Bayesian Econometrics in the Big Data Era

Sylvia Frühwirth-Schnatter

November 28, 2018

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Part I

Big data in Bayesian econometrics

Outline

Part I: Big data in Bayesian econometrics

- Econometrics and big data applications
- Austrian Center for Labor Economics and Welfare

It's exciting times for econometricians . . .

- The world is **changing** also for empirical economists and social scientists
- New types of "big data" create interest in **data science**
- **Interface** of computer science, statistics, and social science/economics
- Apply these tools in **teams** of economists/social scientists/statisticians
- New ways of **visualizing** and presenting information for effective communication

Data sources in modern econometrics

- **Administrative data** collected by national statistical agencies
- **ECB** (European Central Bank), Federal Reserve Banks
- Survey data such as **EU-SILC** (European Union Statistics on Income and Living Conditions)
- Financial time series (Reutters)
- **Scanner data** in marketing
- **Web scraping**, text data, ...

Challenges caused by features of the data

- Important covariates are missing (**unobserved heterogeneity**), because data often collected for a specific purpose
- Data exhibit non-standard features such as **time-varying parameters** (e.g. oil price shock in the 1970s, financial crisis starting in 2008)
- Model miss-specification due to **non-Gaussianity**, heteroscedasticy; higher-order dependence (volatility clusters)
- **Endogeneity** is a very serious issue in econometrics, because covariates are very often correlated with the error. Ignoring endogeneity causes a bias.
- **Big data** can be small for the problem under investigation
- . . . and many more

Bayesian Econometrics

Chris Sims (2007)

“Why econometrics should always and everywhere be Bayesian ”

- **Coherent framework** for estimation, testing/model selection, and forecasting
- Inherent aspect of **learning**
- Probabilistic modelling allows **uncertainty quantification**
- Probabilistic modelling allows a lot of **flexibility**

Bayesian econometrics in the big data era

- Bayesian econometrics is breaking away from the traditional parametric paradigm.
- Bayesian econometrics is a very active area of research, using the latest technique of Bayesian inference such as:
  - **Flexible, highly structured models** (hierarchical Bayes, state space models, time-varying parameter models)
  - **Shrinkage methods and variable selection** (allow for flexibility and shrinkage at the same time to avoid overfitting and guarantee statistical efficiency)
  - **Sparse Bayesian factor models** (low-dimensional representations and sparse covariance estimation)
  - **Mixture analysis** (unobserved heterogeneity/model-based clustering)
  - **Bayesian nonparametric methods** (semi-parametric IV models, semi-parametric volatility models, . . .)
  - . . . and many more

A recent reference . . .

- Special issue of the Journal of Econometrics on **"Complexity and Big Data in Economics and Finance: Recent Developments from a Bayesian Perspective"** (forthcoming):
  - Bayesian Econometrics using Sequential Monte Carlo [Durham and Geweke, 2018] and self-tuning particle filters (Herbst and Schorfheide, 2018)
  - Incomplete econometrics models and prediction pools (McAllin and West, 2018)
  - Flexible models and fast inference for big data using shrinkage (Bitto and SFS, 2018; Kastner, 2018; Kaufmann and Schuhmacher, 2018; Koop etal, 2018)

Outline

Part I: Big data in Bayesian econometrics

- Econometrics and big data applications
- Austrian Center for Labor Economics and Welfare

NRN Austrian Center for Labor Economics and Welfare

- National Research Network **"The Austrian Center for Labor Economics and the Analysis of the Welfare State"** (2008-2015, funded by the Austrian Science Fund)
- Excellent data basis:
  - **Austrian Social Security Database** (ASSD) [Zweimüller et al., 2009]
  - Administrative individual register data that collects information for old-age security benefits.
- Complete individual employment histories since 1972 for the universe of Austrian employees:
  - **daily** wages
  - demographics (age, gender, number of children, . . .)
  - employer; firm characteristics (industry, firm size, . . .)
  - but **important information missing**: education, working hours, . . .

Big data applications within the NRN

- **Effect of labor market entry on earnings dynamics** [Frühwirth-Schnatter et al., 2012]: panel-data set of nearly 50 000 male workers (daily observations over 20 years)
- **Mothers' long-run career patterns after first birth** [Frühwirth-Schnatter et al., 2016]: panel-data set of more than 230 000 female workers (daily observations up to 20 years)
- **Analysing plant closure effects on labour market states** [Frühwirth-Schnatter et al., 2018b]: panel-data set of more than 5 000 male workers (daily observations up to 10 years) and a control sample
- **Earnings Effects of Maternity Leave** [Jacobi et al., 2016]: panel-data set of more than 30,000 mothers (daily observations over up to 6 years)
- **Effect of family size on labor market and health outcomes** [Frühwirth-Schnatter et al., 2018a]: cross-sectional data from 2009; 8 different outcomes for about 100,000 families

Part II
Model-based Clustering of Time Series

Outline

Part II: Model-based Clustering of Time Series

- Effect of labor market entry on earnings dynamics
- Model-based clustering of time-series
- Plant closure data from ASSD Data base
- Business Application in Marketing

Effect of labor market entry on earnings dynamics

Effect of labor market entry on earnings dynamics:

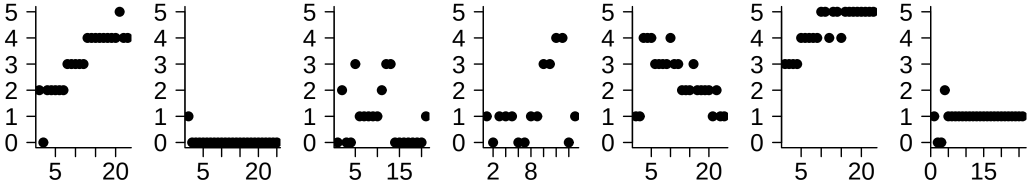
- Data from the **ASSD** (Austrian Social Security Database) [Zweimüller et al., 2009]
- **Cohort study:** entrants into the Austrian labor market in the years 1975 to 1985 (male, Austrian citizenship, at most 25 years)
- Panel-data set of **49 279 time series**; individual lengths range between 2 and 32 years (median equal to 22)

Effect of labor market entry on earnings dynamics

Data reduction:

- gross monthly wages in May of successive years
- wage divided into 6 categories:
  - **0** (zero wage)
  - **1 (lowest) to 5 (highest)**: (population) wage distribution in each year divided according to the quintiles

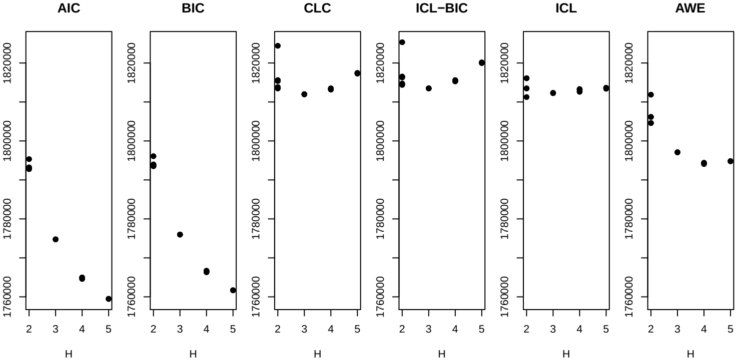
Wage careers of seven randomly selected employees:



Research questions

- Research questions:
  - Identify types of similar wage careers ("clusters")
  - Describe these types in economic terms
  - Explain cluster membership (unemployment rate, skills, . . .)
- Statistical method:
  - Model-based clustering based on finite mixtures of Markov chain models [Pamminger and Frühwirth-Schnatter, 2010]
  - Combined with mixtures-of-experts [Frühwirth-Schnatter et al., 2012]
  - Common criteria (AIC, BIC, . . .) to select the number of clusters

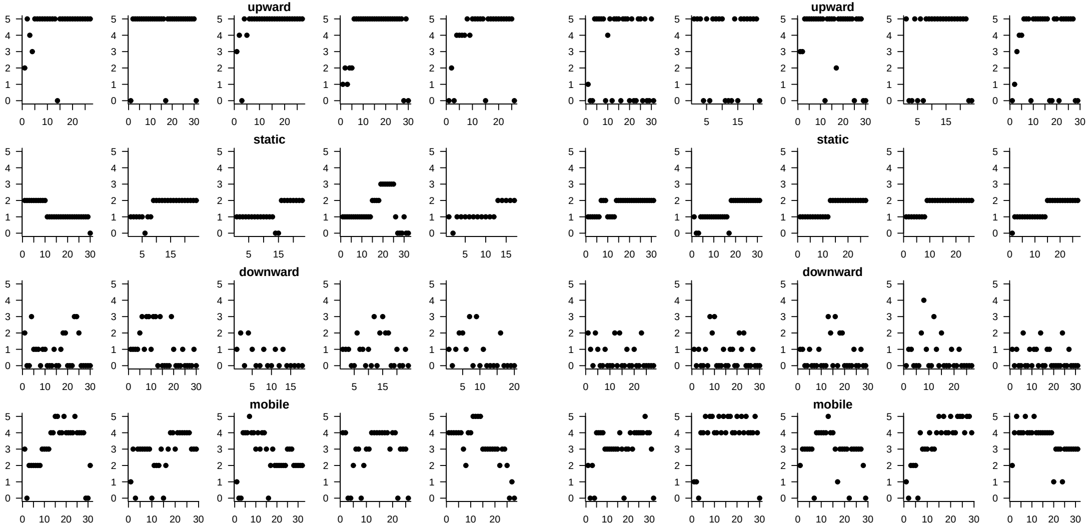
Model selection criteria for various numbers K of clusters (H in the figure)



Economic interpretability led us to choose 4 clusters; confirmed only by AWE

Visualizing the 4 cluster solution

Wage profiles of cluster members ranked 50th, 125th, 250th, 500th, 1000th



Interpretation of the clustering results

Analyze **unobserved heterogeneity in wage mobility**, i.e. the risk and the chance of moving between wage categories using mixtures of Markov chain models:

- Transition matrix ξ_k models wage mobility, e.g. the chance to earn more:

$$\Pr(y_{it} = j + 1 | y_{i,t-1} = j, S_i = k) = (\xi_k)_{j,j+1},$$

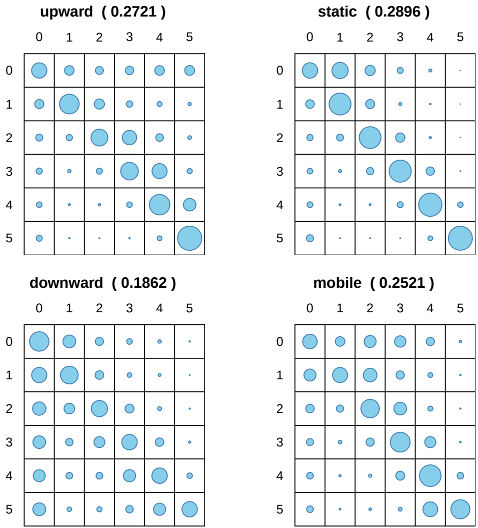
- the risk to earn less:

$$\Pr(y_{it} = j - 1 | y_{i,t-1} = j, S_i = k) = (\xi_k)_{j,j-1},$$

- or to earn the same: $\Pr(y_{it} = j | y_{i,t-1} = j, S_i = k) = (\xi_k)_{j,j}$.

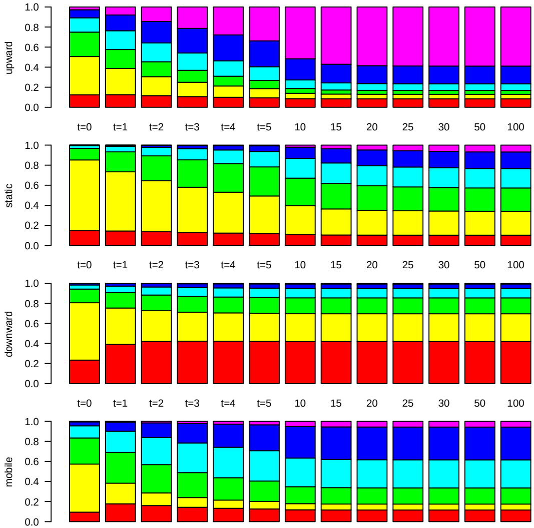
Visualize unobserved heterogeneity in wage mobility

Posterior expectations of transition matrices ξ_1, ξ_2, ξ_3, and ξ_4



Unobserved heterogeneity in the long rum

Posterior expectations of wage distribution $\pi_{k,t} = \pi_{k,0}\xi_k^t$ in cluster k after t years



Mixtures-of-experts approach

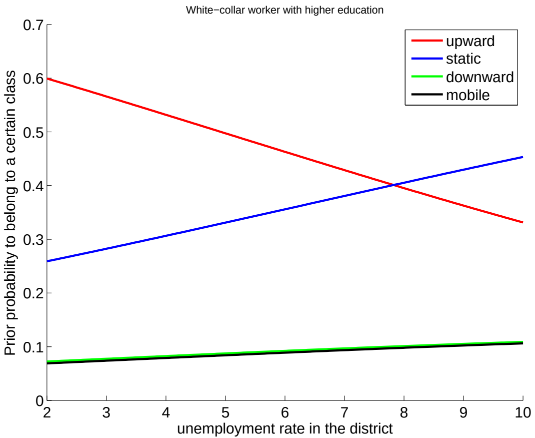
- **Mixture-of-experts model:** multinomial logit model

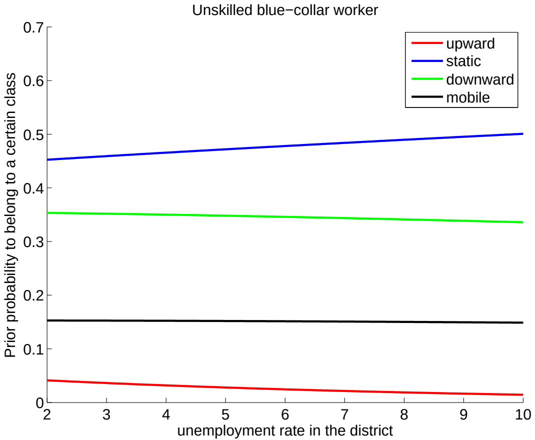
$$\Pr(S_i = k|\mathbf{x}_i) = F(\mathbf{x}_i\beta_k)$$

- Covariates $\mathbf{x}_i$ based on **individual characteristics**:
  - education (unskilled, skilled, higher education)
  - broad type of occupation (white collar/blue collar worker)
  - the initial wage category to handle the initial condition problem
- ... and on **labour market characteristics at the time of entry**:
  - unemployment rate in the region
  - cohort effects, expressed by a set of dummies for the year of labor market entry

What chances have they got?

Comparing prior chances for two young men





Policy implications

Fears caused by an algorithm used by the Public Employment Service Austria (AMS) (gender and nationality play a role)

MONTAG, 15. OKTOBER 2018

Wie der Algorithmus Jobsuchende beim AMS bewerten wird

Das Arbeitsmarktservice hat offengelegt, wie die künftige Einteilung von Arbeitssuchenden in drei Kategorien funktionieren soll. Kritik gibt es, weil auch Staatsbürgerschaft und Geschlecht eine Rolle spielen.

András Szigetvari

Outline

Part II: Model-based Clustering of Time Series

- Effect of labor market entry on earnings dynamics
- Model-based clustering of time-series
- Plant closure data from ASSD Data base
- Business Application in Marketing

Data Structure

Longitudinal data/panel data/**many short time series**:

- Repeated measurements y_{it} are taken on N subjects, indexed by $i = 1, \ldots, N$, at a number of points in time, typically indexed by $t = 0, \ldots, T_i$.
- The outcome y_{it} is a categorical variable with M potential states labelled by $\{1, \ldots, M\}$.
- Subsequently, $\mathbf{y}_i = \{y_{i0}, \ldots, y_{i,T_i}\}$ denotes each individual time series, while $\mathbf{y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$ refers to the whole panel.
- In addition, one may observe exogenous covariates $\mathbf{x}_{it}$ of potential influence on the distribution of the outcome variable y_{it}.

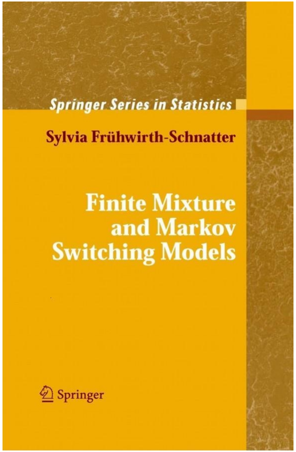
Model-based clustering of time series

General framework for model-based clustering of time series:

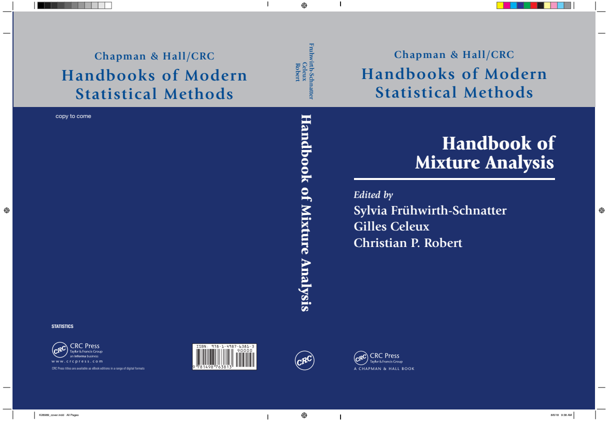
- Each individual time series $\mathbf{y}_i = \{y_{i0}, \ldots, y_{i,T_i}\}$ is treated as a „subject" in a clustering framework.
- Select an appropriate clustering kernel in terms of the sampling density $p(\mathbf{y}_i|\boldsymbol{\vartheta}_k)$ where $\boldsymbol{\vartheta}_k$ is a group-specific parameter
- Use the framework of finite mixtures for estimation and classification using full-blown MCMC

See [Frühwirth-Schnatter and Kaufmann, 2008], [Frühwirth-Schnatter, 2011]

For more details see . . .



2006



2018 (forthcoming)

see [Grün, 2018] for a review of model-based clustering

Capturing serial dependence

Dynamic clustering kernels take account of the serial dependence among the observations $\{y_{i0}, \dots, y_{i,T_i}\}$:

$$p(\mathbf{y}_i|S_i = k) = \prod_{t=1}^{T_i} p(y_{it}|y_{i,t-1}, [\mathbf{x}_{it},]\boldsymbol{\vartheta}_k).$$

Panels of real-valued time series

Clustering kernel: dynamic regression model

Dynamic clustering kernels for panels with real-valued time series observations y_{it} are typically based on

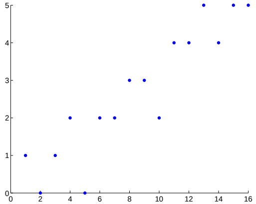
$$y_{it} = \zeta_k + \phi_k y_{i,t-1} + \mathbf{x}_{it}\boldsymbol{\beta}_k + \sigma_k \varepsilon_{it},$$

where $\varepsilon_{it} \sim f(\varepsilon_{it})$ and all parameters of the clustering kernel are cluster-specific, including the variance of the error term.

- The most common choice are $\mathrm{i.i.d.}$ normal errors ε_{it}.
- To achieve robustness against outliers, the Student-t distribution is employed in [Frühwirth-Schnatter and Kaufmann, 2006].
- To capture skewness within each cluster, [Juárez and Steel, 2010] assume that $f(\varepsilon_{it})$ arises from a skew-t distribution.

Panel data of discrete-valued time series

- Each wage career $\mathbf{y}_i$ is a discrete valued time series - individual wage career $\mathbf{y}_i$ of a randomly selected employee (out of 49 279 time series)



- $\mathbf{y}_i$ may be regarded as multivariate categorical vector with dependence among adjacent observations

$$\mathbf{y}_i = (\ 1 \quad 0 \quad 1 \quad 2 \quad 0 \quad 2 \quad 2 \quad 3 \quad 3 \quad 2 \quad 4 \quad 4 \quad 5 \quad 4 \quad 5 \quad 5\ )'$$

Clustering categorical time series

Clustering kernel: First order Markov chain model

$$p(\mathbf{y}_i|S_i = k, \xi_k) = \prod_{t=1}^{T_i} p(y_{it}|y_{i,t-1}, \xi_k)p(y_{i,0}|\xi_k),$$

with initial distribution $p(y_{i,0}|\xi_k)$ and cluster-specific transition matrix ξ_k:

$$(\xi_k)_{hj} = \Pr(y_{it} = j|y_{i,t-1} = h, S_i = k).$$

- Clustering kernel captures the persistence of earning for an individual
- Implies homogeneity within each cluster and stationarity in the long run
- Mixtures of inhomogeneous Markov chain models include covariate information

Outline

Part II: Model-based Clustering of Time Series

- Effect of labor market entry on earnings dynamics
- Model-based clustering of time-series
- Plant closure data from ASSD Data base
- Business Application in Marketing

Plant closure data from ASSD Data base

- Data from the **ASSD** (Austrian Social Security Database) [Zweimüller et al., 2009]
- **Cohort study:** male workers employed in 1982–1988 in firms with more than 5 employees, at least one year of tenure
- Plant closure: an employer identifier ceases to exist, take-over or merger: more than 50% of the employees continue under a new employer identifier
- Workers aged between 35 and 55 at time of plant closure
- Research question:
  - What is the **effect of plant closure** on the employment career?
  - Is there a **difference** between workers facing plant closure and those who did not?

Plant closure data from ASSD Data base

- 5,841 workers displaced by plant closures between 1982 and 1988 (panel with $N = 5,841$ time series with $T_i \leq 40$ quarterly data)
- Outcome variable y_{it}, $i = 1, \ldots, N$, $t = 1, \ldots, 40$ (t is distance from plant closure in quarters):

1	employed
2	sick leave
3	out of labor force
	(registered as unemployed or otherwise out of labor force)
4	retired (claiming government pension benefits)

- Controls: all male workers in the cohort who did not experience a plant closure (more than 1 million workers)

Mixture of inhomogeneous Markov chain models

- Time-inhomogeneity present both for the plant closure data as well as for the marketing application.
- **Generalized transition matrices:** inhomogeneous transition matrix depending on a history $\mathcal{H}_{it}$ [Frühwirth-Schnatter, 2011]:

$$\Pr(y_{it} = j|\mathcal{H}_{it}, S_i = k),$$

  where $\mathcal{H}_{it} = \{y_{i,t-1}, \mathbf{x}_{it}\}$.
- Typically $\mathbf{x}_{it}$ is some discrete covariate, e.g. the year after plant closure:

$$\boldsymbol{\vartheta}_k = (\boldsymbol{\pi}_k, \boldsymbol{\xi}_{k,1}, \boldsymbol{\xi}_{k,2}, \dots, \boldsymbol{\xi}_{k,10})$$

- We could include addition information (age group, ...) in $\mathbf{x}_{it}$

Choosing K for the plant closure time series data

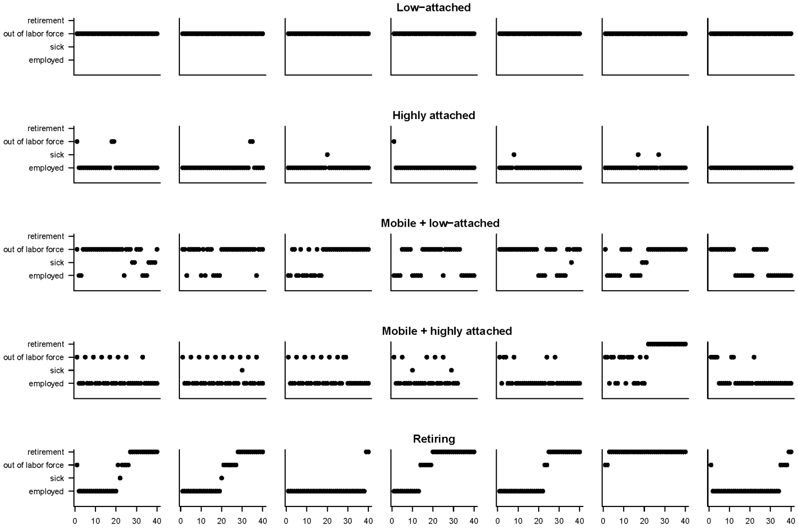
- Statistical model selection criteria for various numbers K of clusters:

K	2	3	4	5	6
AIC	112160.9	110381.0	109113.4	**107567.6**	108057.0
BIC	113575.5	112549.6	112036.0	**111244.2**	112487.7
AWE	114402.1	114188.3	**114159.6**	114539.8	116356.4

- **Economic interpretability led us to choose 5 clusters**; confirmed by AIC and BIC

Model-based clustering of plant closure data

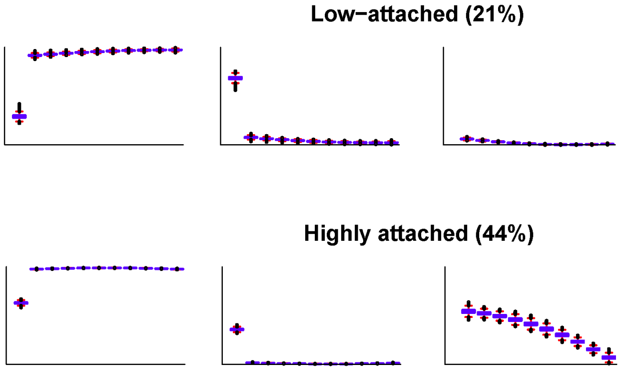
Employment profiles of cluster members ranked 10th, 25th, 50th, 70th, 100th, 200th, 350th



Who suffers and who doesn't?

- Cluster 1 - The ones who really suffer: low level of attachment to the labor market, slow/no recovery after plant closure
- Cluster 2 - The good and lucky ones: high level of attachment to the labor market, quick recovery after plant closure
- Cluster 3 - The less lucky, less fit ones: high mobility between in and out of labor force; low level of attachment to the labor market
- Cluster 4 - The less lucky, but fit ones: high mobility between in and out of labor force; high level of attachment to the labor market
- Cluster 5 - The sick and tired ones: increasing chance to move to retirement, either directly or through the channel of a sick leave

Time-varying transition probabilities

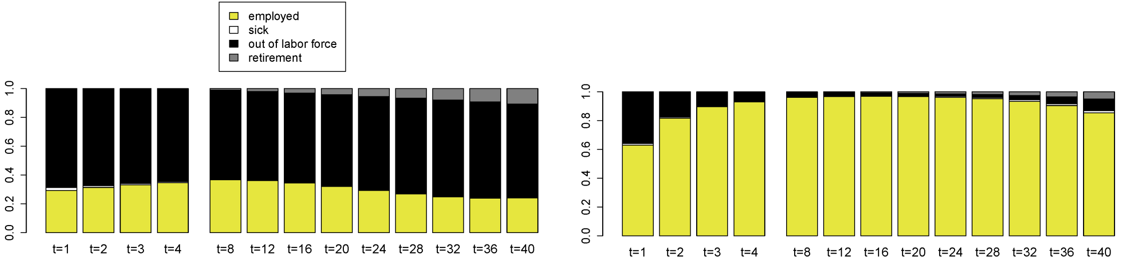


Transition probabilities $1 \to 1$, $1 \to 3$, $3 \to 1$ for cluster 1 (top) and cluster 2 (bottom)

State distribution $\pi_{k,t}$, where

$$\pi_{k,t} = \pi_k \xi_{k,1\to t}, \qquad \xi_{k,1\to t} = \xi_{k,1\to(t-1)} \xi_{ky}; \quad \xi_{k,1\to 2} := \xi_{k1}.$$

over distance $t = 4(y-1) + q$ from plant closure (quarters), for cluster k:



Cluster 1: low attached

Cluster 2: highly attached

Construction of a control group

- **Statistical matching based on characteristics before plant closure** (age, broad occupation, location, industry, employment history in year before closure)
- Experience of a plant closure effects the state distribution $\boldsymbol{\pi}_h$ of displaced workers only in the **first quarter after displacement**, subsequent transition behaviour is the same as for controls.
- **State probability in first quarter after (potential) plant closure is different for controls**:

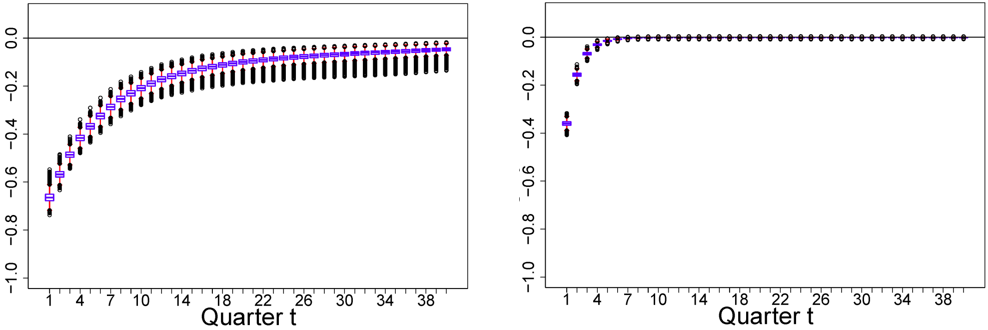
$$\boldsymbol{\pi}_k^c = (\pi_{k,1}^c, \ldots, \pi_{k,4}^c), \quad \pi_{k,j}^c = \Pr(y_{i1}^c = j | S_i^c = k).$$

- $\boldsymbol{\pi}_k^c$ estimated for controls, supervised clustering for the transitions matrices in each cluster k.

Treatment effects

Difference $\pi^c_{k,t} - \pi_{k,t}$ in dynamic probability to be employed, over distance t from plant closure for cluster k:

$$\pi_{k,t} = \pi_k \xi_{k,1\to t}, \quad \pi^c_{k,t} = \pi^c_k \xi_{k,1\to t}.$$



Cluster 1: low attached

Cluster 2: highly attached

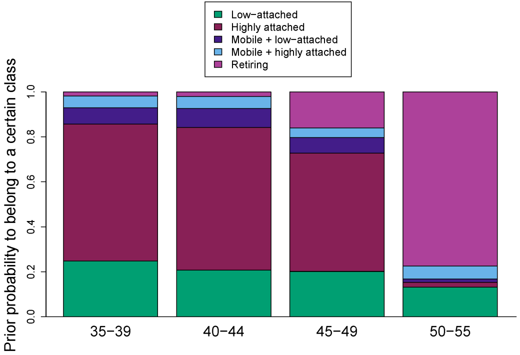
Mixtures-of-experts approach

- **Mixture-of-experts model:** multinomial logit model

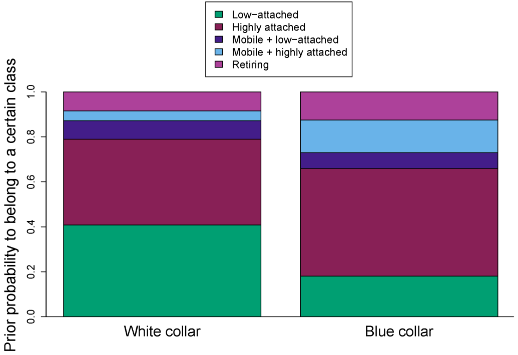
$$\Pr(S_i = k|\mathbf{x}_i) = F(\mathbf{x}_i \beta_k)$$

- Covariates $\mathbf{x}_i$ based on **individual characteristics**:
  - age at the time of plant closure (five age groups: 35-39, 40-44, 45-49, 50-55)
  - levels of experience (low, medium, high)
  - broad occupational status (blue versus white collar)
  - income before plant closure (low, medium, high) based on the tertiles of the general income distribution at time of plant closure
- ... and on **firm characteristics**:
  - three categories of firm size (1-10, 11-100, and more than 100 employees)
  - four broad economic sectors (service, industry, seasonal business outside of hotel and construction, unknown)

Prior probabilities to belong to a cluster

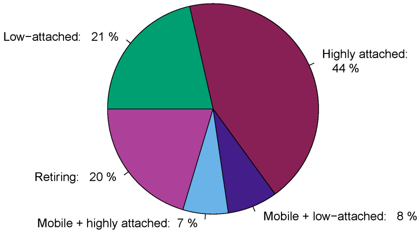


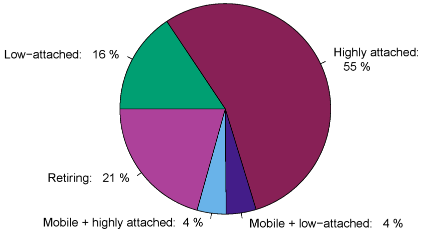
Impact of age



Impact of white versus blue collar

Cluster sizes





Workers experiencing plant closure

Workers not experiencing plant closure

Outline

Part II: Model-based Clustering of Time Series

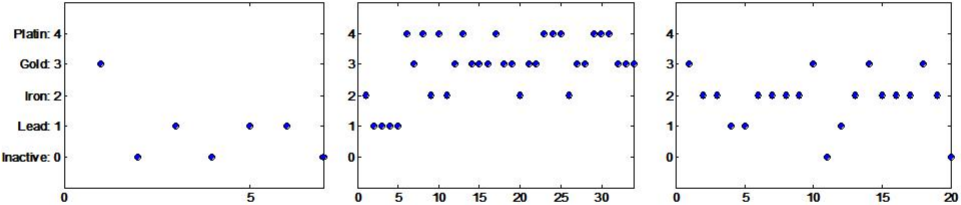
- Effect of labor market entry on earnings dynamics
- Model-based clustering of time-series
- Plant closure data from ASSD Data base
- Business Application in Marketing

Business Application in Marketing

- **B**ayes for **B**ig **B**usiness
- **Customer relationship management (CRM)** (joint work with Thomas Reutterer and Stefan Pittner):
  - Loyal costumers contribute more to a firm's profitability than short-term costumers
  - Analyze monthly purchases of 6601 costumers over 5 years (60 periods from Feb 2000 through Jan 2005) who are involved in a costumer loyalty program with a higher-level clothing retails chain
  - Allocate marketing budgets more profitably

Time series observations

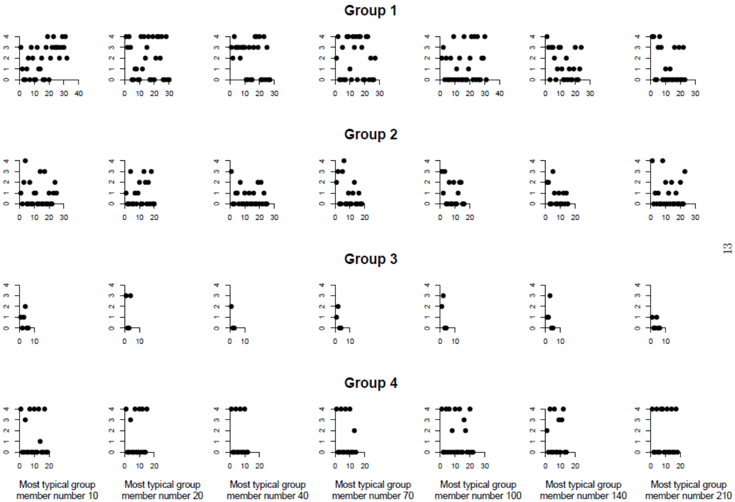
- Examples of evolving costumers:



- Market segmentation to identify loyal costumers early on
- Inhomogeneous Markov chain clustering: allow the **transition matrix** to evolve over time and **account for periods of inactivity**

Resulting cluster solution

4 groups of costumers, holdout analysis shows higher “hit rate” than common models



Part III

Open issues and future challenges

Some thoughts about the prior

- Priors might be informative also in a big data setting (see [Celeux et al., 2018] for high-dimensional mixtures)
- Priors should allow shrinkage to deal automatically with over-fitting models
- **Blessing of informative priors:** in highly overparameterised models, **shrinkage** is introduced through the prior
- **Objective priors** are fine for regular likelihoods, but econometric models often lead to **near-boundary inference and non-regular likelihoods**:
  - Student-t likelihoods, finite mixtures, random-effect models, dynamic (state space) models, IV estimation, . . .
  - Priors should respect the geometry of the likelihood
  - Prior should **down-weight** regions of the parameter space that contain **spurious results**
  - Priors should **always** lead to a **proper posterior** (with finite moments?)

Identifiability and weak identifiability

- Data often are not always informative about parameters - **identical** or **similar likelihood** values with different parameter combinations:
  - IV estimation and cointegration models [Baştürk et al., 2017]
  - overfitting mixtures [Frühwirth-Schnatter and Malsiner-Walli, 2018]
  - mixtures for discrete data [Gormley and Frühwirth-Schnatter, 2018]
  - factor models [Conti et al., 2014, Frühwirth-Schnatter and Lopes, 2018]
- Dealing with **identification problems** within a Bayesian framework:
  - Posterior might be **improper** or posterior moments might not exist; ergodic averages of MCMC draws **useless**
  - diagnosing weak identifiability within MCMC [Koop et al., 2013]

EFaB

Chris Sims (2007)

"*Why econometrics should always and everywhere be Bayesian* "

- Consider joining the **E**conomics, **F**inance, **a**nd **B**usiness (EFaB) Section of ISBA
- It's only *A Fistful of Dollars*, but with high probability it's an investment with high returns

Thank you very much for your attention!

Part IV
References

Baştürk, N., Hoogerheide, L., and van Dijk, H. K. (2017).

Bayesian analysis of boundary and near-boundary evidence in econometric models with reduced rank.

Bayesian Anal., 12(3):879–917.

Celeux, G., Kamary, K., Malsiner-Walli, G., Marin, J.-M., and Robert, C. P. (2018).

Computational solutions for bayesian inference in mixture models.

In Frühwirth-Schnatter, S., Celeux, G., and Robert, C. P., editors, *Handbook of Mixture Analysis*, chapter 5, pages 77–100. CRC Press, Boca Raton, FL.

Conti, G., Frühwirth-Schnatter, S., Heckman, J. J., and Piatek, R. (2014).

Bayesian exploratory factor analysis.

Journal of Econometrics, 183:31–57.

Durham, G. and Geweke, J. (2018).

Adaptive sequential posterior simulators for massively parallel computing environments.

Journal of Econometrics, missing:No, book Jeliazkov.

Frühwirth-Schnatter, S. (2011).

Panel data analysis - a survey on model-based clustering of time series.

Advances in Data Analysis and Classification, 5:251–280.

Frühwirth-Schnatter, S., Halla, M., Posekany, A., Pruckner, G., and Schober, T. (201(8+)a).
Brothers and sisters: Analysing the effect of family size using instrumental variable approaches.
Submitted, missing:XX–XX.

Frühwirth-Schnatter, S. and Kaufmann, S. (2006).
How do changes in monetary policy affect bank lending? An analysis of Austrian bank data.
Journal of Applied Econometrics, 21:275–305.

Frühwirth-Schnatter, S. and Kaufmann, S. (2008).
Model-based clustering of multiple time series.
Journal of Business & Economic Statistics, 26:78–89.

Frühwirth-Schnatter, S. and Lopes, H. (2018).
Sparse Bayesian Factor Analysis when the Number of Factors is Unknown.
arXiv preprint 1804.04231.

Frühwirth-Schnatter, S. and Malsiner-Walli, G. (2018).
From here to infinity - sparse finite versus Dirichlet process mixtures in model-based clustering.
Advances in Data Analysis and Classification, page conditionally accepted.
arXiv:1706.07194.

Frühwirth-Schnatter, S., Pamminger, C., Weber, A., and Winter-Ebmer, R. (2012).
Labor market entry and earnings dynamics: Bayesian inference using mixtures-of-experts Markov chain clustering.
Journal of Applied Econometrics, 27:1116–1137.

Frühwirth-Schnatter, S., Pamminger, C., Weber, A., and Winter-Ebmer, R. (2016).
Mothers' long-run career patterns after first birth.
Journal of the Royal Statistical Society, Series A, 179:707–725.

Frühwirth-Schnatter, S., Pittner, S., Weber, A., and Winter-Ebmer, R. (2018b).
Analysing plant closure effects using time-varying mixture-of-experts markov chain clustering.
Annals of Applied Statistics, 12:1786–1830.

Gormley, I. C. and Frühwirth-Schnatter, S. (2018).
Mixture of experts models.
In Frühwirth-Schnatter, S., Celeux, G., and Robert, C. P., editors, *Handbook of Mixture Analysis*, chapter 12, pages 279–315. CRC Press, Boca Raton, FL.

Grün, B. (2018).
Model-based clustering.

In Frühwirth-Schnatter, S., Celeux, G., and Robert, C. P., editors, *Handbook of Mixture Analysis*, chapter 8, pages 163–198. CRC Press, Boca Raton, FL.

Jacobi, L., Wagner, H., and Frühwirth-Schnatter, S. (2016).
Bayesian treatment effects models with variable selection for panel outcomes with an application to earnings effects of maternity leave.
Journal of Econometrics, 193:234–250.

Juárez, M. A. and Steel, M. F. J. (2010).
Model-based clustering of non-Gaussian panel data based on skew-t distributions.
Journal of Business & Economic Statistics, 28:52–66.

Koop, G., Pesaran, M. H., and Smith, R. P. (2013).
On identification of Bayesian DSGE models.
Journal of Business & Economic Statistics, 31:300–314.

Pamminger, C. and Frühwirth-Schnatter, S. (2010).
Model-based Clustering of Categorical Time Series.
Bayesian Analysis, 5:345–368.

Zweimüller, J., Winter-Ebmer, R., Lalive, R., Kuhn, A., Wuellrich, J.-P., Ruf, O., and Büchi, S. (2009).

The Austrian Social Security Database (ASSD).
Working Paper 0903, NRN: The Austrian Center for Labor Economics and the Analysis of the Welfare State, Linz, Austria.