1 Description of variables.

Dependent Variable	
Lcrashes	Ln of total fatal crashes in 2002
Socioeconomic Variables	
P_pov	Percent of population under poverty in 2000
P16	Percent of population under 16 in 2000
P16_24	Percent of population between 16 and 24 in 2000
P65	Percent of population over 65
Pmales	Percent of males in 2000
P_urban	Percent of urban population in 2000
LDUI	Ln of Driving Under Influence Arrests in 2002
Transportation Related Variables	
LDVMT	Ln of Daily Vehicle-Miles Traveled in 2002
Lfed_aid	Ln of miles of federal aid roads in 2002
Lnonfed_aid	Ln of miles of non-federal aid roads in 2002
Ltravel_t	Ln of Mean travel time to work (minutes), workers age 16+, 2000
Ltotal	Ln of miles of roads (federal and non-federal aid) in 2002
Pfed_aid	Percent of miles of federal aid roads in 2002

The sources of data are:

- Total number of Fatal Crashes by county, 2002: Fatality Analysis Reporting Systems (FARS) Web-Based Encyclopedia, 2004.
- VMT and miles of different types of roads by county 2002: Pennsylvania Department of Transportation Web Page, 2004.
- Population, poverty, and mean travel time by county 2000: US. Census Bureau Web Page, 2004.
- Driver Under Influence (DUI) arrests by county 2002: Pennsylvania State Police, Uniform Crime Reporting System Web Page, 2004

Table 2 shows the summary statistics of variables included in the model.

Table 2 Summary Statistics of variables included in the model

Variables	Mean	Median	Max	Min	S.D.
Dependent Variable					
Lcrashes	2.726	2.639	4.718	0.000	0.898
Socioeconomic Variables					
P_pov	9.775	9.700	18.500	4.400	2.646
P16	19.749	19.914	23.221	15.244	1.679
P16_24	12.466	11.492	27.548	10.188	2.751
P65	16.183	16.066	22.231	10.541	2.227
Pmales	49.103	48.827	55.706	46.538	1.357
P_urban	50.932	52.446	100.000	0.000	27.132
LDUI	5.770	5.746	8.261	2.890	1.203
Transportation Related Variables					
LDVMT	14.805	14.645	17.052	12.218	0.990
Lfed_aid	5.840	5.844	7.363	4.332	0.599
Lnonfed_aid	7.091	7.131	8.328	5.434	0.567
Ltravel_t	3.174	3.153	3.829	2.734	0.181
Ltotal	7.347	7.433	8.651	5.744	0.565
Pfed_aid	22.510	21.999	35.457	13.842	4.088

Results

Table 3 presents the results of the different models. The first model presents the general model testing all the variables of interest. Subsequent models present the variables that resulted significant in the first model or alternative variables being tested.

From the first model it is apparent that socioeconomic variables related to age, sex and urban percent are not significant. It is possible that these variables present high collinearity among them and therefore they were jointly significant but not individually significant. A test of joint significance on these variables was performed to test the later hypothesis and they were found statistically jointly insignificant at 5% level ($F= 0.95$, d.f. = 5,55 p-value = 0.45). As a result of this test, the variables in mention were removed from the subsequent analysis.

In Model 2 the variables describing road miles by type of funding (i.e. federal and non federal) *Lfed_aid* and *Lnonfed_aid* were found individually insignificant at 5% level (see table 3 for details). Here the hypothesis is that roads under Federal Aid have higher design standards and therefore they are expected to be safer. Once again, a test of jointly significance was performed to discard possible collinearity among the variables. The results of the test ($F= 2.80$, d.f. = 2,60 p-value = 0.069) showed that the variables are jointly significant at least at 10% level, which given the small sample size is a good indicator.

Table 3 Linear models for natural log of fatal crashes (sample size = 67)

Variable		MODEL 1	MODEL 2	MODEL 3	MODEL 4
	Estimate / *S.E* / p-value	Estimate	Estimate	Estimate	Estimate
Intercept	Estimate	**-12.2579**	**-9.6169**	**-9.5382**	**-9.6221**
	S.E	*4.7518*	*1.4727*	*1.4641*	*1.4311*
	p-value	0.0126	0.0000	0.0000	0.0000
P_pov	Estimate	**0.0564**	**0.0430**	**0.0423**	**0.0412**
	S.E	*0.0198*	*0.0174*	*0.0173*	*0.0168*
	p-value	0.0062	0.0163	0.0178	0.0177
P16	Estimate	**0.0165**			
	S.E	*0.0601*			
	p-value	0.7851			
P16_24	Estimate	**-0.0095**			
	S.E	*0.0352*			
	p-value	0.7874			
P65	Estimate	**-0.0250**			
	S.E	*0.0485*			
	p-value	0.6072			
Pmales	Estimate	**0.0595**			
	S.E	*0.0433*			
	p-value	0.1753			
P_urban	Estimate	**0.0020**			
	S.E	*0.0044*			
	p-value	0.6545			
LDUI	Estimate	**0.1884**	**0.2473**	**0.2420**	**0.2325**
	S.E	*0.1272*	*0.1064*	*0.1058*	*0.1011*
	p-value	0.1443	0.0238	0.0261	0.0253
LDVMT	Estimate	**0.4555**	**0.4547**	**0.4481**	**0.4376**
	S.E	*0.1701*	*0.1374*	*0.1373*	*0.1326*
	p-value	0.0097	0.0017	0.0019	0.0017
Lfed_aid	Estimate	**0.0717**	**-0.0459**		
	S.E	*0.2758*	*0.2446*		
	p-value	0.7958	0.8516		
Lnonfed_aid	Estimate	**0.2669**	**0.3132**		
	S.E	*0.2195*	*0.1797*		
	p-value	0.2292	0.0869		
Ltravel_t	Estimate	**0.4628**	**0.5705**	**0.5686**	**0.5643**
	S.E	*0.3338*	*0.2239*	*0.2241*	*0.2221*
	p-value	0.1713	0.0137	0.0140	0.0139
Ltotal	Estimate			**0.2869**	**0.3177**
	S.E			*0.1647*	*0.1344*
	p-value			0.0871	0.0216
Pfed_aid	Estimate			**-0.4084**	
	S.E			*1.2435*	
	p-value			0.7442	
R^2		**0.895**	**0.886**	**0.886**	**0.886**
Ajusted R^2		**0.874**	**0.874**	**0.874**	**0.876**

In Model 3 an alternative approach was tested in order to determinate if the functional classification is significant. Instead of using the miles of road by type of funding, total miles of roads (*Ltotal*) and percent of federal Aid roads (*Pfed_aid*) were used. Once more, the variables were found individually insignificant at 5% level. A jointly significance test was performed with virtually he same results as the last test (F= 2.81, d.f. = 2,60 p-value = 0.068). Like in Model 2, the variables representing the amount of roads by type (in this case *Ltotal* and *Pfed_aid*) seem to have high collinearity between them.

Given the shortcomings of Models 2 and 3 an alternative Model 4 was performed using just the total miles of road *Ltotal*. The coefficients for the 5 variables in the model were all highly significant at 5 percent level even for one sided tests. Therefore, Model 5 was selected as the definitive model for the project.

Before discussing Model 4 in detail, it is important to remark the fact that although coefficients for miles of road by type of funding in Model 2 and coefficients for total miles of roads and percent of federal Aid roads in Model 3 were found not individually significant, their sign and values present the expected tendency: higher the risk for non-federal aid roads that have lower design standards.

For Model 4 the interpretation of the coefficients is easy using the fact that the elasticity can be approximated as the $\Delta \log y / \Delta \log x$. Therefore, the coefficients of all the variables represent the elasticities of *number of fatal crashes* with respect to the variables. For this model the variables with higher elasticities are the transportation related variables with 0.564, 0.438, and 0.318 for travel time, DVMT, and total miles of roads respectively. For the socioeconomic variables the elasticities are 0.233 and 0.0412 for DUI arrests and poverty respectively.

Conclusions

In general, it was not found that socioeconomic characteristics related to age, sex, and intensity of land use (percent of urban area) mattered in terms of affecting fatalities. However, this result must be taken carefully given the small sample size. Clearly, bigger sample sizes using cross-sectional data or panel data is necessary to draw better conclusions about these variables.

The socioeconomic variables that presented high significance level are percent of persons under poverty and number of DUI arrests. The finding that counties with higher levels of poverty have relatively higher fatal accident frequencies is also consistent with other findings (Noland and Qudduss, 2004). Additional work should be conducted to identify the characteristics of these areas that may adversely affect road safety. On the other hand, the number of DUI arrests was found positively related to fatal crashes, which is also consistent with previews work (Gary et al, 2003). Here the causal effect is clearest than in the former case: several studies have shown the positive relationship between alcohol and other drugs consumption and increased car accident risk.

Although the confidence levels are relatively low (approximately 94%), the results suggest that the risk of fatal crashes decrease when the miles of Federal Aid roads increases. The equivalent result of decrease of risk of fatal crashes with increase of percent of Federal Aid roads was also found. Again, bigger sample sizes using cross-sectional data or panel data will be better to establish more conclusive statements about these variables.

Using the elasticity as a measure of relevance, the most important variable resulted to be the mean travel time to work following by Daily Vehicle-Miles Traveled, and Total miles of roads. This result suggests that the risk is higher for an increment of travel time than for an increment of travel distance supposing the same proportion of increment in both.

References

Doherty, S.T., Andrey, J.C., MacGregor, C., 1998. The situational risks of young drivers: the influence of passengers, time of the day and day of the week on accident rates. Accident Analysis and Prevention, Vol. 30

FARS Fatal Accident Report System Web-Based Encyclopedia, 2004. Web page http://www-fars.nhtsa.dot.gov.

Gary, S.L.S., Aultman-Hall, L., McCourt, M., Stamatiadis, N., 2003. Consideration of driver home county prohibition and alcohol related vehicle crashes. Accident Analysis and Prevention, Vol. 35.

Kam, B. H., 2003. A disaggregate approach to crash rate analysis. Accident Analysis and Prevention, vol. 35.

Miaou, S., Song, J.J., Mallick, B.K., 2003. Roadway Traffic Crash Mapping: A Space-Time Modeling Approach. Journal of Transportation and Statistics, Vol. 6, No 1.

Murray, C., Lopez, A. (editors), 1996. The Global Burden of Disease. Harvard University Press.

Noland, R.B., Qudduss, M.A., 2004. A spatially disaggregate analysis of road casualties in England. Accident Analysis and Prevention, Article in press.

Pennsylvania Department of Transportation, 2004. Web Page: http://www.dot.state.pa.us/

Pennsylvania State Police, 2004. Uniform Crime Reporting System Web Page: http://ucr.psp.state.pa.us/

United States Census Bureau, 2004. Web Page: http://www.census.gov/

WHO: World Health Organization, 1999. Injury: A Leading Cause of the Global Burden of Disease. Web page http://www.who.int/violence_injury_prevention/injury/gbi/gbi8/en/index1.html