Spatial Elements in Poisson Regression Using Bayesian Methods: Applying the Besag-York-Mollié Model

Wendy Olsen / Diego Perez Ruiz

August 2023

Department of Social Statistics

School of Social Sciences

University of Manchester

https://github.com/dapr12/SpatialRegressionBSPS2023 has this year's activities.
https://github.com/WendyOlsen/SpatialRegressionBayesIndia2022 has last year's activities.

www.socialsciences.manchester.ac.uk/social-statistics/

Contents

- First, an applied regression model.
- Samples of the 'research question'.

METHODS SECTIONS:

- 3rd, a Poisson model as a statistical distribution for the dependent variable, allowing generalised linear modelling with hierarchical elements [HLM refers to hierarchical linear model]
- 4th, a spatial element shown in BYM2 format
  - (Besag-York-Mollié v. 2)
- Conclusion: Summary plus • a tutorial task

‘Labour-Force Active’ People = 1, non-active = 0. Young people join the labour market…

- The debate about productivity has assumed that it is normal for men to work…but what about youths and children and women?
- We make a model of the risk of ‘being labour-market inactive’ vs.

Active in the labour market. Inactive is MUCH more common that Unemp.

- **Making a risk model is similar to looking at any problem,**

and the factors that raise the risk of that injury or illness happening.

  - Debating the components of the model on the right-hand side.
  - $\eta_j=\beta_{0j}+\underline{\mathbf{X\beta}}+$ **sex + social-group**
  - $\eta_j=\beta_{0j}+\underline{\mathbf{X\beta}}+$ **sex + social-group + spatial element (unexplained u_j)**
  - $\eta_j=\beta_{0j}+\underline{\mathbf{X\beta}}+$ **sex + social-group + spatially-varying Z + u_j**
- We call the regression coefficients the ‘slopes **(BOLD)**

Research Questions

- Define youth as age 15-24, or you may use 16-24 or 17-24.
- RQ1 What are the social factors (like religious group or minority-group) associated with being ACTIVE in the labour market as a youth?
  - We have decided to select India's youths up to age 24 for the moment.
  - Age can then be left out of the equation.
- $\eta_j = \beta_{0j} + \underline{\mathbf{X\beta}} +$ **sex + social-group + religious minority status +** $[spatial\ terms]$

Research Questions

- RQ1 First one must adjust for all the child and youth e to cultural norms around doing paid labour.

- $\eta_j=\beta_{0j}+\underline{\mathbf{X\beta}}+$ **+ age + soci**
  $[spatial\ term$

Perhaps this is the broad RQ, but it is **too general** to publish in a journal article!

Narrow down your RQ to generate testable hypotheses

Define ‘youth’ as an age group or a life-stage, eg premarriage. Reduce the scope of the study.

Choose one variable to focus upon.

A Narrower Research Question

- RQ 2 Via what routes does gender affect the risk of a youth in India being active/inactive in the labour market?
- …after taking into account spatial variations in norms
    - You could for example interact SEX with FORMAL EDUCATION

- $^{\eta}{}_{j}=\beta_{0j}+\underline{\mathbf{X\beta}}+$ **gender** + *interactions of gender with* {**social-group + religious minority status**} + $[spatial\ terms]$ + *gender*spatial terms*

Innovative Research Questions (A), Use ICC

- RQ3 What interaction effects occur, a) of gender by social group? After allowing for spatial covariation as an extracted factor; b) of gender by social group without allowing for spatial covariation?
- Model 0 is the empty hierarchical (spatial and s.g.) model, S.g.=SocGroup
- $\eta_j = \beta_{0j} + \underline{\mathbf{X\beta}}$+sex+age+s.g. + sex*s.g. $+ [spatial\ term] + residual \quad (\mathrm{Eq.}\,6)$
  - Gather the Intra-class correlation for the empty model and the above model
- $\eta_j = \beta_{0j} + \underline{\mathbf{X\beta}}$+sex+age+s.g. + sex*s.g. + $residual$ (Eq. 7))

  Now gather the ICC for the revised model without spatial terms

Innovative Research Questions (B): Add additional innovative variables

- You can add more variables
    - For example whether this **youth** is the eldest sibling of 2+ co-resident siblings.
    - And interact this variable with Sex of this youth.

  - Another possibility: interact sex with age, within 15-24 (Females 'down' males 'up')?
  - Risk of **bias/variance** trade-off (Kuhn & Johnson, chapter 5. Avoid Correlation of X with X)

    Increased bias from more parameters?! WATCH OUT.

  - Could add whether it is a single-parent family (BE CAREFUL).
  - Interact that with the sex of the lone parent. 0 = not single. F1=1 if a female headed household with 1+ children. M1 = 1 if male headed household with 1+ children

Summary of the Model and its Spatial Term (A) was to test whether this term *matters.*

- $\eta_j = \beta_{0j} + \underline{\mathbf{X\beta}} + \left[\left(\sqrt{\rho/s}\right)\varphi^* + (\sqrt{1-\rho})\theta^*\right]\sigma$

In general use existing theory to choose the X variates.

- This is **confirmatory regression.**
- Comparison of models is how we draw conclusions about specific hypotheses. Compare via AIC, BIC, or LR test

Use the likelihood-ratio test statistic. (LR test) Fox, chapter 15, section 15.1.1 in 2^{nd} edition.

- $\eta_j = \beta_{0j} + \underline{\mathbf{X\beta}} + [spatial\ term]$

Notation Used Here

- We follow Fox in offering the main linear part of the generalised linear model:
- $\eta_j = \beta_{0j} + \underline{\mathbf{X\beta}}$ where eta is the dependent variable measured suitably.
- B_{0j} is the intercept for risk and may not be interesting.
- **<u>XB</u>** is the estimate for **<u>Xβ,</u>** a vector multiplication. **<u>X</u>** holds several independent variables and **<u>Beta</u>** holds the slope coefficients.
- In an aggregated model at District units j, we will effectively have a randomly distributed District fixed effect – also called a random slope on district.
- In such a model, the intercept B_{0j} drops out.

The Besag-York- Mollié Model (v. 2) - More details of how we interpret it. See Morris et al., 2019. Very helpful.

- …+$\left[\left(\sqrt{\rho/_{s}}\right)\phi^{*} + (\sqrt{1-\rho})\theta^{*}\right]\sigma$
  Rho, ρ, measures the degree of correlation of the data from nearby and contiguous districts, such that when rho is large, the first term is larger and the second term is smaller. Rho runs from [0…1].
  - If the spatially correlated terms are greater, then RHO is larger, and if the spatially uncorrelated parts of the geography are greater, then RHO is closer to zero. (Morris, et al., 2019: 7)
- BYM introduced Phi and Theta. Here, in BYM2, ρ appears twice, with Rho weighting the two parts.
- Phi ϕ is 'spatial effects'.
  - Psi measures how the adjacency matrix is summarised using pairs of locations I and j; when phi is large, there is greater nearness, or greater contiguity, of the pair of locations (districts).
  - Psi is a square matrix, not a single vector, shown with * here.
- Theta θ is the heterogenous spatial effects mop-up term. "Independent Error Terms" (also * matrix)
- Sigma
  - scales-up the spatial part of the model to reflect the spatial terms' standard error. If it is large, then the spatial part plays a **greater part.**

References (1)

- Fox, John (2008), *Applied Regression Analysis and Generalized Linear Models*, London: Sage
- Kuhn, Max, and Kjell Johnson (2013), *Applied Predictive Modelling* (chapter 5 on the variance-bias tradeoff), London: Springer.
- Morris, Mitzi, K. Wheeler Martin, D. Simpson, S J. Mooney, A. Gelman, and C. DiMaggio (2019), Bayesian Hierarchical Spatial Models: Implementing the Besag-York-Mollié model in Stan, *Spatial and Spatio-Temporal Epidemiology*, 31, 100301.

Poisson Model – Used for a “Count”-Dependent Variable.

- In saying “Poisson” regression models, we mean a log Poisson distribution of the Dependent variable is a linear sum of terms.
- Poisson distributions have the feature that the mean of the distribution equals its variance, reducing key parameters from 2 to 1. But we add other parameters.
- A reference work for hierarchical, generalized linear models (GLM) by Fox 2008: Chapter 15) argues that the log Poisson model should be checked for overdispersion. We then have a quasi-Poisson model which many packages can estimate. I will present the version that expresses ‘exposure’ (space or time or population) and ‘risk’ λ.
- Sometimes, a log poisson model needs an overdispersion adjustment.
- An overdispersion adjustment is a multiplicative factor added as $c*\lambda$.
- λ is the risk of the event.

We fit the Poisson, for data X and Y and cases “i” within groups numbered j. Model the risk of Y.

- Dependent variable is count of workers in each social group in each district

    - Each district is marked j

    - You may add an offset (population of District) to get a better estimate.

    - Each individual youth in all households has subscript I initially
    - Aggregation is carried out to create groups with 0 or >0 counts of ‘active workers’
    - Weighted sums are used

There are ny successes in n trials. N(1-y) are fails, ie and the Poisson distribution is uses the factorial of Y.

$$p(y) = \mu^{y} * \frac{e^{x}}{y!} \text{ for } y = 0, 1, 2, \ldots \quad \text{(Eq. 1)}$$

Here when writing code, we inform Stan or R BRMS or lme4 that the <u>canonical link function is log and the family is Poisson</u>.

The expectation is $\mu_i = E(y_i)$ *(Eq. 2 using mu)*

The conditional variance of Y is $V(y_i|\eta_i) = \mu_i$ (Eq. 3 also mu)

Poisson(μ_i) = $\eta_i = \beta_0 + \underline{\mathbf{x\beta}}$ for β from 1 to k for k coefficients

This special situation can be tested for. See Fox 2008: Chap. 15.

Offset written into Poisson Model

- Poisson($c_j\lambda_i$) = η_i = β_0 + $\underline{\mathbf{x\beta}}$ for β from 1 to k for k coefficients,
- Where c_j is exposure in group j either on average or as a total or mean. The units of the count are in logs, so using populations we log the population of the group.

Poisson($c_j\lambda_j$) = η_j = β_{0j} + $\underline{x_j\beta_k}$ for β from 1 to k (Eq. 4) for k coefficients

- Aggregation from i to j groups for districts j for example, might give data as shown for regression:
  - This is a multilevel model with spatial groups. We often also group socially.
- When you remove the log, you get e^{c_j} as an additive offset.

How we could programme the Poisson fit for the count data using R with Stan

Results of a Poisson Model – Theory and Practice

- See our paper "A Bayesian Estimation of Child Labour in India" (*Child Indicators Research, A Bayesian Estimation of Child Labour in India, Kim, J. H., Olsen, W. & Wiśniowski, A., 2020, volume 13*)
- and its supplement.

- Or see more recent appendix Tables A3-A6 for WES paper:

https://github.com/WendyOlsen/normslabourindia

Notation here is from the *Child Indicators Research* paper

Having found the risk of 'child labour' rises rapidly from age 9 upwards, we modelled each year of Age.

Here a refers to the first dataset and b simply refers to a different dataset.

Types of Models	Model 1 IHDS (Poisson)	Model 2 NSS (Poisson)
Likelihood	$y.a_{ij} \sim \text{Poisson}(\mu.a_{ij} * n.a_{ij})$	-
	-	$y.b_{ij} \sim \text{Poisson}(\mu.b_{ij} * n.b_{ij})$
Overdispersion	-	-
	-	-
Prediction	$\hat{y}_{ij} \sim \text{Poisson}(\mu.a_{ij} * N_{ij})$ $\hat{y}_{i+} = \sum_i \hat{\boldsymbol{y}}_{ij}$	$\hat{y}_{ij} \sim \text{Poisson}(\mu.b_{ij} * N_{ij})$ $\hat{y}_{i+} = \sum_i \hat{\boldsymbol{y}}_{ij}$
Model for true child labour rate	$\log(\mu.a_{ij}) = \beta_0 + \beta_1 * x_i + \beta_2 * \log(z_{ij})$	$\log(\mu.b_{ij}) = \beta_0 + \beta_1 * x_i + \beta_2 * \log(z_{ij})$

Notes: i – age; j – state; τ is a precision (inverse variance). The *CIR* paper uses I for integer child—age-groups.

Z_{ij} reflects the population of people in that state, ie Census 2011 data on 'main workers' in that district = an offset.

Besag–York–Mollié terms added to model

- The BYM model adds terms which are built up on geographers' methods of measuring random errors that occur at a spatial level.
  - The correlated random errors are potentially located in contiguous ways.
  - The map of the places can be used to derive measures of contiguity, shared boundary, and distance from the equicentre of each district, but here in BYM, it is simply used for a Queen's matrix. This matrix is square, showing all district codes j on each edge, we do not examine it; from R mapping it is autocreated, with 1 where adjacent, and 0 where not adjacent; it is a sparse matrix.
  - We now add this term to the log Poisson GLM:

$$\eta_j = \beta_{0j} + \underline{X\beta} + \left[\left(\sqrt{\rho/s}\right)\phi^* + \left(\sqrt{1-\rho}\right)\theta^*\right]\sigma$$

(Eq. 5)

$The\ Phi\ matrix\ \phi^*$ is in units of Eta_j and it has a triangle of estimates for districts j with all other districts $k \neq j$, in the J-k matrix. We can sort that matrix.

- **Reminder**: This BYM2 model has spatial smoothing on contiguous and nearby places, as well as a random-effects component which helps with fitting the heterogenous spatial amounts of risk.
- The rho factor varies between 0 to 1, rising with greater correlated local risks in groups of areas j.
- The phi factor ϕ is an ICAR model to take care of correlated risks, spatially, which are contiguous. ICAR means intrinsic conditional autoregressive models. Rho appears twice in the BYM2 formula.
  - 1st term is for the spatial explanation via contiguous similar districts. 2nd term is for the unusual districts, ie heterogeneity.
- The factor s is an adjustment that can scale up/down the first term.
- Lastly, a scaling factor ***sigma*** allows the model to upscale the spatial term.

$$\eta_j = \ldots \left(\sqrt{\rho/s}\right) \phi^* + \cdots$$

(Eq. 5)

AN ILLUSTRATION

- Illustration of the key term in Besag-York-Mollié model
  - 1st term reflects +'ve correlated contiguous similar districts.
  - **Desertification** occurs in contiguous rural districts! SO desertification would cause all those rural areas to have LOW labour-force participation. Rho would be high but so would PHI for those areas. PHI* would have clumps of 5 districts with deserts.

Conclusions

- A series of linear terms embedded in a Poisson model can attribute risk to competing factors. Some can have interaction terms.
- The fit of the model can be assessed using Bayesian Information Criterion (BIC), related measures AIC, nested-models LR test and MCMC type tests.
- The tutorial involves amending an existing model (MC model, ie a MCMC estimation)
  - Install R with packages stan, stanarm and tidyverse;
  - Step 1 run the program using the code provided
    - (see https://github.com/WendyOlsen/https/SpatialRegressionBayesIndia2023)
    - All-India data are also provided here for age 16-31; in the 2022 version in github data are for 15-24 years.
    - You must only do non-commercial work with these Indian Periodic LFS data.
  - Step 2, replace the variable 'rural' with 'your age dummy' which is an R factor or numeric 0/1 binary; and run it again, interpret.
  - Step 3, test a more complex hypothesis. Good luck!

References (2, more inclusive)

- Besag, J.J.Y., Mollié, A. (1991), Bayesian Image Restoration with Two Applications in Spatial Statistics, *Ann. Inst. Stat. Math.* 43, 1-59,10.1007.
- Fox, John (2008), *Applied Regression Analysis and Generalized Linear Models*, London: Sage.
- Kim, Jihye, Olsen, W.K. and A. Wisniowski (2022) Predicting Child-Labour Risks by Norms in India. *Work, Employment and Society*. doi:10.1177/09500170221091886
- Kim, Jihye, Olsen, W.K. and A. Wisniowski (2020), A Bayesian Estimation of Child Labour in India, *Child Indicators Research*, DOI https://doi.org/10.1007/s12187-020-09740-w.
- Kuhn, Max, and Kjell Johnson (2013), *Applied Predictive Modelling* (chapter 5 on the variance-bias tradeoff), London: Springer.
- Morris, Mitzi, K. Wheeler Martin, D. Simpson, S J. Mooney, A. Gelman, and C. DiMaggio (2019), Bayesian Hierarchical Spatial Models: Implementing the Besag-York-Mollié model in Stan, *Spatial and Spatio-Temporal Epidemiology*, 31, 100301.
- Olsen, Wendy, Manasi Bera, Amaresh Dubey, Jihye Kim, Arkadiusz Wisniowski, Purva Yadav (2020). Hierarchical Modelling of COVID-19 Death Risk in India in the Early Phase of the Pandemic, *European Journal of Development Research.* DOI https://link.springer.com/article/10.1057/s41287-020-00333-5

Practice activity suggestions.

- You can carry out a tutorial activity based on data in our github site.
- https://github.com/dapr12/SpatialRegressionBSPS2023
- Note that we keep updating both of these so please, download the whole thing at once. Use the ‘readme’ file to see how to download it.
- Our tutorial documents are aimed at All-India estimates, or a youth age-group.
- You can do all-age estimates or study any age group from 15 to age 100 years.
- The dates of data are 2017/8.

Tutorial tasks – carry out and tick them off.

- ____Preparation: Before or early in our workshop, please choose one of the tasks and run the provided R file. We've named them MC for monte carlo, ICC for intraclass correlation, and BYM for Besag-Yorke-Mollie. Your personal machine may require updating R to R4.2.2 or improvements in your Rstudio or Java updates. We found R4.1 isn't going to work for the ICC, BYM.
    - Note: You may need to get help with folder and file management. You can always email us for advice. Wendy.Olsen@manchester.ac.uk at your service.
    - R-Carpentry is a useful website if you are at an early stage with R.
    - DataCamp is very helpful too. We recommend a small monthly payment if you wish to engage in full broad learning. (£5 student fee)

You may want to find or **create an 'age' factor in R**. Age can be 0=16-19 and 1=20-24, for example. Our hypothesis is that those who are older are more likely to work, but that the association is gender-specific. Men more so, women less so. Many people get married in this period and the gender impact on labour-force participation is reversed. (Discuss). A 'factor' is not a numeric variable in R. So, when you want a correlation matrix, you will have to start again with age as numeric. If we provide ages 16-30 then you might have a factor with four levels 16-19… etc.

- Preferred model: OMIT RURAL/URBAN, INCLUDE AGE 0/1 and age-squared. Here age is numeric, not a factor. (Not dummies – You choose)
- You need to use either a logit regression model or a Poisson model in brms to get initial results (see MC….R program file)
- You may add a household monthly spending variable.
    - Household monthly spending is shown as mpce (monthly per capita in Rupees); log of it using Exp() is logRsincpc. Strongly + correlated with youths working.

- ____ **Run the preferred model** as Poisson model, given in the code files. You can run empty models and those with the given variables (sex and age). Optionally, also include an interaction effect sex*age. You can also run lm from lme4 and the logistic model. Compare results.
- ____ Task 2: Everyone may want to read the ICC program file. See if you can understand the ratio which is the ICC for one model.
- ____ Task 3: Do you have a decent fit overall? **What are your 'fit statistics**', ie measures of goodness of fit? Make a simple table (homework, as this takes time). How do you read and assess these? What is your interpretation?
- ____ Task 4: Now **make a table showing the ICC part** for 4 models. This requires good R skills. There are several ways to do it. We will show some results to you.
- DO THIS LATER ____ Task 5. Run the BYM model program file .R. Your PC folder structure needs to contain the variety of source files: Map files, district codes, our program code file, the 'icar' functions (intrinsic conditional autoregressive model) that stan needs, and the Stan model for the MCMC simulation. In the case of Icar functions, we need to access both the R file and the Stan file, notice the different file-surnames .R and stan. SIMPLY DOWNLOAD THE WHOLE GITHUB REPOSITORY to get the right file structure.
- TASK 5 is likely to work best at home after our workshop. Hence we have a discussion section in the workshop to meet other scholars. Arrange study-buddies! Take down emails!

Not too sure? Read this background paper – takes 2 to 3 hours: Morris, Mitzi, K. Wheeler Martin, D. Simpson, S J. Mooney, A. Gelman, and C. DiMaggio (2019), Bayesian Hierarchical Spatial Models: Implementing the Besag-York-Mollié model in Stan, *Spatial and Spatio-Temporal Epidemiology*, 31, 100301.

Alternative: just gla
day. We make R log
Have fun! In doubt? Use R Cookbook online, or R for Data Science, 2nd ed., also online and both are
free

The definition of being economically ‘Working’ refers to economic activity more broadly than just salaried work and self-employment:

- In India, around 7% to 9% of adults have a job in the sense of a formal or verbal regular contract, and so we use the routine below to gather ‘narrow’ [salaried and daily-cash paid employment], ‘medium’ [adds to that the self-employment, marginal farmer, farmer, business, and helper roles], or ‘wide’ work [includes domestic production]. Our paper by Dubey et al., 2017 affirms this way of labelling it. See Dubey, A., Olsen, W., & Sen, K. (2017). The Decline in the Labour Force Participation of Rural Women in India: Taking a Long-Run View. *The Indian Journal of Labour Economics*. https://doi.org/10.1007/s41027-017-0085-0
- The routine is written in Stata style. See URL
- tab status_principal ; gen narrowwork=0
- replace narrowwork=1 if status_principal ==31 | status_principal == 51 |status_principal== 41 | status_principal==12
- gen medwork=0
- replace medwork=1 if status_principal ==31 | status_principal == 51 |status_principal== 41 | status_principal == 12 ///
- | status_principal ==11 | status_principal== 21

gen widework=0

- replace widework=1 if status_principal ==31 | status_principal == 51 |status_principal== 41 | status_principal == 12 ///
- | status_principal ==11 |status_principal== 21 ///
- | status_principal == 93
- *& IF THE PERSON IS NOT OTHERWISE EMPLOYED,
- replace widework=1 if widework==0 & status_sub==93
- The last line refers to the subsidiary, not principal, status.

*NOTE: status_principal == 81 omitted above, refers to unemployment. Technically, in government and ILO, we refer to the unemployed as economically <u>active</u>, but we have not allowed it here.

Source: Ministry of Statistics and Programme Implementation (MOSPI), Government of India, Annual Report 2017-18, URL https://www.mospi.gov.in/sites/default/files/publication_reports/Annual%20Report%2C%20PLFS%202017-18_31052019.pdf, accessed June 2023. Includes: “Periodic Labour Force Survey *(PLFS) Final Multiplier-posted unit-level data for Schedule- 10.4 of PLFS,* A) Unit level data for the first visit and re-visit of Sch. 10.4 [Periodic Labour Force Survey].; There are 4 data files for each of 4 Quarters (July 2017 - June 2018). Details of data layout is given in Data_LayoutPLFS.XLS.“ “Codes for Block 5.1:

Principal status: worked in h.h. enterprise (self-employed): own account worker 11, employer 12, worked as helper in h.h. enterprise (unpaid family worker) 21; worked as regular salaried/ wage employee 31, worked as casual wage labour: in public works 41, in other types of work 51; did not work but was seeking and/or available for work 81, attended educational institution 91, attended domestic duties only 92, attended domestic duties and was also engaged in free collection of goods (vegetables, roots, firewood, cattle feed, etc.), sewing, tailoring, weaving, etc. for household use 93, rentiers, pensioners , remittance recipients, etc. 94, not able to work due to disability 95, others (including begging, prostitution, etc.) 97. “

Tutorial Table basics from https://github.com/dapr12/SpatialRegressionBSPS2023

• File names	How long it takes to run them
• ‘script1cleanandmerge.R’- optional	Skip ‘script1cleanandmerge.R’ if you wish.

Tutorial Table Guidance (data https://github.com/dapr12/SpatialRegressionBSPS2023)

- Table 1
- Define your models by number.
  - Eg 0 empty Poisson 1 Poisson with variables 2 Poisson with variables with spatial BYM2 term
  - 3 a logistic model is optional
  - Perhaps 4 empty model with spatial BYM2 if you wish.
- Headings:

  Model 0 1 2 3 4 5 6
  - Constant term if any
  - Coeff Age
  - Coeff Sex
  - Coeff Age*sex
  - Mean of the random effects j
  - Variance of the random effects j
  - Covariances must not be in this table!

Use BIC if available

Bayesian Information criterion, Akaike information criteria, Deviance information criterion would all be acceptable goodness-of-fit measures for these non-nested models.

- Table 2
- Use the same models.
- Choose only 4 of them
  - We want to see:
  - Empty Poisson model with BYM
  - Poisson model with BYM without interaction effects
  - Poisson model with BYM with interaction effects
- Table headings:

  Model 0 1 2 3

  Coefficients

  Constant

  X1

  X2

  X1*X2

  Bayesian Info Criterion BIC

  ICC κ

  Variance of the whole model (var(Y-hat)) ν

  Variance not covered in ICC $\nu - \kappa$

Quiz to discuss spatial and BYM research

- Form a small group
- Introduce yourselves (Name, Place)
- What is the Count outcome in your research?
  - Or do you use linear regression?
- What would 3 X variables be? (call this level 1 variables)
- What is the available Spatial Unit?
- Do you have any aggregate variables X_j ? (e g GDP per capita, measured at the spatial-unit level – Call that level 2.

There is only time to develop one person's project! The rest are listening/commenting.

Thank you!