European Review of Agricultural Economics Vol **43** **(4)** (2016) pp. 663–698
doi:10.1093/erae/jbv028
Advance Access Publication 3 November 2015

Analysing farmland rental rates using Bayesian geoadditive quantile regression

Alexander März*, Nadja Klein, Thomas Kneib and Oliver Musshoff

Georg-August-University Göttingen, Germany

Received August 2014; final version accepted August 2015

Review coordinated by Ada Wossink

Abstract

Empirical studies on farmland rental rates so far have predominantly concentrated on modelling conditional means using spatial autoregressive models. While these models only focus on the central tendency of the response variable, quantile regression provides more detailed insight by modelling different points of the conditional distribution as a function of covariates. Based on data from the German agricultural census, this article contributes to the agricultural economics literature by modelling conditional quantiles of farmland rental rates semi-parametrically using Bayesian geoadditive quantile regression. Our results stress the importance of using semi-parametric regression models, as several covariates influence rental rates in an explicit non-linear way. Moreover, our analysis allows us to uncover potential heterogeneities of the estimated effects across the conditional distribution of rental rates. By explicitly modelling and visually presenting the spatial effects, we also provide additional insight into the spatial structure of German farmland rental rates.

Keywords: Bayesian geoadditive quantile regression, component-wise boosting, farmland rental rates, hedonic pricing models, heterogeneity, spatial statistics

JEL classification: C14, C11, Q15

1. Introduction

Farmland is one of the most important production factors in agriculture (Borchers, Ifft and Kuethe, 2014). Based on cash-flow considerations, farmers have to decide whether to buy or lease agricultural land. One advantage of leasing is that farmers can use their cash reserves to invest in new agricultural machinery and equipment, rather than tying up capital in land purchases (Ciaian and Kancs, 2012). This preference might be one of the reasons why Germany is among European countries with a high share of rented farmland; in 2008, on average

*Corresponding author: Department of Agricultural Economics and Rural Development, Georg-August-University Göttingen, Germany, E-mail: amaerz@gwdg.de

70 per cent of the total German agricultural farmland was leased, with the share of rented farmland being considerably higher in East Germany (80 per cent) compared with West Germany (60 per cent) (Ciaian, Kancs and Swinnen, 2010b; Ciaian and Kancs, 2012).[1] Moreover, Ciaian, Kancs and Swinnen (2010b) report that German farmland rental rates exhibit substantial spatial variation, since rental rates are explicitly higher in West Germany than in East Germany. These figures, along with the fact that rental rates have increased considerably over the last few years, imply that the analysis of farmland rental rates is of great importance.

Aside from its relevance for farmers, the analysis of farmland rental rates is an active field of research in agricultural economics. Herriges, Shogren and Barickman (1992), Bierlen, Parsch and Dixon (1999), Lence and Mishra (2003), as well as Roberts, Kirwan and Hopkins (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. Drescher and McNamara (2000), Margarian (2008) and Breustedt and Habermann (2011) investigate the determinants of rental rates in Germany. Kilian *et al.* (2008) as well as Habermann and Ernst (2010) and Habermann and Breustedt (2011) analyse the effects of an increased land use for the production of bioenergy on German rental rates. What emerges from these studies, in particular from those that focus on the German rental market, is that rental rates are influenced by a number of farm-level and regional characteristics. These include – but are not limited to – the return generated by the rented land generally approximated by the net value-added or the per hectare operating income, soil quality, as well as field crops with high profit margins, such as sugar beets and potatoes. In addition, farm size, the share of rented arable land, labour and capital endowments, as well as livestock density per hectare also have an influence on rental rates. Characteristics that reflect local competition and the availability of farmland at the regional level, such as the regional density of operating farmers, or the regional livestock and population densities have also been shown to have an influence on rental rates. More recently, there has been an increasing interest in analysing the potential effects of agricultural subsidies and payments on rental rates, with Patton *et al.* (2008), Swinnen, Ciaian and Kancs (2008), Ciaian, Kancs and Paloma (2010a), Möller *et al.,* (2010), Breustedt and Habermann (2011), Kilian *et al.,* (2012), Feichtinger and Salhofer (2013) and Michalek and Ciaian (2014) empirically analysing the impact of such payments on rental rates for countries within the European Union (EU), while Ciaian and Kancs (2012) and Vranken and van Herck (2013) focus on a selected number of new EU member states only. More recently, Hennig, Breustedt and Latacz-Lohmann (2014) analyse the possible impact of payment entitlements

1 One reason that the share of rental farmland in East Germany exceeds those in West Germany might be that, similarly to Eastern European countries, fragmented land ownership increasingly leads farmers to offer their land for lease and might also encourage larger farms and cooperatives because formal land owners in East Germany might have no desire or ability to cultivate their land.

on German rental rates. The general finding of these incidence studies is that payments that are originally designed to support farmers are actually capitalised into farmland rental rates.

Many empirical studies that analyse farmland and its determinants make use of hedonic pricing models. According to the hedonic pricing theory, the price of a good can be divided into the sum of its attributing values, which are then estimated using regression models (Rosen, 1974). In order to account for spatial dependencies, spatial lag and spatial error models have in recent years evolved as standard tools in hedonic pricing studies. However, even though these models have been widely used in applied work, Nickerson and Zhang (2014) highlight the problem of functional form misspecifications as one crucial issue when estimating hedonic models, since there is no existing theory that guides researchers. Addressing this issue more closely, McMillen (2003) attaches great importance to functional form misspecifications, since they can falsely lead researchers to fit a spatial model to the data. This argument is further supported by the findings of an empirical investigation conducted by Kostov (2009b). In particular, Kostov (2009b) finds that misspecifications with respect to the functional form may be responsible for finding spurious spatial dependencies when hedonic pricing models are used. Basile and Gress (2005) provide another example of the consequences of misspecifying the functional form by demonstrating that when estimating semi-parametric versions of spatial autoregressive models, the parameters that determine the strength of the spatial association are considerably smaller than their counterparts in fully parametric versions of the models. Consequently, functional form misspecifications can adversely influence statistical inference, as they may affect both the magnitude and the significance of the estimated effects (Nickerson and Zhang, 2014). Besides its relevance for statistical inference, Margarian (2008) considers the topic as being equally important from an economic point of view. In particular, Margarian (2008) emphasises that more attention should be dedicated to the analysis of the functional form in applied regression analysis rather than only on quantifying the estimated effects resulting from an assumed linear relationship. These considerations therefore suggest that the assumption of linearity in parametric hedonic pricing models should be relaxed in order to acknowledge the possible non-linearity of some of the covariate effects in the modelling process. Semi-parametric regression models naturally lend themselves to this idea, as this model class allows for a broader class of functional relationships than parametric models (McMillen and Redfearn, 2010).

Although the analysis of farmland is an active field of research in the agricultural economics literature, comparatively few studies go beyond the modelling of conditional means.[2] Consequently, little is known about farmland and the

2 Few exceptions that use quantile regression models for the analysis are Kostov (2009a), Mishra and Moss (2013) and Uematsu, Khanal and Mishra (2013). Kostov (2009a) uses a quantile version of the spatial lag model which allows for varying effects of farmland characteristics, as well as for varying degrees of spatial autocorrelation. Mishra and Moss (2013) use a quantile regression model to analyse the influence of off-farm income of households on farmland values in the United States and also study the effect of farm subsidies, as well as the impact of farm location on farmland values. The scope of analysis of Uematsu, Khanal and Mishra (2013) is to study the effects of

way it is influenced by economic variables for points other than the central part of the response distribution. Since it is reasonable to assume that rents in the upper or lower tail of the distribution are governed by a different data generating process than mean rental rates and therefore may depend on a different set of explanatory variables, the determinants that have so far been found to exhibit an influence on mean rental rates might have no or only little explanatory power for non-central parts of the response distribution. The analysis based on quantile regression can therefore provide additional insight into the data generating process of farmland rental rates, as this model class allows for the analysis of the relationship between covariates and the response at different points in the conditional distribution. Hence, quantile regression models are of special interest for hedonic pricing studies, as interest usually lies in analysing expensive rental rates rather than average rental rates only. Despite these advantages, it has only been recently that quantile regression models have been introduced within spatial econometrics. However, spatial quantile regression models that are commonly used in the applied agricultural economics literature cannot be guaranteed to fully eliminate the problems associated with functional form misspecifications, since a linear relationship between the response and the covariates is usually assumed (Kostov, 2013).

Another topic that has mostly been neglected in the empirical literature on farmland is the issue of heterogeneity (Mishra and Moss, 2013). Despite heterogeneity having many different causes, we consider expectations of economic agents with respect to future conditions as one possible source of such heterogeneity, acknowledging that previous studies have highlighted the importance of conditioning renting decisions on subjective future expectations.[3] Since rental rates are usually mid- to long-term, rental rate agreements are negotiated at different points in time which leads to heterogeneity due to the different sets of expectations with respect to the future profit potential of the farmland under which the agreements were originally formed. Consequently, rental rates in the upper quantiles might have been formed under more optimistic expectations and therefore might reflect a higher willingness to pay, whereas rents in the lower quantiles might reflect a lower willingness to pay as they might have been formed under a more pessimistic outlook for the future. Hence, in a sample of heterogeneous economic agents, we expect the formation of future expectations to differ across farmers so that it is difficult, or even infeasible, to directly and explicitly control for heterogeneity in the estimated effects in conventional hedonic pricing models (Bekkerman, Brester and McDonald, 2013). This is especially true if no information is available with respect to the point in time that rental contracts were negotiated, as is the case with our data. However, even if appropriate proxies exist, the commonly applied

variables representing natural amenity and soil characteristics on US farmland values using a quantile regression model.

3 See, e.g. Harris and Nehring (1976), Lee and Rask (1976), Melichar (1979), Reinsel and Reinsel (1979) or Dunford, Marti and Mittelhammer (1985).

hedonic pricing models only estimate average marginal effects, which may be less conclusive when substantial heterogeneity exists across the conditional distribution of rental rates. Consequently, mean hedonic pricing models may not provide an accurate estimation of the marginal effects, as these models do not allow for identifying heterogeneity that results from factors which are not directly observable (Bekkerman, Brester and McDonald, 2013). Following Barnes and Hughes (2002) and Fitzenberger, Koenker and Machado (2002), we argue that heterogeneity with respect to future expectations of economic agents can be uncovered by looking at different points in the conditional distribution of rental rates and that quantile regression provides a means of empirically uncovering such heterogeneity. In particular, since economic agents' preferences are revealed by their buying and renting decisions, performing a set of regressions for different quantiles of the response distribution allows us to make use of the variation in revealed rental rates. Consequently, quantile regression models exploit the informational content that is implicit in farmland rental rates and provide a means to quantify and identify whether the estimated marginal effects are constant, or whether substantial differences exist for different points in the conditional distribution. Hence, conditional quantiles can be viewed as an approximation of the distinct and unobservable sets of factors that affect the formation of farmland rental rates (Bekkerman, Brester and McDonald, 2013).

Based on the considerations presented above, the purpose of this article is twofold. First, and in contrast to previous studies that were primarily concerned with the analysis of average rental rates, the current article contributes to the agricultural economics literature by modelling conditional quantiles of farmland rental rates using Bayesian geoadditive quantile regression models as introduced by Waldmann *et al.,* (2013). By modelling 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 low, medium and high rents, as well as uncover potential heterogeneities across the conditional distribution of rental rates that would not be revealed if mean regression models were used. This allows us to present a more detailed analysis of the price formation of rental rates and to investigate whether average marginal effects, which have been the main interest in the literature so far, yield an appropriate description of rental rates, or whether these effects differ across the conditional distribution. Second, we address the issue of functional form misspecification by modelling conditional quantiles of farmland rental rates semi-parametrically. In order to alleviate problems that may occur when linear quantile regression models are used, the semi-parametric nature of the regression model relaxes the assumption of linearity and allows for a data-driven selection of the functional form. By additively including a spatial term to the model, we are also able to provide additional insight into the spatial structure of the data. Consequently, our analysis addresses several issues that are currently being discussed in the agricultural economics literature.

The remainder of this article is organised 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 and Section 5 concludes.

2. Methodology

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, Kneib and Lang (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. STAR models include generalised additive models (GAM), varying coefficient models (VCM), generalised additive mixed models (GAMM), geoadditive models, as well as geographically weighted regression models. In order to provide a more detailed description of the conditional distribution of the response variable, Waldmann *et al.,* (2013) extend STAR models to include Bayesian geoadditive quantile regression models, which are presented in this section.[4]

2.1 Bayesian geoadditive quantile regression

Before introducing Bayesian geoadditive quantile regression in a more rigorous sense, we would like to give the reader a rather non-technical explanation of quantile regression first.

Starting from a more traditional perspective, ordinary least squares (OLS) regression models focus on the central location of the response by modelling its conditional mean as a function of covariates. This analysis yields valid results as long as the distributional assumptions hold and as long as the relationship between covariates and the central tendency of the response is the main interest of the analysis. However, if the interest lies with modelling non-central locations, such as the lower or upper tail of the response distribution, or with the analysis of the entire conditional distribution, mean regression models are not well-suited for such an analysis as they do not account for other distributional properties of the response variable. As a means of arriving at a more complete and nuanced view, Koenker and Bassett (1978) have introduced quantile regression models that allows the researcher to model any point of the response distribution within a regression setting. As with mean regression models, where the conditional mean is modelled as a function of covariates, quantile regression relates any specific quantile τ of the response to a set of covariates.

4 Since geoadditive quantile regression models are motivated from a statistical point of view and since many readers may be more familiar with models from spatial econometrics, we illustrate the basic idea of spatial lag and spatial error models that are commonly used in hedonic pricing studies of farmland rental rates in Section A.1 in Appendix A where we also highlight the differences to Bayesian geoadditive quantile regression models that are introduced in Section 2.1.

Consequently, any desired point of the response distribution can be modelled, so that a dense grid of quantile regressions yields a detailed description of the conditional distribution. Therefore, estimating and comparing parameter estimates across a different set of quantiles allows for fully characterising the response distribution and for investigating the differential effect that covariates may have on different points of the conditional distribution. For hedonic pricing studies, quantile regression yields additional insight compared with mean regression models, as it provides a richer description of the relationship between the price of a good and its attributing values for different values of τ. In particular, standard hedonic pricing models disregard important features of the data and yield an incomplete representation of the conditional distribution, as the conventional estimators are fixed for all quantiles so that the estimated effects are averaged out over the response distribution. Considering quantile regression as a means of uncovering heterogeneity across the conditional distribution, we think of τ as an unobservable variable that approximates future expectations of economic agents with respect to the profitability of farmland. Since the vector of regression coefficients $\boldsymbol{\beta}_\tau$ is allowed to vary with τ, latent factors are nested in the quantile regression coefficients in the sense that the estimated effects are allowed to interact with the heterogeneity. Consequently, we are able to capture the heterogeneity that is neglected by traditional OLS regression models, so that performing a set of regressions for different quantiles can be viewed as reflecting the distribution of farmers' expectations regarding the profitability of the agricultural activity (Arias, Hallock and Sosa-Escudero, 2002). While linear quantile regression models assume a linear relationship between covariates and the τ-th quantile of the response, geoadditive quantile regression models allow for this assumption to be relaxed by replacing the linear predictor with a more flexible, semi-parametric predictor, which incorporates possible non-linearities, as well as spatial effects into the model.

Introducing Bayesian geoadditive quantile regression models more formally, we start with the linear quantile regression model as proposed by Koenker and Bassett (1978)

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta}_\tau + \varepsilon_{i,\tau}, \quad i = 1, \ldots, n. \tag{1}$$

where $\boldsymbol{x}_i$ contains both categorical and continuous covariates, $\tau \in (0, 1)$ indicates the quantile of interest, $\boldsymbol{\beta}_\tau$ is a vector of quantile-specific regression coefficients and $\varepsilon_{i,\tau}$ is an unknown error term with cumulative density function $F_{\varepsilon_{i,\tau}}$ that depends on the quantile parameter τ. For traditional quantile regression, no specific assumptions regarding the distribution of the error term are made apart from $\varepsilon_{i,\tau}$ and $\varepsilon_{j,\tau}$ being independent for $i \neq j$, as well as $F_{\varepsilon_{i,\tau}}(0|\boldsymbol{x}) = \tau$, meaning that the τ-quantile of the error term conditional on $\boldsymbol{x}$ is zero. Given these assumptions, the quantile-specific regression coefficients $\boldsymbol{\beta}_\tau$ are estimated by minimising an asymmetrically weighted sum of absolute deviations

$$\boldsymbol{\beta}_\tau^* = \arg\min_{\boldsymbol{\beta}_\tau} \sum_{i=1}^{n} \rho_\tau\big(y_i - \eta_{i,\tau}^{\text{linear}}\big) \tag{2}$$

where $\eta_{i,\tau}^{\text{linear}} = \boldsymbol{x}_i'\boldsymbol{\beta}_\tau$ and

$$\rho_\tau\big(y_i - \eta_{i,\tau}^{\text{linear}}\big) = \begin{pmatrix} \tau|y_i - \eta_{i,\tau}^{\text{linear}}| & \text{if} \quad y_i \geq \eta_{i,\tau}^{\text{linear}} \\ (1-\tau)|y_i - \eta_{i,\tau}^{\text{linear}}| & \text{if} \quad y_i < \eta_{i,\tau}^{\text{linear}} \end{pmatrix} \tag{3}$$

is the check function that defines a suitable loss function for quantile regression. Hence, for a fixed quantile τ and observation i the linear predictor $\eta_{i,\tau}^{\text{linear}}$ models the conditional quantile of the response y_i.

To allow for non-linearities in the relationship between the response and the covariates, as well as to account for spatial dependencies in the data, the model of Waldmann *et al.* (2013) replaces the strictly linear predictor $\eta_{i,\tau}^{\text{linear}}$ with the more flexible geoadditive quantile predictor

$$\eta_{i,\tau} = \boldsymbol{x}_i'\boldsymbol{\beta}_\tau + \sum_{j=1}^{p} f_{j,\tau}(z_{ij}) + f_{\text{geo},\tau}(s_i) \tag{4}$$

that allows the researcher to analyse the influence of the covariates on the response variable semi-parametrically, for each quantile separately. In Equation (4), $\boldsymbol{x}_i'\boldsymbol{\beta}_\tau$ is the parametric part of categorical covariates $\boldsymbol{x}_i$, including the overall intercept $\beta_{0,\tau}$, $f_{j,\tau}(\cdot)$ are non-linear smooth effects of continuous covariates z_{ij} and $f_{\text{geo},\tau}(s_i)$ incorporates the spatial information into the model.[5] Relaxing the assumption of linearity allows for gaining additional insight into the relationship between covariates and the response, as important features in the data might go undetected if linear regression models are used; it also helps to alleviate problems with hedonic pricing models that are associated with

5 The Bayesian formulation of quantile regression relies on the equivalence between posterior mode and maximum likelihood estimation under non-informative priors (Fahrmeir *et al.*, 2013). To establish a link between the minimisation problem in Equation (2) and the Bayesian approach, a distributional assumption for the error term has to be made that allows for setting up a likelihood function that is necessary for Bayesian inference. As suggested by Koenker and Machado (1999) and Yu and Moyeed (2001), the asymmetric Laplace distribution with location parameter $\eta_{i,\tau}^{\text{linear}}$, scale parameter σ^2 and asymmetry parameter τ is particularly suitable for Bayesian quantile regression models, since the minimisation of the check function in the frequentist setting can equivalently be represented as maximising the asymmetric Laplace likelihood function

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

with respect to $\eta_{i,\tau}^{\text{linear}}$, so that posterior mode estimates of the regression coefficients are equivalent to those which minimise the check function in Equation (2). See Waldmann *et al.* (2013) and the references therein for details.

functional form misspecifications. In order to model the non-linear smooth functions $f_j(z_{ij}) = \sum_{k=1}^{m_j} \gamma_{j,k} B_{j,k}(z_{ij}),\ i = 1, \ldots, n;\ j = 1, \ldots, p,$ where the coefficients $\gamma_{j,k}$ can be interpreted as amplitudes that scale the B-spline basis functions $B_{j,k}$ accordingly to fit the data, Lang and Brezger (2004) and Brezger and Lang (2006) introduced a Bayesian analogue to P(enalised)-splines, as originally proposed from a frequentist point of view by Eilers and Marx (1996).[6] In the Bayesian framework, priors are assigned to the coefficients $\boldsymbol{\gamma}_j = (\gamma_{j,1}, \ldots, \gamma_{j,m_j})'$ of the non-linear smooth effect f_j that enforce a smooth function estimation, where the smoothness of f_j is estimated simultaneously with the regression coefficients $\boldsymbol{\gamma}_j$.

Similar to mixed models for longitudinal data, where random effect terms allow to relax the assumption of independence between repeated observations within a group or cluster, the spatial term f_{geo} in Equation (4) accounts for spatial autocorrelation and acts as a surrogate for covariates that are not included in the model (Fahrmeir and Kneib, 2011). In order to arrive at a smooth surface, it has to be ensured that the estimated spatial effects $\boldsymbol{f}_{\text{geo}} = (f_{\text{geo}}(s_1), \ldots, f_{\text{geo}}(s_n))' = \boldsymbol{Z}_{\text{geo}}\boldsymbol{\gamma}_{\text{geo}}$ of neighbouring districts do not differ too strongly from one another, where $\boldsymbol{\gamma}_{\text{geo}} = (\gamma_{\text{geo}}(1), \ldots, \gamma_{\text{geo}}(S))'$ is a vector of regression coefficients that collects all distinct spatial effects for the districts $s = 1, \ldots, S$, and $\boldsymbol{Z}_{\text{geo}}$ is a $n \times S$ design matrix that connects an observation i with the corresponding spatial effect, i.e. $\boldsymbol{Z}_{\text{geo}}[i, s] = 1$ if y_i was observed in district s and 0 otherwise. One way of achieving such spatial smoothness is to assign Gaussian Markov Random Field (GMRF) priors to the coefficients

$$\gamma_{\text{geo}}(s)|\boldsymbol{\gamma}_{\text{geo}}(-s) \sim N\left(\frac{1}{|N(s)|}\sum_{r\in N(s)} \gamma_{\text{geo}}(r), \frac{\phi^2_{\text{geo}}}{|N(s)|}\right), \quad s = 1, \ldots, S. \tag{5}$$

where $\boldsymbol{\gamma}_{\text{geo}}(-s)$ is the vector containing all spatial effects except the one for district s, $|N(s)|$ denotes the total number of neighbours that share a common boundary with district s and the parameter ϕ^2_{geo} controls for the variation of $\gamma_{\text{geo}}(s)$ around its expected value. From Equation (5), it follows that the estimated structured spatial effect $\gamma_{\text{geo}}(s)$ is a weighted average of the neighbouring effects, with region s being conditionally independent of all other regions that do not share a common boundary and where the variability of $\gamma_{\text{geo}}(s)$ around its conditional mean inversely depends on the number of neighbours that surround region s. Consequently, assigning Markov random field priors induces a specific correlation structure and ensures spatial smoothness of the regression coefficients $\boldsymbol{\gamma}_{\text{geo}}$,

6 In the frequentist setting of Eilers and Marx (1996), 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 B-spline basis functions $B_{j,k}$ evaluated at pre-specified knots $z_{j,\min} = \zeta_{j,1} < \zeta_{j,2} < \cdots < \zeta_{j,h_j} = z_{j,\max}$. To ensure a good fit to the data, Eilers and Marx (1996) suggest using a sufficiently high number of equidistant knots, as well simultaneously impose a penalty on adjacent B-spline coefficients $\gamma_{j,k}$ which prevent f_j from being too wiggly.

since parameters of neighbouring districts are not allowed to deviate too strongly from one another. If spatial heterogeneity exists only locally, it is not reasonable to assume that coefficients of neighbouring districts are spatially correlated and an uncorrelated spatial effect should be used instead. To model this unstructured spatial effect, district-specific i.i.d. Gaussian random effects are commonly used (for a more detailed exposition of STAR, we refer the interested reader to Fahrmeir and Kneib (2011) and Fahrmeir *et al.* (2013)).

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).[7] This census 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, labour and employment. We use farmland rental rates per hectare as the response variable, since farmers commonly use this figure for guidance when determining rental agreements.[8] Tenancies that were entered between family members are excluded to obtain a market-based assessment of rental rates. Based on previous studies that analyse German farmland rental rates (see, e.g. Habermann and Ernst (2010) or Habermann and Breustedt (2011)), we use the covariates presented in Table 1 for the analysis that have been shown to exhibit an influence on rental rates.

In order to account for quality differences between arable land and pasture land, as well as to account for the varying levels of productivity of the rented land, we follow Habermann and Ernst (2010) and include the share of rented arable land in the model as a proxy. The share of rye is used as an approximation

7 Most variables in the 2010 census were reported to the Federal Statistical Office at the time of the survey, which was 1 March, 2010. However, some of the variables are also reported for the fiscal and calendar year of 2010. See Statistisches Bundesamt (2010) for details.

8 As far as the rental rates that are used for the analysis are concerned, we chose to select farmland rental rates for tenancies that are older than two years (the so-called 'Bestandspachten') as the response variable. Other studies model rental rates using 'Neupachten' as their dependent variable, which are tenancies that have been signed or for which the terms have been adjusted within the last two years. However, since we want to arrive at a representative picture of the overall situation of the German farmland rental market, our decision to model Bestandspachten is based on the fact that the number of observations is greatly reduced when Neupachten are used, since they are a random sample of the entire sample and are designed so that they are representative on the levels of the federal state and the administrative district only (see Statistisches Bundesamt (2010) for details). The fact that we use 'Bestandspachten' as the response, irrespective of the year the rental contract started, might induce some problems with respect to interpretation, since some farms may have low rents simply because their contracts have been negotiated many years ago. Indeed, a considerable share of the rent contracts are signed for several years in Germany, so that our response variable includes contracts signed in different years (Swinnen, Ciaian and Kancs, 2008). Unfortunately, the data set does not contain any information from which we could deduce the exact point in time when the rental contract started. As such, the fact that we cannot directly account for the time structure of the signed tenancies is related to the heterogeneity argument made in Section 1. Therefore, we make use of quantile regression as a means to account for such heterogeneity present in the data.

Table 1. Description of covariates

Covariate (farm-level)	Description
biogas	Capacity of biogas plant in kWh (continuous)
cattle	Farm-level cattle density in animal units (AU) per hectare (continuous)
east	Dummy variable for East Germany (categorical: 1 = West Germany, 2 = East Germany)
farm_succ	Farm succession (categorical: 1 = yes, 2 = no, 3 = unsettled)
arable_share	Share of rented arable land to total rented agricultural land (continuous)
fulltime	Dummy variable indicating whether the farmer operates his business full-time or part-time (categorical: 1 = no individual enterprise as legal form, 2 = full-time, 3 = part-time)
hog_poultry	Farm-level hog and poultry density in animal units (AU) per hectare (continuous)
labour	Labour force per hectare (continuous)
organic	Dummy variable for organic farming (categorical: 1 = yes, 2 = no)
potato	Share of potato in cropping pattern (continuous)
rent_share	Share of rented agricultural land to total farmed agricultural land (continuous)
rye	Share of rye in cropping pattern (continuous)
size	Total agricultural land of the farmer in hectare (continuous)
sugarbeet	Share of sugar beet in cropping pattern (continuous)
winterwheat	Share of winter wheat in cropping pattern (continuous)
Covariate (district level)	
cattle_district	Average district-level cattle density in animal units (AU) per hectare (continuous)
hhi	Herfindahl-Hirschman index based on the share of rented agricultural land to total agricultural land in each district (continuous)
hog_poultry_district	Average district-level hog and poultry density in animal units (AU) per hectare (continuous)
income	Average per capita income (continuous)
inha	Inhabitants per square kilometre (continuous)
unempl	Unemployment rate (continuous)

Source: Research Data Centres of the Federal Statistical Office and the statistical offices of the federal states, agricultural census 2010, Regionaldatenbank (2010), authors' own calculations.

for varying precipitation levels, as rye can be considered to be comparatively resistant to drought and is predominantly grown on low quality agricultural land in Germany, which makes it a reasonable choice. Cattle density, as well as hog and poultry density in livestock units per hectare of agricultural land serve as measures of the production intensity. To capture local competition for farmland, we include both the average livestock densities of each district and the Herfindahl index as proxies.[9]

Field crops with high profit margins such as the proportion of sugar beet, winter wheat and potato in the cropping pattern are included as well. There are numerous studies that analyse the influence of agricultural subsidies from federal farm programmes on rental rates. For the German rental market, e.g. Breustedt and Habermann (2011) analyse the effects of per-hectare premiums paid to farmers on rental rates. However, since no information on agricultural subsidies is available in our data, we cannot include subsidies for our analysis. To take into account the potential effects of increased land use for the production of bioenergy on rental rates, similar to Hennig, Breustedt and Latacz-Lohmann (2014), we include the capacity of each farmers' biogas plant in kWh in the model.[10] This is an important issue for German farmers, since renewable energy sources have come to the centre of attention for German energy and climate policy in recent years. On the basis of the German Renewable Energy Act (German: Erneuerbare-Energien-Gesetz (EEG)), which guarantees fixed feed-in tariffs for electricity gained through renewable sources, farmers have decided to increasingly use more arable land for the production of bioenergy. Consequently, many farmers and landowners are driven by the question as to how the effects of the EEG affect farmland rental rates. To allow for structural differences in the rental market between East and West Germany, we include a dummy variable in our model.[11] In order to account for spatial heterogeneity that is not captured by farm-level covariates, we additionally include socio-demographic covariates on the district level from the Regionaldatenbank (2010). The discrete spatial information that identifies each farmer and the municipality he or she operates in is provided by an 8-digit code. For ease of visualisation and interpretation, the spatial effects are presented at the district level. The effects of the remaining covariates, however, are presented at the farm level. After removing non-renting farmers, as well as outlying observations, we are left with 107,620 observations for the analysis.[12]

9 As far as the Herfindahl-Hirschman Index is concerned, we follow Habermann and Ernst (2010) and construct the index as follows:

$$\mathrm{HHI}_j = \sum_{i=1}^{N_j} \left(\frac{\mathrm{Rented\_Area\_Farm}_i}{\mathrm{Total\_AgrArea\_District}_j} \right)^2$$

where $\mathrm{Rented\_Area\_Farm}_i$ is the rented area of farmer i and $\mathrm{Total\_AgrArea\_District}_j$ is the total available rented agricultural area within district j for a total of N_j in district j.

10 However, we can only account for the immediate effects of bioenergy production. In order to draw more general conclusions and to allow for a more general assessment of the effects of biogas plants on rental rates, information on the regional plant density or the cumulative installed plant capacity in each district is needed. See also Hennig, Breustedt and Latacz-Lohmann (2014).

11 Due to the structural differences, some studies suggest splitting the data and to analysing farmland rental rates for East and West Germany separately (see, e.g. Habermann and Ernst (2010) or Habermann and Breustedt (2011)). However, splitting the data would mean artificially disconnecting the spatial structure in the data and could bias the estimated effects near the inner German border. Consequently, we have decided not to split the data and to use a dummy variable approach instead, as we are particularly interested in analysing the spatial structure in the data resulting from spatial dependencies.

3.2 Variable selection

Variable selection is a challenging task in regression analysis in general, and for geoadditive quantile regression in particular. The researcher must select a subset of covariates that he or she considers relevant for the analysis and then has to decide whether the spatial information in the data is better described by a structured or unstructured effect. The fact that different quantiles may depend on a different set of covariates, as well as the fact that there is no existing economic theory that dictates the selection of covariates across the entire response distribution further aggravates the problem of variable selection in geoadditive quantile regression. To make the task of variable selection more feasible, we use a systematic and fully data-driven approach based on component-wise functional gradient descent boosting for geoadditive quantile regression, as proposed by Fenske, Kneib and Hothorn (2011). Boosting is a machine learning approach that is aimed towards maximising the prediction accuracy of the response by iteratively combining different model components, called base learners, where in each iteration, only the best-fitting base learner, i.e. the most informative covariate, is selected (for details on the boosting algorithm, see Fenske, Kneib and Hothorn (2011)). For geoadditive quantile regression, boosting is particularly appealing since boosting decides separately which covariates should enter the model for each quantile. For the initial model, we include all covariates presented in Table 1. To cover the entire range of rental rates and to gain detailed insight into low, medium and high rental rates, as well as to reveal non-constant effects across the conditional response distribution, we choose to model conditional quantiles for $\tau = \{0.05, 0.20, 0.50, 0.80, 0.95\}$. Table 2 presents the selected covariates for the quantiles under analysis.

From Table 2, it follows that the set of covariates that influence farmland rental rates is different across quantiles; while the selected variables for $\tau = \{0.20, 0.50, 0.80\}$ are generally similar, differences exist between the selected variables that influence more extreme quantiles. While the average per capita income, as well as the population density have been selected for $\tau = \{0.20, 0.50, 0.80\}$, none of these variables have been selected for $\tau = \{0.05, 0.95\}$. Also note that while the dummy variables *organic* and *farm_succ* have an effect on $\tau = \{0.50, 0.80\}$, they do not influence rents for the 5, 20 and 95 per cent quantiles. Furthermore, animal densities, both at the farm, as well as on the district level, seem to be strong predictors for rental rates, as they have been selected for each individual quantile. From Table 2, it is also apparent that $f_{\mathrm{geo}}(str)$ has been selected for all quantiles in order to model the spatial information in the data. However, the unstructured spatial effect has also been selected for $\tau = \{0.20, 0.50, 0.80\}$ indicating that there seems to be additional small scale, district-specific spatial information in the data for these quantiles. Moreover, the dummy variable that distinguishes

12 To allow for the knots, that are needed for the construction of the B-spline basis functions, to be equidistantly distributed over the domain of the covariates, we remove observations from the data if they exceed the commonly used threshold of $q_{0.5} + 3(q_{0.75} - q_{0.25})$, where q_τ denotes the τ-th quantile of the corresponding variable.

Table 2. Selected covariates across quantiles during boosting iterations

Covariate	$\tau = 0.05$	$\tau = 0.20$	$\tau = 0.50$	$\tau = 0.80$	$\tau = 0.95$
f(*biogas*)		√	√	√	
f(*cattle*)	√	√	√	√	√
east	√				
farm_succ			√	√	
f(*arable_share*)	√	√	√	√	√
fulltime	√	√	√	√	√
f(*hog_poultry*)	√	√	√	√	√
f(*labour*)			√		
organic			√	√	
f(*potato*)		√	√	√	√
f(*rent_share*)	√	√	√	√	√
f(*rye*)	√	√	√	√	√
f(*size*)	√	√	√	√	
f(*sugarbeet*)	√	√	√	√	√
$f_{\mathrm{geo}}(str)$	√	√	√	√	√
$f_{\mathrm{geo}}(unstr)$		√	√	√	
f(*winterwheat*)	√	√	√	√	
f(*cattle_district*)	√	√	√	√	√
f(*hhi*)					
f(*hog_poultry_district*)	√	√	√	√	√
f(*income*)		√	√	√	
f(*inha*)		√	√	√	
f(*unempl*)	√	√	√	√	√

$f(\cdot)$ indicates that the variable is modelled semi-parametrically, whereas the variable name itself indicates that the variable is modelled parametrically. $f_{\mathrm{geo}}(str)$ and $f_{\mathrm{geo}}(unstr)$ denote the structured and unstructured spatial effect, respectively.
Source: Research Data Centres of the Federal Statistical Office and the statistical offices of the federal states, agricultural census 2010, Regionaldatenbank (2010), authors' own calculations.

between East and West German farmers has only been selected for very low rental rates at the 5 per cent quantile. Consequently, differences between East and West German rental markets seem to be captured by the spatial effect terms $f_{\mathrm{geo}}(str)$ and $f_{\mathrm{geo}}(unstr)$ for the other quantiles. 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. Regarding the effect of increased land use for the production of biogas on rental rates, Table 2 shows that boosting has decided to include it as an important covariate for $\tau = \{0.20, 0.50, 0.80\}$. Consequently, the effect of biogas influences medium rental rates, as well as moderately low and high rental rates. However, biogas seems to have no effect on very expensive rents, as boosting has not selected it during any of the iterations for the 95 per cent quantile. 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 average German rental rates, while the results of Habermann and Ernst (2010) do not support this finding. Habermann and Ernst (2010) attribute these various findings to the different granularity of the data that have 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 effects of biogas plants are reflected on rental rates. This inelasticity of rental rates might also apply to our results, which might explain why boosting has decided to include biogas as an important covariate for medium and moderately low and high rental rates, but not for more expensive rents.

Similar to Kostov and Davidova (2013), we view the pattern of the selected covariates across the different quantiles as a way of assessing the relevance of the variables with respect to their influence on forming farmers' expectations. For example, both the average per capita income and the unemployment rate, which have been included to account for the opportunity costs of farming, as well as for the economic prosperity of a region, are not only associated with farmers' expectations at the lower and medium quantiles, but also at the higher ones. This is especially true for the unemployment rate, since this variable exerts influence across the entire set of quantiles. The results therefore suggest that farmers take the overall economic situation into account when forming their expectations. Moreover, farm succession, which allows for a long-term strategic planning, both financially and with respect to the workforce, seems to positively influence future expectations since this variable has been selected for medium and moderately high quantiles. Furthermore, characteristics that reflect local competition and the availability of farmland at the regional level, as indicated by the regional livestock and population densities, also appear to have an influence on the formation of farmers' expectations. Since the scarcity of farmland leads to an increased competition for farmland, farmers seem to anticipate the rise in rental rates when forming their future expectations.

Before we turn to the analysis of farmland rental rates itself, we would like to assess the ability of boosting to identify the most influential covariates across quantiles. To do so, we use the deviance information criterion (DIC) as a way to discriminate between the goodness of fit of the models including all variables of Table 1, and the models that include the variables selected by boosting.

Table 3. DIC of the full models and the models based on boosting

Model	$\tau = 0.05$	$\tau = 0.20$	$\tau = 0.50$	$\tau = 0.80$	$\tau = 0.95$
Full model	−332,043.000	−54,636.200	−18,944.100	−42,466.600	−306,821.000
Boosting	−333,143.000	−54,684.100	−18,947.700	−42,488.200	−307,839.000

Source: Research Data Centres of the Federal Statistical Office and the statistical offices of the federal states, agricultural census 2010, Regionaldatenbank (2010), authors' own calculations.

Table 3 compares the corresponding values of the DIC for the two model alternatives across quantiles.

Table 3 provides evidence that the models using the covariates selected by the boosting algorithm provide a better fit to the data compared with the full models that use all of the variables of Table 1.

4. Analysis of farmland rental rates

After having identified the relevant economic variables that influence farmland rental rates across quantiles, we now present the estimation results of the Bayesian geoadditive quantile regression.[13] Due to the large number of estimated effects, we restrict our presentation and discussion to some selected effects only.[14]

4.1 Parametric and semi-parametric effects

Figure 1 shows posterior mean estimates for the centred semi-parametric effects.[15] For reasons of clarity, and where it is possible due to variable selection,

13 All results are obtained via full Bayesian MCMC simulation based on 120,000 iterations, a burn-in period of 20,000 iterations and a thinning parameter of 100 resulting in a sample of 1,000 samples from the posterior. For all quantiles and for all estimated effects, we investigated the resulting sampling paths for convergence and mixing and found no evidence of remaining autocorrelation. The smooth effects of continuous covariates are estimated via cubic penalised B-splines based on second-order random walk priors. To avoid very rough estimates for the extreme quantiles, we use 10 equidistant inner knots for all model specifications. Hyper-parameters for the smoothing variances ϕ_j^2 are set to $a_j = b_j = 0.001$ as a default, as these parameter values have turned out to be a robust choice in previous studies. To analyse the sensitivity of the estimated effects with respect to the choice of the hyper-parameters, we re-estimated the models for values of $a_j = 1.0$, $b_j = 0.005$ and found no differences in the results. The structured spatial effect is estimated based on a Markov random field prior, whereas for the estimation of the unstructured effect, district-specific i.i.d. Gaussian random effects are used. The estimation is performed using the R-package *BayesXsrc* version 3.0-0 of Adler *et al.* (2015), which is an R-interface to the standalone software *BayesX* of Belitz *et al.* (2015). To assess the robustness of the Bayesian quantile regression results with respect to variable choice and its influence on the estimated effects, we re-estimated the models to include all variables presented in Table 1 and found only marginal differences of the estimated effects compared with the results presented in Section 4.1. To assess the overall robustness of the model, we ran the analysis in a frequentist setting based on iteratively re-weighted least squares. In particular, following the idea of expectiles of Newey and Powell (1987), we re-estimate the effects for each $\tau = \{0.05, 0.20, 0.50, 0.80, 0.95\}$ by least asymmetrically weighted squares (LAWS). In contrast to quantile regression, where an asymmetrically weighted sum of absolute deviations is minimised, expectile regression relies on minimising asymmetrically weighted squared residuals

$$\sum_{i=1}^{n} w_\tau(y_i, \eta_{i,\tau})(y_i - \eta_{i,\tau})^2 \text{ with the weights defined as } w_\tau(y_i, \eta_{i,\tau}) = \begin{pmatrix} \tau & \text{if} \quad y_i - \eta_{i,\tau} \geq 0 \\ (1-\tau) & \text{if} \quad y_i - \eta_{i,\tau} < 0 \end{pmatrix}.$$

Comparing the results from both the quantile and expectile regressions, we generally find only marginal differences in the estimates (for more details on geoadditive expectile regression, we refer to Sobotka and Kneib (2012)).

14 All remaining estimated effects can be found in Figure A1 of Appendix A.

15 Since the parametric part of the geoadditive quantile predictor includes an overall intercept term, each function is centred around zero, i.e. $\sum_{i=1}^{n} f_1(z_{i1}) = \cdots = \sum_{i=1}^{n} f_p(z_{ip}) = 0$ in order to guarantee the identification of the estimated semi-parametric effects. The estimated effects in Figure 1 are plotted in this way. The *x*-axis shows the domain of the corresponding covariate. Moreover, the fact that each non-linear smooth effect is constructed based on B-spline basis functions, where

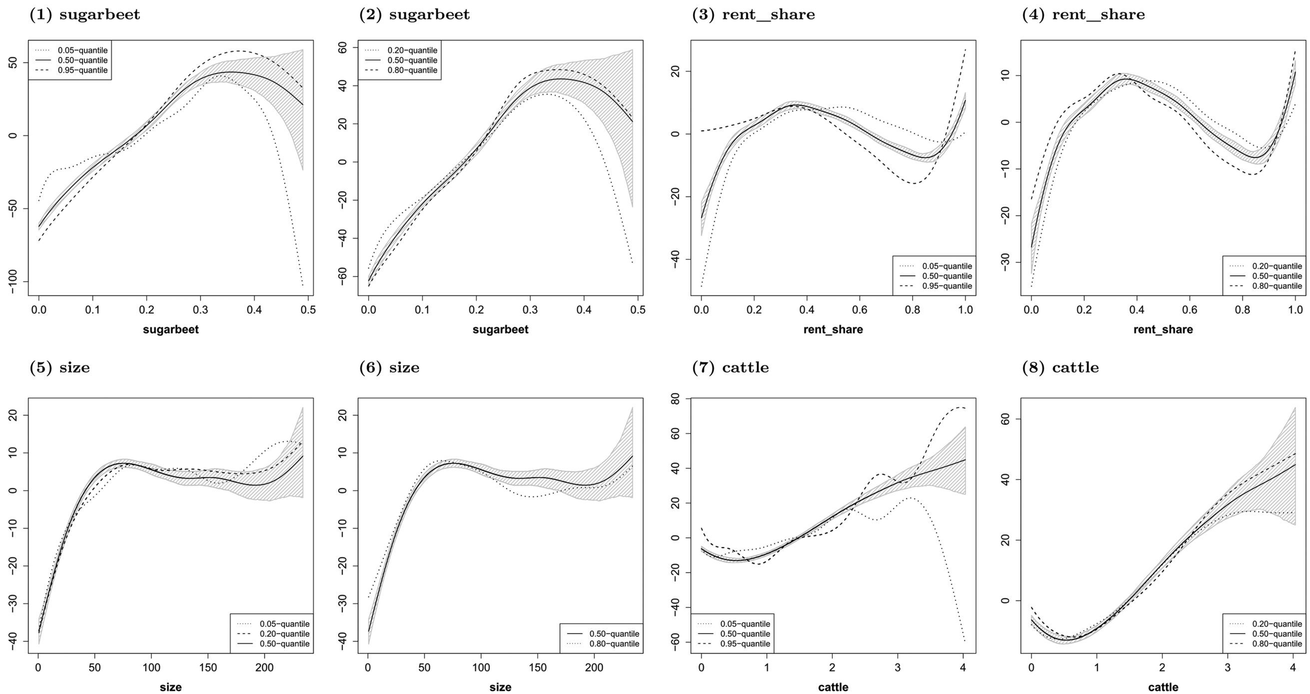

Fig. 1. Semi-parametric covariate effects. For all panels, the solid line indicates the estimated effect for the median (0.50-quantile) of the corresponding covariate combined with its 95 per cent credible interval, as indicated by the shaded area coloured in grey. In addition, the estimated effects for the other quantiles are super-imposed as indicated by the dashed and dotted lines.
Source: Research Data Centres of the Federal Statistical Office and the statistical offices of the federal states, agricultural census 2010, Regionaldatenbank (2010); authors' own calculations.

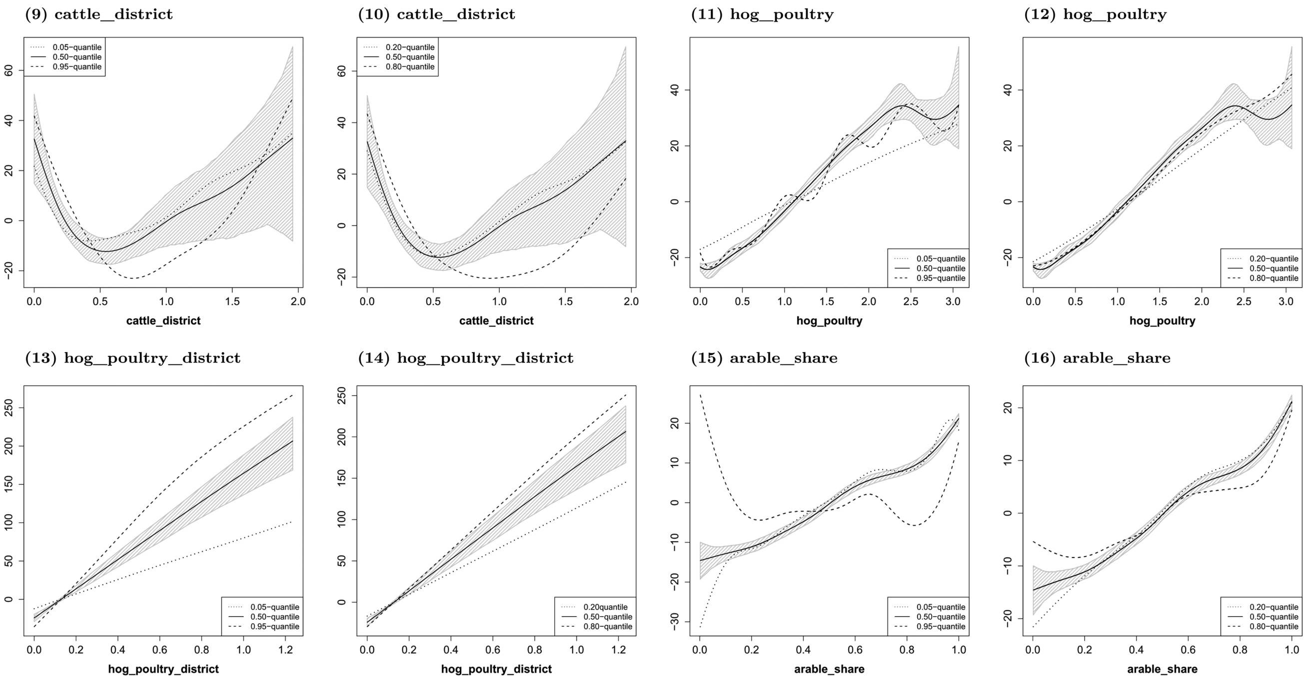

Fig. 1. *Continued*

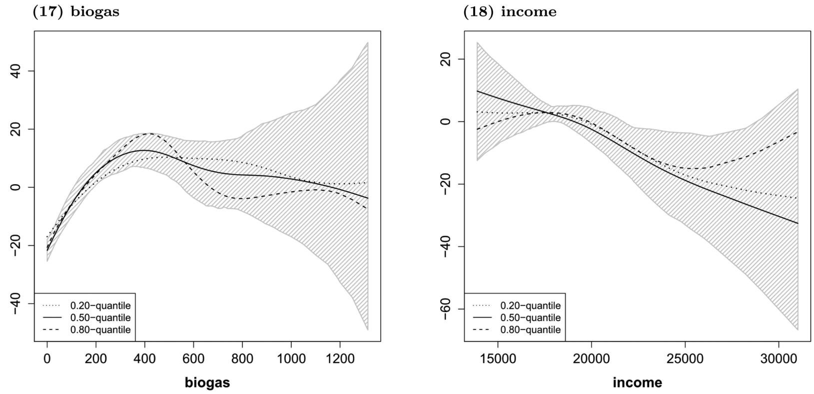

Fig. 1. *Continued*

two non-central quantiles are plotted together with the estimated effect of the median regression. To further facilitate visual inspection, we only plot the 95 per cent credible interval for $\tau = 0.50$.[16]

the basis function for each continuous covariate is constructed based on m_j coefficients $\gamma_{j,k}$ that scale the basis functions $B_{j,k}$ accordingly to fit the data, they have no economic interpretation themselves (e.g. elasticity). Consequently, we do not present the estimated coefficients themselves, as the estimated effect of a continuous covariate can only be represented graphically in a meaningful way. Instead, we graphically investigate the non-linear effects that allows for visually differentiating the way covariates act on the response for different values of τ.

16 One potential issue that is attached to the estimation of the parametric and semi-parametric effects is the lack of a variable that accounts for the effects of subsidies on rental rates from federal farm programmes. According to Kirwan (2009), subsidy measures, if unavailable to the analyst, may result in biased estimates resulting from measurement errors or omitted variables, which according to Wooldridge (2010) are two possible sources of endogeneity. In principle, the Bayesian approach would be to augment the model by simultaneous equations for the endogeneity issues. However, this creates additional complications such as ill-conditioning and additional identification assumptions. In general, one way of dealing with the problem of endogeneity in Bayesian semi-parametric quantile regression would be to follow the proposal of Sobotka *et al.* (2013), who introduce a semi-parametric two-stage instrumental variable approach for expectile regression, which is closely related to semi-parametric quantile regression. However, as pointed out by Sobotka *et al.* (2013), ignoring the additional source of variability in the confidence and credible intervals for the second stage estimates, which is introduced from the residuals calculated in the first stage, would lead to poor coverage probabilities and hence would have negative effects on inference. To take this additional variability of the first stage estimate into account, Sobotka *et al.* (2013) propose a parametric bootstrap. However, the fact that a single estimation run of our model takes several hours (depending on the quantile and the number of iterations, computing times range between 4 and 7h), the parametric bootstrap with several thousands of replications is computationally infeasible. Since none of the aforementioned approaches are applicable in our case, we approach the problem from an omitted variable perspective. Removing unobserved spatial patterns is an important task, especially when spatial unobservables are considered to be potential sources of endogeneity (Basile *et al.*, 2014). This issue can be addressed within the semi-parametric framework of geoadditive quantile regression models by incorporating the discrete spatial location as an additional covariate in the model, which acts as a surrogate for spatially correlated covariates that are not included in the model (Basile *et al.*, 2014). Consequently, adverse effects on inference resulting from endogeneity problems that arise from omitted variables are alleviated in our model, since we directly control for these effects (Basile *et al.*, 2014).

In line with Habermann and Ernst (2010) and Breustedt and Habermann (2011), Figure 1 shows that field crops with high profit margins, such as sugar beets, have a positive effect on rental rates. However, panels (1) and (2) also indicate that the effect of sugar beets on rental rates is non-constant across the conditional response distribution, with both the magnitude and the functional form changing across quantiles. While the effect initially increases across all quantiles, it starts to decrease again for the 5 and 20 per cent quantiles if more than 30 per cent of the farmland is used for cultivation. In contrast, there seems to be a threshold effect for the other quantiles, as the effects level off above a share of 30 per cent for the 50, 80 and 95 per cent quantiles. We further investigate the decrease of sugar beets for the 5 and 20 per cent quantiles by forming the first derivative of the estimated effects and find that the sharp decrease is indeed significant at a nominal level of 5 per cent. For the 95 per cent quantile, the effect of sugar beets shows the highest influence in terms of the magnitude of the estimated effect.

Concerning the effect of farm size on rental rates, previous studies have hypothesised a size-related threshold effect, meaning that the effect of farm size on rental rates may flatten or even change after a given threshold (see Margarian (2008) or Habermann and Ernst (2010)). While it is more difficult to verify such an effect in linear regression models, the semi-parametric nature of geoadditive regression models makes them more suitable for an empirical investigation. In fact, the results presented in panels (5) and (6) of Figure 1 seem to confirm the theoretical conjectures. Initially, the effect of farm size for small- and medium-sized farms is positive, as rental rates increase with growing farm size up until a size of approximately 70 hectares. One explanation for this effect might be that larger farms are more likely to realise economies of scale and are therefore able to pay higher rents. After this threshold, rents decrease for farm sizes between 70 and 180 hectares, where the effect is more pronounced for medium and expensive rents. The decreasing effect on rental rates might be attributed to the fact that, after a given threshold, farms are so large that they may have a comparatively higher market power which allows them to keep rental rates low (Habermann and Ernst, 2010). 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 regarding a farm size greater than 180 hectares cannot be made.

Panels (3) and (4) of Figure 1 further highlight the benefits of semi-parametric quantile regression models as they provide additional insight into the price formation of rental rates that are not captured by parametric mean regression models. Relating our results to the existing literature, Margarian (2008) and Habermann and Ernst (2010) note that the share of rented agricultural land reflects farmers' willingness to pay for rented farmland, with an expected negative sign for the estimated effect. Accordingly, farmers with a low share of rented acreage, who are mainly operating their own land, are able to pay their marginal revenue if they decide to rent additional acreage and are consequently willing to pay higher rental rates. In contrast, farmers who are mainly operating on rented farmland show a lower willingness to pay, since they have to cover additional fixed costs that are associated with rented farmland (Habermann and Ernst, 2010). Indeed, Fuchs (2002) or Habermann and Ernst (2010) find that average rents linearly decrease with an increasing share of rented agricultural land. However, modelling the effects semi-parametrically and allowing them to change across the response distribution allows for a more detailed analysis. Accordingly, panels (3) and (4) show that the functional form of the estimated effects is clearly non-linear, with some degree of heterogeneity of the estimated effects across quantiles also being present. Starting with the mid-range quantiles, panel (4) shows that the share of rented agricultural land initially increases rental rates, then leads to a decrease before the effect increases rental rates again, with this effect becoming more pronounced along increasing quantiles. The increasing effect above a share of 80 per cent is also supported by the narrow credible interval. The decreasing effect, as reported in the previous studies, can therefore only be confirmed for a share of rented agricultural land between 40 and 80 per cent. One explanation for the general finding in the literature of an overall negative relationship between rented agricultural land and farmland rental rates might be that, when averaging across quantiles, the decreasing effect for shares between 40 and 80 per cent might dominate, so that an overall negative relationship results. Turning to the tail quantiles, panel (3) shows that, similar to the mid-range quantiles, the effect up until a share of 40 per cent is increasing. This seems to be especially true for farmers at the lower end of the distribution, since the increase is steepest for the 5 per cent quantile, while farmers at the 95 per cent quantile seem to be indifferent until reaching a rented share of 40 per cent. Beyond a share of 40 per cent, the effect decreases, with the effect being most pronounced for the 95 per cent quantile. Consequently, farmers who increasingly operate on rented farmland show a lower willingness to pay for additional acreage, as the fixed costs might exceed the revenues received from rented farmland. While the effect for the 5 per cent quantile seems to level out for a share of rented farmland exceeding 80 per cent, indicating a continuing decrease in the willingness to pay at the lower end of the distribution, there is again a sharp increase of the effect in the upper tail of the distribution. This rise in rental rates might be attributed to the increasing competition among farmers for rented land, especially at the 95 per cent quantile that shows the steepest ascent.

Panels (15) and (16) show that a higher share of rented arable land leads to an increase in rental rates. Since the share of arable land has been included as a proxy to account for quality differences between arable land and pastureland, as well as for the varying levels of productivity of the rented land, panels (15) and (16) show that the higher the share of rented arable land, the higher the productive use of the rented land and the higher is the rent that farmers are willing to pay. Again, panels (15) and (16) illustrate that parametric OLS results may provide an inaccurate representation of the estimated effects. Even though previous studies find a significantly positive effect of the share of rented arable land on rental rates (Habermann and Ernst, 2010; Habermann and Breustedt, 2011), the results of the quantile regression indicate that the estimated effects are not constant, but change across the conditional distribution. For the 5, 20, 50 and 80 per cent quantiles, the effect of the share of rented arable land is increasing. For very expensive rental rates, however, the effect initially decreases rental rates up until a share of 20 per cent, then remains virtually constant before it starts to increase again after a share of 80 per cent. Consequently, for the 95 per cent quantile, quality differences between arable land and pastureland are only relevant for either very low or high shares of rented arable land, whereas they have a constant effect for shares between 20 and 80 per cent.

From panel (17) of Figure 1 it also follows that the presence of biogas plants increases rental rates across all estimated quantiles, up to a plant capacity of approximately 400 kWh. As a result of the increased uncertainty attached to the estimation, which is reflected in the wide credible interval, reliable statements beyond this capacity cannot be made. Also note that the estimated effects are comparatively similar for the 50 per cent and the 80 per cent quantiles, where the initial increase is more pronounced for these quantiles as indicated by the slope and the magnitude of the estimated effect as compared with the 20 per cent quantile. Furthermore, and in line with Habermann and Ernst (2010), panel (18) shows that rental rates are negatively associated with opportunity costs of farming, as proxied by the average per capita income at the district level. Consequently, in the absence of adequate alternatives, farmers remain in the agricultural sector and are willing to pay higher rents (Habermann and Ernst, 2010).

Since there is little a priori knowledge of whether heterogeneous effects exist, quantile regression can be used in order to investigate whether average marginal effects yield an appropriate description of rental rates, or whether these effects differ across the conditional distribution. Besides allowing the covariates to influence rental rates in a non-linear way, the regression results presented in this section supplement the estimates from traditional models by allowing the marginal effects to vary with conditional quantiles (i.e. expectations) so that a richer description of the data generating process underlying rental rates can be obtained. According to Figure 1, there are some noticeable instances of heterogeneity across quantiles. This is particularly true for the estimated effects of livestock densities, especially for cattle density both at the farm and the county level (panels (7)–(10)), the share of rented agricultural land (panels (3) and (4)) as well as the share of rented arable land (panels (15) and (16)). Even though each of these effects makes an interesting case for highlighting the ability of quantile regression to uncover heterogeneity with respect to future expectations of farmers, we place particular emphasis on the discussion of the effect of livestock density in the following, since this effect has previously been discussed in the literature when it comes to assessing the impact of agricultural policies on farmland rents (see, e.g. Breustedt and Habermann, 2011).

Similar to Breustedt and Habermann (2011) or Habermann and Breustedt (2011), we find that livestock densities have a major impact on rental rates. However, in addition to previous studies that assume a linear relationship using mean regression models, estimating a set of quantile regressions reveals that heterogeneities exist in the functional forms and in the magnitudes of the estimate effects. Accordingly, panels (7)–(10) of Figure 1 show a U-shaped effect for the cattle density at the farm level across all quantiles, where this effect becomes more pronounced for the cattle density at the district level as we move from low to high quantiles. A cattle density of up to approximately 0.5 initially decreases farmland rental rates, which then begin 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 rents on the district level, with the estimated effects being highly significant across all quantiles. One possible explanation with respect to the positive influence of livestock densities on rental rates which is commonly provided by the literature is that environmental regulation restricts the amount of manure farmers are allowed to discharge on their land (Habermann and Ernst, 2010). Farmers with a livestock density that exceeds a certain threshold either have to rent additional farmland or have to register a trade. To avoid tax disadvantages, farmers prefer to rent additional farmland in order to reduce the livestock density.

Regarding the influence of farming premiums on livestock density, Fuchs (2002) notes that since 1992, an increasing share of the agricultural income of farmers has come from investment aids for livestock farming, where between 2000 and 2004, livestock farming premiums amounted to EUR 564.6 million in Germany (Breustedt and Habermann, 2011). Consequently, through a stabilising effect on incomes of livestock farmers, an indirect effect of investment aids for livestock farming on rental rates is to be expected. In particular, even though the literature has generally found a positive relationship between livestock densities and rental rates, Breustedt and Habermann (2011) hypothesise that the effect of livestock density may also be negative, due to the adverse effects that investment aids have on competition for farmland. As indicated by panels (9) and (10), an increase in the cattle density at the district level leads to an increase in rental rates after exceeding a certain number of animal units per hectare. Consequently, despite the fact that investment aids are supposed to stabilise the incomes of livestock farmers, they also increase rental rates for all tenants in that district in the long run due to a higher competition for farmland among livestock and arable farmers. While the traditional mean regression models are not able to identify the differential effects, panel (7) allows for a more nuanced view and seems to support the theoretical considerations of Breustedt and Habermann (2011). Since some farmers (approximated by the 5 per cent quantile) seem to anticipate the increase in the overall level of rental rates in their district, their outlook for future prices results in a lower willingness to pay, while others (as indicated by the 95 per cent quantile) continue to be more optimistic about the future and are willing to pay higher rents due to the effect that investment aids have on their incomes. These considerations are of potential interest for policymakers deciding on investment aid schemes for livestock farming, since the result suggests taking the distributional effects on farmers' expectations into account.

We now turn to the discussion of the parametric effects that are summarised in Table 4, showing posterior means, standard deviations and 95 per cent credible intervals.

Table 4. Summary of estimated posterior of parametric effects

	Mean	Std. Dev.	2.5 per cent	97.5 per cent
	$\tau = 0.05$			
Intercept	169.9840	2.5479	164.7240	175.0510
East-Dummy	−17.1949	8.0001	−32.4506	−1.9856
Part-time	−9.0174	0.8268	−10.6230	−7.3732
Full-time	−4.5236	0.7331	−6.0365	−3.0514
	$\tau = 0.20$			
Intercept	212.3290	11.5614	194.1610	229.0150
Part-time	−8.9851	0.8897	−10.6209	−7.2347
Full-time	−3.3490	0.8085	−4.8647	−1.7232
	$\tau = 0.50$			
Intercept	251.0470	7.7633	237.7560	265.6460
Part-time	−12.4006	1.0310	−14.3351	−10.3478
Full-time	−5.4759	0.9306	−7.2708	−3.6885
Non-organic_farming	−11.4434	1.1887	−13.8433	−9.1778
No_Farm_Successor	−4.7794	0.7131	−6.2617	−3.4215
Farm_Successor_Unsettled	−2.6106	0.5645	−3.7622	−1.4928
	$\tau = 0.80$			
Intercept	378.3920	18.7388	345.8290	408.3470
Part-time	−17.6288	1.1735	−20.1139	−15.4582
Full-time	−6.3772	1.0873	−8.5655	−4.2758
Non-organic_farming	−18.7771	1.3182	−21.3600	−16.2323
No_Farm_Successor	−6.1095	0.8476	−7.8888	−4.4890
Farm_Successor_Unsettled	−4.2972	0.6907	−5.6916	−2.9853
	$\tau = 0.95$			
Intercept	446.6020	3.7128	438.3510	452.9380
Part-time	−29.8321	1.4594	−32.7479	−26.8553
Full-time	−11.5701	1.3398	−14.1691	−8.8309

Source: Research Data Centres of the Federal Statistical Office and the statistical offices of the federal states, agricultural census 2010, Regionaldatenbank (2010), authors' own calculations.

Previous studies, e.g. Habermann and Ernst (2010) and Habermann and Breustedt (2011), have shown that rental rates in East Germany are on average lower compared with West German rental rates. This is supported by the estimated effects for the 5 per cent quantile presented in Table 4. Also note that organic farmers pay higher rents compared with farmers who operate their farms conventionally. Furthermore, differences exist between farmers and the rental rates they are paying depending on whether they operate on a full-time or part-time basis. Similar to Habermann and Breustedt (2011), we find that full-time farmers pay higher rents compared with their part-time counterparts. This difference between part-time and full-time farmers may be attributed to several factors. In order to earn a living, full-time farmers have to have a high production volume and a high production intensity. Consequently, full-time farmers operate businesses that are on average larger than those of part-time farmers, with an average farm size of about 61 hectares, which is the approximate size until which rental rates increase with farm size (compare panels (5) and (6) of Figure 1). Another reason for the difference might be that the proportion of full-time farmers is higher in those districts where the principle income of the farmer is associated with livestock farming and hence in districts where rental rates are high. 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.

Our results with respect to differences between full-time and part-time farmers are also supported by the findings presented in Kostov (2010). Using VCM, Kostov (2010) investigates the influence of buyer and personal relationship characteristics on mean agricultural land prices. By allowing the hedonic attributes to vary with buyer and personal relationship characteristics, Kostov (2010) shows that these characteristics exert a non-uniform effect on the implicit prices of agricultural land characteristics. With respect to differences between full-time and part-time farmers, Kostov (2010) finds that part-time farmers pay a higher price for small acreages, while paying less for larger acreages compared with their full-time counterparts, which is in line with our results. Similar to our line of reasoning, Kostov (2010) offers the explanation that, since farming represents their main source of income, full-time farmers show a higher willingness to pay, especially for larger acreages, whereas part-time farmers are limited in the number of acres they can farm due to time restrictions and consequently show a lower willingness to pay for larger acreages.

4.2 Spatial effects

Our analysis of the spatial effects is motivated from a statistical point of view. In contrast to spatial econometrics, where spatial lag and spatial error models are commonly used, we account for spatial correlation and omitted covariates by adding a spatial term $f_{\mathrm{geo},\tau}$ to the additive predictor $\eta_{i,\tau}$. Consequently, we are mainly interested in investigating spatial patterns that emerge from spatial heterogeneities which remain unexplained after taking covariates into account. Plotting the estimated effects of $f_{\mathrm{geo},\tau}$ 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 insight into the data that were not previously considered. Significance maps shown in Figure 2 further enhance the detection of

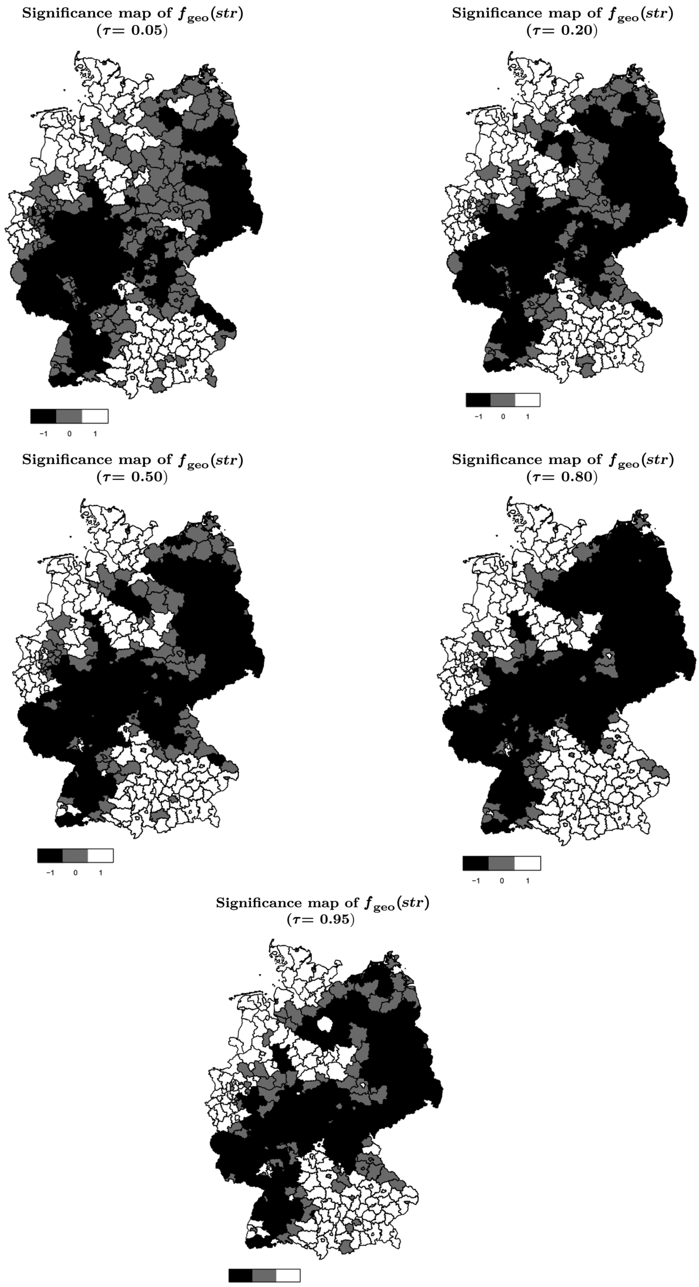

Fig. 2. Significance of the structured spatial effect $f_{\mathrm{geo}}(str)$ based on posterior probabilities using a nominal level of 95 per cent. −1: significantly negative; 0: non-significant; 1: significantly positive. *Source*: Research Data Centres of the Federal Statistical Office and the statistical offices of the federal states, agricultural census 2010, Regionaldatenbank (2010), authors' own calculations.

spatial patterns by classifying the estimated spatial effect into one of three categories: the spatial effect is classified as insignificant at the 95 per cent 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. Figure 2 presents the estimation results for the structured spatial effects only. Moreover, since the estimated structured spatial effects for the 20, 50 and 80 per cent quantiles are generally very similar, we restrict our discussion of the effects to the 5, 50 and 95 per cent quantiles.

Figure 2 shows that rental rates are considerably lower than what can be explained with covariates in the southwest, as well as in large parts of East and Central Germany across all quantiles (black districts), where this effect becomes more pronounced along increasing quantiles. 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 manages agricultural 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 2 shows, the dummy variable leaves the structural differences between East and West German rental markets unexplained, at least for the 20, 50, 80 and 95 per cent quantiles, since boosting has not selected it during any of the iterations of these quantiles, so that differences between East and West German rental markets seem to be captured by the spatial effect terms. Consequently, and in line with Habermann and Ernst (2010) and Habermann and Breustedt (2011), additional covariates other than the dummy variable have to be included in the model for these quantiles in order to account for the differences between East and West German rental markets.

Figure 2 also shows that the covariates are not well-suited to explain expensive rental rates, as the covariates leave heterogeneities unexplained in many districts of Germany across all quantiles (white districts), especially in the South, as well as in western and northern regions of Germany. In accordance with the literature, Figure 2 shows that rental rates for farmland are higher in districts where livestock densities are high. Rental rates are also more expensive in districts in which high livestock densities and high biogas plant densities meet, such as in the southern region of Germany. Also note that the estimated effects are similar with respect to high unexplained rents across all quantiles.

5. Conclusion

In this article, we model and analyse conditional quantiles of farmland rental rates semi-parametrically using Bayesian geoadditive quantile regression. The flexibility of this model class alleviates the problems of functional form misspecifications and allows us to present a more detailed analysis of farmland rental rates. In particular, by allowing different quantiles of the conditional distribution to depend differently on covariates, our study provides additional insight into the data generation process of German rental rates, as the determinants can be separately identified for each quantile. Moreover, since there is usually little a-priori knowledge of whether non-constant covariate effects exist or which of the covariate effects are affected, modelling a dense grid of quantiles allows us to exploit the informational content implicit in farmland rental rates and to reveal heterogeneities of the marginal effects across the conditional response distribution. Our results also stress the importance of making use of semi-parametric regression models, as several covariates influence farmland rental rates in an explicit non-linear way. By explicitly modelling and plotting the spatial effects, we account for spatial autocorrelation and are also 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. The results of our study are of potential interest for farmers and academics alike, since hedonic pricing studies in the agricultural economics literature so far have primarily been concerned with the linear 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 the contract can now be better tailored to the operational characteristics of the farmer. Moreover, since it allows taking into account heterogeneity of the estimated effects, the application of quantile regression is also of potential interest for policy makers in order to evaluate the differential effects that policy measures may have across farmers. Additionally, if desired, the identification of the driving forces behind expensive rents may also serve to assist policy makers in taking corrective action by establishing a ceiling on rental rates in order to prevent an excessive increase of rental rates in the future.

Although the uncovering of heterogeneity with respect to the expectations of economic agents extends the existing literature on modelling farmland rental rates, there are several ways to further develop our analysis in future research. Following Dunford, Marti and Mittelhammer (1985), we consider the explicit analysis of heterogeneity and the economic inferences thereof as an interesting extension. Although quantile regression provides a means to reveal heterogeneities across the conditional distribution, it does not allow us to predict or to provide economic implications in order to gain thorough insight into the formation process of future expectations. Consequently, a promising avenue of future research would be to focus on identifying relevant variables that allows the researcher to establish a link between changes in future expectations and economic factors (such as rates of inflation or interest rates). As far as the estimated marginal effects are concerned, we allow them to change as a function of quantiles. However, formally testing for the presence of heterogeneity across different quantiles allows for a more formal statement of whether the marginal effects are constant or whether substantial differences exist for different points in the conditional distribution and poses another interesting extension (for an application using linear quantile regression models, see e.g. Arias, Hallock and Sosa-Escudero, (2002)). Moreover, it would also 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, it would be interesting to allow the effect of one or more covariates to vary across space. When modelling the covariate effects, we have implicitly assumed that the way in which the covariates act on the response is homogeneous across all districts. However, the effect of one or more covariates may vary from district to district. Geographically weighted regression models allow regression coefficients to differ from their global values at a regional level and assume that the spatial heterogeneity is explained by the spatially varying regression coefficients. This is especially interesting for studying the differences in the rental rate formation between East and West German rental markets, as the spatially varying coefficients allow for a more regionally differenced view of the covariate effects on East and West German rental rates. Similarly, models with non-spatially varying coefficients pose another interesting extension of this study. While the models used in this study are purely main effect models, it might also be of interest to allow one or more covariates to interact with a binary or continuous variable, which allows for an even more nuanced view of the data generating process underlying rental rates (see Kostov (2010) for an empirical investigation of such VCM on agricultural land prices). If interest does not only lie with farmland rental rates themselves, but also with higher moments, such as their (spatial) variability, generalised additive models for location, scale and shape (GAMLSS) also provide an interesting extension of the analysis, as they allow for the modelling of all parameters of a parametric response distribution as additive functions of covariates.

Acknowledgements

The authors thank two anonymous referees and the editors of the *European Review of Agricultural Economics* for their helpful comments and suggestions. They also thank Dr. Alexander Vogel from the FDZ for his outstanding assistance. The authors gratefully acknowledge financial support from the German Research Foundation (DFG).

References

Adler, D., Kneib, T., Lang, S., Umlauf, N. and Zeileis, A. (2015). BayesXsrc: R Package Distribution of the BayesX C++, R package version 3.0-0. http://CRAN.R-project.org/package=BayesXsrc. Access 11 March 2015.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Studies in Operational Regional Science. Dordrecht and Boston: Kluwer Academic Publishers.

Anselin, L. and Lozano-Gracia, N. (2009). Errors in variables and spatial effects in hedonic house price models of ambient air quality. In: G. Arbia and B. H. Baltagi (eds), *Spatial Econometrics: Methods and Applications*, Studies in Empirical Economics. Heidelberg: Physica-Verlag, 5–34.

Arias, O., Hallock, K. and Sosa-Escudero, W. (2002). Individual heterogeneity in the returns to schooling: instrumental variables quantile regression using twins data. In: B. Fitzenberger, R. Koenker and J. Machado (eds), *Economic Applications of Quantile Regression*. Studies in Empirical Economics. Heidelberg: Physica-Verlag, 7–40.

Barnes, M. L. and Hughes, A. T. W. (2002). A quantile regression analysis of the cross section of stock market returns. Working Paper, Federal Reserve Bank of Boston 02-2.

Basile, R., Durbán, M., Mínguez, R., Montero, J. M. and Mur, J. (2014). Modeling regional economic dynamics: spatial dependence, spatial heterogeneity and nonlinearities. *Journal of Economic Dynamics and Control* 48: 229–245.

Basile, R. and Gress, B. (2005). Semi-parametric spatial auto-covariance models of regional growth in Europe. *Region et Developpement* 21: 93–118.

Bekkerman, A., Brester, G. W. and McDonald, T. J. (2013). A semiparametric approach to analyzing differentiated agricultural products. *Journal of Agricultural and Applied Economics* 45: 79–94.

Belitz, C., Brezger, A., Klein, N., Kneib, T., Lang, S. and Umlauf, N. (2015). BayesX: Software for Bayesian Inference in Structured Additive Regression Models. Version 3.0.0. Available from http://www.bayesx.org. Access 11 March 2015.

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.

Borchers, A., Ifft, J. and Kuethe, T. (2014). Linking the price of agricultural land to use values and amenities. *American Journal of Agricultural Economics* 96: 1307–1320.

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.

Ciaian, P. and Kancs, D. (2012). The capitalization of area payments into farmland rents: micro evidence from the new EU member states. *Canadian Journal of Agricultural Economics* 60: 517–540.

Ciaian, P., Kancs, D. and Paloma, S. G. (2010a). Distributional effects of CAP subsidies: micro evidence from the EU. EERI Research Paper Series 5, 2010.

Ciaian, P., Kancs, D. and Swinnen, J. F. M. (2010b). EU land markets and the common agricultural policy. Working Paper, Centre for European Policy Studies (CEPS).

Drescher, K. and McNamara, K. (2000). Analysis of German agricultural land prices. In: P. Tillack and E. Schulze (eds), *Land Ownership, Land Markets and Their Influence on the Efficiency of Agricultural Production in Central and Eastern Europe*. Kiel: Wissenschaftsverlag Vauk KG, 210–228.

Dunford, R. W., Marti, C. E. and Mittelhammer, R. C. (1985). A case study of rural land prices at the urban fringe including subjective buyer expectations. *Land Economics* 61: 10–16.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* 11: 89–102.

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*, 1st edn. Heidelberg, Berlin: Springer-Verlag.

FDZ. (2010). Landwirtschaftszählung 2010. *Forschungsdatenzentren (FDZ) der Statistischen Ämter des Bundes und der Länder*. Research Data Centres of the Federal Statistical Office and the statistical offices of the federal states.

Feichtinger, P. and Salhofer, K. (2013). What do we know about the influence of agricultural support on agricultural land prices? *German Journal of Agricultural Economics* 62: 71–85.

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.

Fitzenberger, B., Koenker, R. and Machado, J. A. F. (2002). *Economic Applications of Quantile Regression*. Heidelberg: Physica-Verlag.

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.

Habermann, H. and Breustedt, G. (2011). Impact of biogas production on farmland rental rates in Germany. *German Journal of Agricultural Economics* 60: 85–100.

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

Harris, D. G. and Nehring, R. F. (1976). Impact of farm size on the bidding potential for agricultural land. *American Journal of Agricultural Economics* 58: 161–169.

Hennig, S., Breustedt, G. and Latacz-Lohmann, U. (2014). The impact of payment entitlements on arable land prices and rental rates in Schleswig-Holstein. *German Journal of Agricultural Economics* 63: 219–239.

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.

Kauermann, G., Haupt, H. and Kaufmann, N. (2012). A hitchhiker's view on spatial statistics and spatial econometrics for lattice data. *Statistical Modelling* 12: 419–440.

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, 20–21 November, 1–15.

Kilian, S., Antón, J., Salhofer, K. and Röder, N. (2012). Impacts of 2003 CAP reform on land rental prices and capitalization. *Land Use Policy* 29: 789–797.

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

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* 48: 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. (2009a). A spatial quantile regression hedonic model of agricultural land prices. *Spatial Economic Analysis* 4: 53–72.

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

Kostov, P. (2010). Do buyers' characteristics and personal relationships affect agricultural land prices? *Land Economics* 86: 48–65.

Kostov, P. (2013). Choosing the right spatial weighting matrix in a quantile regression model. *ISRN Economics* 2013: 1–15.

Kostov, P. and Davidova, S. (2013). A quantile regression analysis of the effect of farmers' attitudes and perceptions on market participation. *Journal of Agricultural Economics* 64: 112–132.

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

Lee, W. F. and Rask, N. (1976). Inflation and crop profitability: how much can farmers pay for land? *American Journal of Agricultural Economics* 58: 984–990.

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.

LeSage, J. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Statistics: Textbooks and Monographs 196. Boca Raton, FL: CRC Press.

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

McMillen, D. P. (2003). Spatial autocorrelation or model misspecification? *International Regional Science Review* 26: 208–217.

McMillen, D. P. and Redfearn, C. L. (2010). Estimation and hypothesis testing for nonparametric hedonic house price functions. *Journal of Regional Science* 50: 712–733.

Melichar, E. (1979). Capital gains versus current income in the farming sector. *American Journal of Agricultural Economics* 61: 1085–1092.

Michalek, J. and Ciaian, P. (2014). Capitalization of the single payment scheme into land value: generalized propensity score evidence from the European Union. *Land Economics* 90: 260–289.

Mishra, A. K. and Moss, C. B. (2013). Modeling the effect of off-farm income on farmland values: a quantile regression approach. *Economic Modelling* 32: 361–368.

Möller, L., Henter, S. H., Kellermann, K., Röder, N., Sahrbacher, C. and Zirnbauer, M. (2010). Impact of the introduction of decoupled payments on functioning of the German land market: Country report of the EU tender study on the functioning of land markets in those EU member states influenced by measures applied under the Common Agricultura. Discussion Paper. Leibniz Institute of Agricultural Development in Central and Eastern Europe.

Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society* 55: 819–847.

Nickerson, C. J. and Zhang, W. (2014). Modeling the determinants of farmland values in the United States. Chapter 5 In: J.M. Duke and J. Wu (eds), *The Oxford Handbook of Land Economics*, Oxford: Oxford University Press, 111–137.

Patton, M., Kostov, P., McErlean, S. and Moss, J. (2008). Assessing the influence of direct payments on the rental value of agricultural land. *Food Policy* 33: 397–405.

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

Reinsel, R. D. and Reinsel, E. I. (1979). The economics of asset values and current income in farming. *American Journal of Agricultural Economics* 61: 1093–1097.

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. *The Journal of Political Economy* 82: 34–55.

Sobotka, F. and Kneib, T. (2012). Geoadditive expectile regression. *Computational Statistics & Data Analysis* 56: 755–767.

Sobotka, F., Radice, R., Marra, G. and Kneib, T. (2013). Estimating the relationship between women's education and fertility in Botswana by using an instrumental variable approach to semiparametric expectile regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 62: 25–45.

Statistisches Bundesamt (2010). Methodische Grundlagen der Landwirtschaftszählung 2010. *Statistisches Bundesamt* 3: 1–345.

Swinnen, J., Ciaian, P. and Kancs, D. (2008). Study on the Functioning of Land Markets in the EU Member States under the Influence of Measures applied under the Common Agricultural Policy. Unpublished Report to the European Commission, Centre for European Policy Studies, Brussels.

Uematsu, H., Khanal, A. R. and Mishra, A. K. (2013). The impact of natural amenity on farmland values: a quantile regression approach. *Land Use Policy* 33: 151–160.

Vranken, L. and van Herck, K. (2013). Direct payments and land rents: evidence from new member states. Centre for European Policy Studies Factor Markets Working Paper No. 62.

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

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge MA, USA: MIT Press.

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

Appendix A

A.1. Spatial lag and spatial error models

To account for spatial dependence in agricultural hedonic pricing studies, the general spatial model

$$\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \varepsilon \tag{A.1}$$

is most commonly used, where

$$\varepsilon = \lambda \boldsymbol{W}\varepsilon + \boldsymbol{u}$$

with the error term $\boldsymbol{u} \sim N(0, \sigma^2 \boldsymbol{I}_n)$ and $\boldsymbol{I}_n$ being the identity matrix. In Equation (A.1), $\boldsymbol{y}$ is a $n \times 1$ vector of the response variable, $\boldsymbol{X}$ is a $n \times p$ design matrix that carries the covariate information, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients, ρ and λ are the spatial lag and spatial error parameters that measure the strength of spatial spillover effects and $\boldsymbol{W}$ is a $n \times n$ spatial weight matrix for the spatial lag and spatial error part that reflects the spatial neighbourhood structure in the data, with $W_{ij} = 1$ if observations i and j are neighbours and $W_{ij} = 0$ otherwise (see Anselin (1988) and LeSage and Pace (2009) for further details). Depending on the spatial structure of the data, the model in Equation (A.1) can be reduced to the spatial lag or the spatial error model by setting $\lambda = 0$ or $\rho = 0$, respectively.

From Equation (A.1), it follows that for the spatial lag model $\boldsymbol{y} = \rho\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \varepsilon,$ the spatially lagged term $\rho\boldsymbol{W}\boldsymbol{y}$ accounts for the spatial association within the data and corresponds to a weighted average of neighbouring observations, where $\boldsymbol{W}_i\boldsymbol{y}$ is a linear combination of all y_j to which the i-th observation is a neighbour. Moreover, the spatial lag model allows for modelling spatial multiplier effects. This can be shown by rewriting the model in its reduced form

$$\boldsymbol{y} = (I - \rho\boldsymbol{W})^{-1}\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\varepsilon$$

where the inverse $(\boldsymbol{I} - \rho\boldsymbol{W})^{-1}$ can be expressed as

$$\boldsymbol{y} = (\boldsymbol{I} + \rho\boldsymbol{W} + \rho^2\boldsymbol{W}^2 + \cdots)\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\nu}$$

using the power expansion and with $\boldsymbol{\nu} = (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\varepsilon$ being a spatially correlated and heteroscedastic error term (Anselin and Lozano-Gracia, 2009). In this expression, the value of $\boldsymbol{y}$ at one location does not only depend on the characteristics in the same location ($\boldsymbol{X}$), but also on the values of neighbouring locations ($\rho\boldsymbol{W}\boldsymbol{X}$, $\rho^2\boldsymbol{W}^2\boldsymbol{X}, \ldots$), where the neighbouring effects decay as the distance increases due to the powering of the spatial autoregressive parameter ρ and the spatial weights matrix $\boldsymbol{W}$. With respect to motivating the spatial error model, a common reason for spatial correlation in cross-sectional regression studies is the presence of spatially correlated variables that are omitted in the model, which are then absorbed by the error term so that the covariance matrix

$$\mathrm{Cov}(\boldsymbol{y}|\boldsymbol{X}) = \mathrm{Cov}(\varepsilon) = \sigma_\varepsilon^2\boldsymbol{\Omega}$$

is no longer diagonal (Anselin and Lozano-Gracia, 2009). One way of addressing spatial correlation that results from omitted variables is through the introduction of spatial fixed effects to the model, i.e. including an intercept for each discrete spatial unit (LeSage and Pace, 2009). In such a regression framework, the constant is allowed to vary between spatial units, where the spatial fixed effects are expressed as a measure of the difference of the level or the mean of spatial unit j relative to the reference group. Such an approach is reasonable if the omitted effects can unambiguously be associated with administrative districts, such that the unobserved heterogeneity is specific to each spatial unit. However, in the case of neighbourhood effects, where the omitted variables do not follow the spatial structure defined by the administrative boundaries, such an approach is suspect to error (Anselin and Lozano-Gracia, 2009). Instead, a common specification in spatial econometrics is to allow the effects of the omitted variables to spill over through the spatial units by assuming a spatial autoregressive structure for the error term

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon \quad \text{with} \quad \varepsilon = \lambda \boldsymbol{W}\varepsilon + \boldsymbol{u}$$

From the spatial error model, it follows that the error terms of neighbouring regions are spatially correlated, where the value of ε_i in one region is dependent on the weighted average of the neighbouring values $\boldsymbol{\varepsilon}_{-i}$. Compared with the spatial lag model, the spatial error model implicitly accounts for the spatial dimension in the data by imposing a certain covariance structure on the error term (Kauermann, Haupt and Kaufmann, 2012).

In contrast, geoadditive regression models are affiliated with spatial statistics, where the spatial process is additively included as an additional covariate in the model which allows for gaining additional insight into the spatial structure of the data. Similar to spatial error models, models in spatial statistics can be motivated from the point of view of modelling unobserved heterogeneity, i.e. modelling the structure in the data that remains unexplained after all available covariates have been included in the model. However, they differ from spatial econometric models with respect to the way that the spatial structure is treated and modelled. While the main interest of models in spatial econometrics lies with measuring the strength of the spatial dependence and spillover effects, either in the response variable $\boldsymbol{y}$, in the error term $\boldsymbol{\varepsilon}$, or in both, models in spatial statistics explicitly model and plot the unobserved heterogeneity, so that spatial patterns that would otherwise go unnoticed can be detected (Kauermann, Haupt and Kaufmann, 2012). In their most basic form, spatial statistical models can be represented as follows

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon$$

where the error term $\varepsilon = \boldsymbol{s} + \boldsymbol{e}$ can be additively decomposed into a spatial effects term $\boldsymbol{s} \sim N(0, \sigma_s^2\boldsymbol{\Omega})$ and into uncorrelated residuals $\boldsymbol{e} \sim N(0, \sigma_e^2\boldsymbol{I}_n)$. The covariance structure is then given by[17]

$$\text{Cov}(\boldsymbol{y}|\boldsymbol{X}) = \text{Cov}(\boldsymbol{\varepsilon}) = \sigma_s^2\boldsymbol{\Omega} + \sigma_e^2\boldsymbol{I}_n$$

Since the covariance structure of the error term can be additively decomposed into two components, and since spatial statistical models can be cast in the framework of mixed models, the spatial effects term $\boldsymbol{s}$ can be estimated using mixed model procedures and has the interpretation of a random effects term which accounts for unobserved heterogeneity and spatial correlation that result from spatially correlated variables that are omitted from the model. Consequently, the explicit modelling, as well as the visual presentation of the estimated spatial effects within the framework of spatial statistical models, poses an interesting extension when compared with spatial econometric models that treat the spatial structure in the data as a nuisance, holding little interest of its own. Comparing the spatial effects of the geoadditive model with the general spatial model in

17 See (Kauermann, Haupt and Kaufmann, 2012) for details on the exact structure of $\boldsymbol{\Omega}$.

Equation (A.1) shows that the unstructured spatial effect of the geodadditive re-

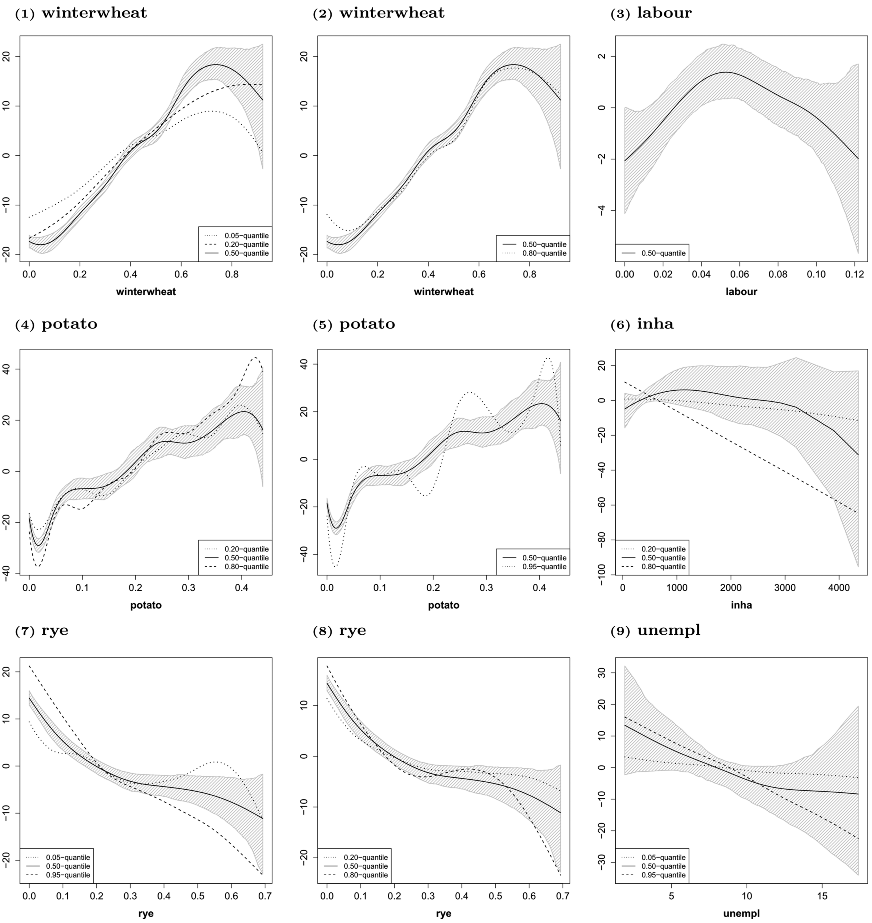

Fig. A1. Semi-parametric covariate effects. For all panels, the solid line indicates the estimated effect for the median (0.50-quantile) of the corresponding covariate combined with its 95 per cent credible interval, as indicated by the shaded area coloured in grey. In addition, the estimated effects for the other quantiles are super-imposed as indicated by the dashed and dotted lines.
Source: Research Data Centres of the Federal Statistical Office and the statistical offices of the federal states, agricultural census 2010, Regionaldatenbank (2010), authors' own calculations.

gression model is similar to the spatial error model in the case of spatial fixed effects, whereas the structured spatial effect generally resembles the spatial lag model. However, note the differences between the models, both with respect to the dimension of the matrices that capture the neighbourhood structure, with