Predicting agricultural structural change using census and sample 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 2012 AAEA Annual Meeting, Seattle, Washington, August 12-14, 2012

Predicting agricultural structural change using census and sample data

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

Problem & Policy relevancy

- One important characteristic of the state of the agricultural sector is information about the number of farms in different size or farm specialization classes
- To evaluate and adjust policies appropriately it is important that this information is available for the most recent years
- In the EU this information is only available every 2-3 years from the Farm Structural Survey (FSS)
- Additionally, there is yearly information about the movement of individual farms between classes from a sample of farms from the Farm Accountancy Data Network (FADN)

Objective

- Predict number of farms for years after the last FSS year for which FADN data is available, using these FADN data in combination with all FSS and FADN data from previous years
- Evaluate the proposed approach in comparison of naïve prediction methods in different out-of sample predictions

Background information on data sources

	Farm Structural Survey (FSS)	Farm Accountancy Data Network (FADN)
Coverage	All agricultural holdings	Sample of farms available on a yearly bases
Availability	Every 2-3 Years (census every 10, 3 surveys in-between)	Every year
Level of aggregation	Individual farms can not be identified (aggregated information)	Individual farms can be indentified, development of a farm can be track over time
Information provided	Aggregated information about the total number of farms in different size or farm type's classes (macro data)	Information about the movement of farms in the sample between size or farm type classes (micro data)

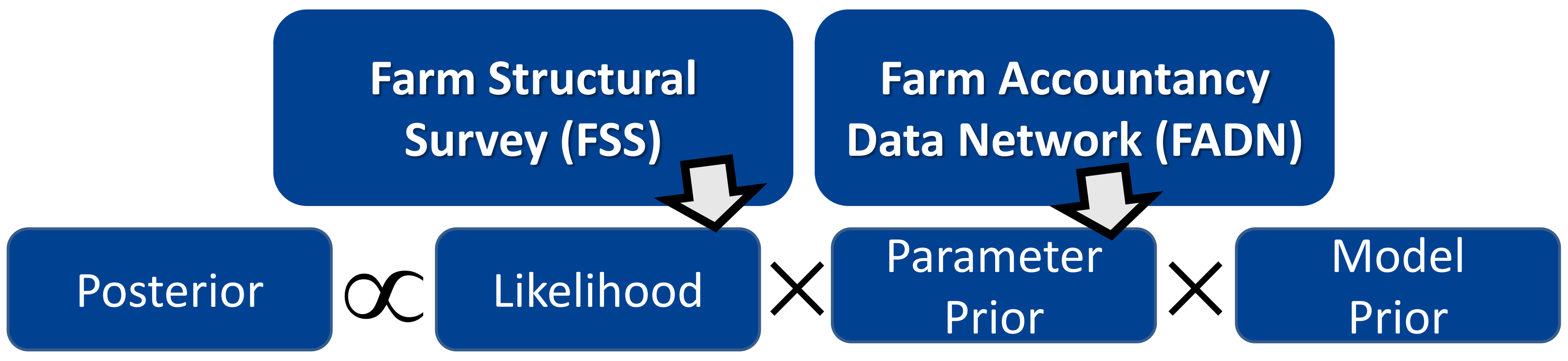
Method

Modeling of farm movement

- Movement of farms between classes is modeled as a Markov process

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

- with $\mathbf{P}_t$ being a function of explanatory variables and unknown coefficients $\boldsymbol{\beta}$
- Prediction of farm numbers can directly be based on the estimated transition probabilities



$$h\left(\boldsymbol{\beta},\boldsymbol{\gamma}\,|\,\mathbf{d}_{FSS},\mathbf{d}_{FADN}\right) \propto f\left(\mathbf{d}_{FSS}\,|\,\boldsymbol{\beta},\boldsymbol{\gamma}\right) f\left(\mathbf{d}_{FADN}\,|\,\boldsymbol{\beta},\boldsymbol{\gamma}\right) f\left(\boldsymbol{\gamma}\right)$$

MCMC sampling algorithm

- A reversible jump MCMC algorithm is employed (Green 1995) to sample from the joint posterior
- The sampler allows jumps between different models specifications, potentially of different dimension
- Implemented version builds on Fouskakis et al. 2009 using several parallel chains, heated at different simulated temperatures

Estimation framework

- Estimation builds on a Bayesian framework (Storm et al. 2011) to combine yearly FADN data $\mathbf{d}_{FADN}$ and FSS data $\mathbf{d}_{FSS}$, available every 2-3 years, in the estimation of yearly transition probabilities
- The framework is extended to considering explicitly uncertainty in the model selection by specifying a joint posterior density of model $\boldsymbol{\gamma}$ and model parameter $\boldsymbol{\beta}$

Bayesian model averaging and Bayesian prediction approach

- Instead of selecting one single model or one parameter point estimate the full posterior sample is used for prediction
- Farm numbers are predicted for each posterior sample outcome
- Resulting prediction distribution reflects uncertainty in the model selection and parameter estimation

Implementation & Results

Design of the three out-of-sample predictions

Prediction period	2000-2003	2003-2005	2005-2007
FSS Data used	1990,1993,1995, 1997,2000	... +2003	... +2005
FADN data used	1989 to 2003	... to 2005	... to 2007
Regional coverage	7 West German regions		
Classes	(1) Entry/exit; (2) Small; (3) Medium; (4) Large		
Measure for prediction quality	Absolute Scaled Error (ASE)		
Reference prediction methods	Naïve constant, linear and geometric prediction		

Steps to model selection and estimation

1. Explanatory variables are identified based on theoretical grounds and empirical findings
2. Two preliminary runs of the sampler are used to restrict the parameter space by excluding all parameters with a inclusion probability <50% in both runs
3. Two runs of the sampler are preformed in the restricted parameters space. A comparison of the marginal inclusion probabilities and the specification with the highest model probability in both runs provides a check for convergence

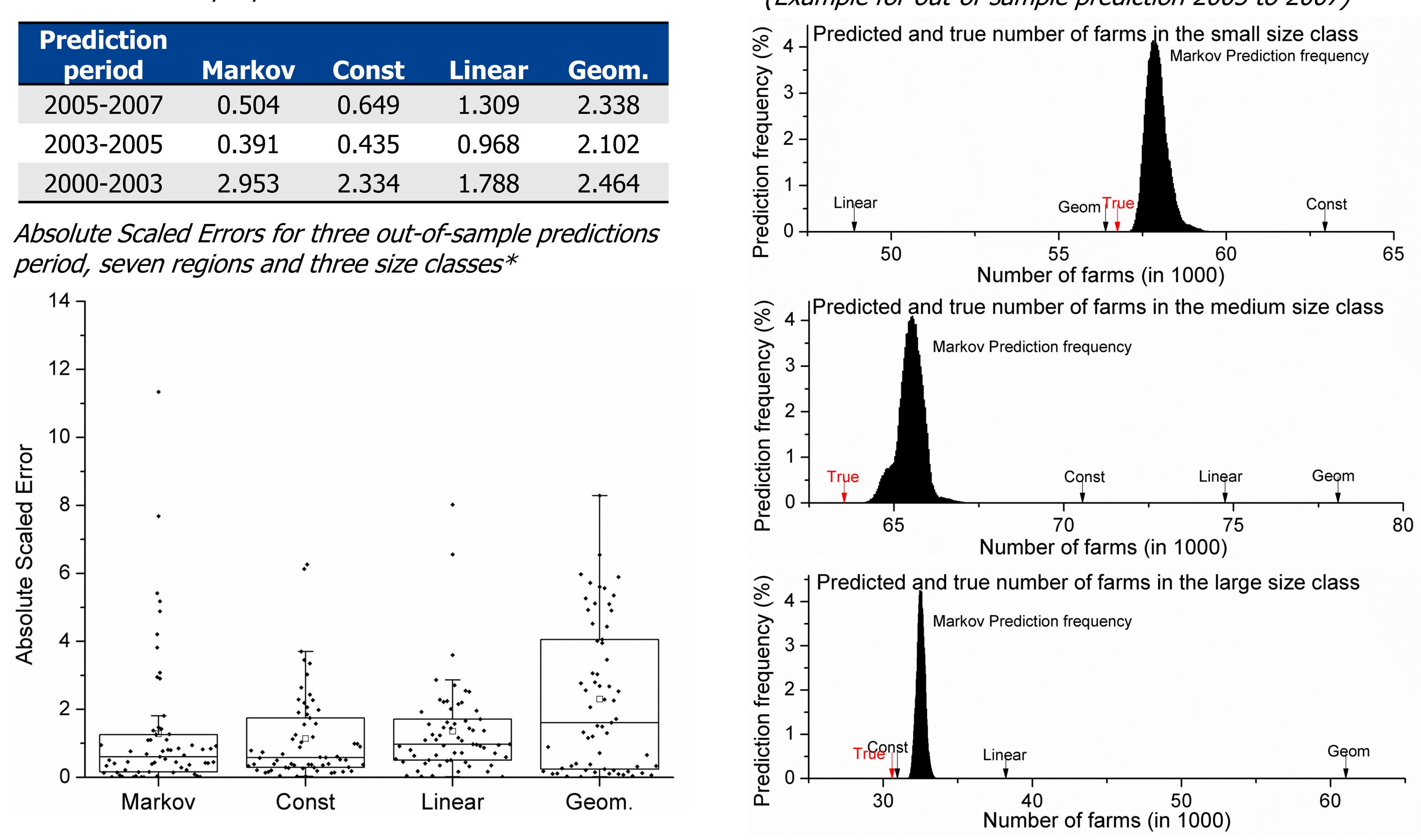
Results for prediction of farm numbers in different size classes

*Mean Absolute Scaled Error of different prediction methods in the out-of-sample predictions**

Prediction period	Markov	Const	Linear	Geom.
2005-2007	0.504	0.649	1.309	2.338
2003-2005	0.391	0.435	0.968	2.102
2000-2003	2.953	2.334	1.788	2.464

*Absolute Scaled Errors for three out-of-sample predictions period, seven regions and three size classes**

Comparison of different prediction methods results (Example for out-of-sample prediction 2005 to 2007)



*For the Markov prediction the mean of the prediction distribution is used as a point prediction

Conclusion & Outlook

- First results indicate that the proposed estimation framework outperformed naïve prediction methods in most of the cases
- Results from the prediction period 2000-2003 indicating problems of the approach when estimation is based on only a few observations
- In further steps the approach will be adopted for the prediction of farm numbers in farm specializations to broaden the bases of comparison

References

Fouskakis D, Ntzoufras I, Draper D. 2009. Population-based reversible jump Markov chain Monte Carlo methods for Bayesian variable selection and evaluation under cost limit restrictions. Journal of the Royal Statistical Society: Series C (Applied Statistics) 58: 383–403.

Green PJ. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82: 711–732.

Storm H, Heckelei T, Mittelhammer RC. 2011. Bayesian estimation of non-stationary Markov models combining micro and macro data. Discussion Paper 2011:2, Institute for Food and Resource Economics, University of Bonn.

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

Acknowledgements
This research is partially funded by the German Research Foundation (DFG), grant No. 585064 and The Institute for Prospective Technological Studies (IPTS), grand No. 2010-J05-23-NC