Road Traffic Fatalities: The Effect of the Implementation of the 'accompanied driving from the age of 17' in Germany

Alexander Dalheimer

Introduction

Most of young people desire to be authorised to make their driver's license. Especially for those living in more rural areas, the driving license is an important step towards autonomy and mobility. However, countries differ in the age which is required to start the driver's license. While for many countries an age of eighteen is necessary, some introduced the opportunity to get a license already with an age of seventeen or even earlier (e.g. Germany or France). In this project, Germany is considered with its policy 'Begleitetes Fahren mit 17' (BF17) or 'accompanied driving from the age of 17'. BF17 enabled the new drivers to drive with an age of 17 accompanied by a registered and experienced driver.

This project has the aim to analyse if the implementation of BF17 had a significant influence on the road traffic fatalities (RTF) under new drivers (18-24 years old). Thus, the research question of this project is: Did the impelemtation of BF17 lead to a decrease of the RTF under the 18-24 years old driver? This technical memo provides detailed information about the data, method, results, robustness and limitations of the analysis.

Data

Suprisingly, the availability of longitundial data about road traffic fatalities is not as comprehensive as expected. Data about the absolute number of RTFs is provided by Eurostat / CARE Data (2020) for the period 1999 to 2018 as anual data. For the data collection the single countries are responsible. In order to have a variable which can be compared across countries the number of RTF per 100,000 inhabitants was computed ($\frac{RTF}{population\ size/100{,}000}$). This fraction is calculated for each age group separately using the corresponding population sizes for each age group. The age groups are pre-defined by Eurostat: *(1)* 18-24 years old, *(2)* 25-49 years old, *(3)* 50-64 years old and *(4)* older than 65.

Another data source which was used is OECD.Stat / International Transport Forum (2020) which provided information about the GDP per capita, the employment rate, the amount of road infrastructure spending, the amount of accidents with injured persons and population data. GDP per capita is already transformed in constant Euro which makes the variable comparable across countries. Also the employment rate for working population did not need any transformation. The infrastructure spending is provided as absolut values in constant Euro. Since absolute values are not appropriate to compare across countries, the spending was calculated per 100,000 inhabitants - as it was done for the RTF before. Also the amount of accidents with injured people were transormed in that way.

Table 1 shows descriptive statistics for all variables used in the main analysis. Due to the space restrictions only overall descriptive statistics are presented. For example, the second column / variable RTF (18-24) provides information about the road traffic fatalities per 100,000 inhabitants (18-24 years old) across all countries and the whole time period (1999-2018). For each variable the minimum, maximum, median and mean is listed.
Furthermore, table 1 shows partly the structure of the missing values. All the RTF variables have 11 NAs each. Thus, the missing values are somehow systematic. A more detailed look in the data reveals that the NAs are in the same rows across all RTF variables. This means that the NAs depend crucially on the country and on time. Thus, the NAs are cleary 'not missing at random' (NMAR). Furthermore, the missing values are always at the beginning and/or the end of the time series considering each country seperately.

	RTF (18-24)	RTF (18-24) Imputed	RTF (25-49)	RTF (50-64)	RTF (65+)	Employment Rate	GDP per capita	Road Spending	Accident injured
Min.	2.077	2.077	0.6358	1.073	1.215	48.80	11302	521577	114.0
Median	10.163	10.464	3.3129	6.147	7.354	64.97	33934	17229842	326.2
Mean	11.362	11.696	3.9912	7.816	8.705	64.68	35765	19996803	373.0
Max.	27.412	27.412	11.3802	24.168	25.552	80.10	66179	82529512	844.8
NA's	11	–	11	11	11	16	–	10	1

Table 1: Descriptive Statistics
Overview of all of the used variables from the analysis. Note that the variable 'RTF (18-24) Imputed' is the variable 'RTF (18-24)' with imputed NAs by using a time series extrapolation technique which is explained in detail further below. Note that the mean and median has not changed that much after imputation

The used synthetic control method - which is explained in method section - requires only complete information for the dependent variable. Thus, those missing values needed to be imputed. There are a lot of different approaches how to impute NAs. The decision which imputation method is appropriate depends crucially on the data structure and the analysis method. The data used for the synthetic control method is longitudinal data in the long format. That means that the data includes information for each country at different points in time. One popular approach is multiple imputation. However, this approach is based on a model, including the covariates from the main analysis (synthetic control), which predicts the NAs. Consequently, the dependent variable would depend on the covariates by construction due to the imputation. Thus, I decided to use an univariate time series approach. Since the missing values are at the beginning and/or the end of the series, statistical interpolation is not appropriate because it needs the NAs within the series. Thus, statistical extrapolation is used in order to estimate values outside the known range of values. Before the imputation, it was necessary to transform the data from long to wide format. In the wide format each column represents a country, where each cell represents the RTF per 100,000 inhabitants at a special point in time. The wide format has the advantage that each column/country can be seen as a univariate time series. This makes it possible to impute NAs for each county seperately. Figure 1 shows the approximately linear trend of RTF for the 18-24 years old drivers (dependent variable) for Germany. The trends are also approximately linear for the other countries but only Germany is presented due to reasons of visualisation and recognizability. In order to capture the approximately linear trends, a time series model in its structural form is used fitted by maximum likelihood. The popular Kalman smoother algorithm is not used since it is recommendet by the author of the R package *imputeTS* (Moritz & Bartz-Beielstein 2017) that the KalmanRun algorithm should be used for extrapolation. Also the Kalman filter algorithm is used to get valid results which take statistical noise into account.

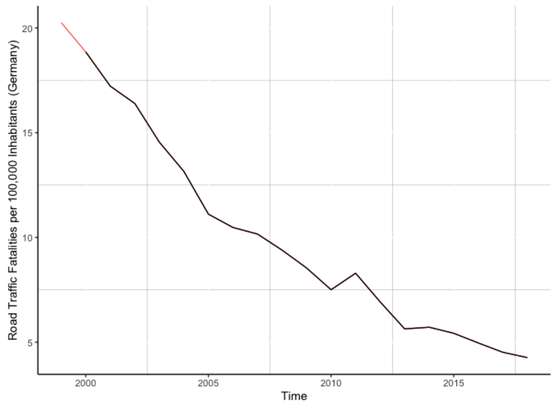

Figure 1: Trend of the Dependent Variable
The black line shows the avaliable data while the red line highlights the imputed data (NA for 1999).

Figure 1 shows the RTF in Germany over the whole period. Obviously, there is a negative trend which implies that the fatalities decreased from 1999 to 2018. The black line shows the data which is available, while the red line shows the imputed value for 1999 which is missing in the original data. What can be seen is that the imputation has succesfully recognized the linear trend and extrapolated it. Figure 2 shows a more general result for the implications of the imputation for the dependent variable across all countries (variable in long format). The black line represents the density of the dependent variable (RTF for 18-24 years old driver) across all countries and the whole time period. The red line represents the density after imputation. Generally, the closer the densities the better, since this implies that the imputation has not changed the density dramatically. If the density would have been changed dramatically, there would be the risk of false statistical inferences due to the imputation. According to figure 2 there is no such risk, since the density after imputation is very close to the one before.

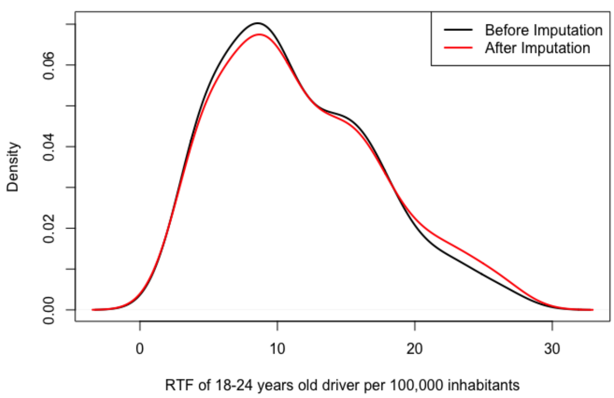

Figure 2: Effect of Imputation on the Distribution of the Dependend Variable
The black line shows densitiy for the original data. In contrast to figure 1, here the dependent variable in long format is used. Thus, the variable contains the information for the whole time series across all countries. The red line shows the density of the dependent variable after the imputation.

Method

The method which is used in this project is *synthethic control*. The aim of synthetic control is to find out how the trend of the treated unit would have looked like if it would have not been treated. In this study Germany is the treated unit and the treatment is the policy 'Begleitetes Fahren mit 17 (BF17)' which was implemented in 2004 in Lower Saxony. After the Bundesrat revised the StVO (Straßenverkehrsordnung) in 2005, Lower Saxony implemented the new law in July 2006 as the first federal state. One by one, the other federal states implemented it. The last state Baden-Württemberg implemented the law in 2008. The point in time of the treatment is chosen to be 2007, since at this point in time more than two-thirds of the states have implemented the new policy. Thus, the years before 2007 are defined as pre-treatment and those after are considered as post-treatment. Important to mention is that placebos in time have proven that setting an earlier intervention date does not change the results.

The goal of synthetic control is to create a synthetic Germany that is as similar as possible to the observed pre-treatment trend of RTF of the 18-24 years old drivers in Germany. Synthetic control can be seen as a generalization of the difference-in-difference approach since it relies not only on one country, which is assumed to be similar to the treated one, but on a pool of countries - called donor pool. Countries in the donor pool are Belgium, Czech Republic, Finland, Greece, Latvia, Poland, Portugal, Spain and Switzerland. These countries have all in common that they have no similar policy as the BF17 in Germany. The weighted average of these countries is the synthetic Germany. The synthetic Germany consists of 12.9% Belgium, 10% Finland, 41.1% Portugal and 36% Switzerland. Additionally, covariates are used to create a synthetic control which is as similar as possible to Germany. For the covariates *employment rate*, *GDP per capita*, *number of passanger cars per 100,000 inhabitants* the mean for the pre-treatment period for each country is used to predict the dependent variable. These variable have been proven to have a significant influence on the RTF (see Ali et al. 2018). Furthermore, the special predictors for which the mean is only computed for the two years pre-treatment are *RTF for the 25-49 years old driver*, *RTF (50-64)*, *RTF (65 and over)* and *number of accidents where people got injured per 100,000 inhabitants*. Furthermore, the dependent variable itself is included as special predictor for two pre-treatment periods. Many alternatives were tested but this model showed the best results.

Since this is a technical memo, only a few words about the assumptions of synthetic control. One assumption is the one of *no interference*. Countries in the donor pool have still not implemented some policy similar to BF17 in the year 2018. Thus, the assumption of no interference can be seen as fulfilled. The assumption of *anticipation effects* is not an issue in this application since they are unlikely. An assumption which could be an issue is the one of *unconfoundness*. It is assumed that the only difference in covariates between the treated unit and the synthetic control, post-treatment, is the incidence of the intervention. However, there are omitted variables e.g. how many police controls a country conduct, how costly traffic ofences are etc. Influences like these could violate the assumption of unconfoundness. In order to capture as many omitted variables as possible, the traffic fatalities of all age groups are included since they can be assumed to be correlated with the omitted variables as well.

Table 2 shows the balance table for the synthetic control. Across all covariates except for the *accidents with injured people* the synthetic control is closer to the real Germany as the simple sample mean. Furthermore, the last column shows the weights each covariate was assigned by the algorithm. The bad balance for the accidents can be neglected since the one of 2006 got a weight of 0% and the one of 2005 0.1%. The most relevant covariates are *number of passanger cars* (22,6%), *RTF 25-49 years old (2006)* (14.9), *RTF 50-64 years old (2005)* (13%) and *RTF 65 and older (2006)* (18.7%).

Covariates	**Treated**	**Synthetic**	**Sample Mean**	**Weight**
Employment Rate	67.192	70.584	63.831	0.081
GDP per Capita	41311.911	42239.868	32729.088	0.094
Number of Passanger Cars per 100,000 inhabitants	543.375	499.715	411.008	0.226
RTF 18-24 years old (2006)	10.477	9.511	13.466	0.032
RTF 18-24 years old (2005)	11.111	12.899	15.070	0.059
RTF 25-49 years old (2006)	3.374	3.147	4.904	0.149
RTF 25-49 years old (2005)	3.566	4.274	5.500	0.005
RTF 50-64 years old (2006)	6.683	6.159	9.697	0.087
RTF 50-64 years old (2005)	7.075	8.651	11.126	0.13
RTF 65 and older (2006)	6.285	6.761	10.995	0.187
RTF 65 and older (2005)	6.889	9.299	12.509	0.038
Accident with injured people (2006)	596.262	484.359	362.762	0
Accident with injured people (2005)	613.228	497.644	373.922	0.001

Table 2: Balance Table

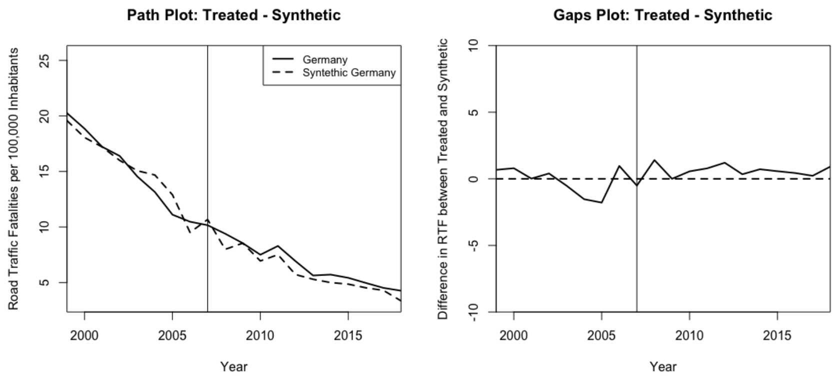

Figure 3: Treatment-Effect-Plots
In both plots the vertical line represents the invervention, the black line represents Germany and the dashed line the synthetic Germany. The plot on the left hand side shows the trend of Germany and its the approximated trend by the sythetic control pre-treatment. After 2007 the treatment effect can be seen. The plot on the right hand side shows the difference between Germany (black line) and synthetic control (dashed line)

Results and Robustness

The results show clearly no difference before and after the implementation of the policy. Figure 3 shows the treatment effect plots. On the left, the trend for both German (black line) and synthetic Germany (dashed line) is shown. The vertical line represents the year 2007 when the policy was implemented. In the years 2003 to 2005 the model fit is not perfect. However, different model specification have not changed the goodness fit. Nevertheless, the model fit looks fine overall. In the post-treatment period it can be seen that there is no significant gap at any point in time. Thus, the plot shows no significant treatment effect. On average the treatment effect is 0.55. The plot on the right side shows the differences between Germany and synthetic Germany. Again, the non-optimal pre-treatment model fit can be seen. Additionally, the neglectable differences betwenn Germany and synthetic control in the post-treamtent period can be seen.

Furthermore, placebo tests are conducted in order to ensure robustness of the results. Figure 4 shows the results of those. The left plot presents the difference between Germany and its snythetic control (black line) while the grey lines represent the control units and their difference from their respective synthetic controls over time. If the treatment effect of Germany would be significant, the black line would be clearly distiguishable from the grey line considering the post-treatment period since there would be a gap.

However, the plot shows that there is no difference between Germany and the other units. Here it can be even seen that other countries do much greater differ from their synthetic control than Germany does. This result is also be shown in the plot on the right hand side which shows the post/pre MSPE ratio for each country. Again, Germany does not really differ from its synthetic control. This is also

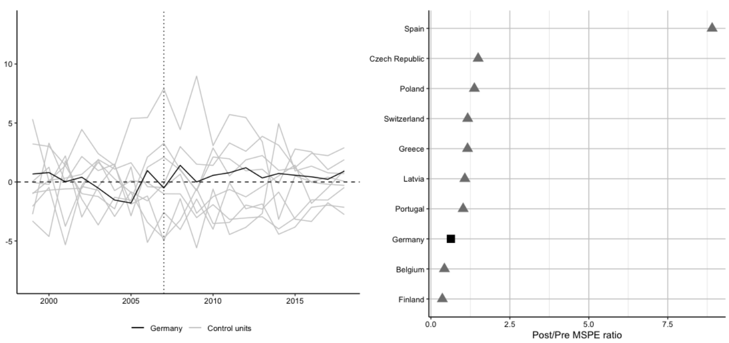

Figure 4: Placebo Tests
Left Plot: Black line represents the difference between Germany and its synthetic control, while the grey lines represents the difference between the control units and their respective synthetic control.
Right Plot: Post/pre MSPE ratio for each of the countries.

true for all other countries except Spain. Furthermore, placebo in time tests were conducted where the invervention point in time was set to 2006 and 2005. However, the previous result which indicate no meaningful difference between Germany and its synthetic control maintained.

Conclusion

The main finding of this project is that the implementation of BF17 - the ability to get ones driver's license with an age of seventeen - did not had an significant influence on the road traffic fatalities under the 18-24 years old drivers. The limitations of this study are the restricted data availability which results, on the one hand, in only few countries in the donor pool and on the other hand in only few pre-treatment periods. Nevertheless, figure 3 shows that the trend of Germany could have been approximated but more data would have improven the results. Furthermore, individual data would allow better and more detailed inferences. If there would have been data about the drivers, e.g. if they used the opportunity to get the driver's license in the age of 17 or not would have improven the inference.

References

Ali, Q., Yaseen, M. R., & Khan, M. T. I. (2018). Road traffic fatalities and its determinants in high-income countries: a continent-wise comparison. *Environmental Science and Pollution Research*, *26*, 19915–19929.

Eurostat / Care Data (2020). *Persons killed in road accidents by age.* data retrieved from Eurostat database, `https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=tran_sf_roadag&lang=en`

Moritz S, Bartz-Beielstein T (2017). "imputeTS: Time Series Missing Value Imputation in R." *The R Journal*, *9*(1), 207–218. doi: 10.32614/RJ-2017-009.

OECD.Stat / International Transport Forum (2020). Level of GDP per capita and productivity. data retrieved from OECD.Stat, `https://stats.oecd.org/index.aspx?DataSetCode=PDB_LV#`

OECD.Stat / International Transport Forum (2020). Transport infrastructure investment and maintenance spending. data retrieved from OECD.Stat, `https://stats.oecd.org/Index.aspx?DataSetCode=ITF_INV-MTN_DATA`

OECD.Stat (2020). Historical population. data retrieved from OECD.Stat, `https://stats.oecd.org/Index.aspx?QueryId=88956#`

OECD (2020), Road accidents (indicator). doi: 10.1787/2fe1b899-en, `https://data.oecd.org/transport/road-accidents.htm#indicator-chart`

OECD (2020), Working age population (indicator). doi: 10.1787/d339918b-en, `https://data.oecd.org/pop/working-age-population.htm`