UNIVERSITÄT SIEGEN	**Fakultät Wirtschaftswissenschaften, Wirtschaftsinformatik und Wirtschaftsrecht**	**Fakultät III:** Lehrstuhl für BWL, insbesondere Unternehmensnachfolge

Lecture work: Markets & Strategies II

Examiner: Prof. Dr. Matthias Hunold, Dr. Jannika Schad

Semester: WS 2021/2022

Additional note to the presentation on the Term Paper I on the topic “What factors can explain the excess mortality in Germany during COVID-19?

Surname, First name: Daria.Kharitonova 1609765

E- Mail: daria.kharitonova@student.uni-siegen.de

Date of submission: 15.02.2022

Grade:

Date:

Signature:

Feedback after the presentation

- Collect the data on temperature and rain on Bundesland level → **done**
- Remove IV regression with instrumentation on temperature and rain, use them as independent variables in the main regression → **done**
- COVID-19 deaths don't happen immediately after COVID-19 infection → introduce lag values for COVID-19 cases → **done**
- Interpret coefficients in standard deviation terms → **done**

Empirical strategy

Data

I focused on the weekly data on Bundesland and Germany level due to the highest data availability and quality vs. county level. Time period: 10th week 2020 – 43th week 2021

In this section I would like to give you an overview how the data were structured for the better understanding how the results were obtained.

Overview of the Bundesland level data

Data head

year	week	Bundesland	age_group	exc_mortality	vacc_per_100_cum	cases_per_100	temperature	rain
2020	10	Baden-Württemberg	60+	0.9705987	0.0	0.0019225522	6.0	50.5
2020	10	Baden-Württemberg	0-59	0.8885449	0.0	0.0090384069	6.0	50.5
2020	10	Bayern	60+	0.9335797	0.0	0.0012280499	5.8	12.5
2020	10	Bayern	0-59	0.9328264	0.0	0.0057733747	5.8	12.5
2020	10	Berlin	60+	0.9145533	0.0	0.0011632417	5.5	10.0
2020	10	Berlin	0-59	0.9857434	0.0	0.0054686951	5.5	10.0
2020	10	Brandenburg	60+	0.9242298	0.0	0.0005959691	5.4	25.7
2020	10	Brandenburg	0-59	0.7139785	0.0	0.0028018021	5.4	25.7
2020	10	Bremen	60+	1.0000000	0.0	0.0013410377	6.2	47.5
2020	10	Bremen	0-59	1.6078431	0.0	0.0063045594	6.2	47.5
2020	10	Germany	60+	0.9262235	0.0	0.0012807319	5.3	60.0
2020	10	Germany	0-59	0.8917264	0.0	0.0060210461	5.3	60.0
2020	10	Hamburg	60+	0.9607390	0.0	0.0016948433	5.7	43.2
2020	10	Hamburg	0-59	1.0558376	0.0	0.0079678895	5.7	43.2
2020	10	Hessen	60+	0.9508704	0.0	0.0008472953	7.4	47.9
2020	10	Hessen	0-59	0.8151659	0.0	0.0039833508	7.4	47.9

Data structure:

```
'data.frame':	2618 obs. of  9 variables:
 $ year           : int  2020 2020 2020 2020 2020 2020 2020 2020 2020 2020 ...
 $ week           : int  10 11 12 13 14 15 16 17 18 19 ...
 $ Bundesland     : chr  "Baden-Württemberg" "Baden-Württemberg" "Baden-Württemberg" "Baden-Württemberg" ...
 $ age_group      : chr  "0-59" "0-59" "0-59" "0-59" ...
 $ exc_mortality  : num  0.889 0.894 1.019 1.048 1.125 ...
 $ vacc_per_100_cum: num  0 0 0 0 0 0 0 0 0 0 ...
 $ cases_per_100  : num  0.00904 0.02982 0.05497 0.06042 0.03866 ...
 $ temperature    : num  6 6 6 6 11.8 11.8 11.8 11.8 11.8 13.2 ...
 $ rain           : num  50.5 50.5 50.5 50.5 14.9 14.9 14.9 14.9 14.9 52.1 ...
```

Data summary:

```
      year           week        Bundesland         age_group         exc_mortality    vacc_per_100_cum cases_per_100          temperature         rain       
 Min.   :2020   Min.   :10.00   Length:2618        Length:2618        Min.   :0.5217   Min.   : 0.000   Min.   :-0.002834   Min.   : 2.20   Min.   :  0.90  
 1st Qu.:2020   1st Qu.:19.00   Class :character   Class :character   1st Qu.:0.9768   1st Qu.: 0.000   1st Qu.: 0.002355   1st Qu.: 7.70   1st Qu.: 23.50  
 Median :2020   Median :29.00   Mode  :character   Mode  :character   Median :1.0451   Median : 0.000   Median : 0.010209   Median :12.30   Median : 41.90  
 Mean   :2020   Mean   :29.01                                         Mean   :1.0635   Mean   : 8.287   Mean   : 0.028291   Mean   :12.85   Mean   : 47.46  
 3rd Qu.:2021   3rd Qu.:38.00                                         3rd Qu.:1.1219   3rd Qu.:16.075   3rd Qu.: 0.031298   3rd Qu.:17.80   3rd Qu.: 61.40  
 Max.   :2021   Max.   :52.00                                         Max.   :2.3895   Max.   :79.700   Max.   : 0.411574   Max.   :22.00   Max.   :174.10  
```

Source: author's project

Overview of Germany level data

Data head

year	week	Bundesland	age_group	exc_mortality	exc_mortality_demo	vacc_per_100_cum	cases_per_100	temperature	rain
2020	10	Germany	0-59	0.8917264	0.9000024	0.0	0.0060210461	5.3	60
2020	10	Germany	60+	0.9262235	0.8967776	0.0	0.0012807319	5.3	60
2020	11	Germany	0-59	0.9070832	0.9155213	0.0	0.0213044781	5.3	60
2020	11	Germany	60+	0.9767442	0.9453944	0.0	0.0045316583	5.3	60
2020	12	Germany	0-59	0.9546209	0.9634863	0.0	0.0334555396	5.3	60
2020	12	Germany	60+	1.0134730	0.9807373	0.0	0.0071163008	5.3	60
2020	13	Germany	0-59	0.9811961	0.9903426	0.0	0.0372922477	5.3	60
2020	13	Germany	60+	1.0440627	1.0101314	0.0	0.0079324039	5.3	60

Data structure:

```
'data.frame':	154 obs. of  10 variables:
 $ year              : int  2020 2020 2020 2020 2020 2020 2020 2020 2020 2020 ...
 $ week              : int  10 10 11 11 12 12 13 13 14 14 ...
 $ Bundesland        : chr  "Germany" "Germany" "Germany" "Germany" ...
 $ age_group         : chr  "0-59" "60+" "0-59" "60+" ...
 $ exc_mortality     : num  0.892 0.926 0.907 0.977 0.955 ...
 $ exc_mortality_demo: num  0.9 0.897 0.916 0.945 0.963 ...
 $ vacc_per_100_cum  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ cases_per_100     : num  0.00602 0.00128 0.0213 0.00453 0.03346 ...
 $ temperature       : num  5.3 5.3 5.3 5.3 5.3 5.3 5.3 5.3 10.4 10.4 ...
 $ rain              : int  60 60 60 60 60 60 60 60 45 45 ...
```

Data summary:

```
      year           week        Bundesland         age_group         exc_mortality    exc_mortality_demo vacc_per_100_cum cases_per_100        temperature   
 Min.   :2020   Min.   :10.00   Length:154         Length:154         Min.   :0.8605   Min.   :0.8461     Min.   : 0.000   Min.   :0.0004155   Min.   : 3.00  
 1st Qu.:2020   1st Qu.:19.00   Class :character   Class :character   1st Qu.:0.9714   1st Qu.:0.9656     1st Qu.: 0.000   1st Qu.:0.0032212   1st Qu.: 6.20  
 Median :2020   Median :29.00   Mode  :character   Mode  :character   Median :1.0108   Median :1.0049     Median : 0.000   Median :0.0121332   Median :11.90  
 Mean   :2020   Mean   :29.01                                         Mean   :1.0270   Mean   :1.0140     Mean   : 9.714   Mean   :0.0299907   Mean   :12.12  
 3rd Qu.:2021   3rd Qu.:38.00                                         3rd Qu.:1.0701   3rd Qu.:1.0460     3rd Qu.:17.250   3rd Qu.:0.0305290   3rd Qu.:16.90  
 Max.   :2021   Max.   :52.00                                         Max.   :1.4002   Max.   :1.3536     Max.   :79.700   Max.   :0.1748441   Max.   :20.00  
      rain       
 Min.   : 30.00  
 1st Qu.: 45.00  
 Median : 60.00  
 Mean   : 69.74  
 3rd Qu.: 94.00  
 Max.   :140.00  
```

Source: author's project

Introduction of the main definitions

Excess mortality ($exc\_mortality_{itk}$ in regression) - number of people died in particular week during COVID-19 divided by average number of people died in 2016-2019 in bundesland i in period t. Allows to account for seasonal variation.

Example:

- number of people died in NRW in week 48, 2020 in the age group 60+: 3927
- average number of people died in NRW in week 48, 2016-2019 in the age group 60+: ((3296 + 3374 + 3325 + 3494):4) = 3372.25
- excess mortality in NRW in week 48 2020 would be the ratio of these numbers:

$$exc\_mortality_{itk} = \frac{3927}{3372.25} = 1.16$$

It means that 16% more people ((1.16-1.0) * 100) died in NRW in week 48 in 2020 than in the same time period 2016-2019.

Excess mortality accounting for demographics ($exc\_mortality\_demo_{tk}$ in regression) number of actual deaths / number of forecasted deaths in Germany in period $\boldsymbol{t}$ in particular age group. Allows to account for seasonal variation and demographic shift, but can only be calculated on Germany level.

Example:

- actual deaths: number of people died in Germany in week 48, 2020 aged 60+: 20 867
- death ratio in the age group 60+ in period $\boldsymbol{t}$:

$$death\ ratio_{it} = \frac{\text{number of people died in Germany in week 48, 2016 aged 60 +}}{\text{number of people aged 60 + in Germany}}$$
$$= \frac{16985}{22\ 775\ 976} x\ 100 = 0.075$$

- forecasted deaths for week 48, 2020 in the age group 60+:

$$forecasted\ deaths_{it}$$
$$= \frac{(\text{average of death ratio in week 48 in } 2016 - 2019 \text{ aged 60 x number of people aged 60 in Germany})}{100}$$
$$= (((0.075 + 0.072 + 0.071 + 0.074): 4 * 24031804 / 100 = 0.073 * 24031804 / 100$$
$$= 17\,488$$

- Excess mortality accounting for demographics:

$$exc\_mortality\_demo_{tk} = \frac{\text{actual number of people died in week } 48, 2020 \text{ aged } 60 +}{\text{forecasted number of deaths in week } 48, 2020 \text{ aged } 60 +}$$
$$= \frac{20\,867}{17\,488} = 1.19$$

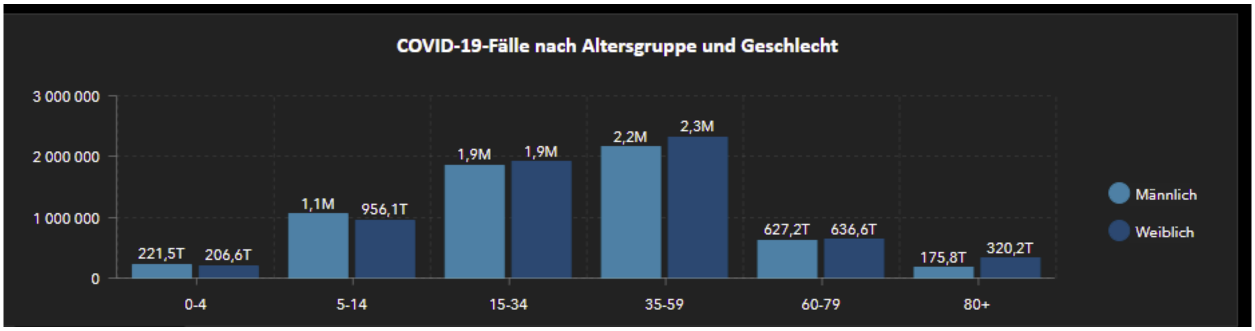
It means that 19% more people ((1.19-1.0) * 100) died in Germany in week 48 in 2020 than forecast based on 2016-2019.

Excess mortality accounting for demographics vs. Excess mortality allows to capture the change the demographics that occurred in Germany over 2016-2021. Number of people in the age group 60+ has increased, which presumably means that absolute number of deaths within this age group has also increased. In the case of $exc\_mortality_{itk}$ deaths due to demographic change will also be attributed to the COVID-19 deaths, because this is the ratio of absolute numbers. In case of $exc\_mortality\_demo_{tk}$ it doesn't happen, because it forecasts number of deaths for 2020-2021 bases on average death ratio in 2016-2019, i.e. it looks at relative numbers.

Model 1 Bundesland level: assumptions regarding variables

1. $cases\_per\_100_{itk}$ = number of COVID-19 confirmed cases per 100 in bundesland i in period t of the age group k. Note: there are no data of COVID-19 cases by the age group, so this number was approximated based on the infection split by the age groups from the

pandemic till 10.01.2022 from Robert-Koch Dashboard (https://experience.arcgis.com/experience/478220a4c454480e823b17327b2bf1d4)



The assumption is that this ratio represents the true value based on the LLN. Example:

1. 82.46% of all COVID-19 cases are observed within age group 0-59, the rest 17.56% - within the age group 60+.
2. In week 48, 2020 in NRW there were 26 740 COVID-19 cases revealed.

- Number of cases within age group 0-59 in week 48, 2020 in NRW:

$$\frac{26\ 740 * 82.46\%}{100} = 22\ 049.84$$

- Number of cases within age group 60+ in week 48, 2020 in NRW:

$$\frac{26\ 740 * 17.56\%}{100} = 4\ 690.196$$

1. age_group_60 = dummy variable, equal to 1, if people are older than 60. Note: there is a mismatch of the age groups in the data. For the **vaccinations** age split is: 0-65, 65+, for **excess mortality** and **cases** it is 0-59, 60+. This problem exists only on Bundesland level, in the data on Germany level there is no mismatch. As the solution I assigned the people in the age group 0-65 to the age group 0-59 and people in the age group 65+ to the age group 60+. The assumption is that this mismatch doesn't affect the final result, which can be tested by estimating different models.

Model 2 & 3 for Germany level: assumptions regarding variables

- $cases\_per\_100_{tk}$ = number of COVID-19 confirmed cases per 100 in period t of the age group k. Note: the same problem as in the Model 1 (please see above for the details)

Results (slide 14)

- The table represents the regression results for all 3 model specifications (slides 11-13)
- $vacc\_per\_100\_cum$ is economically and statistically insignificant within all model specifications → once controlled for other variables vaccination seems to have no effect on excess mortality;
- Variables $cases\_per\_100$ and age_group_60 + have positive effect on excess mortality, meaning that the higher COVID-19 incidents and the older person is, the higher is the excess mortality. The result is significant at 10% or 1% level in all model specifications;
- The higher is $temperature$ the higher is the excess mortality. The result is a bit counterintuitive; the potential explanation might be that the higher is the temperature, the more people are going out in general;
- $rain$ doesn't have any effect on excess mortality;
- The strongest result is demonstrated by the interaction term; $cases\_per\_100_{tk} * age\_group\_60$. So the higher COVID-19 incidents are within people aged 60+, the higher is the excess mortality. The coefficients are significant 1% level in all model specifications;
- The models are able to explain from 27.5% till 43% variations in the data;
- Comparable results within different models with different level of data quality and precision might imply, that assumption made regarding the data and variables hold;
- The detailed example of coefficients interpretation may be found on the slide 15.

Robustness check: COVID-19 cases with lags

- It is quite intuitive that after the infection with COVID-19 the death doesn't happen immediately, but rather after some time after the infection. That is why I re-evaluated the models with lag values for the variable $cases\_per\_100$ in order to test model robustness;
- The main question was what lags to chose to correctly evaluate effect of COVID-19 infections on the excess mortality. I used $lags \in [0, 4]$, where $l = 0$ is the baseline model

without lag and $l = 1, 2, 3, 4$ represents how many weeks before death COVID-19 infection occurred;

- What lag is the true lag? There is no the answer to this question and no precise statistics on how many days after infection the death occurs. If I am to believe the publicly available information in the internet, the highest probability to die occurs from 6th till 16th day after the COVID-19 infection, which corresponds to lag values from 0 to 2.

Robustness check: COVID-19 cases with lags Bundesland data (1)

- The table on slide 17 represents the regression results of the Model (1) and other models with different sizes of lags on the Bundesland level;
- It can be noted that the results for all variables except $temperature$ and $vacc\_per\_100\_cum$ in the Models (2)-(5) are analogous in sign and significance to the base Model (1), however, starting with lag = 3 the magnitude of these coefficients fall → the main result is robust.
- The maximum R^2 is achieved with lag value = 2 (R^2 = 39.6), starting from lag = 3 R^2 falls
- $vacc\_per\_100\_cum$ coefficient is significant is Models (4) and (5), however it might be driven solely by weakened influence of dependent variable with other independent variables due to too big lag value.

Robustness check: COVID-19 cases with lags Germany data (3)

- the table on slide 18 represents the regression results of the Model (3) and other models with different sizes of lags on the Germany level with account for the demographic change
- The result for all variables except interaction term in the Models (2)-(5) are analogous in sign and significance to the base Model (1) → the main result is robust.
- The coefficient for the interaction term is positive in Models 1,3,5 and negative in 2,4, but significant at least at 5% in all the models. This ambiguity of the results might be driven by smaller data sample vs. data sample used to estimate models on the slide 17.
- The maximum R^2 is achieved with lag value = 2 (R^2 = 31).
- In the case of the data on Germany level lag values significantly affect the final result. The relationships between COVID-19 cases, age dummy, vaccinations, rain and temperature

and excess mortality still hold, but interaction term interpretation depends on the lag value chosen.

Robustness check: Conclusions

1. The main relationship between variables still holds after replacing COVID-19 cases with lagged COVID-19 cases, so it can be claimed that the models' results are robust. So, independent of the model specifications the data indicate that the more there are COVID-19 cases and the older people are, the higher is the excess mortality. The effect of the vaccination rate on excess mortality is not statistically significant when controlling for other variables.
2. Open question is what lag value is optimal. Judging from the coefficients obtained and R^2 it makes sense to look at the lags from 0 to 2 weeks, starting from the 3d week the influence of COVID-19 on the excess mortality weakens which is reflected in R^2 drop.