universitätbonn

# Bayesian estimation of non-stationary Markov models combining micro and macro data

*Hugo Storm, Thomas Heckelei*

*Institute for Food and Resource Economics (ILR), University of Bonn*
*Contact: hugo.storm@ilr.uni-bonn.de*

***Poster prepared for presentation at the Agricultural & Applied Economics Association's 2011 AAEA & NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania, July 24-26, 2011***

# Bayesian estimation of non-stationary Markov models combining micro and macro data

*Hugo Storm, Thomas Heckelei, Institute for Food and Resource Economics (ILR), University of Bonn*

## Objective:

(1) Develop a Bayesian estimation framework for non-stationary Markov models that allows combining micro and macro data based estimation techniques previously considered as alternatives only

(2) Evaluate the derived estimator using Monte Carlo Simulations

## Relevance: Farm Structural Change

***General aims:*** (1) Indentify and quantify factors that determine farm structural change (2) Predict structural change in response to these factors

- Often modeled as a Markov Process (Zimmermann et. al 2009)
- But current estimation techniques do not allow using available micro and macro data in a satisfying way

## Background on Markov processes

- A Markov process allows to model the movement of individuals between a finite number of predefined states, i=1,...,k , as a stochastic process
- A Markov process is characterized by a transition probability matrix $\mathbf{P}_t$
- The vector $\mathbf{n}_t$ denotes the number of individuals in each state and develops over time, t, according to a (first order) Markov process:

$$\mathbf{n}_t = \mathbf{P}_t' \mathbf{n}_{t-1}$$

## Specification of $\mathbf{P}_t$

- $\mathbf{P}_t$ is assumed to be a function of explanatory variables
- We propose two different specifications for ordered and unordered Markov states based on the multinomial logit model and the ordered logit model
- Main differences: (1) Multinomial logit model requires assumption of iid errors which might be inappropriate with ordered Markov states
  (2) Ordered logit model requires less parameter to be estimated

# Combining...

## macro data...

***Definition:***

- The number of individuals in each class, $\mathbf{n}_t$, is observed over time
- Individual transitions are not observed and many different transitions could result in the observed data
- $\mathbf{P}_t$ needs to be estimated

***Data availability:***

- (Usually) good
- Example: For the analysis of EU farm structural change it is available from the Farm Structure Survey at population level

**Example: Macro data**

| State | A | B | C |
|---|---|---|---|
| Size | Small | Medium | Large |
| Farms in t=0 | 60 | 30 | 10 |
| Farms in t=1 | 40 | 40 | 20 |

## with micro data...

***Definition:***

- The movements between classes is observed for each individual over time
- The micro transition matrix give the number of individuals transiting from a specific state in t-1 to a state in t
- $\mathbf{P}_t$ can be calculated directly

***Data availability:***

- (Usually) limited
- Example: For the analysis of EU farm structural change a sample is available from the Farm Accountancy Data Network (FADN)

**Example:**
**Micro transition matrix**

| from \ to | A | B | C | Σ |
|---|---|---|---|---|
| A | 30 | 10 | 10 | 60 |
| B | 10 | 10 | 10 | 30 |
| C | 0 | 10 | 0 | 10 |
| Σ | 40 | 40 | 20 | |

**Posterior** ∝ **Likelihood** × **Prior**

### Likelihood

Macro data based likelihood: $\mathbf{n}_t$ are distributed as weighted sum of independent multinomials (MacRae 1977)

$$L\left(\mathbf{n}_1,\ldots,\mathbf{n}_T \mid \boldsymbol{\beta}\right) = \prod_{t=1}^{T} \sum_{\mathbf{H}_t \in \mathbb{H}_t} \prod_{i=1}^{k} n_{i\,t-1}! \left( \prod_{j=1}^{k} P_{ijt}^{\eta_{ijt}} / \eta_{ijt}! \right)$$

A large sample approximation is employed (Brown and Payne 1986)

### Prior

A micro data based likelihood function is specified for the prior weights

Micro transitions are multinomial distributed with size equal the number of individuals in the corresponding class in t-1 and probabilities of $\mathbf{P}_t$

$$p\left(\boldsymbol{\beta}\right) = \prod_{t=1}^{T} \prod_{i=1}^{k} n_{i\,t-1}! \left( \prod_{j=1}^{k} \mathbf{P}_{ijt}^{n_{ijt}} / n_{ijt}! \right)$$

# in Bayesian estimation of Markov models

## Computation

- A sample from the posterior density is obtained via a random walk Metropolis Hastings algorithm
- The posterior mean, which is the optimal Bayesian estimator under squared error loss, is approximated by the mean of the posterior sample

## Monte Carlo Simulation

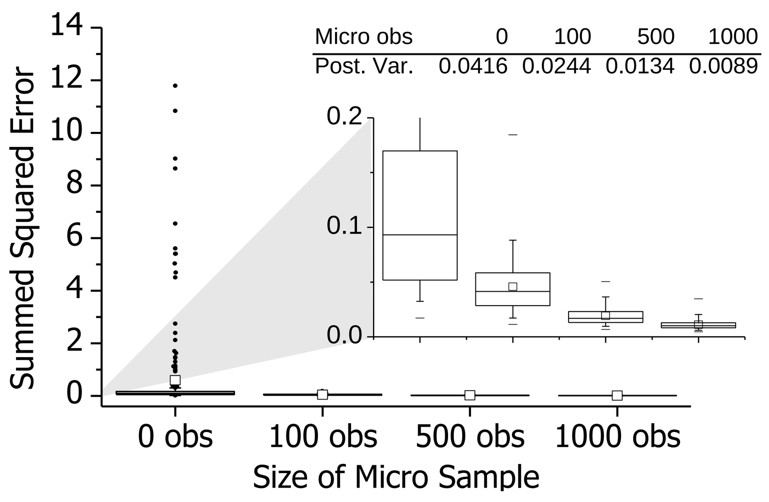
- ***Aim:*** Analyse influence of prior on posterior and estimator performance
- Separate simulation for ordered/unordered Markov states
- 10 *true* models with 20 repetitions each

| Scenarios | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Size of micro sample | 0 | 100 | 500 | 1000 |
| Markov States (k) | 3,4,5 | | | |
| Size of macro sample | 10.000 | | | |
| Time periods | 100 | | | |
| Expl. variables ($n_z$) | 6 (incl. a constant) | | | |

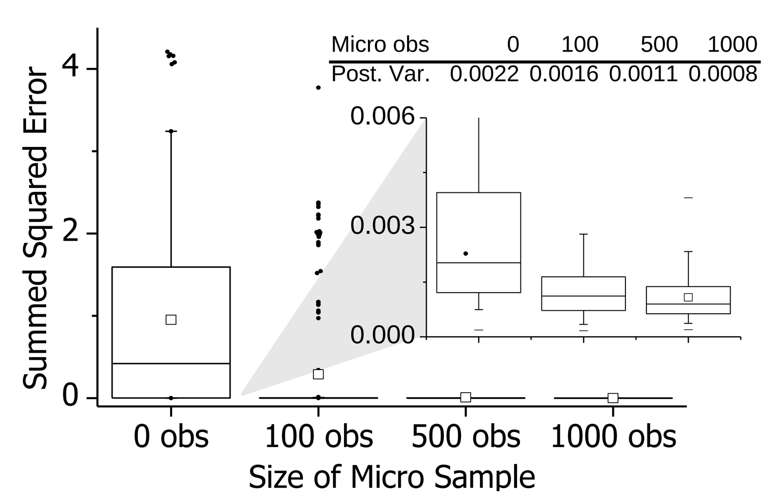
## Monte Carlo Results

Box-Whisker plots of the (summed) squared deviation from the true values in the 200 simulation runs and (summed) variance of the posterior density

Exemplary result




Four unordered Markov states




Four ordered Markov states

## Conclusions

- Inclusion of micro data as prior information reduces Mean Square Error (MSE) and posterior variance
- Improvement stronger the more Markov states are considered
- Prior information increases numerical stability of the estimation

## References


Brown, P.J. and C. D. Payne (1986): Aggregate data, ecological regression, and voting transitions. In: Journal of the American Statistical Association 81(394):452–460.

MacRae, E.C. (1977): Estimation of Time-Varying Markov Processes with Aggregate Data. In: Econometrica 45(1):183–198.

Zimmermann, A.; T. Heckelei and I. P. Domínguez (2009): Modelling farm structural change for integrated ex-ante assessment: review of methods and determinants. In: Environmental Science & Policy 12(5):601–618.


**Contact:**
Hugo Storm, e-mail: hugo.storm@ilr.uni-bonn.de Tel: +49-228 - 73 2323
Institute for Food and Resource Economics, University of Bonn, Nussallee 21, 53115 Bonn, Germany

**Acknowledgements**
This research is funded by the German Research Foundation (DFG), grant No. 585064