AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

ANALYSING FARMLAND RENTAL RATES USING BAYESIAN GEOADDITIVE QUANTILE REGRESSION

Alexander März[1], Nadja Klein[2], Thomas Kneib[2], Oliver Mußhoff[1]

[1] Georg-August-University Göttingen, Faculty of Agricultural Sciences, Department of Agricultural Economics and Rural Development, Farm Management Group, Platz der Göttinger Sieben 5, 37073 Göttingen, Germany.
alexander.maerz@agr.uni-goettingen.de (corresponding author)

[2] Georg-August-University Göttingen, Chair of Statistics, Platz der Göttinger Sieben 5, 37073 Göttingen, Germany.

Paper prepared for presentation at the EAAE 2014 Congress
'Agri-Food and Rural Innovations for Healthier Societies'

August 26 to 29, 2014
Ljubljana, Slovenia

Copyright 2014 by Alexander März, Nadja Klein, Thomas Kneib and Oliver Mußhoff. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

ANALYSING FARMLAND RENTAL RATES USING BAYESIAN GEOADDITIVE QUANTILE REGRESSION

Abstract

Empirical studies on farmland rental rates have predominantly concentrated on modelling conditional means using spatial autoregressive models, where a linear functional form between the response and the covariates is assumed. This paper extends the hedonic pricing literature by modelling conditional quantiles of farmland rental rates semi-parametrically using Bayesian geoadditive quantile regression models. The flexibility of this model class overcomes the problems associated with functional form misspecifications and allows us to present a more detailed analysis. Our results stress the importance of making use of semi-parametric regression models as several covariates influence farmland rental rates in an explicit non-linear way.

Keywords: *Bayesian Geoadditive Quantile Regression, Farmland Rental Rates, Hedonic Pricing Models, Penalized Splines, Structured Additive Regression.*

1 Introduction

Farmland is one of the most important production factors in agriculture. Based on cash-flow considerations, farmers have to decide whether to buy or lease agricultural land. Amongst others, one advantage of leasing is that farmers can use their cash reserves to invest in new agricultural machinery and equipment, rather than to tie up capital in land purchases (Ciaian et al., 2012). This preference might explain why Germany is among European countries with a high share of rented farmland: in 2008, on average 70% of the total German agricultural farmland was leased, where the share of rented farmland was considerably higher in East Germany (80%) compared to West Germany (60%) (Ciaian et al., 2010, 2012). Moreover, Ciaian et al. (2010) report that German farmland rental rates exhibit substantial spatial variation with rental rates being almost twice as high in West Germany than in East Germany. These figures, and the fact that farmland rental rates have increased considerably over the last years, imply that the analysis of rental rates is of great importance in practice.

Besides its relevance for farmers, the analysis of farmland rental rates and their determinants is an active field of research in agricultural economics. Herriges et al. (1992), Bierlen et al. (1999), Lence and Mishra (2003), as well as Roberts et al. (2003) and Kirwan (2009) analyse the determinants of price formations on agricultural rental markets in the United States. Fuchs (2002) analyses rental rates of farmland and their determinants in Belgium, Denmark, France, Germany and the Netherlands. Doll and Klare (1995), Drescher and McNamara (2000), Brümmer and Loy (2001), Breustedt and Habermann (2008), Margarian (2008), Breustedt and Habermann (2009) and Breustedt and Habermann (2011) investigate the determinants of rental rates in Germany. Kilian et al. (2008), as well as Breustedt and Habermann (2010) and Habermann and Ernst (2010) analyse the effects of increased land use for the production of bioenergy on German rental rates.

Most empirical studies that analyse farmland rental rates and their determinants make use of hedonic pricing models, as originally proposed by Court (1939) and popularized by Griliches (1961), Lancaster (1966) and Rosen (1974). According to hedonic pricing theory, farmland rental rates can be divided into the sum of its attributes‘ values which are then estimated using regression models. In order to avoid biased estimates and misleading inference resulting from spatial dependencies in the data, spatial autoregressive models have evolved as a standard tool in hedonic pricing studies of farmland rental rates. However, a remaining problem with hedonic pricing models is related to the choice of an appropriate functional form, since there is no theory that guides the researcher (Martins-Filho and Bin, 2005). A common workaround is to use data transformations as proposed by Box and Cox (1964). The functional form chosen by the Box-Cox technique may, however, not adequately approximate the true relationship between covariates and the response. In spatial modeling, misspecifications of the dependence structure have particularly severe consequences; as an illustration, Kostov (2009) reanalyses the data of Patton and McErlean (2003) using semi-parametric regression models, and concludes that misspecifications regarding the functional form may be responsible for spuriously finding spatial dependencies when hedonic pricing models are used. Consequently, hedonic models should be extended to semi-parametric regression models that allow for a broader class of functional relationships than parametric models (Ekeland et al., 2004).

While empirical studies on farmland rental rates have predominantly concentrated on modeling conditional means, extending the analysis to the modeling of conditional quantiles can provide valuable insights into the price formation of rental rates. It seems reasonable to assume that some covariates have an effect on the mean, while they may have no influence on more extreme quantiles; even if the same covariates are used during the analysis, the manner in which they affect rental rates may change across quantiles. Therefore, the analysis based on quantile regression models can provide a more detailed picture of the conditional distribution of the response variable. For this reason, linear spatial quantile regression models have recently been introduced within spatial econometrics (see McMillen (2013) for a recent overview). However, this model class cannot fully avoid the problems resulting from functional form misspecifications (Kostov, 2013).

The purpose of this paper is to extend the hedonic pricing literature of farmland rental rates by semi-parametrically modeling conditional quantiles of German farmland rental rates using Bayesian geoadditive quantile regression models. The flexibility of this model class frees the researcher from choosing the underlying functional form a-priori and allows for the modeling of a variety of covariates: linear effects of categorical covariates, smooth non-linear effects of continuous covariates as well as spatial effects to account for spatial autocorrelation and unobserved heterogeneity. In contrast to previous studies that were primarily concerned with the analysis of average rental rates, the current study presents an important extension as the modeling of conditional quantiles allows to gain deeper insights into the data generation process of rental rates. By modeling different quantiles of the response distribution, we are able to separately identify the determinants of rental rates for each quantile and can therefore uncover the driving forces behind both, expensive and low rents, as well as for medium rents. We also avoid the problems associated with linear spatial quantile regression models since Bayesian geoadditive quantile regression models account for non-linearities in the relationship between rental rates and their determinants and, hence, allow for more informative conclusions.

The remainder of this paper is organized as follows: Section 2 introduces the reader to the methodology. Section 3 gives an overview of the data and is concerned with variable selection. Section 4 presents the results. Section 5 concludes.

2 Structured Additive Regression Models

In recent years, statistical research on semi-parametric regression models that go beyond traditional linear regression has brought forward a powerful toolkit that allows for a more realistic treatment of a variety of real data problems. Structured Additive Regression Models (STAR), originally proposed by Fahrmeir et al. (2004) and Brezger and Lang (2006), have turned out to be a very powerful model class as they cover the most prominent model extensions as special cases (we refer the interested reader to Fahrmeir and Kneib (2011) or Fahrmeir et al. (2013) for further details on Bayesian STAR Models, e.g., derivation of the full conditionals and the MCMC-sampling algorithm). We first introduce the reader to STAR models for the conditional mean in Section 2.1, since this model class forms the basis for Structured Additive Quantile Regression Models introduced in Section 2.2.

2.1 STAR Models: Getting the mean right

As with the usual linear regression framework, we assume that observations $(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i), i = 1, \ldots, n$ are given, where y_i is a continuous response, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{iq})$ is a vector of categorical covariates and $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{ip})$ is a vector of continuous covariates. In the Generalized Linear Model (GLM) framework of Nelder and Wedderburn (1972), the conditional mean of the response is modeled via

$$\mathbb{E}(y_i|\boldsymbol{x}_i, \boldsymbol{z}_i) = h(\eta_i^{linear}), \quad \text{with} \quad \eta_i^{linear} = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{\gamma} \tag{1}$$

where $h(\cdot)$ is a response function that links the conditional mean of y_i with the linear predictor η_i^{linear}. To allow the response to depend non-linearly on continuous covariates, GLMs can be extended to Generalized Additive Models (GAMs) by replacing the strictly linear predictor in Equation (1) with a more flexible semi-parametric predictor

$$\eta_i = \boldsymbol{x}_i'\boldsymbol{\beta} + f_1(z_{i1}) + \ldots + f_p(z_{ip}) \tag{2}$$

where $f_1, \ldots, f_p$ are non-linear smooth effects of the continuous covariates and $\boldsymbol{x}_i'\boldsymbol{\beta}$ is the usual parametric part. For modeling the unknown functions f_j, we follow Lang and Brezger (2004) and Brezger and Lang (2006) who introduce a Bayesian analogue to P(enalized)-splines originally proposed from a frequentist point of view by Eilers and Marx (1996). In order to illustrate the basic principles of P-splines, we present the frequentist approach first, where it is assumed that the unknown function f_j can be approximated by a polynomial spline of degree l_j. The spline is then represented as a linear combination of $m_j = h_j + l_j - 1$ B-spline basis functions $B_{j,k}$ evaluated at pre-specified knots $z_{j,min} = \zeta_{j,1} < \zeta_{j,2} < \ldots < \zeta_{j,h_j} = z_{j,max}$

$$f_j(z_{ij}) = \sum_{k=1}^{m_j} \gamma_{j,k} B_{j,k}(z_{ij}) \quad , \quad i = 1, \ldots, n. \quad ; \quad j = 1, \ldots, p. \tag{3}$$

where the coefficients $\gamma_{j,k}$ can be interpreted as amplitudes that scale the basis functions $B_{j,k}$ accordingly to fit the data. To ensure a good fit to the data, Eilers and Marx (1996) suggest using a sufficiently high number of equidistant knots (usually between 20 and 40) as well as to simultaneously impose a penalty $\lambda_j \sum_{k=d+1}^{m_j} (\Delta^d \gamma_{j,k})^2$ on adjacent B-spline coefficients $\gamma_{j,k}$ that prevents f_j from being too wiggly. The penalty depends on the smoothing parameter λ_j that balances the trade-off between a good fit to the data and the amount of smoothness of f_j, and Δ^d denotes the d-th order difference operator, i.e., $\Delta^1 = \gamma_{j,k} - \gamma_{j,k-1}$ for d=1. Rewriting the smooth functions in matrix form $\boldsymbol{f}_j = (f_j(z_{1j}),\ldots,f_j(z_{nj}))' = \boldsymbol{Z}_j\boldsymbol{\gamma}_j$ leads to the semi-parametric predictor $\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Z}_1\boldsymbol{\gamma}_1 + \ldots + \boldsymbol{Z}_p\boldsymbol{\gamma}_p$, where $\boldsymbol{Z}_j = B_{j,k}(z_{ij})$ is a $(n \times m_j)$ design matrix and $\boldsymbol{\gamma}_j = (\gamma_{j,1},\ldots,\gamma_{j,m_j})'$ is a vector of regression coefficients. The penalized least squares criterion is then given by

$$\mathrm{PLS}(\boldsymbol{\lambda}) = (\boldsymbol{y}-\boldsymbol{\eta})'(\boldsymbol{y}-\boldsymbol{\eta}) + \sum_{j=1}^{p} \lambda_j \boldsymbol{\gamma}_j' \boldsymbol{K}_d \boldsymbol{\gamma}_j \tag{4}$$

where $\boldsymbol{K}_d$ is a penalty matrix based on d-th order differences. In the Bayesian framework, the vector of regression coefficients $\boldsymbol{\gamma}_j$ and $\boldsymbol{\beta}$ are considered as random variables so that appropriate prior distributions have to be assigned. For the parameters $\boldsymbol{\beta}$ of the parametric part, non-informative priors are assumed, i.e., $p(\beta_r) \propto const, r = 1,\ldots,q$. Priors for the regression parameters $\boldsymbol{\gamma}_j$ of the smooth curves are defined by replacing the difference penalty by first or second order random walks, respectively. From a Bayesian point of view, the quadratic penalty $\lambda_j \boldsymbol{\gamma}_j' \boldsymbol{K}_j \boldsymbol{\gamma}_j$ in Equation (4) can then be replaced with a Gaussian smoothing prior for the regression coefficients $\boldsymbol{\gamma}_j$

$$p(\boldsymbol{\gamma}_j|\tau_j^2) \propto \frac{1}{(\tau_j^2)^{rank(\boldsymbol{K}_j)/2}} \exp\left(-\frac{1}{2\tau_j^2}\boldsymbol{\gamma}_j'\boldsymbol{K}_j\boldsymbol{\gamma}_j\right) \quad , \quad j = 1,\ldots,p. \tag{5}$$

The amount of smoothness of f_j is controlled by the variance parameter τ_j^2, which corresponds to the inverse of the smoothing parameter λ_j in the frequentist setting.

Besides categorical and continuous covariates, spatially referenced data also contain information about the location where the observations have been collected. To include this information into the model, an additional spatial term f_{geo} is added to the predictor from Equation (2)

$$\eta_i = \boldsymbol{x}_i'\boldsymbol{\beta} + f_1(z_{i1}) + \ldots + f_p(z_{ip}) + f_{geo}(s_i) \tag{6}$$

yielding a geoadditive model as proposed by Kammann and Wand (2003). The spatial effect f_{geo} acts as a surrogate for unobserved covariates that are not included in the model and also accounts for spatial autocorrelation (Fahrmeir and Kneib, 2011). In the case that the spatial effect originates from both spatially correlated and uncorrelated unobserved covariates, it is advisable to split up the spatial effect $f_{geo} = f_{str} + f_{unstr}$ into a structured, correlated effect f_{str} and an unstructured, district specific effect f_{unstr}. This partition allows the researcher to assess the complete spatial information in the data. The estimation of the correlated spatial effect f_{str} can be represented in a Bayesian framework by assigning Gaussian Markov Random Field (GMRF) priors for the regression coefficients $\boldsymbol{\gamma}_{str}$, reflecting the restriction that neighbouring districts should have similar effects

$$\gamma_{str}(s)|\boldsymbol{\gamma}_{str}(-s) \sim N\left(\frac{1}{|N(s)|}\sum_{r\in N(s)} \gamma_{str}(r), \frac{\tau_{str}^2}{|N(s)|}\right) \quad , \quad s = 1,\ldots,S. \tag{7}$$

where $\boldsymbol{\gamma}_{str}(-s)$ is the vector containing all spatial effects except the one for district s and $|N(s)|$ denotes the total number of neighbours that share a common boundary with district s. GMRF assume that, given the effects of all other districts, the expected value of $\gamma_{str}(s)$ equals the average of the neighbouring districts. The amount of how much an effect for district s is allowed to vary around its expected value depends on the variance τ_{str}^2 and inversely on the number of neighbours $|N(s)|$. The joint distribution of all spatial effects can be derived from the conditional distributions and can be represented as

$$p(\boldsymbol{\gamma}_{str}|\tau_{str}^2) \propto \frac{1}{(\tau_{str}^2)^{rank(\boldsymbol{K}_{str})/2}} \exp\left(-\frac{1}{2\tau_{str}^2}\boldsymbol{\gamma}_{str}'\boldsymbol{K}_{str}\boldsymbol{\gamma}_{str}\right) \tag{8}$$

The matrix $\boldsymbol{K}_{str}$ contains the neighbourhood information in its non-zero entries, i.e., $\boldsymbol{K}_{str}[s,r] = -1$ if districts s and r are neighbours and $\boldsymbol{K}_{str}[s,r] = 0$ otherwise. If spatial heterogeneity exists only locally, it is not reasonable to assume that coefficients of neighboring districts are spatially correlated and an uncorrelated spatial effect should be used instead. To model f_{unstr}, district specific i.i.d. Gaussian random effects $\gamma_{unstr}(s)|\tau_{unstr}^2 \sim N(0,\tau_{unstr}^2), s = 1,\ldots,S$ are commonly used. In the Bayesian framework, the joint multivariate prior distribution of the unstructured effect can be represented as in Equation (8), with $\boldsymbol{K}_{unstr} = \boldsymbol{I}$.

2.2 Structured Additive Quantile Regression: Going beyond the mean

The regression models discussed so far focus on modeling conditional means. Since these models only allow limited insights into the way covariates influence the response, Waldmann et al. (2013) extend STAR models from the previous section to semi-parametric additive quantile regression models that allow for the modeling of conditional quantiles.

While in the frequentist setting of Koenker and Bassett (1978) no distributional assumptions regarding the error term have to be made, the Bayesian formulation of quantile regression relies on assuming an asymmetric Laplace distribution (ALD) as an auxiliary error distribution, as suggested by Koenker and Machado (1999) and Yu and Moyeed (2001). This assumption allows for the specification of a likelihood function that is needed for Markov chain Monte Carlo (MCMC) inference. The asymmetric Laplace distribution with location parameter η_{i_τ}, scale parameter σ^2 and asymmetry parameter τ is particularly suitable for geoadditive quantile regression models, since the minimization of the check function in the frequentist setting can equivalently be represented as maximizing the asymmetric Laplace likelihood function

$$\prod_{i=1}^{n} p(y_i|\eta_{i_\tau}, \sigma^2, \tau) \propto \exp\left(-\sum_{i=1}^{n} \rho_\tau \frac{(y_i - \eta_{i_\tau})}{\sigma^2}\right) \tag{9}$$

with respect to η_{i_τ}, where τ denotes the quantile of interest. In the Bayesian framework of Waldmann et al. (2013), the strictly linear predictor in Equation (9) is replaced with the more flexible geoadditive quantile predictor

$$\eta_{i_\tau} = \beta_{0_\tau} + \beta_{1_\tau} x_{i1} + \ldots + \beta_{q_\tau} x_{iq} + f_{1_\tau}(z_{i1}) + \ldots + f_{p_\tau}(z_{ip}) + f_{geo_\tau}(s_i) \tag{10}$$

that allows the researcher to analyze the influence of the covariates on the response variable in a non-linear way, for each quantile separately (we refer to Fahrmeir et al. (2013) and Waldmann et al. (2013) for the derivation of the full conditionals and the MCMC-sampling algorithm).

3 Data description and variable selection

3.1 Data description

For our analysis, we use farm-level data based on the 2010 German agricultural census (FDZ, 2010). It is the most comprehensive survey since 1999 and gives a representative picture of the agricultural situation in Germany. The focus of the census is on questions regarding land use and livestock, property and leasing agreements, organic-farming, labor and employment. We use farmland rental rates per hectare as the response variable, since this number can be interpreted more easily and farmers use this figure for guidance when determining appropriate rental agreements. We exclude tenancies that were entered between family members to obtain a market based assessment of rental rates. Based on previous studies that analyse German farmland rental rates (see Breustedt and Habermann (2010) or Habermann and Ernst (2010) among others), we use the covariates presented in Table A1 for the analysis. To adjust for unobserved spatial heterogeneity that is not accounted for by farm-level covariates, we additionally include socio-demographic covariates on the district level from the Regionaldatenbank (2010). After removing non-renting farmers, as well as outlying observations, we are left with 107,620 observations for the analysis.

3.2 Variable selection

Variable selection is a challenging task in geoadditive quantile regression. The researcher has to select a subset of covariates that he or she considers relevant for the analysis and has to decide whether the spatial information in the data is best described by an unstructured or structured effect. To make the task of variable selection feasible, we use a systematic and fully data-driven approach based on componentwise functional gradient descent boosting for Structured Additive Quantile Regression, as proposed by Fenske et al. (2011). Boosting is a machine learning approach that is aimed towards maximizing the prediction accuracy of the response by iteratively combining different model components, called base learners, where in each iteration step only the best-fitting base learner, i.e., the most informative covariate, is selected. For the starting model, we include all covariates presented in Table A1. Table 1 presents the final covariates and their selection frequencies for different quantiles. To cover the entire range of rental rates and, in particular, to gain detailed insights into very low and very high as well as into medium rental rates, we choose to model conditional quantiles of $\tau = \{0.05, 0.50, 0.95\}$. Variable selection is performed using the R-package *mboost* of Hothorn et al. (2013).

Table 1. Covariates and their selection frequencies during boosting iterations.

$\tau = 0.05$		$\tau = 0.50$		$\tau = 0.95$	
Covariate	Freq.	Covariate	Freq.	Covariate	Freq.
f_{str}	0.3998	f_{str}	0.7162	f_{str}	0.3208
$f(cattle\_district)$	0.1453	$f(cattle\_district)$	0.0735	$f(hog\_poultry\_district)$	0.1239
$f(size)$	0.0711	$f(cattle)$	0.0422	$f(sugarbeet)$	0.1103
$fulltime$	0.0675	$f(rent\_share)$	0.0222	$f(cattle\_district)$	0.1056
$f(farmland\_share)$	0.0564	f_{unstr}	0.0217	$fulltime$	0.0762
$f(rent\_share)$	0.0547	$f(size)$	0.0216	$f(farmland\_share)$	0.0752
$f(hog\_poultry\_district)$	0.0398	$f(hog\_poultry\_district)$	0.0199	$intercept$	0.0414
$f(sugarbeet)$	0.0395	$f(sugarbeet)$	0.0149	$f(potato)$	0.0334
$f(winterwheat)$	0.0347	$f(farmland\_share)$	0.0107	$f(inha)$	0.0313
$f(unempl)$	0.0239	$fulltime$	0.0083	$f(rye)$	0.0281
$f(cattle)$	0.0234	$f(winterwheat)$	0.0078	$f(hhi)$	0.0213
$f(hog\_poultry)$	0.0179	$f(unempl)$	0.0076	$f(hog\_poultry)$	0.0196
$intercept$	0.0152	$f(hog\_poultry)$	0.0068	$f(rent\_share)$	0.0130
$f(rye)$	0.0108	$f(potato)$	0.0053		
		$f(rye)$	0.0053		
		$f(biogas)$	0.0048		
		$f(income)$	0.0040		
		$f(rent\_lag)$	0.0032		
		$f(labour)$	0.0022		
		$f(inha)$	0.0019		
Σ	1.0000	Σ	1.0000	Σ	1.0000

Source: own calculations based on data from the 2010 German agricultural census and from the Regionaldatenbank (2010). $f(\cdot)$ indicates that the variable is modeled semi-parametrically, whereas the variable name itself indicates that the variable is modeled parametrically.

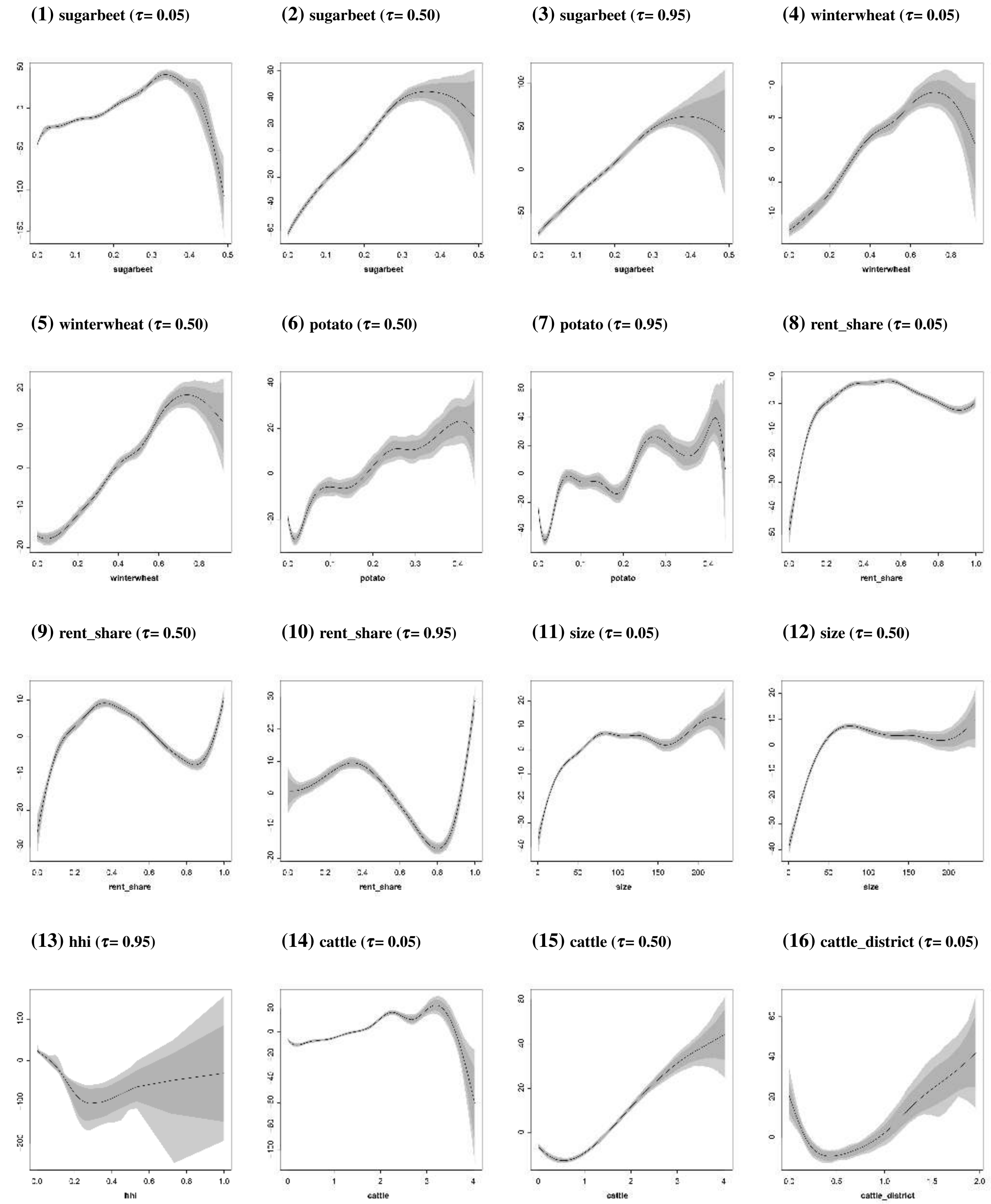
Table 1 shows that the way in which covariates affect farmland rental rates varies across quantiles: while some covariates have an influence on low and medium rents, they have no effect on more expensive ones. From Table 1 it is also apparent that only f_{str} has been selected for the 5% and 95% quantile in order to model the spatial information in the data. The spatial effect can therefore be assumed not to exist only locally, but to be correlated across districts for these quantiles. However, the unstructured spatial effect has also been selected for the 50% quantile indicating that there seems to be additional small scale, district specific spatial information in the data for medium rents. Also note that, for expensive rents, mainly those covariates that reflect local competition for farmland or field crops with high profit margins have been selected. Consulting the literature about the influence of biogas on rental rates shows that it is a subject of great controversy. Kilian et al. (2008) find that, in general, higher concentrations of biogas plants lead to an increase in rental rates, while the results of Habermann and Ernst (2010) do not support this effect. Habermann and Ernst (2010) attribute these varying findings to the different granularity of the data that has been used for the analysis as Kilian et al. (2008) use data on the community level, while Habermann and Ernst (2010) use district averages. More importantly, Habermann and Ernst (2010) argue that, due to the long duration of the leasing contracts, it may take some time before the effect of biogas is reflected in rental rates. This inelasticity of rental rates might also apply to our results, since boosting has decided to include biogas as an important covariate for medium rents only.

4 Analysis of farmland rental rates

After having identified the relevant economic variables that determine farmland rental rates, we now present the estimation results of the Bayesian geoadditive quantile regression. The estimation is performed using the R-packages *BayesXsrc* of Adler et al. (2013) and *R2BayesX* of Lang et al. (2013), which is an R interface to the standalone software *BayesX* of Belitz et al. (2013).

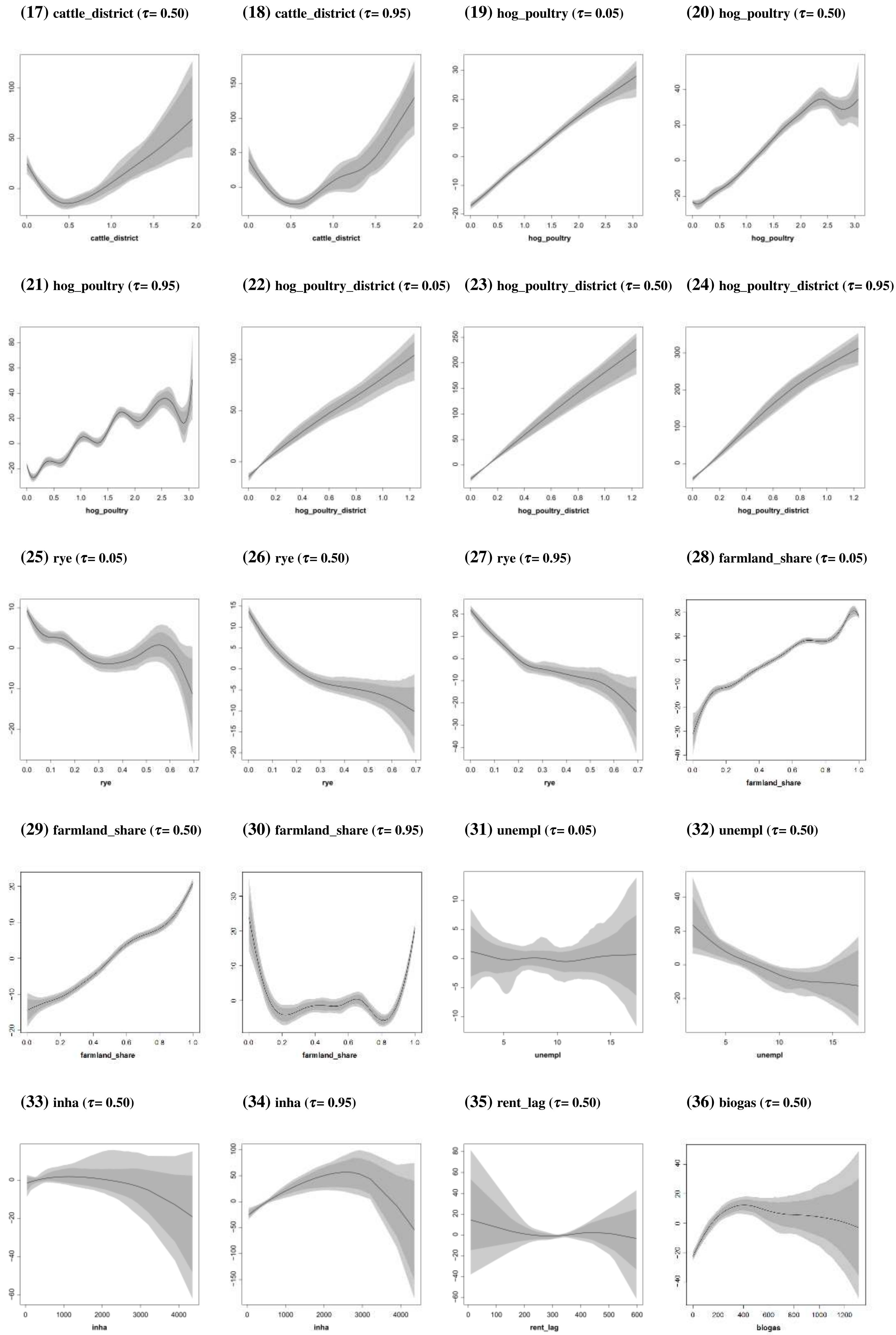
4.1 Parametric and semi-parametric effects

Figure 1 shows posterior mean estimates for the semi-parametric effects together with pointwise 80% (dark grey) and 95% (light grey) credible intervals.



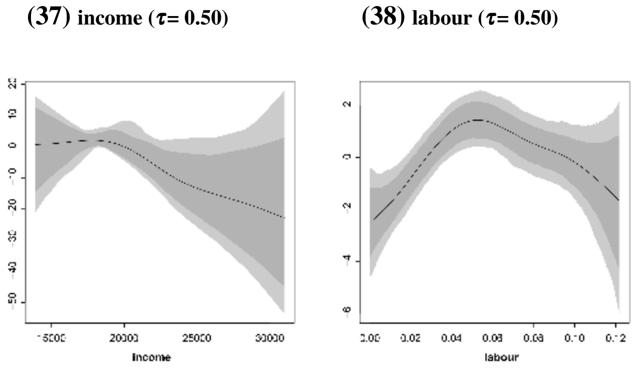
Source: own calculations based on data from the 2010 German agricultural census and from the Regionaldatenbank (2010).

Figure 1. Estimated semi-parametric effects with pointwise 80% and 95% credible intervals.



Source: own calculations based on data from the 2010 German agricultural census and from the Regionaldatenbank (2010).

Figure 1 continued.



Source: own calculations based on data from the 2010 German agricultural census and from the Regionaldatenbank (2010).

Figure 1 continued.

In line with Breustedt and Habermann (2009) and Habermann and Ernst (2010) Figure 1 shows that field crops with high profit margins, like sugar beets or potatoes, have a positive effect on rental rates. However, due to the non-linearities of the effects, there are some differences to previous studies that have to be noted. While the effect of sugar beet initially increases across all quantiles, it starts to decrease again for low rents if more than 30% of the farmland is used for the cultivation. We further investigate the decrease of sugar beets for the 5% quantile by forming the first derivative of the estimated effect and find that the sharp decrease is indeed significant at a nominal level of 5%. In contrast, there seems to be a threshold effect for medium and high rents as the effects level off above a share of 30% for the 50% and 95% quantile. The effect of farm size for small and medium-sized farms is a positive one, as rental rates increase initially with growing farm size. Larger farms are more likely to realize economies of scale and are therefore able to pay higher rents. However, note that the size only has an increasing effect until a farm size of approximately 70 hectares. After this threshold, low rents remain almost constant with increasing farm size, while medium rents are negatively affected for farm sizes between 70 and 180 hectares. A possible explanation might be that, after a given threshold, farms may be so large that they may have a comparatively higher market power which allows them to keep rental rates low. The effect seems to increase again for farm sizes of 180 hectares and above. However, due to the increased uncertainty attached to the estimated effects, reliable statements above a farm size of 180 hectares cannot be made.

While Fuchs (2002), Margarian (2008) and Habermann and Ernst (2010) find that average rents decrease with an increasing share of rented agricultural land, panels (8)-(10) of Figure 1 allow for a more detailed analysis. The decreasing effect, that is reported in previous studies, can only be confirmed for a share of rented agricultural land between 40% and 80%. For shares other than this range, rental rates clearly increase with rented agricultural land. The incresing effects for medium and high rates are also supported by the narrow credible intervals. While the increasing effect is more pronounced for medium and high rates, there is virtually no effect for low rental rates above a threshold of 80%. The general finding in the literature of an overall negative relationship between rented agricultural land and farmland rental rates might be attributed to the fact that, when averaging across quantiles, the decreasing effect for shares ranging between 40% and 80% might dominate, so that an overall negative relationship results. This example illustrates the usefulness of semi-parametric regression, as important features in the data go undetected if linear regression models are used. Similar to Drescher and McNamara (2000), Fuchs (2002) or Breustedt and Habermann (2011), we find that livestock densities have a major impact on rental rates. However, in addition to previous studies, panels (16)-(18) of Figure 1 show a pronounced U-shaped effect for the cattle density at the district level across all quantiles: a cattle density of up to 0.5 first decreases farmland rental rates, before rental rates start to increase beyond this livestock density. The increasing effect of livestock density on rental rates is also very pronounced for hog and poultry densities at the farm level, and even more so for medium and high rents on the district level. The positive influence of livestock densities on rental rates might be explained by a statutory framework within which farmers are restricted in the amount of manure they are allowed to discharge on their land. Farmers with a livestock density that exceeds a certain threshold either have to rent additional acreage or have to register a trade. To avoid tax disadvantages, farmers prefer to rent additional farmland instead in order to reduce the livestock density (Habermann and Ernst, 2010). The strong effects for regional livestock density may also reflect the heavy competition for farmland in certain districts. From Figure 1, it also appears that biogas increases medium rental rates, at least up to a plant capacity of approximately 380 kWh. As a result of the increased uncertainty attached to the estimation, which is reflected in the wide credible bands, reliable statements beyond this capacity cannot be made.

We now turn to the analysis of parametric effects that are summarized in Table 2, showing posterior means, standard deviations and 95% credible intervals. Table 2 shows that differences exist between farmers and the rental rates they have to pay depending on whether they operate their farm full-time or part-time, since full-time farmers have to pay higher rents compared to their part-time counterparts. This difference may be attributed to several reasons. In order to earn a living, full-time farmers have to have a high production volume and a high production intensity. As a consequence, full-time farmers are on average larger than part-time farmers, with an average farm size of about 61 hectares. This is about the size until which rental rates increase with farm size (compare Panels (11)-(12) of Figure 1). Another reason for the difference might be due to the fact that the proportion of full-time farmers is high in those districts where the principle income of the farmer is associated with livestock farming, and hence, in districts where rental rates are high (compare Panels (14)-(24) of Figure 1). Due to the high demands with respect to capital intensity and the employment of labour, livestock farming on a larger scale can be operated successfully only as a full-time farmer.

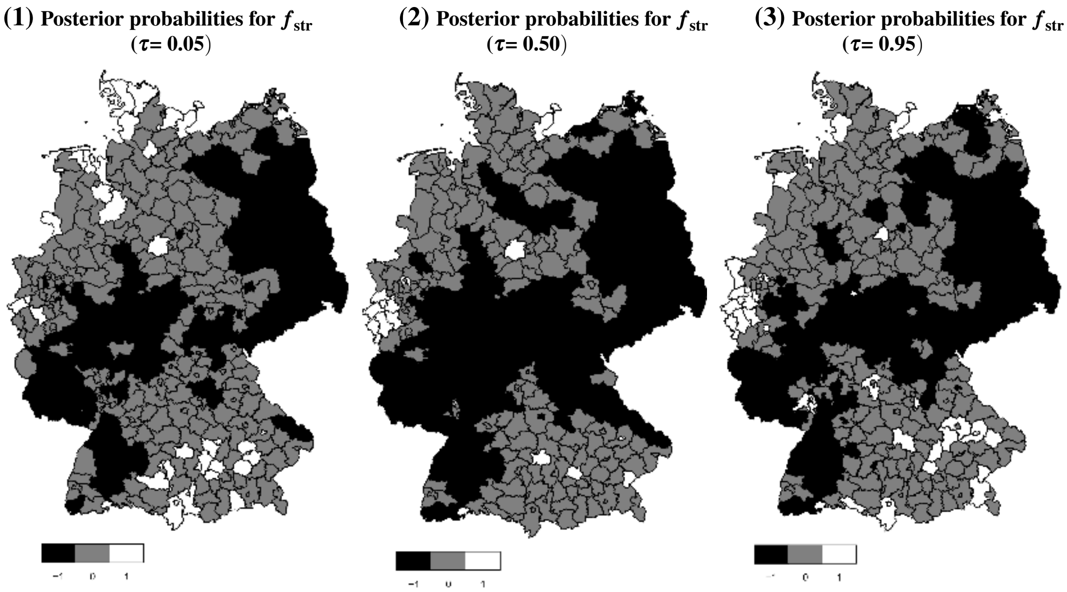
Table 2. Estimated parametric effects.

	$\tau = 0.05$			
	Mean	Std. Dev.	2.5%	97.5%
Intercept	167.3150	1.5635	164.3530	170.2770
full-time	-4.4783	0.7025	-5.8641	-3.0617
part-time	-9.0094	0.8166	-10.5819	-7.4320
	$\tau = 0.50$			
	Mean	Std. Dev.	2.5%	97.5%
Intercept	324.0170	3.0758	317.5430	329.7030
full-time	-5.3913	0.9449	-7.2346	-3.5716
part-time	-12.1409	1.0729	-14.2923	-10.1269
	$\tau = 0.95$			
	Mean	Std. Dev.	2.5%	97.5%
Intercept	445.6230	4.8821	436.3380	455.0340
full-time	-11.3771	1.4211	-14.1269	-8.6229
part-time	-29.8125	1.4229	-32.5452	-27.0207

Source: own calculations based on data from the 2010 German agricultural census and from the Regionaldatenbank (2010).

4.2 Spatial effects

Our analysis of the spatial effects is motivated from a statistical point of view. In contrast to econometrics, where spatial autoregressive models are commonly used, we account for spatial correlation and non-observable farmland characteristics by adding a spatial term f_{geo_τ} to the additive predictor η_{i_τ}. As a consequence, we are mainly interested in investigating spatial patterns that emerge from spatial heterogeneities that are left unexplained after taking covariates into account. Plotting the estimated effects of f_{geo_τ} allows us to graphically investigate these spatial patterns and assists in identifying additional covariates that capture the remaining heterogeneity in the data. A careful visual inspection of the distribution of these spatial effects can also provide new insights into the data that were not previously considered. Significance maps shown in Figure 2 further enhance the detection of spatial patterns by classifying the estimated spatial effect into three categories; the spatial effect is classified as insignificant at the 80% level and the corresponding district is coloured in grey, if the credible interval includes zero. Districts with significant positive effects are coloured in white, whereas districts with significantly negative effects are coloured in black.



Source: own calculations based on data from the 2010 German agricultural census and from the Regionaldatenbank (2010).

Figure 2. Posterior probabilities (80%) of f_{str} for different quantiles.

Figure 2 shows that rental rates are considerably lower what can be explained with covariates in the southwest, as well as in large parts of East Germany across all quantiles (black districts). It is reasonable to assume that the pattern in East Germany results from structural differences between East and West German rental markets, and in particular from the way rental rates were set by the Bodenverwertungs- und -verwaltungs GmbH (BVVG), a company that managed state-owned land in East-Germany. In order to account for the differences between East and West German rental markets, we have included a dummy variable. However, as Table 1 shows, the dummy variable leaves the structural differences between East and West German rental markets unexplained since boosting has never selected it during any of the iterations. Consequently, additional covariates other than the dummy variable have to be included in the model in order to account for the differences between East and West German rental markets. The patterns in the southwest of Germany may be attributed to the wine-growing districts. Since rental rates are grouped by the type of use of the agricultural land, rental rates of vineyards are recorded separately in the data. As a consequence, although winegrowers had to pay an average rent of approximately 828 EUR per hectare for vineyards in certain wine-growing districts, these high rents do not contribute to the estimation and the map only shows below average farmland rental rates.

Figure 2 also reveals that the covariates are better suited to explain expensive rental rates, as the covariates leave heterogeneities unexplained only in very few districts of Germany across all quantiles (white districts). While the pattern for the median and 95% quantile are similar with respect to high unexplained rents, the 5% quantile identifies some additional districts in the far north. In accordance with the literature and with the results from the semi-parametric effects, Figure 2 shows that rental rates for farmland are more expensive in districts where livestock densities are high. Rental rates are also more expensive in districts in which high livestock densities and high biogas densities meet, such as in the southern part of Germany.

5 Conclusion

In this paper we model and analyse conditional quantiles of farmland rental rates semi-parametrically using Bayesian geoadditive quantile regression models. The flexibility of this model class overcomes the problems of functional form misspecifications and allows us to present a more detailed analysis of farmland rental rates and their determinants. In particular, by allowing different quantiles of the distribution to depend differently on covariates, our study provides additional insights into the data generation process of rental rates as the possible determinants can be separately identified for each quantile. By explicitly modeling and plotting the spatial effects, we account for spatial autocorrelation and are able to detect spatial patterns in the data that can be used in future studies in order to identify additional covariates that capture the remaining spatial heterogeneity. Our results also stress the importance of making use of flexible, semi-parametric models as some of the covariates clearly have a non-linear influence on farmland rental rates. The results of our study are of potential interest for both, practitioners and academics, since hedonic pricing studies in the agricultural economics literature have primarily been concerned with the analysis of average rental rates. For instance, our results can serve as a basis for negotiating new tenancies or designing rent adjustment clauses, as the terms of contract can now be better tailored to the operational characteristics of the farmer. In addition, if desired, the identification of the driving forces behind expensive rents may also serve to assist policy makers in taking corrective actions by setting a ceiling on rental rates in order to prevent an excessive rise of rental rates in the future.

There are several ways to extend the current analysis. From an agricultural point of view, it would be interesting to investigate whether the increased investment in agricultural land by both agricultural and non-agricultural investors, or the regionally high demand for zoning, traffic and compensation areas also have an effect on farmland rental rates. From a statistical point of view, the data could be re-analyzed using Generalized Additive Models for Location, Scale and Shape (GAMLSS), originally proposed by Rigby and Stasinopoulos (2005) and extended to Bayesian Structured Additive Distributional Regression by Klein et al. (2013). This model class allows the researcher to model all parameters of an assumed response distribution as additive functions of covariates. This is important in the case of (spatial) heteroscedasticity, where interest does not only lie with farmland rental rates themselves, but also with their (spatial) variation.

References

Adler, D., Kneib, T., Lang, S., Umlauf, N. and Zeileis, A. (2013). BayesXsrc: R package distribution of the BayesX C++ Sources, R package version 2.1-2, URL: http://CRAN.R-project.org/package=BayesXsrc.

Belitz, C., Brezger, A., Kneib, T., Lang, S. and Umlauf, N. (2013). BayesX: Software for Bayesian inference in structured additive regression models, URL: http://www.BayesX.org/.

Bierlen, R., Parsch, L. D. and Dixon, B. L. (1999). How cropland contract type and term decisions are made: evidence from an Arkansas tenant survey. *The International Food and Agribusiness Management Review* 2: 103–121.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)* 26: 211–252.

Breustedt, G. and Habermann, H. (2008). Determinants of Agricultural Cash Rents: Emprical Insights from Farms in Lower Saxony. *Contributed Paper prepared for presentation at the Leibniz Institute of Agricultural Development in Central and Eastern Europe (IAMO) Forum, Halle (Saale) Germany, 25-27 June* : 1–11.

Breustedt, G. and Habermann, H. (2009). Determinants of agricultural cash rents in Germany: A spatial econometric analysis for farm-level data. *Contributed Paper prepared for presentation at the International Association of Agricultural Economists Conference, Beijing, China, August 16-22, 2009* : 1–20.

Breustedt, G. and Habermann, H. (2010). Einfluss der Biogaserzeugung auf landwirtschaftliche Pachtpreise in Deutschland. *Contributed Paper prepared for presentation at the German Association of Agricultural Economists (GEWISOLA), 50th Annual Conference, Braunschweig, Germany, September 29-October 1, 2010* : 1–12.

Breustedt, G. and Habermann, H. (2011). The Incidence of EU Per–Hectare Payments on Farmland Rental Rates: A Spatial Econometric Analysis of German Farm–Level Data. *Journal of Agricultural Economics* 62: 225–243.

Brezger, A. and Lang, S. (2006). Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics & Data Analysis* 50: 967–991.

Brümmer, B. and Loy, J. P. (2001). Der Einfluss staatlicher Ausgleichszahlungen auf Landpreise in Schleswig-Holstein. *Contributed Paper prepared for presentation at the German Association of Agricultural Economists (GEWISOLA), 41th Annual Conference, Braunschweig, Germany, October 8-October10, 2001* : 1–10.

Ciaian, P., Kancs, d., Swinnen, J., Herck, K. van and Vranken, L. (2012). Key Issues and Developments in Farmland Rental Markets in EU Member States and Candidate Countries. *Centre for European Policy Studies, Factor Markets Working Paper* 13: 1–25.

Ciaian, P., Kancs, D. and Swinnen, J. F. M. (2010). EU Land Markets and the Common Agricultural Policy. *Centre for European Policy Studies (CEPS) Paperbacks* : 1–282.

Court, A. (1939). Hedonic Price indexes with automotive examples. *The Dynamics of Automobile Demand, New York, General Motors* : 99–117.

Doll, H. and Klare, K. (1995). Empirische Analyse der regionalen landwirtschaftlichen Bodenmärkte in den neuen Bundesländern. *Landbauforschung Völkenrode* 4: 205–217.

Drescher, K. and McNamara, K. (2000). Analysis of German agricultural land prices. *In: Land ownership, land markets and their influence on the efficiency of agricultural production in Central and Eastern Europe, Wissenschaftsverlag Vauk Kiel* : 210–228.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical science* 11: 89–121.

Ekeland, I., Heckman, J. J. and Nesheim, L. P. (2004). Identification and estimation of hedonic models. *Journal of Political Economy* 112: 60–109.

Fahrmeir, L. and Kneib, T. (2011). *Bayesian smoothing and regression for longitudinal, spatial and event history data*, Oxford statistical science series *36*. Oxford and New York: Oxford University Press.

Fahrmeir, L., Kneib, T. and Lang, S. (2004). Penalized structured additive regression for space-time data: a Bayesian perspective. *Statistica Sinica* 14: 731–762.

Fahrmeir, L., Kneib, T., Lang, S. and Marx, B. (2013). *Regression: Models, methods and applications*. Berlin: Springer, 1st ed.

FDZ (2010). Landwirtschaftszählung, 2010. *Forschungsdatenzentren (FDZ) der Statistischen Ämter des Bundes und der Länder* .

Fenske, N., Kneib, T. and Hothorn, T. (2011). Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. *Journal of the American Statistical Association* 106: 494–510.

Fuchs, C. (2002). The influence of per-hectare premiums on prices for rented agricultural area and on agricultural land prices. *German Journal of Agricultural Economics* 51: 396–403.

Griliches, Z. (1961). Hedonic price indexes for automobiles: An econometric of quality change. In *The Price Statistics of the Federal Goverment*. NBER, 173–196.

Habermann, H. and Ernst, C. (2010). Entwicklungen und Bestimmungsgründe der Landpachtpreise in Deutschland. *Berichte über Landwirtschaft* 88: 57–85.

Herriges, J. A., Shogren, J. F. and Barickman, N. E. (1992). The implicit value of corn base acreage. *American Journal of Agricultural Economics* 74: 50–58.

Hothorn, T., Buehlmann, P., Kneib, T., Schmid, M. and Hofner, B. (2013). mboost: Model-based boosting, R package version 2.2-2, URL http://CRAN.R-project.org/package=mboost.

Kammann, E. E. and Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 52: 1–18.

Kilian, S., Anton, J., Röder, N. and Salhofer, K. (2008). Impacts of 2003 CAP reform on land prices: from theory to empirical results. *In 109th EAAE Seminar" The CAP After the Fischler Reform: National Implementations, Impact Assessment and the Agenda for Future Reforms". Viterbo, Italy, November* : 1–15.

Kirwan, B. E. (2009). The incidence of US agricultural subsidies on farmland rental rates. *Journal of Political Economy* 117: 138–164.

Klein, N., Kneib, T. and Lang, S. (2013). Bayesian Structured Additive Distributional Regression. *University of Innsbruck - Working Papers in Economics and Statistics* 23: 1–62.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society* : 33–50.

Koenker, R. and Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association* 94: 1296–1310.

Kostov, P. (2009). Spatial dependence in agricultural land prices: does it exist? *Agricultural Economics* 40: 347–353.

Kostov, P. (2013). Choosing the Right Spatial Weighting Matrix in a Quantile Regression Model. *ISRN Economics* 2013: 1–15.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy* 74: 132–157.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics* 13: 183–212.

Lang, S., Kneib, T., Umlauf, N. and Zeileis, A. (2013). R2BayesX: Estimate structured additive regression models with BayesX, R package version 0.3-1, URL: http://CRAN.R-project.org/package=R2BayesX.

Lence, S. H. and Mishra, A. K. (2003). The impacts of different farm programs on cash rents. *American Journal of Agricultural Economics* 85: 753–761.

Margarian, A. (2008). Sind die Pachten im Osten zu niedrig oder im Westen zu hoch? *Arbeitsberichte aus der vTI-Agrarökonomie - Johann Heinrich von Thünen Institute, Federal Research Institute for Rural Areas, Forestry and Fisheries* : 1–42.

Martins-Filho, C. and Bin, O. (2005). Estimation of hedonic price functions via additive nonparametric regression. *Empirical economics* 30: 93–114.

McMillen, D. P. (2013). *Quantile regression for spatial data*. Springer Briefs in regional science. Berlin and New York: Springer.

Nelder, J. and Wedderburn, R. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)* 135: 370–384.

Patton, M. and McErlean, S. (2003). Spatial effects within the agricultural land market in Northern Ireland. *Journal of Agricultural Economics* 54: 35–54.

Regionaldatenbank (2010). Regionaldatenbank Deutschland. *Statistische Ämter des Bundes und der Länder* .

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54: 507–554.

Roberts, M. J., Kirwan, B. and Hopkins, J. (2003). The incidence of government program payments on agricultural land rents: The challenges of identification. *American Journal of Agricultural Economics* 85: 762–769.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy* 82: 34–55.

Waldmann, E., Kneib, T., Yue, Y. R., Lang, S. and Flexeder, C. (2013). Bayesian semiparametric additive quantile regression. *Statistical Modelling* 13: 223–252.

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters* 54: 437–447.

A Appendix

Table A1. Description of Covariates.

Covariate (farm-level)	**Description**
farm_succ	Farm succession (categorical: 1=yes, 2=no, 3=unsettled)
east	Dummy Variable for East Germany (categorical: 1=West Germany, 2=East Germany)
organic	Dummy Variable for organic farming (categorical: 1=yes, 2=no)
fulltime	Dummy Variable indicating whether the farmer operates his business in full-time or part-time (categorical: 1=no individual enterprise as legal form, 2=full-time, 3=part-time)
rent_share	Share of rented agricultural land to total agricultural land (continuous)
farmland_share	Share of rented farmland to total rented agricultural land (continuous)
cattle	Farm-level cattle density in animal unit (AU) per hectare (continuous)
hog_poultry	Farm-level hog and poultry density in animal unit (AU) per hectare (continuous)
biogas	Capacity of biogas plant in kWh (continuous)
winterwheat	Share of winter wheat in cropping pattern (continuous)
sugarbeet	Share of sugar beet in cropping pattern (continuous)
potato	Share of potato in cropping pattern (continuous)
rye	Share of rye in cropping pattern (continuous)
labour	Labour force per hectare (continuous)
irrigation	Share of agricultural land that could have been irrigated (continuous)
size	Total agricultural land of the farmer in hectare (continuous)
f_{str}	Structured spatial effect
f_{unstr}	Unstructured spatial effect
Covariate (district-level)	**Description**
hhi	Herfindahl-Hirschman index based on the share of rented agricultural land to total agricultural land in each district (continuous)
inha	Inhabitants per square kilometre (continuous)
unempl	Unemployment rate (continuous)
income	Average income per inhabitant (continuous)
rent_lag	Spatially lagged farmland rental rate (continuous)
dist2cc	Distance to next city center in kilometres (continuous)
cattle_district	Average district-level cattle density in animal unit (AU) per hectare (continuous)
hog_poultry_district	Average district-level hog and poultry density in animal unit (AU) per hectare (continuous)

Source: own calculations based on data from the 2010 German agricultural census and from the Regionaldatenbank (2010).