Bayesian Nonparametric Priors for Model-Based (Partial) Clustering

with Applications in Macroeconomics, Land-use Subsidies, and Health

Gregor Kastner

Department of Statistics, U Klagenfurt, Austria

January 28, 2026

Outline

- Part I: Unemployment in Italy – Spatio-Temporal Clustering (Mozdzen et al., 2022)
- Part II: Partial Clustering for EU Agricultural Subsidies (Mozdzen et al., 2024, major update incoming)
- Part III: Repulsive Mixture Weights and the Sparsity-Inducing Partition Prior (Mozdzen et al., 2025)

Part I: Spatio-Temporal Clustering

Panel Regression

Model the relationship between a response variable y and $k = 1, .., K$ explanatory variables in $\boldsymbol{x}$:

$$y = \beta_0 + \boldsymbol{x}\boldsymbol{\beta} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

observing $i = 1, ..., N$ areal units (states, provinces):

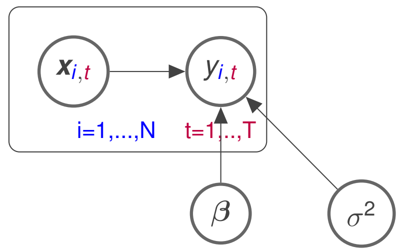
$$y_i = \beta_0 + \boldsymbol{x}_i\boldsymbol{\beta} + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

for $t = 1, ..., T$ years:

$$y_{i,t} = \beta_0 + \boldsymbol{x}_{i,t}\boldsymbol{\beta} + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim \mathcal{N}(0, \sigma^2)$$

$\rightarrow$ collection of time series of areal units

$$\begin{bmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,T} \end{bmatrix} = \begin{bmatrix} x_{1,i,1} & x_{2,i,1} & \cdots & x_{K,i,1} \\ x_{1,i,2} & x_{2,i,2} & \cdots & x_{K,i,2} \\ \vdots & & & \\ x_{1,i,T} & x_{2,i,T} & \cdots & x_{K,i,T} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} \epsilon_{i,1} \\ \epsilon_{i,2} \\ \vdots \\ \epsilon_{i,T} \end{bmatrix}$$



Conditional Autoregressive (CAR) Distribution

How can we incorporate the spatial and temporal structure?

CAR distribution:

- Random effects of neighbours are correlated, while non-neighboring areas are modeled as independent, conditionally on the remaining $\boldsymbol{w}$
- Define precision Q according to Leroux et al. (2000) as a weighted average of spatially dependent and independent correlation structures:

$$Q(\rho, W) = \rho(diag(W\mathbf{1}) - W) + (1-\rho)\mathbb{I}_N$$

where ρ is a spatial autoregressive parameter

$$\boldsymbol{w}_t | \boldsymbol{w}_{t-1} \sim N_I\big(diag(\xi)\boldsymbol{w}_{t-1}, \tau^2 Q(\rho, W)^{-1}\big), \quad t = 2, \dots, T$$

- The joint distribution $p(\boldsymbol{w})$ is Gaussian,
- Depends only on its neighbors

$\rightarrow$ **Sparse Gaussian Markov Random Field (GMRF)**

Adding the Spatio-temporal Random Effects

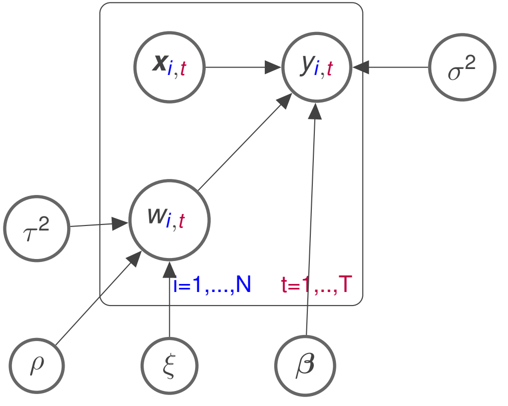
Adding $w_{i,t}$ we obtain the following model:

$$y_{i,t} = \beta_0 + \boldsymbol{x}_{i,t}\boldsymbol{\beta} + w_{i,t} + \epsilon_{i,t}$$

Together with $w_{i,t}$ we added:

- spatial autoregressive parameter ρ
- temporal autoregressive parameter ξ
- scale parameter τ^2

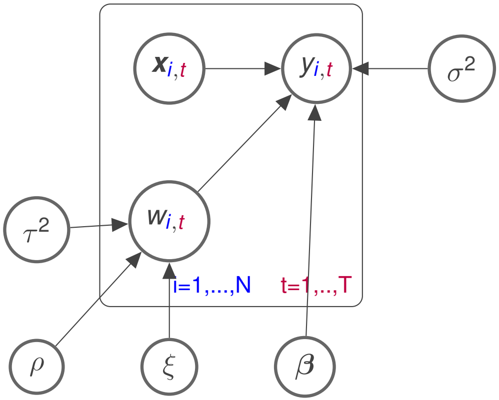
How to estimate all these parameters?



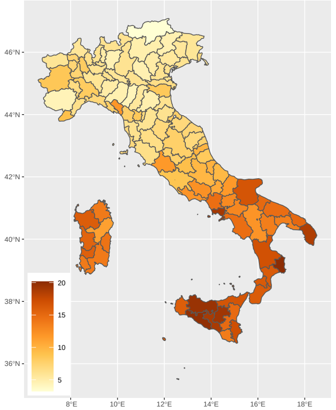
Bayesian Hierarchical Model

Typical priors:

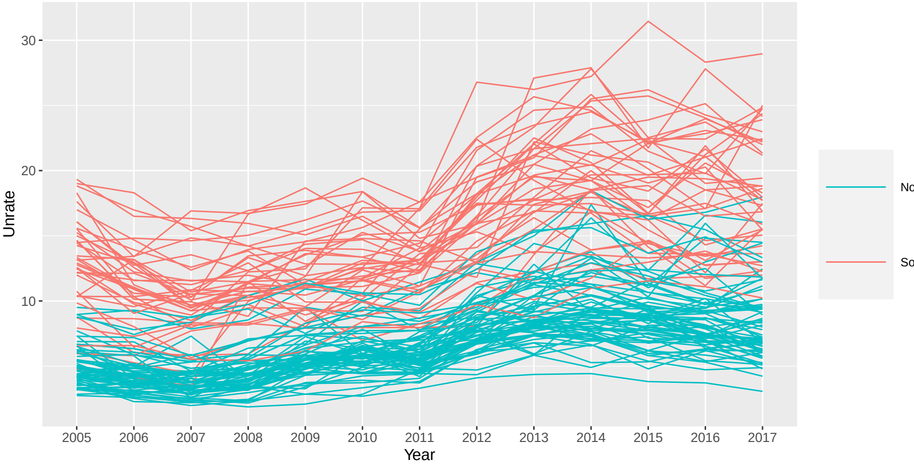
$$\sigma^2, \tau^2 \sim \mathit{InverseGamma}(3, 2)$$
$$\rho \sim \mathit{Gamma}(6, 1)$$
$$\boldsymbol{w_t} | \boldsymbol{w_{t-1}}, \xi, \tau^2, \rho \sim \mathcal{N}\left(\xi \boldsymbol{w_{t-1}}, \tau^2 Q(\rho, W)^{-1}\right)$$
$$\boldsymbol{w_1} | \tau^2, \rho \sim \mathcal{N}\left(\mathbf{0}, \tau^2 Q(\rho, W)^{-1}\right)$$
$$\beta \sim N(\mathbf{0}, \mathbb{I}_K)$$
$$\xi \sim \mathit{Beta}(1, 1)$$



MOTIVATION: ITALIAN UNEMPLOYMENT RATES



Averages over time



Time series plot

There appears to be potential to cluster the regressors (including the intercept) as well as the temporal effects!

THE CHINESE RESTAURANT PROCESS (CRP)

- Customers enter and choose a table, their seating pattern defines a partition $\pi_{[n]}$.

$$P(\text{customer i joins table } k|\pi_{[n]}) = \begin{cases} \frac{n_{k,-i}}{\alpha+n-1} & \text{if } k \in \pi_{[n]}, \\ \frac{\alpha}{\alpha+n-1} & \textit{otherwise}. \end{cases}$$

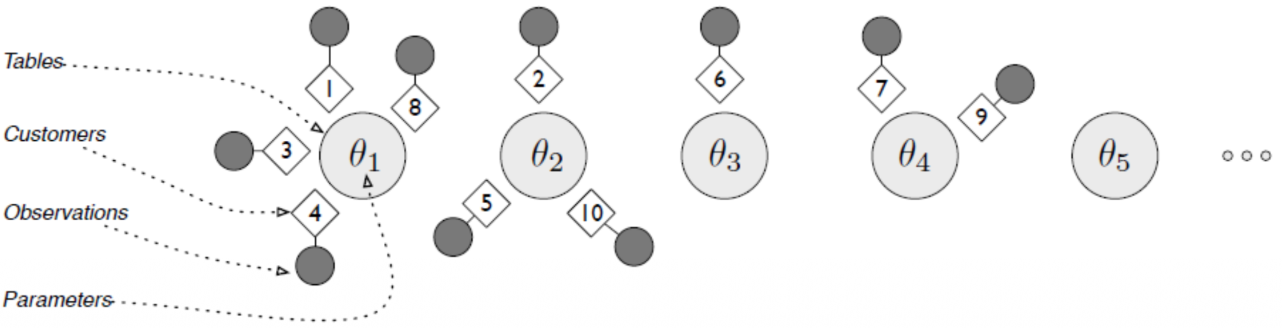


Figure adopted from Gershman and Blei (2012)

The Chinese restaurant process (CRP)

- Each table k is associated with a cluster and a cluster parameter $\boldsymbol{\theta}_k$ drawn from prior G_0
- Since the partition $\pi_{[n]}$ is random, the CRP defines a distribution on partitions

$$\boldsymbol{\pi}_{[N]} \sim CRP(\alpha, N)$$
$$P(\boldsymbol{\pi}_{[N]}) = \frac{\alpha^K}{\alpha^{(N)}} \prod_{k \in \boldsymbol{\pi}_{[N]}} (|k| - 1)!, \quad \alpha^{(N)} := \alpha(\alpha+1)\dots(\alpha+N-1)$$

- We can use the CRP to define a mixture model:

$$\pi_{[N]} \sim CRP(\alpha, N)$$
$$\theta_k | \pi_{[N]} \sim G_0 \quad \text{for } k \in \pi_{[N]}$$
$$y_i | \theta, \pi_{[N]} \sim F(\theta_k) \quad \text{for } k \in \pi_{[N]}, i \in k$$

- The CRP distribution is invariant to ordering (probability only depends on the size of the clusters and α) $\rightarrow$ *exchangeability* property

What is the random measure underlying the CRP?

The Dirichlet Process (DP)

The DP is a completely random measure with

- atoms ϕ_k drawn from a base distribution G_0:

$$\phi_k \stackrel{i.i.d.}{\sim} G_0 \quad k = 1, 2, ...$$

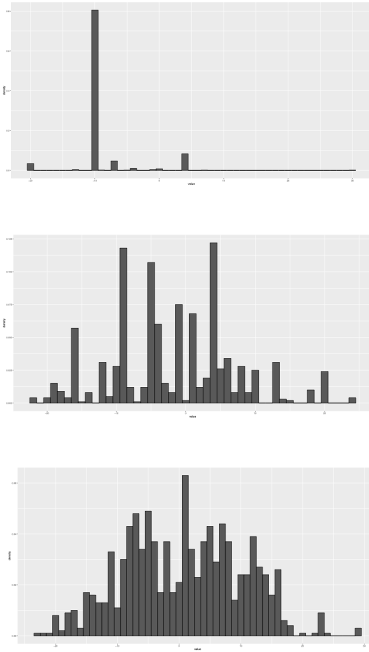
- weights w_k drawn according to the stick-breaking process by Sethuraman (1994).

$$G = \sum_{k=1}^{\infty} w_k \delta_{\phi_k}$$

Ferguson (1973) defined the DP as random measure $G \sim DP(\alpha, G_0)$ that satisfies:

$$(G(A_1), ...G(A_K)) \sim \mathcal{D}(\alpha G_0(A_1), ...\alpha G_0(A_K))$$

- α is the concentration parameter
- G_0 is the base measure

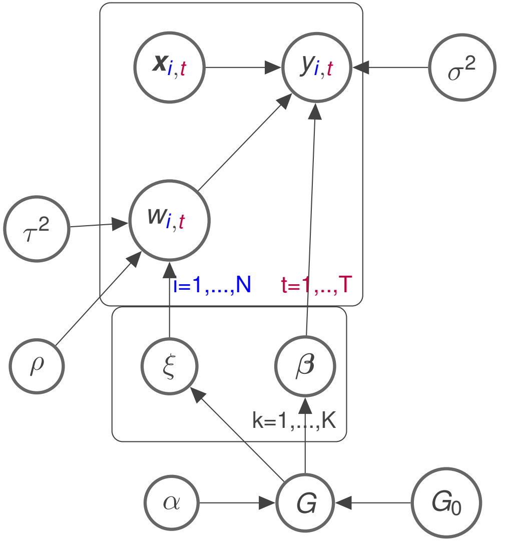


Draws from a DP for small, medium and high α

The complete model

New Model:

$$\sigma^2, \tau^2 \sim InverseGamma(3, 2)$$
$$\rho \sim Gamma(6, 1)$$
$$\boldsymbol{w_t} | \boldsymbol{w_{t-1}}, \boldsymbol{\xi}, \tau^2, \rho \sim \mathcal{N}\left(\boldsymbol{\xi} \boldsymbol{w_{t-1}}, \tau^2 Q(\rho, W)^{-1}\right)$$
$$\boldsymbol{w_1} | \tau^2, \rho \sim \mathcal{N}\left(\mathbf{0}, \tau^2 Q(\rho, W)^{-1}\right)$$
$$G \sim DP(\alpha, P_0)$$
$$\alpha \sim Gamma(3, 1)$$
$$P_0 = \mathcal{N}\left(\boldsymbol{\mu}_0, \Sigma_0\right) \times Beta(\alpha_0, \beta_0),$$
$$\boldsymbol{\beta}, \boldsymbol{\xi} \sim G$$



Navigating a complex Clustering Landscape

How to choose a single partition from our posterior draws $\boldsymbol{s}^{(1)}, \ldots, \boldsymbol{s}^{(M)}$?

- Variable number of clusters and label switching pose challenges for posterior analysis
- Past solutions include
  - Using identifiability constraints (Frühwirth-Schnatter, 2001; Richardson & Green, 1997)
  - Reordering algorithms (Papastamoulis & Iliopoulos, 2010; Rodríguez & and, 2014)

Decision-theoretic approach

- Summarize posterior by minimizing suitable loss function L

$$\boldsymbol{s}^{\star} = \underset{\hat{\boldsymbol{s}}}{\operatorname{argmin}}\, \mathbb{E}\left(L(\boldsymbol{s}, \hat{\boldsymbol{s}}) \mid \boldsymbol{X}\right) \approx \frac{1}{M} \sum_{h=1}^{M} L(\boldsymbol{s}^{(m)}, \hat{\boldsymbol{s}} \mid \boldsymbol{X})$$

  where M is the number of MCMC iterations, $\boldsymbol{s}^{(m)}$ a sample of the allocation vector
- Prominent choices for L include the Binder loss (Binder, 1978) and the Variation of Information (Meilă, 2007) as well as generalizations therof (e.g., Dahl et al., 2022)

Italian Unemployment Data

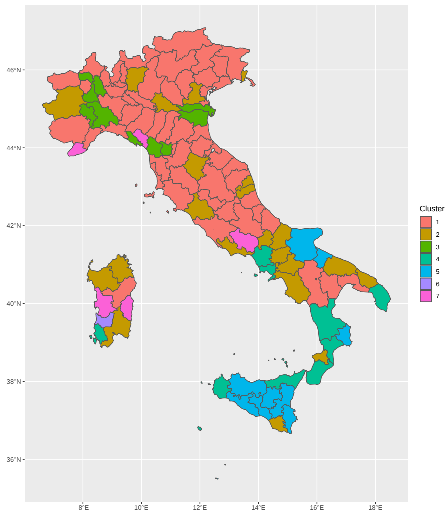
Data

- Obtained from the Italian National Institute of Statistics (ISTAT)
- Areal units: Italian provinces (i.e., $I = 110$)
- Time period from 2005 to 2017 (i.e., $T = 13$)
- Quantity of interest: unemployment rate

Covariates:

- *agri*: employment in agriculture sector
- *ind*: employment in industrial sector
- *serv*: employment in service sector
- *cons*: employment in construction sector
- *popdens*: population density
- *partrate*: indicator of willingness to work
- *empgrowth*: employment growth

Some Results



Chosen cluster allocation

Posterior means of the regression coefficients β within each cluster conditional on the chosen cluster allocation

Cluster (size)	intercept	agri	ind	cons	serv	partrate	empgrowth	lpopdens
Cluster 1 (55)	-0.282	-0.402	-0.947	-0.201	-0.784	-0.090	-0.056	0.057
Cluster 2 (22)	0.056	0.433	0.620	0.157	0.755	0.022	-0.030	-0.002
Cluster 3 (9)	-0.536	-0.150	-0.092	0.036	-0.050	0.764	-0.212	0.114
Cluster 4 (9)	1.924	0.042	0.590	-0.226	0.459	0.633	-0.171	0.130
Cluster 5 (9)	2.786	0.119	0.299	0.067	0.318	1.034	-0.166	-0.251
Cluster 6 (1)	1.268	-1.710	-0.108	0.227	0.491	-0.064	-0.314	-1.448
Cluster 7 (5)	0.481	-0.687	-1.151	-0.344	-1.021	0.658	-0.163	-0.319

Spatio-temporal Random effects

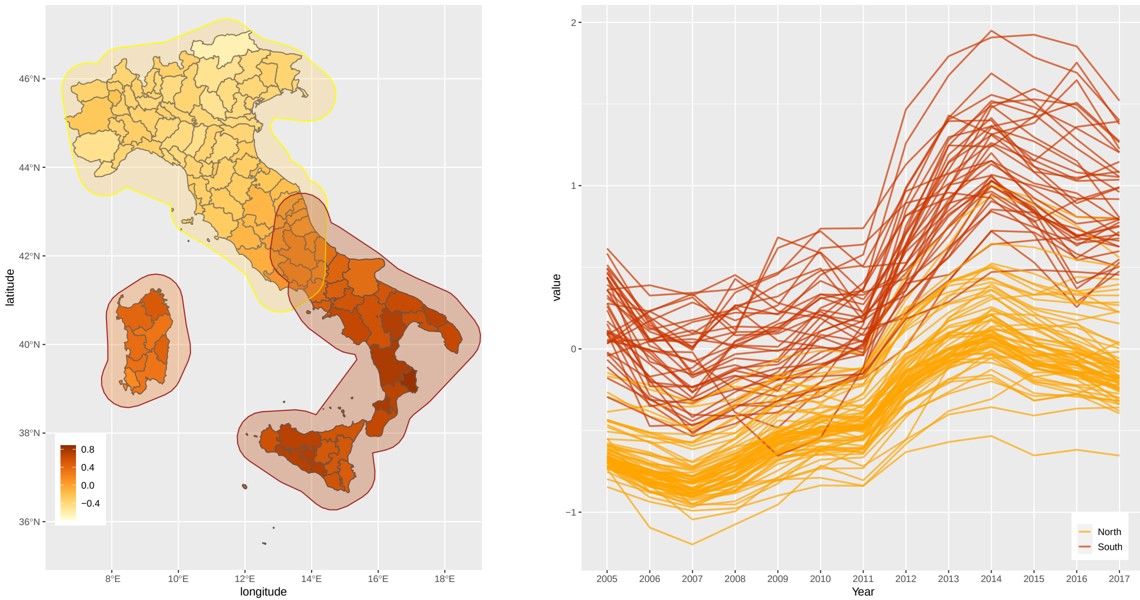


Figure. Average estimated spatio-temporal random effects (left panel) and the estimated time series for each province (right panel). The yellow and orange outlines on the map delineate the northern and southern Italian provinces, respectively.

Model Comparison

Out-of-sample MAE of the competing models and the Bayesian spatio-temporal clustering model

Model	2009	2010	2011	2012	2013	2014	2015	2016	2017	Average
Pooled	0.957	0.865	0.775	**0.428**	0.386	0.524	0.450	0.477	0.482	0.594
IFE	0.254	0.238	**0.196**	0.474	0.554	0.578	0.400	0.373	0.421	0.387
Pooled-SAR	0.403	0.369	0.327	0.672	0.712	0.756	0.523	0.557	0.513	0.537
IFE-SAR	0.371	0.280	0.254	0.498	0.524	0.502	0.320	0.328	0.345	0.380
PS-ANOVA	0.287	0.241	0.216	0.463	0.379	**0.317**	0.306	0.334	0.329	0.319
PS-ANOVA-SAR	0.287	0.241	0.216	0.463	0.379	**0.317**	0.306	0.334	0.329	0.319
ST.CARar	0.295	0.268	0.323	0.617	0.730	0.832	0.264	0.267	0.596	0.466
BSTC	**0.237**	**0.232**	0.216	0.499	**0.350**	0.356	**0.239**	**0.241**	**0.284**	**0.295**

Logarithms of the one-step-ahead predictive likelihoods, their sum and the WAIC of the Bayesian models

Model	2009	2010	2011	2012	2013	2014	2015	2016	2017	Sum	WAIC
ST.CARar	-63	-54	-43	**-89**	-134	-181	-70	-86	-140	-861	2013
BSTC	**-56**	**-37**	**-42**	-172	**-66**	**-125**	**-63**	**-69**	**-107**	**-737**	**-737**

Infinite vs. Finite Mixture Models

Computationally powerful representations:

- Chinese Restaurant Process
- Pólya Urn Model (Blackwell & MacQueen, 1973; Sethuraman, 1994)

Paving the way for a myriad of efficient sampling algorithms:

- Conjugate sampler (Escobar & West, 1995; Lo, 1984; Neal, 2000)
- No-Gaps Sampler by MacEachern and Müller (1998)
- Neal's Algorithm 8 (Neal, 2000)

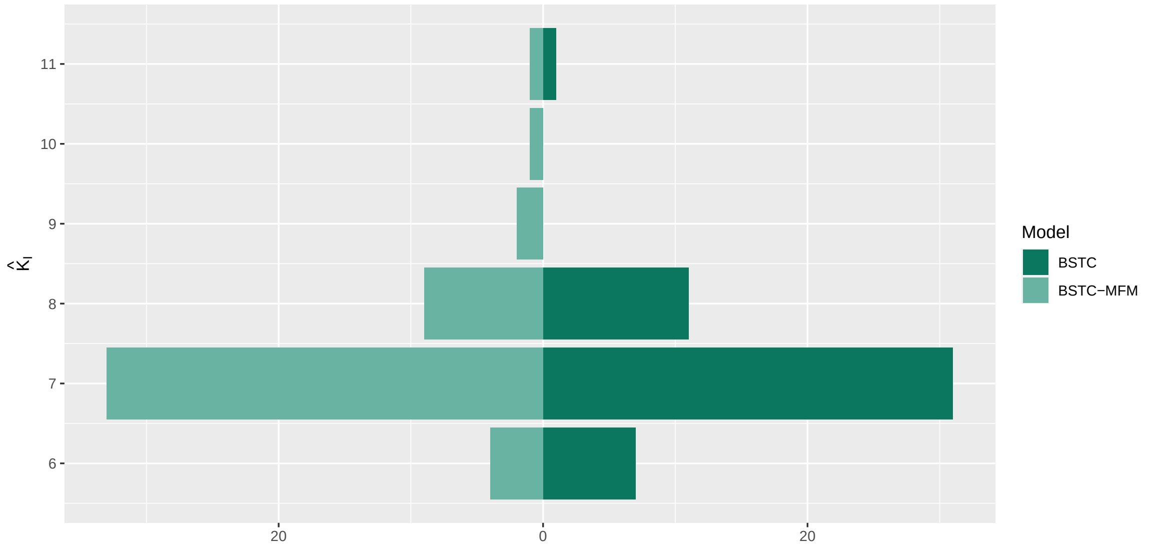
Infinite vs. finite, in a nutshell:

- Miller and Harrison (2014) showed DPM's inconsistency (the number of clusters in a DPM does not concentrate around the truth)
- Ascolani et al. (2023) and Giordano et al. (2023) showed that the DPM is consistent when a suitable prior is placed on the concentration parameter α
- Miller and Harrison (2018) translated the BNP algorithms into the framework of mixtures of finite mixtures
- Frühwirth-Schnatter and Malsiner-Walli (2019) showed similarities between finite and infinite mixtures

$\rightarrow$ Here, we try to be agnostic and go nonparametric only for computational convenience

Infinite vs. Finite Mixture BSTC

For our model (even though the prior distributions for the number of clusters are not identical, cf. Frühwirth-Schnatter & Malsiner-Walli, 2019), we tend to get similar results

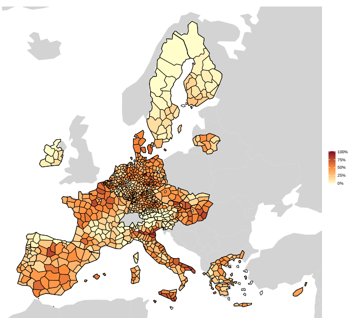


Barplots for $\hat{K}_l$ obtained from the 50 simulated datasets for the BSTC-MFM (left) and BSTC (right) models. The true value for the simulated data is $K = 7$.

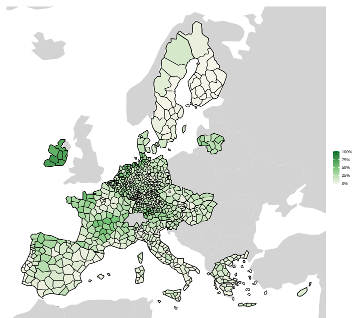
Part II: Partial Clustering

Motivation

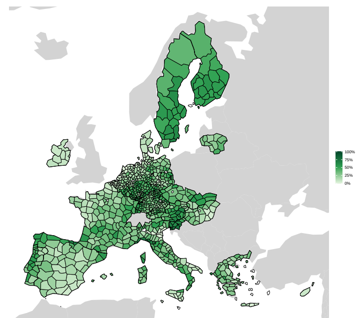
- The **Common Agricultural Policy (CAP)** is the EU's largest budget item (roughly €50B annually), aiming to balance:
  - Agricultural productivity
  - Environmental sustainability
  - Rural development
- **Highly heterogeneous across Europe:** Varies by region, climate, farm structure, socio-economic context (Kuemmerle et al., 2016; Levers et al., 2018; Reidsma et al., 2006)



Cropland density



Grassland density



Forest density

Land-Use Decisions and Environmental Trade-Offs

- Land-use decisions are **compositional**: increasing one land type reduces others
  - CAP subsidies often favor **arable expansion**, sometimes at the **expense of grasslands or forests** (Giannakis & Bruggeman, 2015; Overmars et al., 2013; Rega et al., 2019)
  - Example: maintaining land for sheep farming vs. converting it to forests for carbon sequestration (O'Neill et al., 2020)
- **Multinomial Logit models** naturally model such compositional outcomes
  - Widely used in land-use analysis (Debella-Gilo & Etzelmüller, 2009; Hao et al., 2015; Temme & Verburg, 2011)
- Satellite data from **Land Use and Coverage Area Frame Survey** and the **Corine Land Cover** databases
  - $N = 912$ EU regions (NUTS-3) across 21 countries
  - $T = 10$ years from 2008 to 2018
  - Consolidated into $J = 5$ defined land-use categories Cropland, Grassland, Forest, Urban, Other Natural Land

Bayesian estimation for the Multinomial Logit Model

- Observations $\boldsymbol{Y}_{it\bullet}$ can be seen as a sample from a J-dimensional multinomial distribution with probabilities p_{itj}, $j = 1, \ldots, J$

$$\boldsymbol{Y}_{it\bullet} \overset{\text{ind}}{\sim} \text{Multinomial}(1, p_{it1}, \ldots, p_{itJ})$$

- Probabilities link covariates $\boldsymbol{X}$ and regression parameters β with the log odds

$$log\left(\frac{p_{itj}}{\sum_{j=1}^{J} p_{itj}}\right) = \boldsymbol{X}_{it\bullet}^{T}\beta_{i\bullet j} \to p_{itj} = \frac{exp(\boldsymbol{X}_{it\bullet}^{T}\beta_{i\bullet j})}{\sum_{j=1}^{J} exp(\boldsymbol{X}_{it\bullet}^{T}\beta_{i\bullet j})}$$

- **No direct conjugacy** exists for the logit model due to the nonlinearity of the logistic link
- Early strategies (Frühwirth-Schnatter & Frühwirth, 2007; Held & Holmes, 2006; Scott, 2011) used a **utility representation** (McFadden, 1974)

Posterior Sampling using Pólya-Gamma Variables

Polson et al. (2013) propose a breakthrough via **Pólya-Gamma augmentation**:

- Let $\omega \sim \mathrm{PG}(b, 0)$, with $b > 0$, then

$$\frac{(e^{\psi})^a}{(1+e^{\psi})^b} = 2^{-b} e^{\kappa\psi} \int_0^{\infty} e^{-\omega\psi^2/2}\, p(\omega)\, d\omega$$

  where $\kappa = a - b/2$ for all $a \in \mathbb{R}$
- Applied to the likelihood of a logistic regression:

$$p(y_i \mid \boldsymbol{\beta}) = \frac{\left(\exp(\boldsymbol{X}_{it\bullet}^T \boldsymbol{\beta}_{i\bullet j})\right)^{y_i}}{1+\exp(\boldsymbol{X}_{it\bullet}^T \boldsymbol{\beta}_{i\bullet j})} \propto \exp\left(\kappa_i \boldsymbol{X}_{it\bullet}^T \boldsymbol{\beta}_{i\bullet j}\right) \int_0^{\infty} \exp\left\{-\omega_i \left(\boldsymbol{X}_{it\bullet}^T \boldsymbol{\beta}_{i\bullet j}\right)^2 /2\right\} p\left(\omega_i \mid 1, 0\right) d\omega_i$$

$$\Rightarrow \quad p(\mathbf{y} \mid \boldsymbol{\beta}, \omega) \propto \exp\left\{-\frac{1}{2}(z - X\boldsymbol{\beta})^T \Omega (z - X\boldsymbol{\beta})\right\}$$

where $z = (\kappa_1/\omega_1, \ldots, \kappa_n/\omega_n)$ and $\Omega = \mathrm{diag}\,(\omega_1, \ldots, \omega_n)$

- Transforms the logistic likelihood into a conditionally Gaussian form
- Enables Gibbs sampling without utility latent variables
- Choi and Hobert (2013) show corresponding Markov chain is uniformly ergodic

Further improvement: Zens et al. (2024)

Multinomial logistic regression

- Held and Holmes (2006) write the likelihood conditional on β_{-j}

$$p(\boldsymbol{Y}_{\bullet\bullet j} \mid \beta_{-j}) = \prod_{i=1}^{N}\prod_{t=1}^{T}\left(\frac{\exp(\eta_{itj})}{1+\exp(\eta_{itj})}\right)^{y_{itj}}\left(\frac{1}{1+\exp(\eta_{itj})}\right)^{1-y_{itj}}$$

$$\eta_{itj} = \boldsymbol{X}_{it\bullet}^{T}\beta_{i\bullet j} - C_{itj} \quad C_{itj} = \log\sum_{l\neq j}\exp \boldsymbol{X}_{it\bullet}^{T}\beta_{i\bullet l}$$

$\rightarrow$ desired binary logistic form

$$p(\boldsymbol{Y}_{\bullet\bullet j} \mid \boldsymbol{\omega}, \boldsymbol{X}, \boldsymbol{\beta}) \propto \prod_{t=1}^{T}\prod_{i=1}^{N}\exp\left\{\frac{\omega_{itj}}{2}\left(\eta_{itj} - \kappa_{itj}/\omega_{itj}\right)^2\right\}$$

$$\propto \prod_{t=1}^{T}\exp\left\{-\frac{1}{2}\left(\boldsymbol{X}_{\bullet t\bullet}^{T}\boldsymbol{\beta}_{\bullet\bullet j} - C_{\bullet tj} - \boldsymbol{z}_{tj}\right)^{T}\boldsymbol{\Omega}_{tj}\left(\boldsymbol{X}_{\bullet t\bullet}^{T}\boldsymbol{\beta}_{\bullet\bullet j} - C_{\bullet tj} - \boldsymbol{z}_{tj}\right)\right\}$$

where

$$\kappa_{itj} = \boldsymbol{Y}_{itj} - \frac{1}{2}, \quad \boldsymbol{z}_{tj} = \left(\kappa_{1tj}/\omega_{1tj}, \ldots, \kappa_{Ntj}/\omega_{Ntj}\right), \quad \boldsymbol{\Omega}_{tj} = \mathrm{diag}\left(\omega_{1tj}, \ldots, \omega_{Ntj}\right).$$

$\rightarrow$ Straightforward Gaussian posterior for $\boldsymbol{\beta}_{\bullet\bullet j}$

$\rightarrow$ requires sampling categories j in a loop

To Pool or not to Pool?

- Krisztin et al. (2022):
  - Spatial multinomial logit with autoregressive dependence
  - Spatial structure via fixed neighborhood matrix
  - **Limitation:** common regression coefficients across all areal units:

$$\beta_{1\bullet j} = \beta_{2\bullet j} = \cdots = \beta_{N\bullet j}$$

  $\rightarrow$ **Complete pooling**
- Temme and Verburg (2011):
  - Region-specific multinomial logit models
  - Separate coefficients for each region

$$\beta_{1\bullet j} \neq \beta_{2\bullet j} \neq \ldots \neq \beta_{N\bullet j}$$

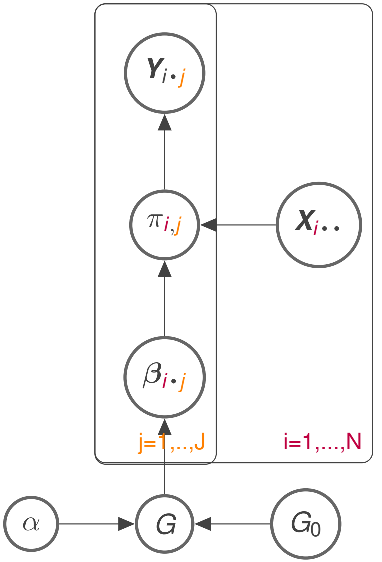
  - **Limitation:** no information sharing across regions and $N \times (K+1) \times (J-1) = 51072$ parameters in our case

  $\rightarrow$ **No pooling**

$\rightarrow$ Cluster the areal units!
  - The discreteness of DP samples implies that multiple areal units may share the same parameters $\rightarrow$ considered part of the same cluster
  - We encode this clustering through $\boldsymbol{s} = (s_1, \ldots, s_N)$, where $s_i \in \{1, \ldots, M\}$

The Bayesian Nonparametric Clustering Model

$$
\begin{aligned}
\boldsymbol{Y}_{i\cdot\cdot} \mid \boldsymbol{X}_{i\cdot\cdot}, \boldsymbol{\beta}_{i\cdot\cdot} &\overset{\text{ind}}{\sim} \mathcal{M}\left(\boldsymbol{Y}_{i\cdot\cdot} \mid \boldsymbol{X}_{i\cdot\cdot}, \boldsymbol{\beta}_{i\cdot\cdot}\right), \quad i = 1, \ldots, N, \\
\boldsymbol{\beta}_{i\cdot\cdot} \mid G &\overset{\text{iid}}{\sim} G, \\
G \mid \alpha &\sim DP\left(\alpha, G_0\right), \\
\alpha &\sim \mathcal{G}(\alpha_0, \beta_0) \\
G_0 &= \mathcal{N}(\boldsymbol{\beta}_0, \Sigma_0)
\end{aligned}
$$



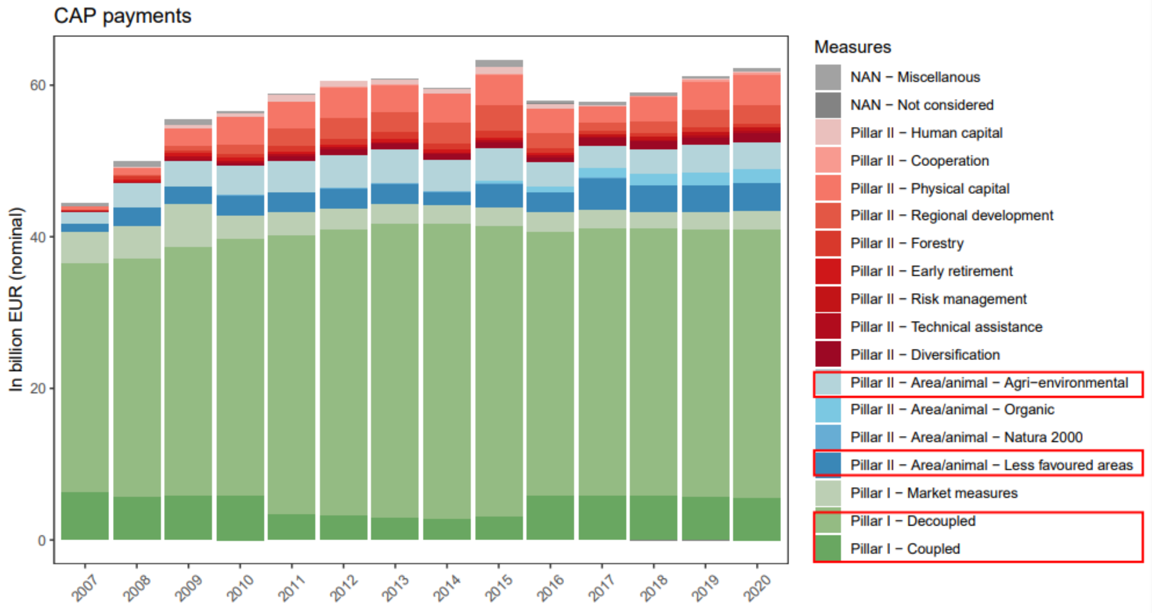
Hyperparameters

- $\boldsymbol{\beta}_0 = \mathbf{0}$ and $\Sigma_0 = I \to$ fairly general and enable good mixing
- Loss function parameters $a_\alpha = 1$ and $b_\alpha = 2 \to$ imply an interpretable number of clusters $\mathbb{E}[M \mid \alpha] \approx 4$ a priori (Teh, 2010).

Data

Variable	Description	Source
Employment primary	Share of employment in the primary sector.	ARDECO
Employment tertiary	Share of employment in the tertiary sector.	ARDECO
Log GDP per capita	Logarithm of the gross domestic product divided by the population.	ARDECO
Population density	Population per square kilometer.	ARDECO
Elevation	Average elevation in meters.	EU-DEM
Slope	Average slope in degrees.	EU-DEM
Farm output	Total farm output in EUR divided by utilized agricultural area.	
Rent	Total rent paid in EUR divided by the total rented area in ha.	
Pillar I - Coupled (crops)	Total of Pillar I coupled payments for crops in EUR divided by the total agricultural area in hectare.	
Pillar I - Coupled (livestock)	Total of Pillar I coupled payments for livestock in EUR divided by the total agricultural area in hectare.	FADN
Pillar I - Decoupled payments	Total of Pillar I decoupled payments in EUR divided by the total agricultural area in hectare.	
Pillar II - Environmental payments	Total of Pillar II environmental payments for crops in EUR divided by the total agricultural area in hectare.	
Pillar II - Least favored area payments	Total of Pillar II least favored area payments for crops in EUR divided by the total agricultural area in hectare.	

DATA

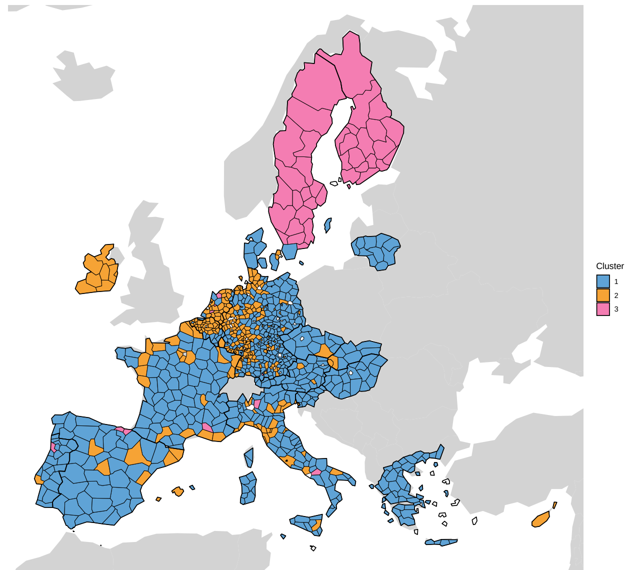


CAP payments from 2007 to 2020 in billion EUR. Included variables are marked with a border

Clustering using **all** covariates – a Boring Result?

- Except Sweden, Finland and Ireland mostly just one, large cluster
- Homogenous impact of covariates like slope and elevation encourages algorithm to place most NUTS-3 regions in the same cluster.

$\rightarrow$ Exclude homogenous variables?

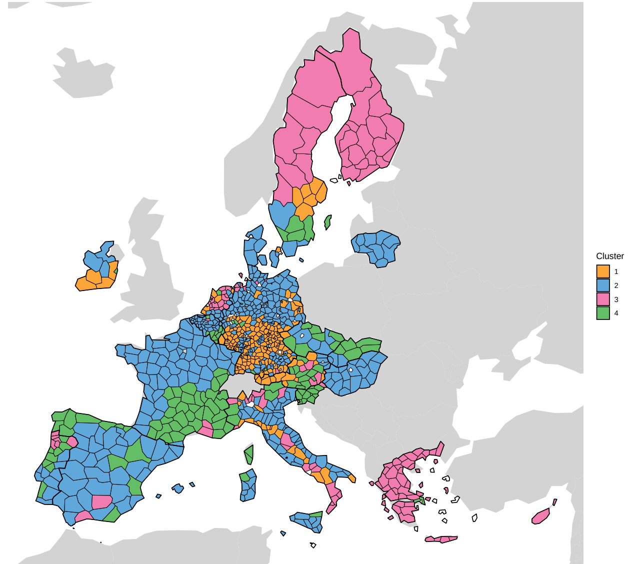


Posterior estimate using the generalized Binder loss with $a_{cost} = 1.5$ including all covariates

Clustering Using Only Farm Related Covariates – Better?

- North-South split in Germany and France, one additional cluster
- Cherry-picking variables
- Loss of available information
- Fewer controls

→ **Goal**: exclude global variables from the clustering process



Posterior estimate using the generalized Binder loss with $a_{cost} = 1.5$ including only farm subsidies

Partial Clustering

Split the likelihood into:

- covariates used to cluster the areal units X^c
- global control variables X^{nc}

$$p_{itj} = \frac{exp(\boldsymbol{X}_{it\bullet}^T \beta_{i\bullet j})}{\sum_{j=1}^J exp(\boldsymbol{X}_{it\bullet}^T \beta_{i\bullet j})} \rightarrow \frac{exp(\boldsymbol{X}_{it\bullet}^{c,T} \beta_{i\bullet j} + \boldsymbol{X}_{it\bullet}^{nc,T} \boldsymbol{\theta}_{\bullet j})}{\sum_{j=1}^J exp(\boldsymbol{X}_{it\bullet}^{c,T} \beta_{i\bullet j} + \boldsymbol{X}_{it\bullet}^{nc,T} \boldsymbol{\theta}_{\bullet j})}$$

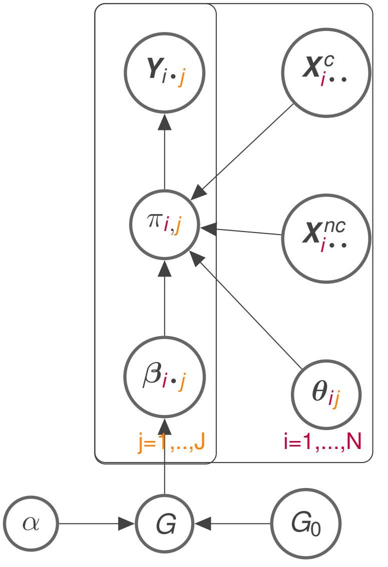
How to sample $\boldsymbol{\theta}$?

- Compute the joint full conditional of all regression coefficients of the jth category
- Construct a new design matrix $\boldsymbol{X}^L \in \mathbb{R}^{N \times T \times (C \times K^c + K^{nc})}$ and stack all regression coefficients

$$\boldsymbol{X}^L = \begin{pmatrix} \boldsymbol{X}^c_{s_i=1\bullet\bullet} & 0 & 0 & 0 & \boldsymbol{X}^{nc}_{s_i=1\bullet\bullet} \\ 0 & \boldsymbol{X}^c_{s_i=2\bullet\bullet} & 0 & \cdots & \boldsymbol{X}^{nc}_{s_i=2\bullet\bullet} \\ 0 & 0 & \ddots & 0 & \vdots \\ 0 & 0 & 0 & \boldsymbol{X}^c_{s_i=M\bullet\bullet} & \boldsymbol{X}^{nc}_{s_i=M\bullet\bullet} \end{pmatrix}, \quad \phi = \begin{bmatrix} \beta^\star_{s_i=1\bullet\bullet} \\ \beta^\star_{s_i=2\bullet\bullet} \\ \vdots \\ \boldsymbol{\theta}_{\bullet\bullet} \end{bmatrix}$$

The Bayesian Nonparametric Partial Clustering Model

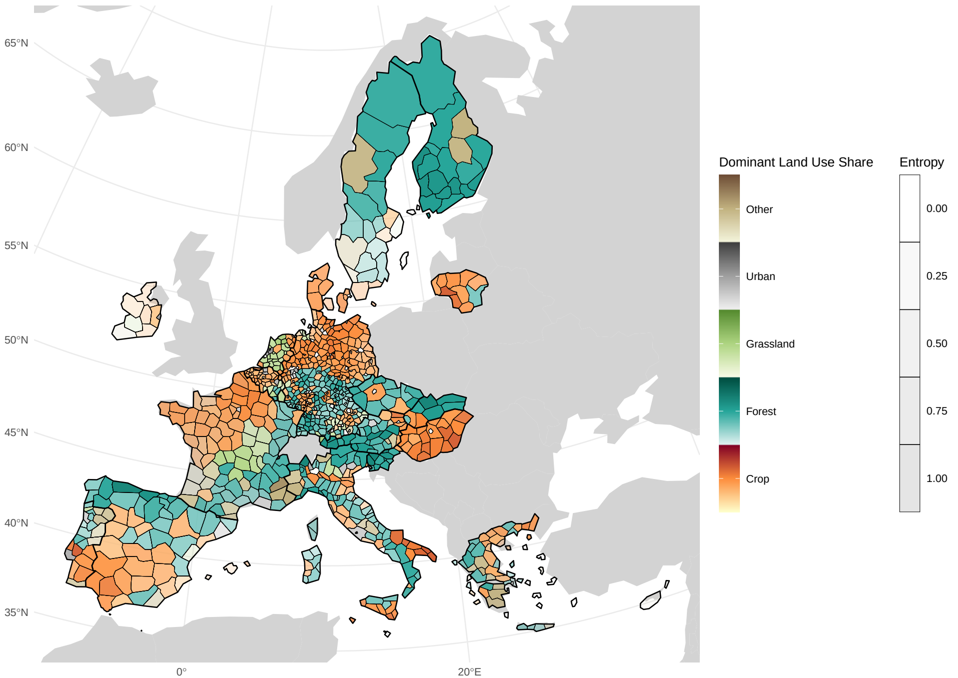
1. For every $i = 1, \ldots, N$ sample s_i according to Algorithm 8 by Neal (2000)
2. For every category $j = 1, \ldots J - 1$:
   2.1 For every cluster $c = 1, \ldots, M$:
      2.1.1 Sample auxiliary variables $\omega_{s_i=c,tj}$ using those i such that $s_i = c$
      2.1.2 Sample $\beta^{\star}_{c\bullet j}$ from their full conditional using the data augmentation by Polson et al. (2013)
   2.2 Sample $\boldsymbol{\omega}^L_{itj}$
   2.3 Sample $\boldsymbol{\theta}_{\bullet j}$ using the stacked matrices
3. Sample concentration parameter α



Partial clustering using farm subsidies and global covariates

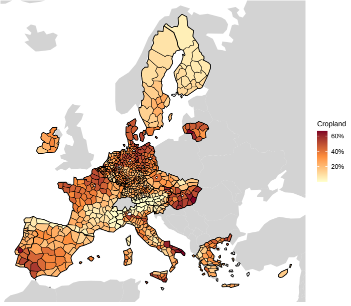
- Cluster covariates X^c
  - farm output
  - farm rent
  - coupled payments crops
  - coupled payments livestock
  - decoupled payments
  - environmental payments
  - least favored area payments
- Global Covariates X^{nc}
  - employment primary sector
  - employment tertiary sector
  - log GDP per capita
  - population density
  - elevation
  - slope

Results – Predicted Values

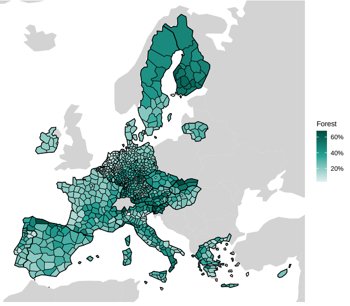


Units colored corresponding to the land-use class with the highest probability, i.e., $\arg\max_{j=1,\dots,J}(p_{i,T+1,j} \mid \boldsymbol{Y}, \boldsymbol{X})$, additionally shaded depending on the respective entropy and exemplified for year $T+1=11$

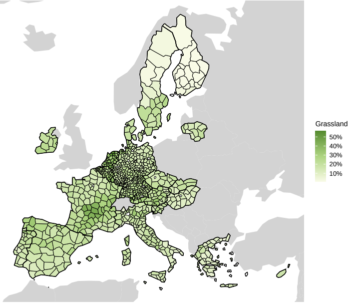
Results – Predicted Values



Cropland



Forest

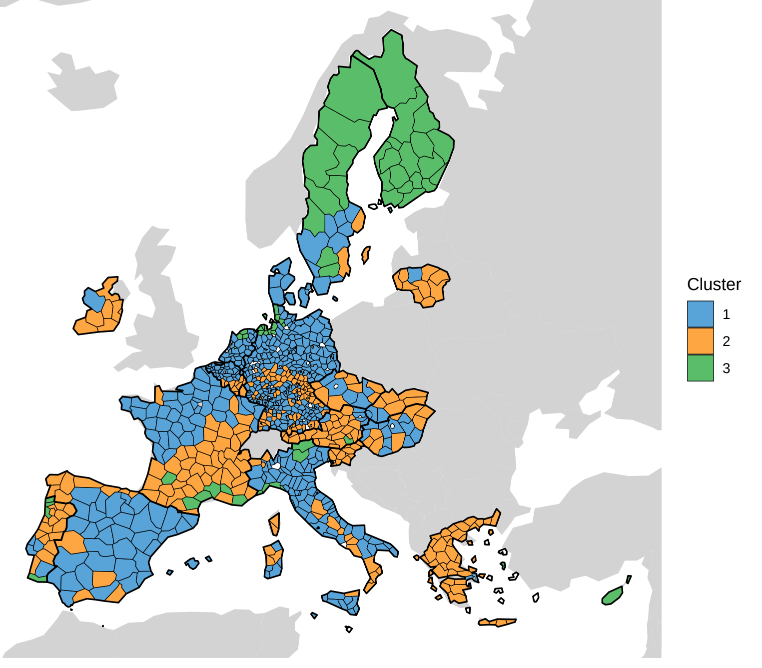


Grassland

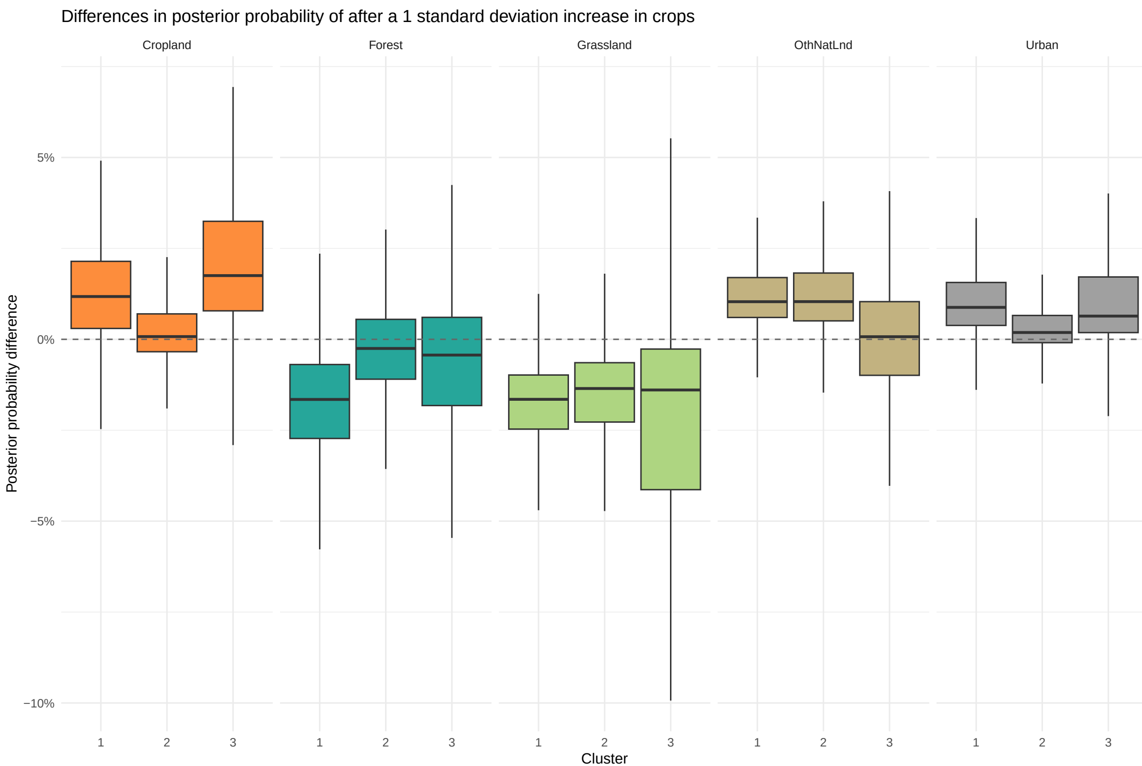
Posterior mean $E(p_{i,T+1,j}|\boldsymbol{Y}, \boldsymbol{X})$ by land-use category (excluding Rural and Other Natural Land), exemplified for year $T + 1 = 11$.

Results – Clustering (Binder Loss, $a_{cost} = 1.5$)

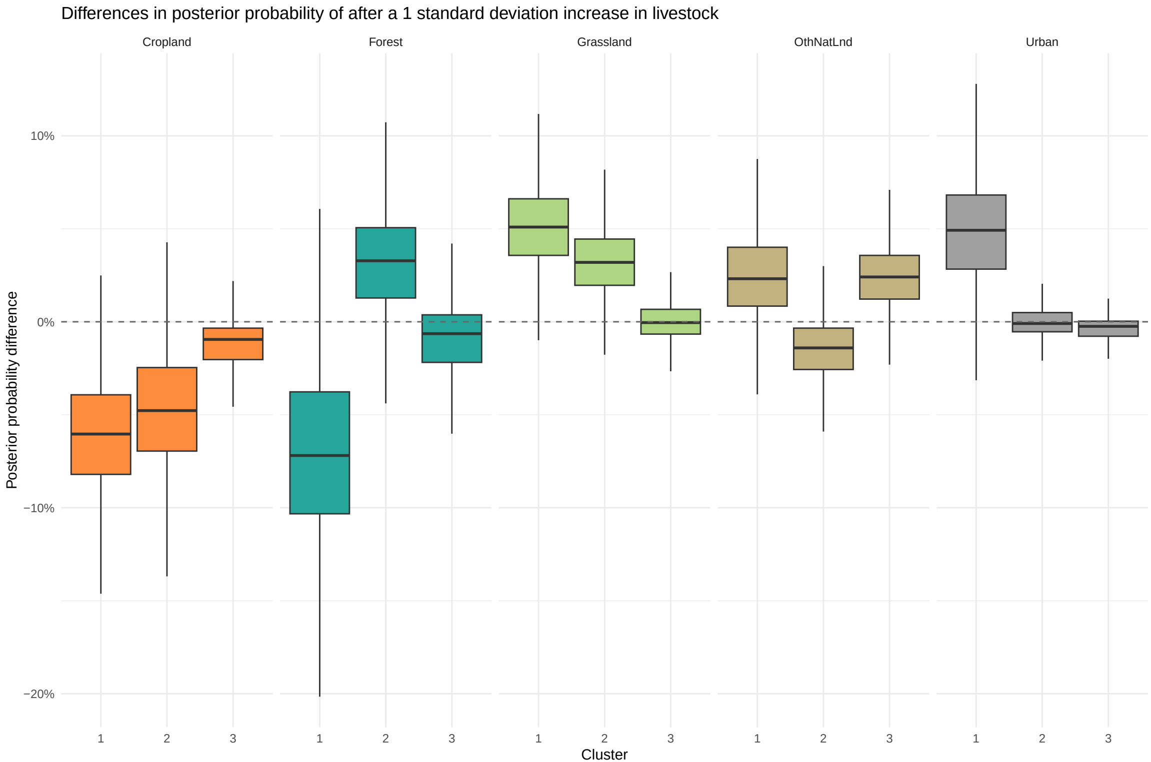
- Country specific clustering of policies evident
- Clusters align closely with the EEA biogeographical regions
- Regional fragmentation in France and Spain



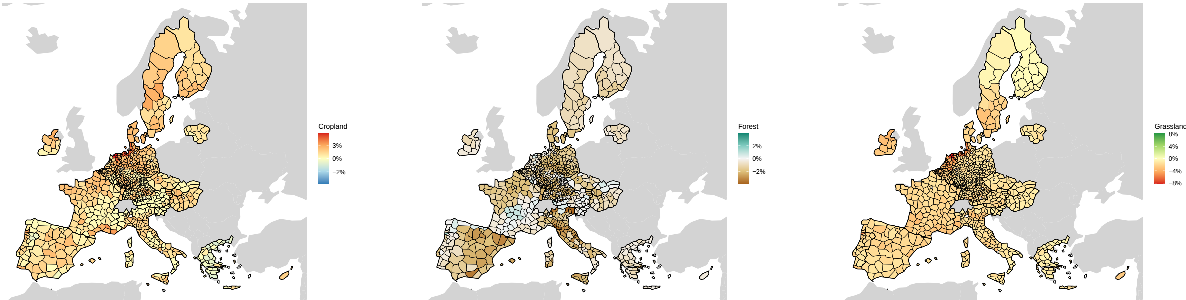
Crop Subsidy Increase – Cluster-specific Effects



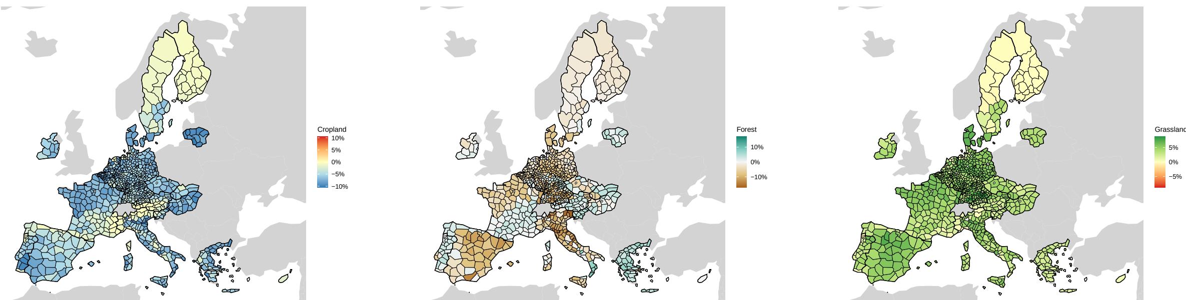
Livestock Subsidy Increase – Cluster-specific Effects



Results – Policy Changes



Impact of a one standard deviation **increase in crop subsidies** on Cropland, Forest and Grassland



Impact of a one standard deviation **increase in livestock subsidies** on Cropland, Forest and Grassland

Part III: Repulsive Mixture Weights

Motivation

- Prior on component locations in a standard mixture model:

$$\boldsymbol{\mu} \overset{\text{iid}}{\sim} \pi_{\boldsymbol{\mu}}(\boldsymbol{\mu}), \quad \boldsymbol{\mu} = (\mu_1, \ldots, \mu_M)$$

- Problem: overlapping, redundant clusters
- Idea: include distance between parameters $d\left(\boldsymbol{\mu}_i, \boldsymbol{\mu}_j\right)$ into prior
- Repulsive mixture model:

$$\boldsymbol{\mu} \sim \pi_{\boldsymbol{\mu}}(\boldsymbol{\mu}), \quad \boldsymbol{\mu} = (\mu_1, \ldots, \mu_M)$$
$$\pi(\boldsymbol{\mu}) \propto \left(\prod_{m=1}^{M} g(\mu_m)\right)\left(\prod_{1\leq i<j\leq M} d\left(\mu_i, \mu_j\right)\right)$$

- Origin in statistical mechanics

How and What to Penalize?

- Distance measures for repulsive mixture models:

Publication	$\boldsymbol{d}\left(\boldsymbol{\mu_i}, \boldsymbol{\mu_j}\right)$
Petralia et al. (2012)	$\exp\left\{-g/\left(\mu_i-\mu_j\right)' A^{-1}\left(\mu_i-\mu_j\right)\right\}$
Quinlan et al. (2021)	$-\log\left(1-e^{-((\mu_i-\mu_j)/\sigma)^2}\right)$
Non-local Priors (Fúquene et al., 2019)	$\left(\mu_i-\mu_j\right)' A^{-1}\left(\mu_i-\mu_j\right)/g$
Cremaschi et al. (2023)	$\lvert(\mu_i-\mu_j)\rvert^{\zeta}$

- Other methods for repulsive priors in recent literature: determinantal point processes (DPP) (Bianchini et al., 2020; Xu et al., 2016), normalized random measures (Beraha et al., 2020)

- Our idea: repulsive prior on the component weights $\boldsymbol{w}$

Sparsity-inducing partition (SIP) prior

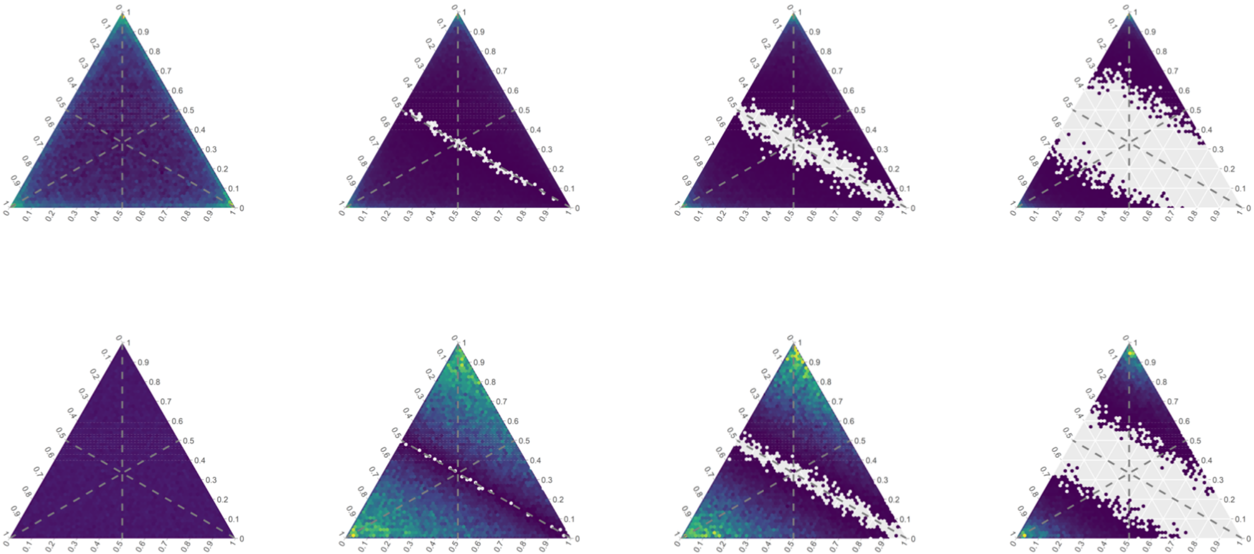
Selberg Dirichlet (*SDir*) distribution

$$SDir(\boldsymbol{w}, \alpha, \gamma, M) = \frac{1}{D(\alpha, \gamma, M)} \left(\prod_{m=1}^{M} w_m^{\alpha - 1} \right) |\triangle \boldsymbol{w}|^{2\gamma}$$

$$D(\alpha, \gamma, M) = \frac{\Gamma(\alpha)}{\Gamma(M\alpha + \gamma(M-1)(M-2))} \prod_{j=1}^{M-1} \frac{\Gamma(\alpha + (j-1)\gamma)\Gamma(1 + j\gamma)}{\Gamma(1+\gamma)}$$

- Repulsion parameter $\gamma \geq 0$
- Concentration parameter $\alpha > 0$
- The term $\triangle \boldsymbol{w} = \prod_{1 \leq i < j \leq M-1} |w_i - w_j|$ represents the product of the absolute pairwise differences among the components of $\boldsymbol{w}$.
- Introduced in Pham-Gia (2009)

Selberg Dirichlet (equal α-elements)



Simulation-based densities of the Selberg Dirichlet distribution for varying values of the repulsion parameter γ and concentration parameter α. From left to right: $\gamma = (0, 0.5, 1, 3)$. From top to bottom: $\alpha = (0.5, 1)$. For illustration purposes, the colors are scaled separately for each plot.

Generalized Selberg Dirichlet

- Posterior update of mixture weights increments α by the corresponding cluster size

$\rightarrow$ need for varying concentration parameters α_m

Generalised Selberg Dirichlet

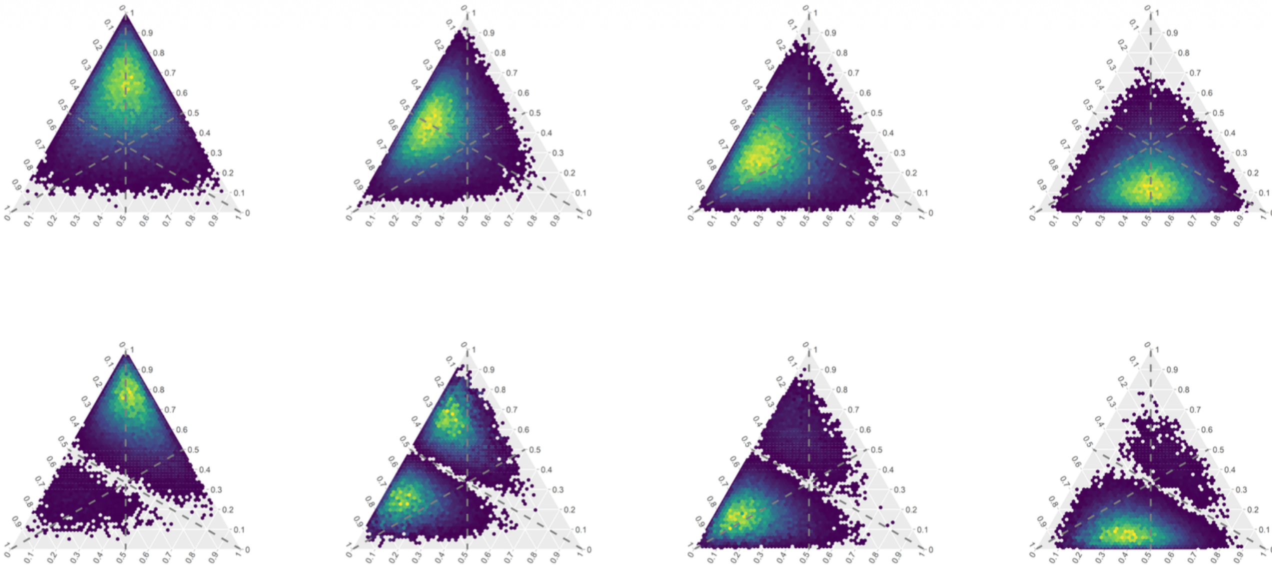
The M-dimensional generalised Selberg Dirichlet has a density given by:

$$GSDir(\boldsymbol{w}, \boldsymbol{\alpha}, \gamma, M) = \frac{1}{GD(\boldsymbol{\alpha}, \gamma, M)} \left(\prod_{m=1}^{M} w_m^{\alpha_m - 1} \right) |\triangle \boldsymbol{w}|^{2\gamma}$$

$$GD(\boldsymbol{\alpha}, \gamma, M) = \int_0^1 \cdots \int_0^1 \left(\prod_{m=1}^{M} w_m^{\alpha_m - 1} \right) |\triangle \boldsymbol{w}|^{2\gamma} dw_1 \dots dw_M$$

- Repulsion parameter $\gamma \geq 0$
- Potentially different concentration parameters $\alpha_m > 0$ for $m = 1, \dots, M$
- Avoid unknown normalization constant *GD* using Metropolis-Hastings steps

Generalized Selberg Dirichlet (unequal α-elements)

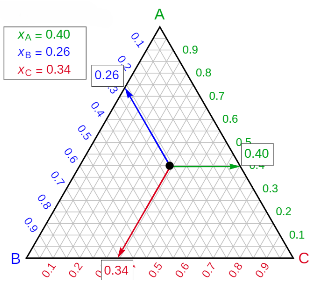


Simulation-based densities of the Dirichlet and Selberg Dirichlet distribution for varying values of the concentration parameter α and $\gamma = 1$. From left to right: $\alpha = (2, 5, 2), (5, 5, 2), (5, 5, 2), (5, 3, 2)$. The first row shows the Dirichlet distribution and the second the generalized Selberg Dirichlet with $\gamma = 1$. For illustration purposes, the colors are scaled separately for each plot.

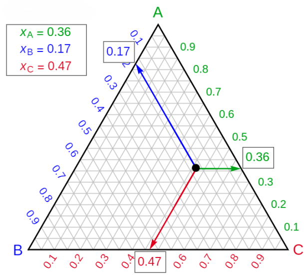
Pairwise Differences

Distance measured as $\left|\prod_{1 \leq i < j \leq M-1} x_i - x_j\right|$

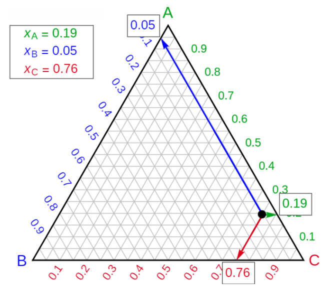
- "Clear"for $\boldsymbol{\mu} \in \mathbb{R}$
- What about $w_1, ..., w_p : 0 \leq w_i \leq 1, \quad \sum_{i=1}^{M} w_i = 1$?



$|\triangle \boldsymbol{w}| = 0.00067$



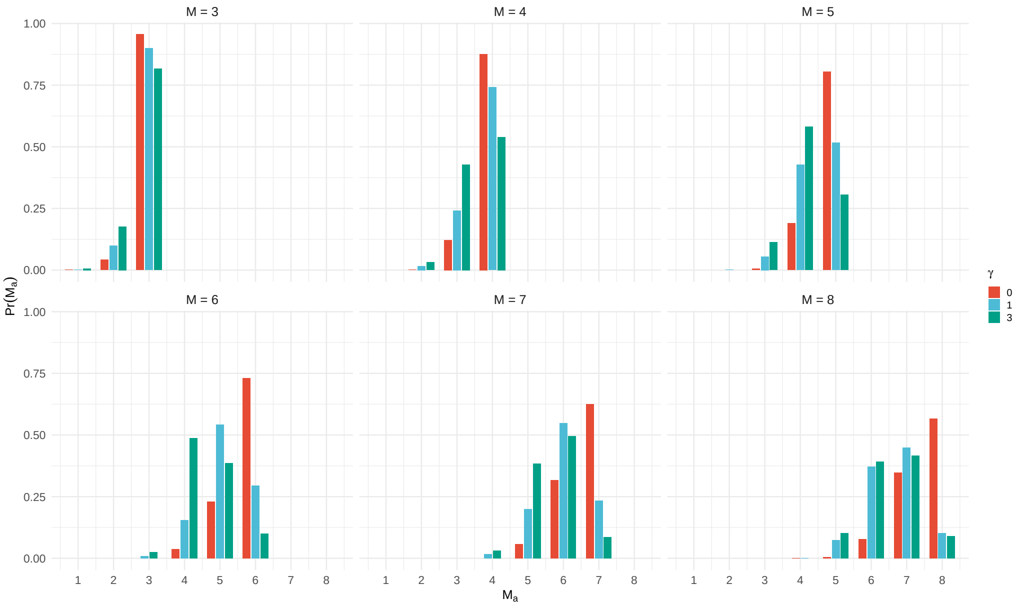
$|\triangle \boldsymbol{w}| = 0.00627$



$|\triangle \boldsymbol{w}| = 0.05666$

- Smaller $M \rightarrow$ larger $\triangle \boldsymbol{w}$
- Smoothing over redundant components

IMPLIED PRIOR ON THE NUMBER OF CLUSTERS M_a



Simulation-based, implied prior on M_a for $\alpha_0 = 1$ and varying values of M and γ.

A Fully Repulsive Mixture Model

Fully Repulsive Mixture Model

$$
\begin{aligned}
\boldsymbol{y}_i \mid c_i, \boldsymbol{\mu}_{c_i}, \Sigma_{c_i} &\overset{\text{ind}}{\sim} \mathcal{N}(\boldsymbol{\mu}_{c_i}, \Sigma_{c_i}), \quad i = 1, \ldots, N \\
\mu_{1,d}, \ldots, \mu_{M,d} &\sim GE(\zeta), \quad d = 1, \ldots, D \\
\boldsymbol{\Sigma}_m &\overset{\text{iid}}{\sim} \mathcal{IW}(\boldsymbol{V}_0, \nu_0), \quad m = 1, \ldots, M \\
c_1, \ldots, c_N \mid \boldsymbol{w} &\overset{\text{iid}}{\sim} \mathcal{C}(1, \boldsymbol{w}) \\
\boldsymbol{w} &\sim \text{SDir}(\alpha, \gamma, M) \\
M &\sim \text{Poi}_1(\lambda) \\
\gamma &\sim \mathcal{G}(\alpha_{\gamma,0}, \beta_{\gamma,0}) \\
\zeta &\sim \mathcal{G}(\alpha_{\zeta,0}, \beta_{\zeta,0})
\end{aligned}
$$

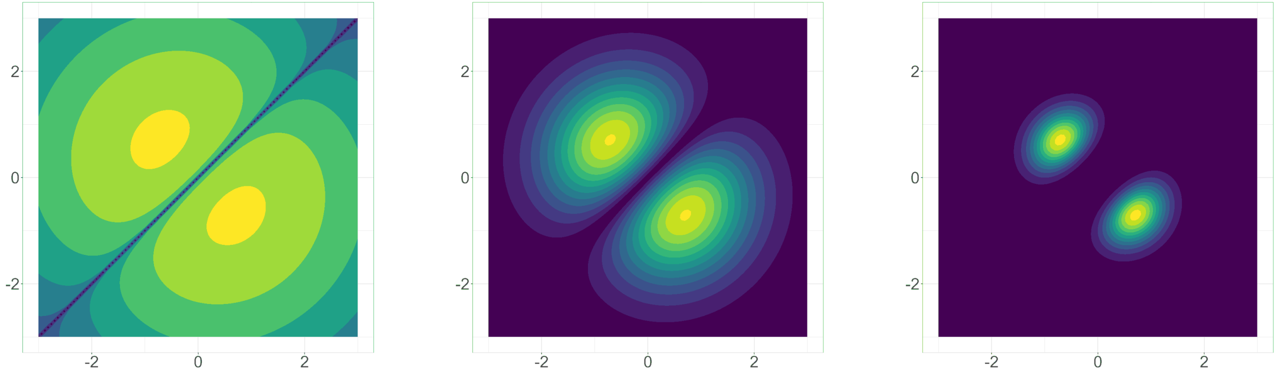
- **Gaussian ensemble** distribution as prior for locations
  - $\rightarrow$ Higher repulsion supports separation
- **Selberg Dirichlet distribution** as prior for weights
  - $\rightarrow$ Higher repulsion leads to smoothing effect

Repulsive prior on cluster locations

Gaussian Ensemble (*GE*) distribution (Cremaschi et al., 2023; Forrester, 2010)

$$\pi(\boldsymbol{\mu}) = GE(\boldsymbol{\mu} \mid M, \zeta) = \frac{1}{\mathcal{G}(M,\zeta)} \prod_{k=1}^{M} e^{-\frac{\zeta}{2}\mu_k^2} |\triangle \boldsymbol{\mu}|^{\zeta}$$

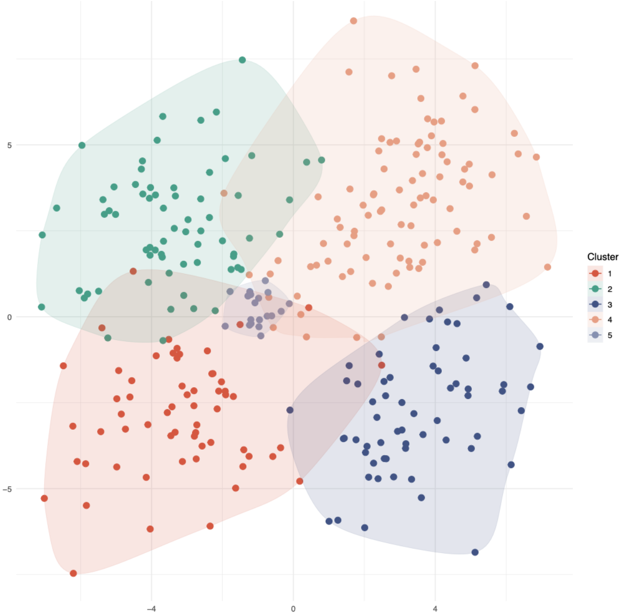
- $\triangle \boldsymbol{x} = \prod_{1 \le i < j \le M} x_i - x_j$
- Repulsion parameter ζ

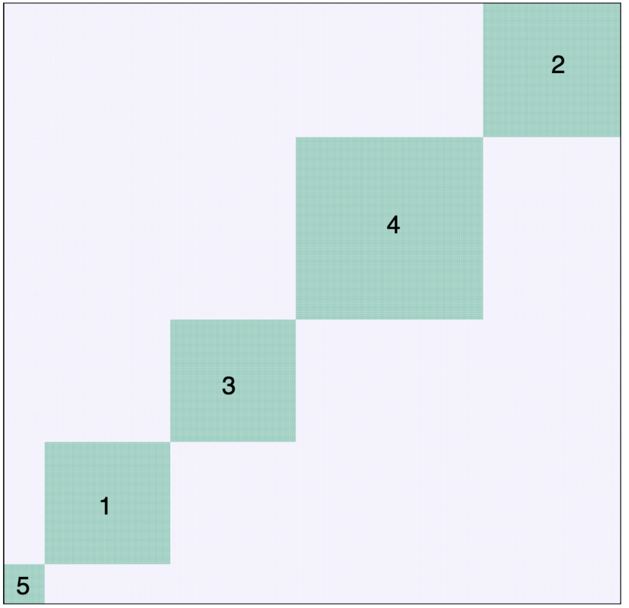


GE prior for $M = 2$ on $\boldsymbol{\mu}$, repulsion parameter $\zeta = 0.1, 1, 5$

Simulation Study

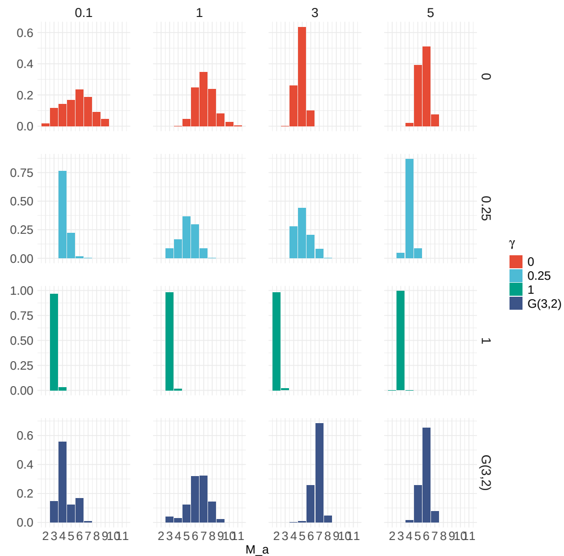
- Simulate $n = 300$ observations from 5-component bivariate Gaussian mixture
- Intentionally substantial cluster overlap and redundant component



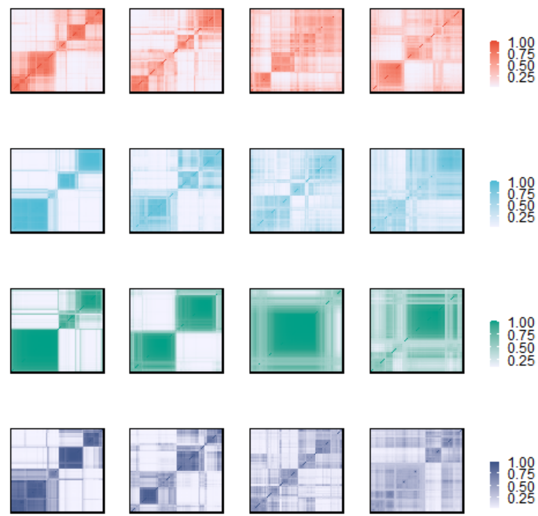


Posterior number of clusters and similarity matrices

- Increasing γ reduces both the posterior mean and variance of M_a, whereas increasing ζ leads to a higher estimated number of clusters
- Posterior similarity matrices show low variability for $\gamma = 1$, except when both γ and ζ are large, in which case cluster assignments become more uncertain



(a)



(b)

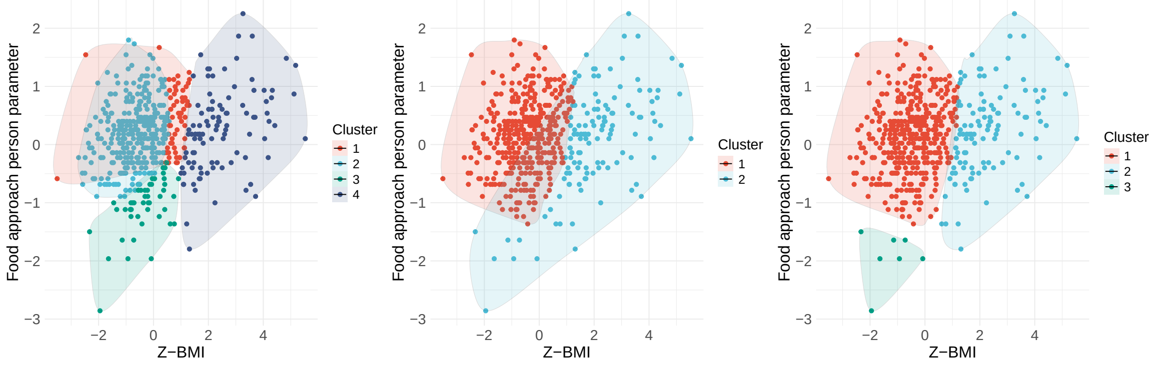
Application to data on children's BMI and eating behavior

- 537 children from a highly phenotyped prospective cohort (Soh et al., 2013)
- Standardized BMI (Z-BMI)
- Eating behavior: CEBQ questionnaire (Wardle et al., 2001)
- Food-approach subscales:
  - food responsiveness
  - enjoyment of food
  - emotional overeating
  - desire to drink
- Latent trait recovery using the Partial Credit Model (Masters, 1982)
- Positive correlation between latent food-approach and Z-BMI ($r = 0.24$), cf. Fogel et al. (2017)

Z-BMI vs. latent food-approach for 537 GUSTO children.

Clustering results based on the Binder loss function for the SIP mixture

- Three repulsion settings (γ, ζ) are evaluated based on 5,000 thinned samples of the SIP repulsive mixture model

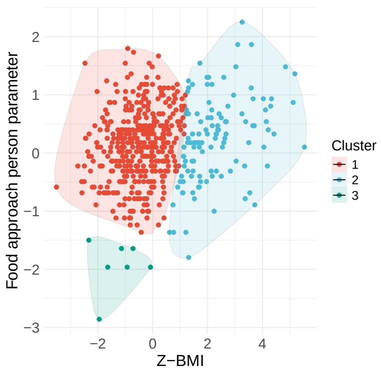


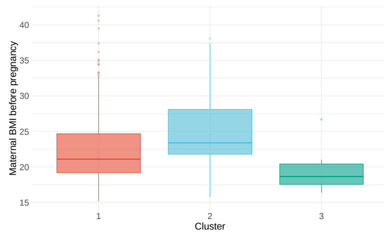
(a) $\gamma = 0.1$, $\zeta = 0.1$ **(b)** $\gamma = 2$, $\zeta = 3$ **(c)** $\gamma = 1$, $\zeta = 3$

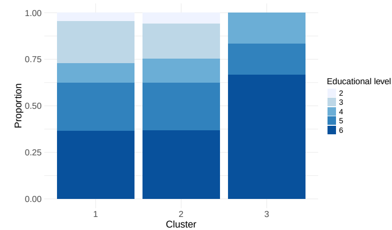
- The preferred specification identifies three clusters: average Z-BMI and food-approach; high Z-BMI with more dispersed traits; and low Z-BMI with low food-approach
- Maternal pre-pregnancy BMI differs significantly across clusters, while maternal education does not

Cluster analysis of pre-pregnancy maternal BMI and education

- Maternal pre-pregnancy BMI varies across clusters: Cluster 1 shows highest variability, Cluster 2 has higher median, and Cluster 3 has lowest median and variability
- Maternal education (6-level scale) distribution is similar for Clusters 1 and 2; Cluster 3 has more mothers with university degrees but lacks lower education levels
- Due to small sample size in Cluster 3, no significant differences in maternal education were found across clusters







References I

Ascolani, F., Lijoi, A., Rebaudo, G., & Zanella, G. (2023). **Clustering consistency with Dirichlet process mixtures.** *Biometrika*, *110*(2).

Beraha, M., Argiento, R., Møller, J., & Guglielmi, A. (2020). **MCMC computations for Bayesian mixture models using repulsive point processes.** *Journal of Computational and Graphical Statistics*, *31*, 422–435.

Bianchini, I., Guglielmi, A., & Quintana, F. A. (2020). **Determinantal point process mixtures via spectral density approach.** *Bayesian Analysis*, *15*(1), 187–214.

Binder, D. A. (1978). **Bayesian cluster analysis.** *Biometrika*, *65*(1), 31–38.

Blackwell, D., & MacQueen, J. B. (1973). **Ferguson distributions via Pólya urn schemes.** *The Annals of Statistics*, *1*(2), 353–355.

Choi, H. M., & Hobert, J. P. (2013). **The Polya-Gamma Gibbs sampler for Bayesian logistic regression is uniformly ergodic.** *Electronic Journal of Statistics*, *7*(none), 2054–2064.

Cremaschi, A., Wertz, T. M., & Iorio, M. D. (2023). **Repulsion, chaos and equilibrium in mixture models.**

Dahl, D. B., Johnson, D. J., & Müller, P. (2022). **Search algorithms and loss functions for bayesian clustering.** *Journal of Computational and Graphical Statistics*, *31*(4), 1189–1201.

Debella-Gilo, M., & Etzelmüller, B. (2009). **Spatial prediction of soil classes using digital terrain analysis and multinomial logistic regression modeling integrated in GIS: Examples from Vestfold County, Norway.** *CATENA*, *77*(1), 8–18.

Escobar, M. D., & West, M. (1995). **Bayesian density estimation and inference using mixtures.** *Journal of the American Statistical Association*, *90*(430), 577–588.

Ferguson, T. S. (1973). **A Bayesian analysis of some nonparametric problems.** *The Annals of Statistics*, 209–230.

References II

Fogel, A., Goh, A. T., Fries, L. R., Sadananthan, S. A., Velan, S. S., Michael, N., Tint, M.-T., Fortier, M. V., Chan, M. J., Toh, J. Y., Chong, Y.-S., Tan, K. H., Yap, F., Shek, L. P., Meaney, M. J., Broekman, B. F. P., Lee, Y. S., Godfrey, K. M., Chong, M. F. F., & Forde, C. G. (2017). **Faster eating rates are associated with higher energy intakes during an ad libitum meal, higher BMI and greater adiposity among 4·5-year-old children: Results from the growing up in singapore towards healthy outcomes (GUSTO) cohort.** *Br J Nutr*, *117*(7), 1042–1051.

Forrester, P. (2010). ***Log-gases and random matrices (lms-34).*** Princeton University Press.

Frühwirth-Schnatter, S. (2001). **Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models.** *Journal of the American Statistical Association*, *96*(453), 194–209.

Frühwirth-Schnatter, S., & Frühwirth, R. (2007). **Auxiliary mixture sampling with applications to logistic models.** *Computational Statistics and Data Analysis*, *51*(7), 3509–3528.

Frühwirth-Schnatter, S., & Malsiner-Walli, G. (2019). **From here to infinity: Sparse finite versus Dirichlet process mixtures in model-based clustering.** *Advances in Data Analysis and Classification*, *13*, 33–64.

Fúquene, J., Steel, M., & Rossell, D. (2019). **On choosing mixture components via non-local priors.** *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *81*(5), 809–837.

Gershman, S. J., & Blei, D. M. (2012). **A tutorial on Bayesian nonparametric models.** *Journal of Mathematical Psychology*, *56*(1), 1–12.

Giannakis, E., & Bruggeman, A. (2015). **The highly variable economic performance of European agriculture.** *Land Use Policy*, *45*, 26–35.

Giordano, R., Liu, R., Jordan, M. I., & Broderick, T. (2023). **Evaluating sensitivity to the stick-breaking prior in Bayesian nonparametrics (with discussion).** *Bayesian Analysis*, *18*(1), 287–366.

References III

Hao, C., Zhang, J., Li, H., Yao, F., Huang, H., & Meng, W. (2015).**Integration of multinomial-logistic and Markov-chain models to derive land-use change dynamics.** *Journal of Urban Planning and Development*, *141*(3), 05014017.

Held, L., & Holmes, C. C. (2006).**Bayesian auxiliary variable models for binary and multinomial regression.** *Bayesian Analysis*, *1*(1), 145–168.

Krisztin, T., Piribauer, P., & Wögerer, M. (2022).**A spatial multinomial logit model for analysing urban expansion.** *Spatial Economic Analysis*, *17*(2), 223–244.

Kuemmerle, T., Levers, C., Erb, K., Estel, S., Jepsen, M. R., Müller, D., Plutzar, C., Stürck, J., Verkerk, P. J., Verburg, P. H., et al. (2016).**Hotspots of land use change in Europe.** *Environmental research letters*, *11*(6), 064020.

Leroux, B. G., Lei, X., & Breslow, N. (2000). **Estimation of disease rates in small areas: A new mixed model for spatial dependence.** In M. E. Halloran & D. Berry (Eds.), *Statistical models in epidemiology, the environment, and clinical trials* (pp. 179–191). Springer New York.

Levers, C., Schneider, M., Prishchepov, A., Estel, S., & Kuemmerle, T. (2018).**Spatial variation in determinants of agricultural land abandonment in Europe.** *The Science of the total environment*, *644*, 95–111.

Lo, A. Y. (1984).**On a class of Bayesian nonparametric estimates: I. density estimates.** *The Annals of Statistics*, *12*(1), 351–357.

MacEachern, S. N., & Müller, P. (1998).**Estimating mixture of Dirichlet process models.** *Journal of Computational and Graphical Statistics*, *7*(2), 223–238.

Masters, G. N. (1982).**A Rasch model for partial credit scoring.** *Psychometrika*, *47*(2), 149–174.

McFadden, D. (1974). **Conditional logit analysis of qualitative choice behavior.** In P. Zarembka (Ed.), *Fontiers in Econometrics* (pp. 105–142). Academic press.

References IV

Meilă, M. (2007).**Comparing clusterings – an information based distance.** *Journal of multivariate analysis*, *98*(5), 873–895.

Miller, J. W., & Harrison, M. T. (2014).**Inconsistency of Pitman-Yor process mixtures for the number of components.** *The Journal of Machine Learning Research*, *15*(1), 3333–3370.

Miller, J. W., & Harrison, M. T. (2018).**Mixture models with a prior on the number of components.** *Journal of the American Statistical Association*, *113*(521), 340–356.

Mozdzen, A., Addo, F., Krisztin, T., & Kastner, G. (2024). **Bayesian nonparametric partial clustering: Quantifying the effectiveness of agricultural subsidies across europe.**

Mozdzen, A., Cremaschi, A., Cadonna, A., Guglielmi, A., & Kastner, G. (2022).**Bayesian modeling and clustering for spatio-temporal areal data: An application to Italian unemployment.** *Spatial Statistics*, *52*, 100715.

Mozdzen, A., Wertz, T., Iorio, M. D., Cremaschi, A., Kastner, G., & Eriksson, J. (2025). **Repulsive mixtures via the sparsity-inducing partition prior.**

Neal, R. M. (2000).**Markov chain sampling methods for Dirichlet process mixture models.** *Journal of Computational and Graphical Statistics*, *9*(2), 249–265.

O'Neill, C., Lim, F. K. S., Edwards, D. P., & Osborne, C. P. (2020).**Forest regeneration on European sheep pasture is an economically viable climate change mitigation strategy.** *Environmental Research Letters*, *15*(10), 104090.

Overmars, K. P., Helming, J., van Zeijts, H., Jansson, T., & Terluin, I. (2013).**A modelling approach for the assessment of the effects of common agricultural policy measures on farmland biodiversity in the EU27.** *Journal of environmental management*, *126*, 132–141.

Papastamoulis, P., & Iliopoulos, G. (2010).**An artificial allocations based solution to the label switching problem in Bayesian analysis of mixtures of distributions.** *Journal of Computational and Graphical Statistics*, *19*(2), 313–331.

REFERENCES V

Petralia, F., Rao, V., & Dunson, D. (2012). **Repulsive mixtures.** In F. Pereira, C. Burges, L. Bottou, & K. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 25). Curran Associates, Inc.

Pham-Gia, T. (2009).**The multivariate Selberg beta distribution and applications.** *Statistics*, *43*(1), 65–79.

Polson, N. G., Scott, J. G., & Windle, J. (2013).**Bayesian inference for logistic models using Pólya-gamma latent variables.** *Journal of the American Statistical Association*, *108*(504).

Quinlan, J. J., Quintana, F. A., & Page, G. L. (2021).**On a class of repulsive mixture models.** *Test*, *30*, 445–461.

Rega, C., Helming, J., & Paracchini, M. L. (2019).**Environmentalism and localism in agricultural and land-use policies can maintain food production while supporting biodiversity. findings from simulations of contrasting scenarios in the EU.** *Land Use Policy*, *87*, 103986.

Reidsma, P., Tekelenburg, T., Van den Berg, M., & Alkemade, R. (2006).**Impacts of land-use change on biodiversity: An assessment of agricultural biodiversity in the European Union.** *Agriculture, ecosystems & environment*, *114*(1), 86–102.

Richardson, S., & Green, P. J. (1997).**On Bayesian analysis of mixtures with an unknown number of components (with discussion).** *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *59*(4), 731–792.

Rodríguez, C. E., & and, S. G. W. (2014).**Label switching in bayesian mixture models: Deterministic relabeling strategies.** *Journal of Computational and Graphical Statistics*, *23*(1), 25–45.

Scott, S. L. (2011).**Data augmentation, frequentist estimation, and the Bayesian analysis of multinomial logit models.** *Statistical Papers*, *52*(1), 87–109.

Sethuraman, J. (1994).**A constructive definition of Dirichlet priors.** *Statist. Sinica*, *4*(2), 639–650.

References VI

Soh, S.-E., Tint, M. T., Gluckman, P. D., Godfrey, K. M., Rifkin-Graboi, A., Chan, Y. H., Stünkel, W., Holbrook, J. D., Kwek, K., Chong, Y.-S., Saw, S. M., & the GUSTO Study Group. (2013).**Cohort profile: Growing up in singapore towards healthy outcomes (GUSTO) birth cohort study.** *International Journal of Epidemiology*, *43*(5), 1401–1409.

Teh, Y. W. (2010). **Dirichlet process.** In *Encyclopedia of machine learning* (pp. 280–287). Springer US.

Temme, A., & Verburg, P. (2011).**Mapping and modelling of changes in agricultural intensity in Europe.** *Agriculture, Ecosystems & Environment*, *140*(1), 46–56.

Wardle, J., Guthrie, C. A., Sanderson, S., & Rapoport, L. (2001).**Development of the children's eating behaviour questionnaire.** *Journal of Child Psychology and Psychiatry*, *42*(7), 963–970.

Xu, Y., Müller, P., & Telesca, D. (2016).**Bayesian inference for latent biologic structure with determinantal point processes (DPP).** *Biometrics*, *72*(3), 955–964.

Zens, G., Frühwirth-Schnatter, S., & Wagner, H. (2024).**Ultimate Pólya gamma samplers – efficient MCMC for possibly imbalanced binary and categorical data.** *Journal of the American Statistical Association*, *119*(548), 2548–2559.