Spatial Economics – Assignment 3

Elia Di Gregorio (h12039313@s.wu.ac.at) Max Heinze (h11742049@s.wu.ac.at)
Sophia Ludescher (h11801892@s.wu.ac.at)

May 3, 2024

Contents

The executable code that was used in compiling the assignment is available on GitHub at https://github.com/maxmheinze/spatial.

Task A

Naïve Panel Model and Specification Search

We begin by fitting the spatially naïve panel model using the `plm` function. The output is printed below.

```
cm1 <- plm(logc ~ logp + logy,
           data = cigs,
           effect = "twoway",
           model = "within",
           index = c("state", "year"))
```

	Dependent variable:
	logc
logp	−1.035***
	(0.042)
logy	0.529***
	(0.047)
Observations	1,380
R^2	0.394
Adjusted R^2	0.359
F Statistic	424.344*** (df = 2; 1303)
Note:	*p<0.1; **p<0.05; ***p<0.01

As the true spatial econometricians we are, we now ignore everything that is reminiscent of economic theory and venture on a standard specification testing path. We begin by performing a Lagrange Multiplier test for spatial lags in both the error and the dependent. The results are printed below.

```
slmtest(cm1, cigm, test = "lme")
```

```
##
##  LM test for spatial error dependence
##
## data:  formula (within transformation)
## LM = 54.655, df = 1, p-value = 1.437e-13
## alternative hypothesis: spatial error dependence
```

```
slmtest(cm1, cigm, test = "lml")
```

```
##
##  LM test for spatial lag dependence
##
## data:  formula (within transformation)
## LM = 46.901, df = 1, p-value = 7.468e-12
## alternative hypothesis: spatial lag dependence
```

As we can see, both tests return significant p-values, meaning that we reject the null hypotheses of spatial dependence being present in the errors or the dependent, respectively. We thus continue by performing robust LM tests for spatial dependence in both the errors and the dependent which account for the respectively other type of spatial dependence potentially being present. The results of these two tests are printed below.

```
slmtest(cm1, cigm, test = "rlme")
```

```
##
##  Locally robust LM test for spatial error dependence sub spatial lag
##
## data:  formula (within transformation)
## LM = 8.9106, df = 1, p-value = 0.002835
## alternative hypothesis: spatial error dependence
```

```
slmtest(cm1, cigm, test = "rlml")
```

```
##
##  Locally robust LM test for spatial lag dependence sub spatial error
##
## data:  formula (within transformation)
## LM = 1.1563, df = 1, p-value = 0.2822
## alternative hypothesis: spatial lag dependence
```

Since this time, testing for spatial dependence in the error given spatial dependence in the dependent returns a significant result, and testing for spatial dependence in the dependent given spatial dependence in the errors does not, we conclude that based on these test results, a **SEM model** is the most appropriate specification.

Estimating a SEM Model and Comparing it to an SLX Model

We use the `spml` function from the `splm` package to estimate a spatial panel model using Maximum Likelihood. Consistent with our previous answer, we estimate a **SEM model**, and we do this by setting `lag = FALSE` as well as `spatial.error = "b"`. Written down, the model equation is identical to the one presented in the assignment question, except for the important distinction that

$$\boldsymbol{\varepsilon} = \rho \boldsymbol{W} \boldsymbol{\varepsilon} + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}).$$

```
cm2 <- spml(
  logc ~ logp + logy,
  data = cigs,
  listw = cigm,
  effect = "twoways",
  model = "within",
  index = c("state", "year"),
  lag = FALSE,
  spatial.error = "b"
)
```

The results of this estimation are printed in the following.

	Dependent variable:
	logc
logp	-1.004^{***}
	(0.040)
logy	0.554^{***}
	(0.049)
Spatial error rho	0.240^{***}
	(0.033)
Observations	1,380
Note:	*p<0.1; **p<0.05; ***p<0.01

Next, we estimate an **SLX model** using the `plm` function in combination with the `slag` function to create spatial lags. Since there neither is a lag of the dependent nor are there spatial errors, there is no need to use a function for a dedicatedly spatial panel model. The model we estimate below amounts to

$$\boldsymbol{y}_t = \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}_t\boldsymbol{\gamma} + \boldsymbol{\mu} + \phi_t\boldsymbol{\iota} + \boldsymbol{\varepsilon}_t,$$

where, diverting from the original notation for ease of reading, $\boldsymbol{y}_t$ is a $n \times 1$ vector of stacked $\log(C_{it})$, $\boldsymbol{X}_t$ is a $n \times 2$ matrix consisting of stacked $\log(P_{it})$ and $\log(I_{it})$ as columns, $\boldsymbol{W}$ is the provided weights matrix, $\boldsymbol{\beta} = (\beta_1, \beta_2)'$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$, and both $\boldsymbol{\mu}$ and $\boldsymbol{\varepsilon}$ are stacked over individuals and thus $n \times 1$ vectors.

```
cm3 <- plm(
  logc ~ logp + logy + slag(logp, listw = cigm) + slag(logy, listw = cigm),
  data = cigs,
  effect = "twoways",
  model = "within",
  index = c("state", "year")
)
```

Again, the results are printed below.

	Dependent variable:
	logc
logp	-1.017^{***}
	(0.042)
logy	0.608^{***}
	(0.060)
slag(logp, listw = cigm)	-0.220^{***}
	(0.077)
slag(logy, listw = cigm)	-0.219^{***}
	(0.080)
Observations	1,380
R^2	0.400
Adjusted R^2	0.364
F Statistic	217.105^{***} (df = 4; 1301)
Note:	*p<0.1; **p<0.05; ***p<0.01

We can see that there is positive spatial autocorrelation of errors in the SEM model, and a negative spillover effect of price changes in the SLX model. The **positive spatial autocorrelation of errors** can be interpreted as that states that are contiguous are likely to experience similar shocks. This can be explained as spillover effects of an unobserved variable affecting cigarette demand; however, it is not possible to deduce information about the spillover effect of price changes, one of the explanatory variables in the model, from this error autocorrelation.

In contrast, the SLX model allows for directly examining and interpreting spillover effects, which could be the rationale for choosing it over the SEM (or a SAR) model. The coefficient we receive for **the effect of adjacent states' price changes**, γ_1, is \$-\$0.220. This can directly be interpreted as that given a one-unit increase in log price of cigarettes, log demand in adjacent states decreases by an average of 0.22 (or, if you want to be imprecise, the price of cigarettes increasing by one percent is associated with demand in adjacent states decreasing by an average of 0.22 percent), ceteris paribus.

Of course, this is really weird, or should at least strike us as such if we hold any belief in our original bootlegging hypothesis of people traveling to other states if cigarettes are cheaper there. This is because the above mentioned result indicates the exact opposite, i.e., if the price of cigarettes in State A increases, *less* people will demand cigarettes in neighboring State B, where the price has *not* increased.

Distance Decay SLX

Since we know that results depend quite strongly on our choice of $\boldsymbol{W}$, it makes sense that **halleckvega2015<empty citation>** try a different $\boldsymbol{W}$ and we are asked to follow after them. So, in the following, we estimate an **SLX model with a distance decay specification**, where $w_{ij} = d_{ij}^{-\gamma}$ and $\gamma = 3$.

```
cigd <- read_excel("./assignment3/data/cigarettes/cigar_states.xls") %>%
  select(longitude, latitude) %>%
  as.matrix() %>%
  `rownames<-`(cign)

cigi <- distm(cigd, fun = distVincentyEllipsoid)

cigj <- (cigi/1000000)^(-3) %>%
  `diag<-`(0) %>%
  mat2listw(style = "B")
```

```
cm4 <- plm(
  logc ~ logp + logy + slag(logp, listw = cigj) + slag(logy, listw = cigj), #SLX w/ distance
decay
  data = cigs,
  effect = "twoways",
  model = "within",
  index = c("state", "year")
)
```

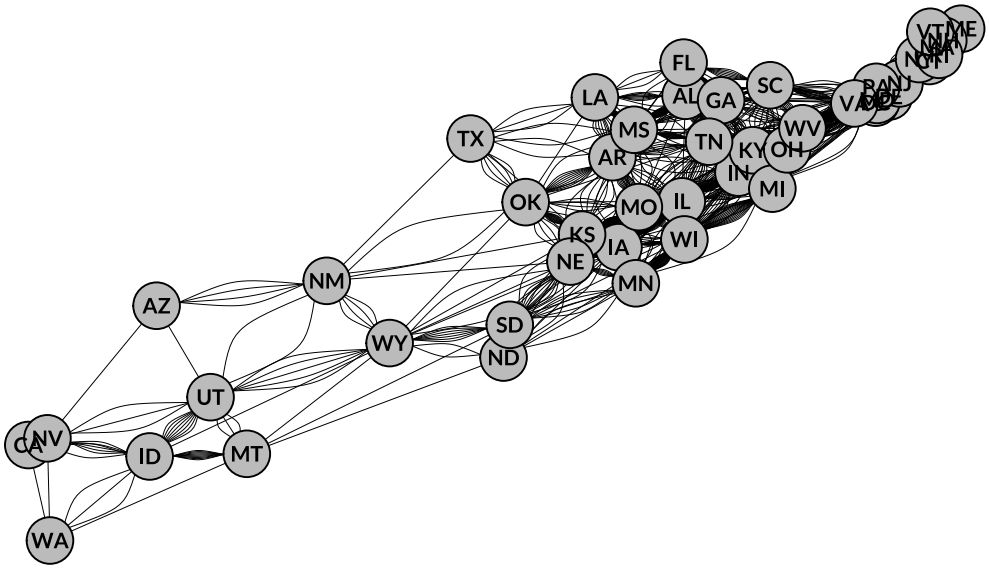
Again, the model results are printed below.

	Dependent variable:
	logc
logp	-0.900^{***}
	(0.038)
logy	0.642^{***}
	(0.042)
slag(logp, listw = cigj)	0.00004^{***}
	(0.00001)
slag(logy, listw = cigj)	-0.0001^{***}
	(0.00001)
Observations	1,380
R^2	0.520
Adjusted R^2	0.491
F Statistic	351.699^{***} (df = 4; 1301)
Note:	*p<0.1; **p<0.05; ***p<0.01

We can see that this time, we actually get the suspected bootlegging effect, as γ_1 is positive. This may indicate that ~~you get your desired results if you just try a sufficiently large number of specifications~~ the distance decay matrix fits the true DGP better.

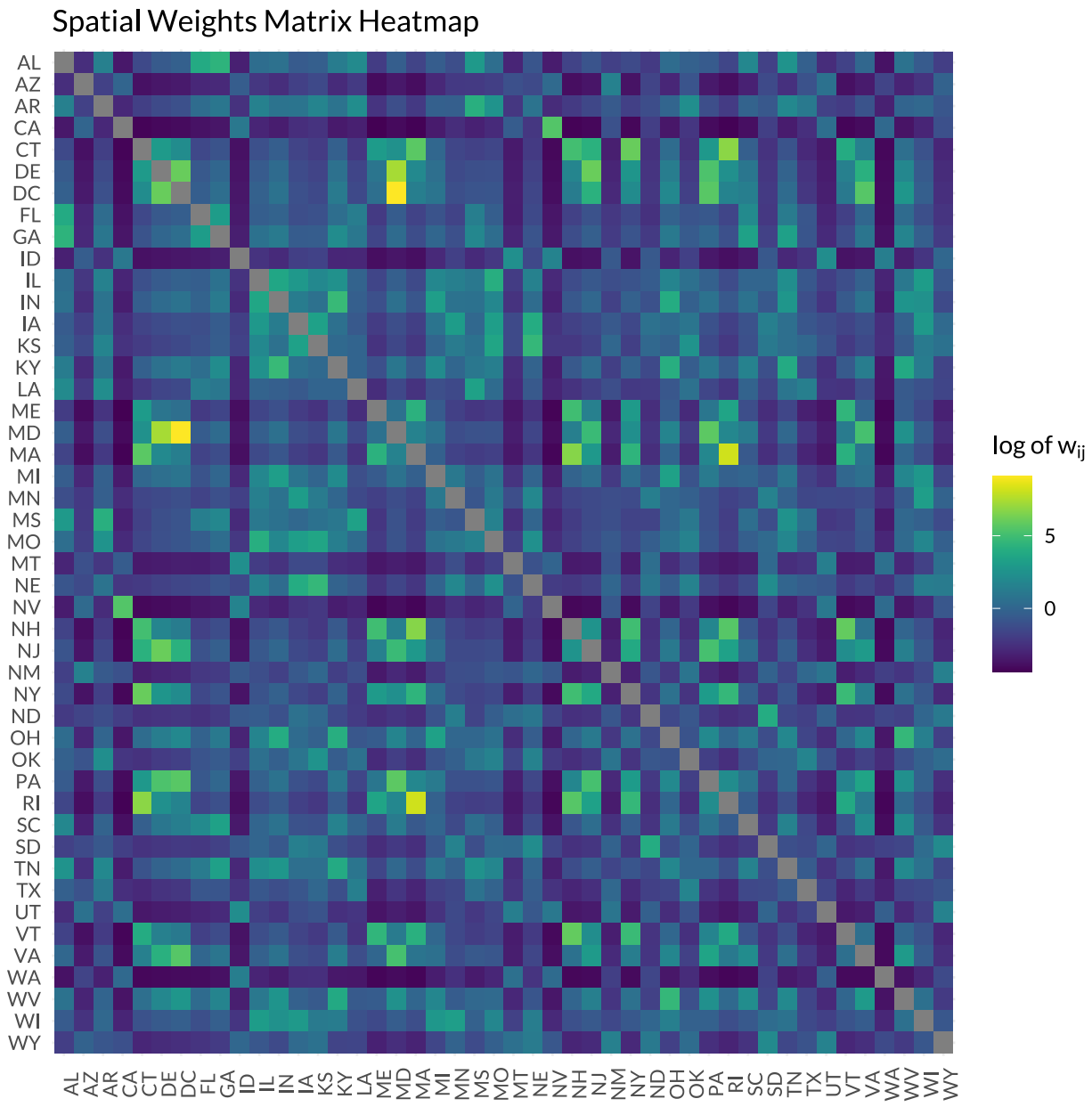
From a network perspective, this distance decay matrix has interesting implications. Our idea was that if there is a large number of edges, and these edges are weighted by the inverse of a power of the distance, then plotting these network must yield something that vaguely looks like the U.S. And it actually did:

```
plot(cigg,
     vertex.size = 10,
     vertex.color = "#BBBBBB",
     vertex.label.cex = 0.7,
     vertex.label.font = 2,
     vertex.label.family = "Lato",
     vertex.label.color = "black",
     edge.color = "black",
     edge.width = 0.5,
     asp = 0)
```



Since the plot is constructed anew every time we compile the document, we do not know how exactly it looks in our final submission. However, it should always resemble the true spatial position of the states recognizably—it may be flipped, it may be mirrored, it may not have north at the top, but the result should resemble the continental United States.

Then we thought about printing the weights matrix, only to find out that this is all but impossible on an A4 sheet of paper. So instead, here is a heatmap of the matrix (with logged values so that the scale appears more clearly):



Regarding centrality, we calculated two well-known measures: row sums (since our $\boldsymbol{W}$ is not row-stochastic) and eigenvector centrality. The results are printed below.

```
rowSums(cigk) %>%
  sort(decreasing = TRUE) %>%
  knitr::kable(col.names = "Row Sum")
```

	Row Sum
MD	11026.67198
DC	9945.46534
RI	4756.35048
MA	4728.87454
DE	2663.12139
CT	1913.64955
NH	1878.45853
PA	1082.00099
NJ	983.52610

	Row Sum
NY	968.47870
VT	746.96390
VA	615.20596
ME	386.75814
KY	333.66481
OH	308.21679
IN	303.99804
WV	271.93286
NV	253.46415
CA	248.41830
KS	215.91668
NE	209.03832
IL	207.81238
GA	201.84045
AL	193.97210
MO	191.43001
IA	185.93943
TN	176.79008
MS	160.89378
AR	140.08363
WI	127.73114
MI	123.99357
SC	100.11099
FL	98.10254
SD	90.44541
MN	85.93133
ND	71.63010
LA	70.08663
OK	60.44297
WY	38.15529
ID	35.44483
UT	32.25951
TX	25.43502
MT	25.15952
NM	24.47703
AZ	14.19491
WA	10.75984

	Eigenvector Centrality
MD	1.0000000
DC	0.9871728
DE	0.2112788
PA	0.0682453
VA	0.0486044
NJ	0.0310696
CT	0.0033553
WV	0.0030710
RI	0.0028691
NY	0.0025882
MA	0.0023642
OH	0.0013980
NH	0.0013860
VT	0.0008475
SC	0.0007672
KY	0.0006332
IN	0.0003951
MI	0.0003824

	Eigenvector Centrality
ME	0.0003685
GA	0.0002656
TN	0.0002394
AL	0.0000095
IL	0.0000043
WI	0.0000021
FL	0.0000017
MO	0.0000013
MS	0.0000012
IA	0.0000005
AR	0.0000005
MN	0.0000003
KS	0.0000002
LA	0.0000001
NE	0.0000001
OK	0.0000000
SD	0.0000000
TX	0.0000000
ND	0.0000000
WY	0.0000000
NM	0.0000000
MT	0.0000000
UT	0.0000000
AZ	0.0000000
ID	0.0000000
CA	0.0000000
NV	0.0000000
WA	0.0000000

Using both measures of centrality, Maryland is the most central state, followed by DC as the second-most central entity. The then following positions differ a bit between the two measures of centrality.

Regarding the question where spillover effects occur, let us take another look at the model equation of the SLX model:

$$\boldsymbol{y}_t = \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}_t\boldsymbol{\gamma} + \boldsymbol{\mu} + \phi_t\boldsymbol{\iota} + \boldsymbol{\varepsilon}_t,$$

The **spillover effect** we are interested in is described by γ_1, which in the equation is pre-multiplied by the first column of $\boldsymbol{W}\boldsymbol{X}_t$. In a sense, the first column of $\boldsymbol{W}\boldsymbol{X}_t$ determines where, or how strongly, spillover effects arise. Given a one-unit increase in every state's $\log(P_{it})$, this means that spillover effects of state i on state j will arise where w_{ij} is large. For an illustration, you can thus refer to the heatmap above.

The **average partial effect** is the effect that a (say, one-unit) change in an explanatory variable exerts on the dependent, averaged across all units. In a normal regression setting without spatial lags, this would be represented by the coefficient $\boldsymbol{\beta}$, which is constant over all i. However, in this case, changing the log income vector by one unit for all states i would exert influence over the demand for cigarettes via two channels: the "direct" channel, represented by the parameter β_2, and the "indirect" channel, represented by γ_2 times the i-th row of $\boldsymbol{W}$. The partial effect of a one-unit change in $\log(I_{it})$ for a certain i is thus

$$\mathrm{PE}_i = \beta_2 + \gamma_2 \sum_j w_{ij},$$

and the average partial effect of a one-unit change in $\log(I_{it})$ is then

$$\mathrm{APE} = \beta_2 + \gamma_2 n^{-1} \sum_i \sum_j w_{ij}.$$

Bonus question

First of all, it should be noted that the bonus question largely refers to the paper by **kuschnig2022<empty citation>**. Thereby, parts of it are replicated in this exercise to answer the question asked in the task assignment. In order to have the same initial data, the first step is therefore to reload the data used in this exercise in a way that is identical to the specifications used in the paper.

The following code is used to replicate the results, and therefore also the distance decay parameter, from the $SLX(\delta)$ model in the paper. δ thereby refers to the distance decay parameter γ. In order to maintain parity with the paper, we stick to naming the distance decay parameter as δ in the bonus question.

```
years_numb <- length(unique(cigs$year))

W_cont <- kronecker(diag(years_numb), cigw / rowSums(cigw))

dist <- as.matrix(dist(cigd))
diag(dist) <- Inf

Psi <- function(delta) {
  W_dist <- dist ^ (-delta) # Build
  W_dist <- W_dist / max(eigen(W_dist, symmetric = TRUE)$values) # Scale
  kronecker(diag(years_numb), W_dist)
}

y <- cbind(logc = cigs$logc)
X <- model.matrix(logc ~ logp + logy + year + state, data = cigs)
X_lag <- X[, c("logp", "logy")]
colnames(X_lag) <- c("w_logp", "w_logy")

X_cont <- W_cont %*% X_lag

n_save <- 500L
n_burn <- 100L

out_slxdx <- bslx(y ~ X, W = Psi, X_SLX = X_lag,
                  n_save = n_save, n_burn = n_burn, options = set_options(
                    SLX = set_SLX(delta = 3, delta_scale = 0.05, delta_a = 2, delta_b = 2)))
```

Following the paper and the R code provided on github, we calculate the number of unique years in the dataset to then create a spatial contiguity matrix by applying the Kronecker product to the identity matrix of the years (diag(years_numb)) and normalizing the weights. To create the inverse-distance decay function, we compute the pairwise distances between the states using the Euclidean distance, whereby the diagonal is set to infinity (diagonal elements will be 0) to prevent self-connections. The function *Psi* then takes delta as a parameter and defines the weights matrix $1/dist^{\delta}$, scales the matrix by the maximum eigenvalue to standardize it and applies the Kronecker product with the identity matrix of the years to expand for the panel data.

We then create a response variable matrix *y* with the log of cigarette consumption as well as construct the model matrix *X*, including logged cigarette prices, logged income, and fixed effects for year and state (as stated in the task assignment). Creating *X_lag*, we extract the relevant columns for the spatial lags of prices and income and calculate the spatially lagged explanatory variables by performing a matrix multiplication of *W_cont* with *X_lag*. In a next step, we fit the SLX model using the bslx function. We define the model formula, provide the function for the distance decay weights matrix, specify the explanatory variables for spatial lags, set the number of posterior draws, as well as provide settings/otpions for the SLX model following the specifications from the github code. In a next step, we fit the SLX model using the bslx function. We define the model formula, provide the function for the distance decay weights matrix, specify the explanatory variables for spatial lags, set the number of posterior draws, as well as provide settings/otpions for the SLX model (including prior parameters for the distance decay parameter) following the specifications from the github code. We specify the starting value for the decay parameter in the Markov Chain Monte Carlo sampler, which then iteratively updates this value based on the posterior distribution obtained through sampling. We also specify the prior distribution for δ, indicating the shape of the inverse-gamma prior distribution.

The following table shows the estimate (mean) for the distance decay parameter δ as well as a 99% confidence interval derived with the *coda* package:

```
## # A tibble: 1 x 4
##   Parameter              Estimate `Lower Limit (99%)` `Upper Limit (99%)`
##   <chr>                     <dbl>               <dbl>               <dbl>
## 1 Distance Decay (Delta)     3.04                2.76                3.52
```

The distance decay parameter derived using the out_slxdx output and therefore Bayesian inference quantifies the rate at which the spatial influence of a spatial variable decreases/decays over distance. This parameter is critical for modeling how changes in one location affect changes in another. In the context of the cigarette consumption model, the distance decay parameter indicates how cigarette sales or consumption in one state influence neighboring states, considering the spatial distribution of these states. A higher δ means that spatial influence decays more rapidly with distance, meaning that spatially close states have a much stronger influence on each other than those further away. We cannot be 100% sure what the exact value of the estimate will be when knitting the final Rmd file. The estimate, however, matches quite well the value from the paper by **kuschnig2022<empty citation>**. The estimated mean distance decay parameter ($\delta = 3.005$ when we last ran code and therefore when rounded equal to the 3.01 from the paper) thereby indicates the average of the posterior distribution of the parameter. The distance decay estimate signifies a relatively steep decline in spatial influence with distance, indicating that regions closer to each other have a significantly stronger effect on one another. This also makes sense intuitively in the context of cigarette consumption.

Task B

Describe the Unit of Observation

The unit of observation is defined in the paper as *grid cells* at a sub-national level in Africa, with each single cell as one unit of observation. Each of these cells within the grid system represents a certain area on the African map. For the analysis, a panel data set is used that comprises around 2700 cells in 46 African countries from 1997 to 2011, including various geo-referenced covariates. Within the units of observation (cells), different types of data are aggregated (conflict data, climate data, as well as supplementary data on infrastructure and ethnic fractionalization, to give just a few examples).

What are their areas?

Each of these cells has a resolution of 1 degree latitude $\times$ 1 degree longitude (which corresponds to about 110 km $\times$ 110 km at the equator, see also the next point on relative differences in sizes of the cells). The validation of the resolution (1 degree resolution) is carried out by the authors through a robustness analysis with higher and lower spatial scales, in which they look at both a higher resolution 0.5 $\times$ 0.5 degree grid as well as a lower resolution with a 2 $\times$ 2 grid in addition to the chosen 1 degree resolution. The aim is to find a good balance between the degree of localization of agricultural shocks and the detection of the extent of conflict spillovers. As can be seen from the online appendix, and what is interesting for the interpretation of the area of the units of observations, the authors in their analysis also calculate the area of each cell in square kilometers corresponding to the land, excluding seas and lakes. Furthermore, cell-level controls such as elevation, roughness, or altitude are also taken into account.

Why and to which (relative) extent do they differ?

One possible reason for potential relative differences in the areas of each cell is the spherical shape of the Earth. Since the earth is a sphere and not a two-dimensional surface, a 1 $\times$ 1-degree grid cell does not correspond to an area of 110 km $\times$ 110 km everywhere on the globe. As the Earth is a globe, the longitudes converge at the poles. The horizontal distance (from East to West, i.e. the width of a cell) from one longitude to the next longitude becomes smaller the further you move from the equator to the poles (to the North or South). However, the distance between the lines of latitude, which run parallel, remains roughly constant across the globe. So while around the equator a 1 degree $\times$ 1 degree cell area has about 12100 squared kilometers (110 km $\times$ 110 km), the area of a 1 degree $\times$ 1 degree cell deviates and decreases the further you move towards the poles (as the East-West distance of the cell decreases). As the equator runs through Africa, the areas of 1 $\times$ 1 degree cells in Central Africa are larger than those in North and South Africa (as these cells are not located on/around the equator but located further towards the poles, where eventually the width of a cell would converge to zero).

Furthermore, as taken into account in the analysis, the area of each cell in squared kilometers of land (and thus excluding seas and lakes), depends on the occurrence/proportion of standing or flowing waters in a cell. The presence of roughness, hills, or mountains can also influence the area surface. The areas can therefore differ depending on various conditions and environmental factors of a cell. The area of a 110 km $times$ 110 km square can, additionally and to mention a third reason, in general vary depending on the map and projection used. This is because distances and areas can deviate according to the projection used in an analysis due to distortion.

Un-normalized weights matrix and its effects on interpretations

The weights matrix *W* used in the paper is a binary contiguity matrix, which is not normalized. All cells surrounding the cell of interest (neighboring cells being defined as being within a distance threshold of 180km) are assigned a value/weight of 1 and all other cells are assigned a value/weight of 0. In this case, the 8 adjacent cells are effectively considered neighbors (this is also called queen contiguity). More precisely, the average cell in the sample has 7.4 neighbors (since the cells on the edges no longer have 8 neighbors). Due to the binarity of the matrix, there is no further variation in influence based on the exact distance (either a cell is a neighbor of the cell of interest and thus has influence or not). As the matrix is not normalized, the sum of the values/weights in each row is not equal to 1, as is the case in the usual row-normalized matrix. The influence of each adjacent cell is taken into account equally and the cumulative effects of all neighbors are summed directly in the analysis. The influence of a certain cell on the cell of interest therefore does not depend on how many other neighbors the

cell of interest has. The decision to use a non-normalized weights matrix was made by the authors to simplify the interpretation, as the coefficients of the spatial lags can be directly interpreted as the effects of a marginal change in a particular variable in one of the neighbors.

The spatial econometric model thus captures the total influence of those units that are considered neighboring, whereby each coefficient of a neighboring cell with a change can be interpreted as the total additional effect of this one other neighboring cell. Therefore, the coefficients on spatially lagged variables show cumulative and not averaged effects of changes in neighboring cells. In contrast, a normalized weights matrix (more precisely, a row-normalized weights matrix) shows the average influence of all neighbors. The coefficients for spatial lags can thus be interpreted directly as the effects of a change in a neighbor on the dependent variable in the cell of interest, and hence the authors' decision to use the chosen matrix W for simplicity reasons. Specifically, the coefficient for spatially lagged variables gives a direct interpretation for the effect on the cell of interest of a change in, say, conflict frequency or weather shocks in one or more neighboring cells.

What is also important to mention is that this un-weighted weights matrix tends to be sensitive to the number of identified neighbors. The cumulative influence of neighboring cells can potentially be more significant and the coefficients larger if the cell of interest has a higher number neighbors (more neighbors means higher cumulative influence, which then says very little about the average effect of a change in a neighbor). However, since the weights matrix W is binary (no further weighting) and most cells have exactly 8 neighbors (with the exception of border cells) as well as the fact that the average number of neighbors is quite close to the 8 neighbors at 7.4, this possible distortion due to different numbers of neighbors is moderate, a comparison of the cells is largely possible or the issue only needs to be taken into account, if at all, when a comparison is made with border cells.