Abstract.

Deaths caused by car accidents vary across countries and it is important and interesting to find out what might be the reason for accident rates to differ cross-sectional. This paper is to examine blood alcohol content(bac), percentage of people leaving in urban areas, percentage of population that is in the labor force and my last variable is a categorical dummy variable if a country is a high or low/medium income country. The data have been collected for 2013 year from different sources and for 121 countries around the world. I treat I have collected data from different sources online: the World Bank, The Census Bureau and mostly from the 2013 road safety report. Regression analysis have been performed to obtain the results.

Results.

Although bac was expected to have the biggest influence on traffic roads because one would think that if a country has a higher alcohol content allowance then it would have higher accident rates. My results show that bac is actually not very significant and with this data it does not influence the accident rates. Regression analysis show that the higher the population that lives in urban areas the smaller the probability of getting into an accident. Also if country is a high income country, chances of dying on the road are smaller, probably mostly due to better infrastructure and better constructed and respected laws than in low income countries. Also in countries that have higher youth population number, the risk of fatality accident is much greater then in countries where there is relatively smaller number of drivers.

Introduction.

Injures that are caused by road traffic are among leading ones globally and might become fifth one by 2030[2]. FiA foundation has proposed in 2011 the United Nations

Decade of Action for Road Safety 2011-2020. They partner with the World Health Organization, the UN Environment Programme, The World Bank and the Federation Interationale de I ‘Automobile to through multiple activities and campaign promote safer environment on the road.[4]Moreover road traffic deaths are the very first leading cause of death for population aged 15-29 with around 320,000 deaths in 2012. Suicide is the second leading cause with 250,000 deaths[3]. What the report on the road deaths found is that road traffic death rates are twice higher in low and middle class countries than they are in high-income countries. [3] In the paper countries are also examined based on the level of income for a particular country. Looking more into the US particularly I found that approximately 35,000 deaths were due to road accidents in 2015 in the US. It is slightly more than in Europe but taking into account that there have been increasing numbers of cars on the road the fatality rate per 100million vehicles-miles traveled is about 2 percent in 2015 which went down from 25 in 1921.[8]

Findings and concerns above motivated me to use regression analysis and data to discover the influence of particular variables on the road accident rates within the countries that are reported in the Road Safety Report.

Literature review:

Almost 90% of road related deaths occur in low-income economies and middle-income economies. They nevertheless bear a top-heavy number of deaths relative to their level of motorization, as they account for only 54% of the world’s registered vehicles. [3] It is important to notice the percentage of the population that lives in urban areas since according to the report almost half of the deaths are due to carless behavior of cyclist and pedestrians.[3] Also it is important to notice that young population is the most affected, especially in poor countries like Zimbabwe which are "disproportionately affected by road traffic injuries and fatalities, which place undue pressure on inadequate health systems and on victims and their families, but also on society, since half of those who die or are disabled are young adults, its most productive segment” [6]. Moreover, the more cars we have on the road the bigger the probability of getting into an accident. Research showed that

16 percent of fatal accidents happened between 3pm and 6pm and freeways when people were driving back from work and it was busy time of the day.[5]

Empirical Model and Data.

Accidents have been a cause of death for many people and from theoretical point of view a lot of variables can cause an accident rate to be higher in some countries then in the other. In my primary regression I used variables that I think are theoretically important: blood alcohol content (bac) because the higher the blood alcohol content the bigger risk of getting into an accident since even one glass of wine can affect your vision and your reaction on the road[8]. The expected sign of the coefficient bac is positive because higher blood alcohol content should increase accident rates. In addition to bac, percentage of the population that lives in urban areas and percentage of the population that is in labor force should be taken into account. I expected both to have a positive sign. People commute every day to work, school, gym, mall etc. I also came up with variable that takes into account young population. This involves using more cars. In addition to that young people do not have as much experience like old drivers do and very often try to show off or get distracted easily. Distracted driving is also the leading cause of accidents in the United States [9]. The more young people in the population the higher the risk of accidents. Moreover, I categorized countries based on low/medium income per capita and high incomer per capita. Countries with higher income have better infrastructure and better quality of life which should decrease interest rates.

Primary regression(Refer to Table.1) :

$$TRD= \beta_0 + \beta_1 URPOP_i + \beta_2 YOUTHPOP + \beta_3 LFOR + \beta_4\ BAC\_ - \beta_4\ INCOME \quad (1.1)$$

$$\underset{t=}{TRD=} .028 - \underset{\substack{(.02)\\-1.77}}{.052URPOP_i} + \underset{\substack{(.23)\\2.89}}{.657YOUTHPOP_i} + \underset{\substack{(.06)\\2.66}}{.158LFOR_i} + \underset{\substack{(.19)\\.30}}{.056BAC_i} \underset{\substack{(1.77)\\-4.26}}{-7.51INCOME_i}$$

Estimation Results.

For the estimated primary regression from above the sample size is 119 countries, adjusted R^2=.6005 which is relatively high. It means that 60 percent variation in dependent variable 'TRD' can be explained by explanatory variables blood alcohol content per country, urban population, percentage of the total population that is 15-24 years old, percentage of total population that is in labor force and high vs. low income countries.

Hypothesis testing for the overall significance of the regression:

$H_0 = \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 = 0$

$H_A = H_0$ is not true

The F- stat test for this regression is 12.80 and the probability of F exceeding is .0005 which tells us that the we can reject the null hypothesis at the 5% significance level and declare that the equation is statistically significant at 5 percent level of significance.

The slope coefficient value, 0.052, for the explanatory variable, URPOP, means that a 1 percentage-point increase in the urban population is associated with a

0.052 people per thousand people decrease in the traffic death rate per thousand people, ceteris paribus.

The slope coefficient value, 0.657, for the explanatory variable, YOUTHPOP, means that a 1 percentage-point increase in the population that is between 15-24 years old is associated with a 0.657 people per thousand people increase in the traffic death rate per thousand people, ceteris paribus.

The slope coefficient value, 0.157, for the explanatory variable, LFOR, means that a 1 percentage-point increase in the labor force is associated with a 0.157 people per thousand people increase in the traffic death rate per thousand people, ceteris paribus.

The slope coefficient value, 0.0566, for the explanatory variable, BAC, means that a 1 percentage-point increase in the blood alcohol level is associated with a 0.0566 people per thousand people increase in the traffic death rate per thousand people, ceteris paribus. The variable is not significant at 5% significance level but from theoretical point of view it belongs in the equation and the expected sign is right.

The slope coefficient value, 7.509, for the explanatory variable, INCOME, means that if a country is a high income country it is associated with a 7.509 people per thousand people increase in the traffic death rate per thousand people, ceteris paribus.

The only variable that ended up with a different sigh of the coefficient than expected is urban population. Like I mentioned above data shows that traffic and living in noisy urban areas matter. The negative sign could be explained by omitted variable bias that is causing the negative sign of the coefficient, like for example number people who use public transportation over a personal car. In this case the urban population will decrease accident rates because if in big cities more people uses for example metro to get to school/work then the accident rate will drop.

Clearly instead of blood alcohol content that I would expect to be the most significant at the beginning of paper other variables seemed to play first drums. Variable YOUTHPOP and INCOME are the most significant for accident rates.

To test my regression if it healthy I took one variable out of the regression to see if it belongs to the equation. I decided to take the YOUTHPOP out of the equation because from my theoretical point of view it belongs it the equation. After taking the YOUTHPOP out of the equation the adjusted R^2 dropped to .57 and the value of INCOME increases significantly from 7.5 to 10.8 which confirms that the YOUTHPOP belongs in the regression. Also I took out the INCOME variable of the regression, the adjusted R^2 dropped to .55. It confirms my theory that INCOME belongs in the equation.

Although one would think that there would be multicollinearity between variables because for example urban population, labor force population and young people population would overlap, the results show the severity is low between variables. The variance inflection factor is smaller than 5 for every single variable in the data set and R^2 is smaller then .8.

The data does not have heteroskedasticity. The Hal White test performed on the primary regression has showed that the variance of the error term is constant. The NR^2 of 19 degrees of freedom is 23.16 and the critical value at 5 percent significance level is 30.1. We fail to reject the null hypothesis that there is homoscedasticity therefore we conclude that there is no heteroskedasticity.

The biggest problem in my data set is another omitted variable bias. I have not included in my regression one more variable that I thought was theoretically significant which is how many vehicles there is on a road in a particular country. This variable could change the coefficient of the independent variable URPOP because the more people that live in the urban areas the more vehicles there should be on the road. Number of vehicles would have a positive expected sign. The more vehicles there are on the road the higher probability of getting into an accident. Data for vehicles per capita for a given country is restricted and I do not have access to it.

Also the variable BAC could be insignificant in my primary regression because people above 24 years old do not drink as much as people below 24 years old. A new variable OLD_COUNTRY(see Table1) has been created to test if BAC in primary regression would become significant after putting restriction on older population. After running the regression with OLD_COUNTRY in, the adjusted R^2 is gets smaller and BAC remains insignificant and OLD_COUNTRY is not significant either, at 5 percent level of significance.

Conclusion.

Accidents injuries that resulted in death can be explained in 60% by country's wealth, urbanization, labor force participation, number of young population and blood alcohol content. Income per capita and if a country is classified as a high income country matters for its accident rate. Also a country's urbanization or labor force participation does not matter as much as a country's youth. People between 15-24 are the most likely to get into accidents and my regression confirms the theory. Surprisingly, blood alcohol content has almost zero or very little influence on accident rates but this might be caused by an omitted variable that I failed to find or recognize. The next step for this paper is to collect more data on accident rates from next years and run time series regression with the data as well. Also the variable that I mentioned above, vehicles per capita for a given country should be included in the regression. The data is restricted but having numbers for this variable could be crucial to obtain even better results. The are many more factor that can be associated with accident rates like a country culture or a driver's character that cannot be measured here or be minimized.

Table1.	Summary statistics			
VARIABLE	Description	Mean	Std.Dev.	Exp. Sign
TRD	Accident Rates that resulted in death(per 100,000 population)	15.84	8.9400	.
BAC_	Blood alcohol content per country	7.739	3.0041	+
URPOP	Urban population(% of total population)	59.63	22.996	+
YOUTHPOP	Percentage of population that is 15-24 years old	16.64	3.7893	+
LFOR	Labor Force(% of total population)	63.61	10.090	+
INCOME	Dummy variable for high income (1 if a country is classified as high income country and 0 otherwise)	.3361	.47438	-
OLD_COUNTRY	Dummy variable,1, if a country's young population(% of the total population) is below 17% of total population.	.5126	.5109	-

Table2. Primary Regression

Source	SS	df	MS			
				Number of obs	=	119
				F(5, 113)	=	36.47
Model	5822.98695	5	1164.59739	Prob > F	=	0.0000
Residual	3608.03944	113	31.9295525	R-squared	=	0.6174
				Adj R-squared	=	0.6005
Total	9431.02639	118	79.9239524	Root MSE	=	5.6506

trdeaths_13	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
urpop_13	-.0516204	.0290964	-1.77	0.079	-.1092656	.0060248
youthpop_13	.6569914	.2271465	2.89	0.005	.2069731	1.10701
lfor_13	.157891	.0594535	2.66	0.009	.0401029	.2756792
bac_	.0565163	.1867835	0.30	0.763	-.3135354	.426568
income_13	-7.509501	1.764651	-4.26	0.000	-11.00559	-4.01341
_cons	.0283884	5.038924	0.01	0.996	-9.954629	10.01141

```
. reg trdeaths_13 urpop_13 youthpop_13 bac_ income_13
```

Table 3. Regression with a new variable OLD_COUNTRY.

```
. reg trdeaths_13 urpop_13 youthpop_13 bac_ income_13 lfor_13 old_country
```

Source	SS	df	MS			
				Number of obs	=	119
				F(6, 112)	=	30.22
Model	5829.76771	6	971.627951	Prob > F	=	0.0000
Residual	3601.25868	112	32.1540954	R-squared	=	0.6181
				Adj R-squared	=	0.5977
Total	9431.02639	118	79.9239524	Root MSE	=	5.6705

trdeaths_13	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
urpop_13	-.0539672	.0296424	-1.82	0.071	-.1126997	.0047653
youthpop_13	.7715148	.3378641	2.28	0.024	.1020804	1.440949
bac_	.0500631	.1879651	0.27	0.790	-.3223657	.4224919
income_13	-7.522749	1.77108	-4.25	0.000	-11.03192	-4.013582
lfor_13	.157851	.0596623	2.65	0.009	.0396379	.2760641
old_country	-1.037475	2.259211	-0.46	0.647	-5.513811	3.438861
_cons	-1.14891	5.669376	-0.20	0.840	-12.38205	10.08423

Table 3. Regression without YOUTHPOP in.

```
. reg trdeaths_13 urpop_13 bac_ income_13 lfor_13
```

Source	SS	df	MS		
Model	5555.87095	4	1388.96774	Number of obs =	119
Residual	3875.15544	114	33.9925916	F(4, 114) =	40.86
				Prob > F =	0.0000
Total	9431.02639	118	79.9239524	R-squared =	0.5891
				Adj R-squared =	0.5747
				Root MSE =	5.8303

trdeaths_13	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
urpop_13	-.054869	.0299993	-1.83	0.070	-.1142974	.0045593
bac_	.2075042	.1850436	1.12	0.264	-.1590657	.574074
income_13	-10.85123	1.376257	-7.88	0.000	-13.57759	-8.124878
lfor_13	.2252767	.056438	3.99	0.000	.1134734	.33708
_cons	6.824949	4.599223	1.48	0.141	-2.286077	15.93598

Table4. *Regression without INCOME in*.

```
. reg trdeaths_13 urpop_13 youthpop_13 bac_ lfor_13
```

Source	SS	df	MS		
Model	5244.76102	4	1311.19025	Number of obs =	119
Residual	4186.26537	114	36.7216261	F(4, 114) =	35.71
				Prob > F =	0.0000
Total	9431.02639	118	79.9239524	R-squared =	0.5561
				Adj R-squared =	0.5405
				Root MSE =	6.0598

trdeaths_13	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
urpop_13	-.0995589	.02877	-3.46	0.001	-.1565521	-.0425658
youthpop_13	1.289866	.1841263	7.01	0.000	.9251134	1.654619
bac_	-.0564443	.1982771	-0.28	0.776	-.4492297	.3363411
lfor_13	.0759413	.0603218	1.26	0.211	-.0435557	.1954383
_cons	-4.083399	5.303561	-0.77	0.443	-14.58971	6.422914

Bibliography:

1. Global status report on road safety: Supporting a decade of action. (2013). Geneva: World Health Organization.
2. Global status report on road safety: Time for action. (2009). Geneva: World Health Organization.
3. Global status report on road safety: Summary (2015). Geneva: World Healt Organization.
4. "About Us." - FIA Foundation. Accessed May 01, 2016. http://www.fiafoundation.org/about-us.
6. Holody, Elena. "Traffic Fatalities in the US Have Been Mostly Plummeting for Decades." Business Insider. 20 Apr. 2016. Web. 23 Apr. 2016
7. Sauti, Lazarus. "Zimbabwe: Poor Roads, Human Error to Blame for Traffic Accidents." Allafrica.com. 25 Apr. 2016. Web. 23 Apr. 2016.
8. "Where Car Accidents Happen Most." Esurance.com. Allstate. Web. 2 May 2016.
9. "How Does Alcohol Affect Driving?" Drug News. Web. 01 May 2016.
10. "Facts and Statistics." Distracted Driving : Facts And Statistics. Web. 26 Apr. 2016

Data:

World Bank http://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS
http://data.worldbank.org/indicator/SL.TLF.CACT.ZS

United States Census Buearau
http://www.census.gov/population/international/data/idb/region.php?N=%20Results%20&T=4&A=separate&RT=1&Y=2013&R=58,60,61,64&C=AF

Global Road Report 2013
http://www.who.int/violence_injury_prevention/road_safety_status/2015/Country_profiles_combined_GSRRS2015_2.pdf?ua=1

World Health Organization http://apps.who.int/gho/data/node.main.A997