European Review of Agricultural Economics Vol **49** **(3)** (2022) pp. 696–718
doi:https://doi.org/10.1093/erae/jbab052
Advance Access Publication 10 December 2021

The distribution of the rent–price relationship of agricultural land in Germany

Henning Schaak[†,*] **and Oliver Musshoff**[‡]

[†]*Department of Economics, Swedish University of Agricultural Sciences, Uppsala, Sweden;* [‡]*Department of Agricultural Economics and Rural Development, University of Goettingen, Germany*

Received September 2020; final version accepted November 2021

Abstract

This paper studies the profitability of investments in agricultural land, using the rent–price ratio (RPR) as a profitability measure. In order to allow for district-level heterogeneity, the full conditional distribution of the RPR is modelled using a generalised additive model for location, shape and scale. The analysis is based on data from Lower Saxony, Germany. The profitability of investments in land varies between and within districts. The variation can be explained by differences in the farming structure, the production programme and economic indicators. Further, differences in the distribution of the RPR between arable land and grassland are found.

Keywords: agricultural land market, GAMLSS, heterogeneity, spatial statistics

JEL classification: C21, C46, Q15

1. Introduction

Land is the most important factor for agricultural production. Price developments for agricultural land and their determinants are therefore an important research topic in agricultural economics. A sharp increase in land price in Germany, as well as other European countries (Eurostat, 2019) has motivated intensive research activity in this area. Latruffe and Le Mouël (2009) report that land prices were positively affected by agricultural support policy instruments. Likewise, Hennig, Breustedt and Latacz-Lohmann (2014) find a positive effect of payment entitlements on land rental prices. The effect of biogas subsidies on rental rates is studied by Habermann and Breustedt (2011) and Hennig and Latacz-Lohmann (2017). Feichtinger and Salhofer (2013) provide a meta-analysis on the impact of subsidies on agricultural land prices. Regarding the spatial dynamics of land prices, Yang, Ritter and Odening (2017) find

*Corresponding author: E-mail: henning.schaak@slu.se

clusters of regional convergence in Germany. Following the 2008 financial crisis, the hypothesis that additional demand from non-agricultural investors accelerated price increases has drawn much attention (Hüttel, Wildermann and Croonenbroeck, 2016; Plogmann *et al.*, 2020; Tietz, Forstner and Weingarten, 2013). However, as land can be bought and rented, research ideally would not only consider sales prices, but also rental rates for land, as well as their dependencies.

While there are multiple ways of approaching the analysis of farmland values, a commonly applied approach is rooted in the theory of valuing financial assets which is dependent on income capitalisation, or the net present value (NPV) model (Burt, 1986). In this context, theoretical land values are often derived by using cash rents as a proxy for returns from agricultural activities. In an efficient market, the sales price should equal the capitalised returns and therefore only further depend on the interest rate. These theoretical farmland values can then be compared with the observed values. Alternatively, the ratio between observed rental and sale prices (rent–price ratio (RPR)) can serve as an indicator of the profitability of an investment in land. Likewise, under the assumption of a static economic environment, the ratio can be interpreted as the capital recovery factor of an investment in land. Using a variance decomposition approach (Campbell and Shiller, 1988), Plogmann *et al.* (2020) recently undertook a study of the RPR at the federal state level in the German land market. The authors find a substantial variation of the RPR between federal states, which remains unexplained. The study concludes 'that differences regarding price formation on land markets are in place, which might be the result of different farm structures in the various federal states' (Plogmann *et al.*, 2020: 12). However, the farming structure can also vary substantially within a federal state or region (see, e.g., Destatis, 2021). At the same time, it has also been shown that agricultural land markets are only integrated regionally (Yang, Ritter and Odening, 2017). Therefore, the question arises, whether the RPR is also heterogeneous at finer spatial levels and whether it can be explained by the local farm structure.

The relationship between agricultural land prices and rental rates has been extensively studied in economics research (see, e.g., Alston, 1986; Burt, 1986; Falk, 1991; Hyder and Maunder, 1974; Phipps, 1984; Traill, 1979). Generally, it is assumed that cash rents should vary in congruence with farmland values, with a strong positive relationship in their respective trends (Gutierrez, Westerlund and Erickson, 2007). Ibendahl and Griffin (2013) find that rent costs lag behind changes in land prices when they are increasing, but not when they are decreasing. Saguatti, Erickson and Gutierrez (2014) find that the long-run elasticity of cropland values with respect to net cash rents is close to unity. This can be interpreted as evidence for the validity of the NPV assumption. However, the literature also shows some conflicting results. Although farmland price and rental rate movements are highly correlated, price movements are not always in accordance with the expected relationship (Clark, Fulton and

Scott, 1993; Falk, 1991; Hallam, Machado and Rapsomanikis, 1992). Therefore, the real options approach has also been applied in order to account for uncertainty in future growth and capital gains (Turvey, 2003).

An important related issue is the (cross-sectional) heterogeneity on the land markets, as land remains an important cost factor in agricultural production. Thus, the understanding of the land market is important in order to fully understand other production-related developments. Nevertheless, heterogeneity of agricultural land sale prices and farmland rental rates is an issue that has rarely been considered in the literature (März *et al.*, 2016; Mishra and Moss, 2013). Even if the NPV approach holds, farmland values could obviously vary due to the different natural conditions. Still, at a given interest rate, the RPR should be identical between all regions. However, a recent study by Plogmann *et al.* (2020) finds that this is not the case in Germany. This result represents the starting point for the present paper. The objective is to study the relationship between rental rates and sale prices on the basis of the RPR. In contrast to Plogmann *et al.* (2020), the focus of the paper is at the cross-sectional distribution of the RPR on the regional level. Based on observations at the district level, this paper is the first to explicitly model all parameters of the district level distribution of the RPR, using a unique dataset, combining data from the German agricultural census and data collected by the expert committees for land evaluation Lower Saxony (*Oberer Gutachterausschuss für Grundstückswerte Niedersachsen*, OGA Lower Saxony). The study area is well-suited to the research topic as local farming structures in Lower Saxony are heterogeneous, with areas of intensive dairy, livestock and crop production. Other structural parameters, such as the average farm size, also differ at the local level in Lower Saxony (cf. Destatis, 2021; NMELV, 2017).

In order to model the distribution of the RPR, the paper relies on the 'generalized additive models for location shape and scale' (GAMLSS; Rigby and Stasinopoulos, 2005) framework. In this framework, not only the mean, but also the higher moments of a distribution can be modelled by generalised additive models (GAMs; Hastie and Tibshirani, 1990). Thus, the response distribution is completely characterised by one joint model (Umlauf and Kneib, 2018). In this context, the spatial heterogeneity at the district level can be modelled by effect specifications simply accounting for (i) the mere presence of district-level heterogeneity (unstructured spatial effects), as well as (ii) effects taking the neighbourhood structure of the district into account (structured spatial effects, cf. Fahrmeir and Kneib, 2011). Modelling the mean and the scale parameter of the RPR's distribution allow for identifying factors which influence the average profitability of land investments, as well as its heterogeneity. In order to avoid an overly complex model and overfitting, a variable selection procedure is used to define the final model.

The remainder of this paper is structured as follows. Section 2 outlines the methodological basis of the paper. In section 3, the data sets used and the preparation of these are described, followed by a motivation of the applied variable selection procedure. The results of the analysis are presented and discussed in section 4. The paper ends with conclusions (section 5).

2. Methodology

In order to model the distribution of the RPR, all parameters of its conditional distribution are considered. In the context of regression methods, this can be achieved by using a generalisation of the GAM framework, referred to as GAMLSS (Rigby and Stasinopoulos, 2005). The GAMLSS-framework makes way for more flexibility than more traditional regression frameworks. This is achieved by (i) not only modelling the mean (or location) parameter of the dependent variable's conditional distribution but also other parameters (e.g. the variance) and (ii) allowing for distributions which do not belong to the exponential family. The aim of this section is not to give a comprehensive presentation of the framework, but rather to outline the overall concept and the specification used in the present study. For general discussions of the GAMLSS framework, the reader is referred to the canonical references (e.g. Rigby and Stasinopoulos, 2005; Stasinopoulos and Rigby, 2007; Stasinopoulos *et al.*, 2017).

Generally, within the GAMLSS framework, parameters $\boldsymbol{\theta}^{\mathrm{T}} = (\theta_1, \theta_2, \ldots, \theta_p)$ of the dependent variable Y's distribution are individually modelled by a GAM. As mentioned earlier, the distribution of Y is not limited to the exponential family and can be chosen from a more general family (see Rigby *et al.* (2019) for a comprehensive discussion). Distributions from this family have up to four parameters which can be modelled individually, implying that $1 \leq p \leq 4$. Depending on the specific distribution, the parameters represent the distribution's location (e.g. the mean), scale (e.g. the variance) and shape (skewness and kurtosis). The following description of the GAMLSS closely follows Rigby and Stasinopoulos (2005).

It is assumed that for $k = 1, 2, \ldots, p$; $i = 1, 2, \ldots, n$ independent observations y_i have the probability density function (PDF) $f(y_i|\boldsymbol{\theta}^i)$, where $\boldsymbol{\theta}^i$ denotes the i^{th} row of the matrix $\boldsymbol{\theta} = (\theta_{ik}) \in \mathbb{R}^{n \times p}$. Further, a known monotonic link function $g_k(\cdot)$ (e.g. a log-link function) is used to relate the k^{th} column of $\boldsymbol{\theta}$, denoted as $\boldsymbol{\theta}_k$, to the explanatory variables and random effects. The corresponding additive model, the GAMLSS, is given by

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \mathbf{Z}_{jk} \boldsymbol{\gamma}_{jk}. \tag{1}$$

Here, $\boldsymbol{\eta}_k$ is a vector of length n, $\mathbf{X}_k$ is a known design matrix, $\boldsymbol{\beta}_k^{\mathrm{T}} = \left(\beta_{1k}, \ldots, \beta_{J'_k k}\right)$ is a parameter vector of length J'_k, $\mathbf{Z}_{jk}$ is a known design matrix and $\boldsymbol{\gamma}_{ik}$ is q_{ik}-dimensional random variable. If $J_k = 0$ for all k, equation (1) reduces to a fully parametric model. If, for all combinations of j and k, $\mathbf{Z}_{jk} = \mathbf{I}_n$, where $\mathbf{I}_n$ is an $n \times n$ identity matrix, and $\boldsymbol{\gamma}_{ik} = \mathbf{h}_{jk} = h_{jk}(\mathbf{x}_{jk})$ the model reduces to a semiparametric GAMLSS, where $\mathbf{h}_{jk}$ is the vector which evaluates an unknown function h_{jk} at $\mathbf{x}_{jk}$ and $\mathbf{x}_{jk}$ is a vector of length n. In practice, h_{jk} can be approximated using smoothing splines in the estimation. If $\mathbf{Z}_{jk} = \mathbf{I}_n$ and $\boldsymbol{\gamma}_{ik} = \mathbf{h}_{jk} = h_{jk}(\mathbf{x}_{jk})$ only for specific combinations of j and k, the resulting model incorporates parametric, semiparametric and random-effect terms.

One frequent concern in agricultural research is the potential presence of spatial dependencies among observations. For data where observations originate from different spatial areal units, De Bastiani *et al.* (2018) extend the standard GAMLSS. Building on Gaussian Markov Random Fields (GMRF), the authors show how random effects are expressed in a way to account for the neighbourhood structure of the observations. Generally, a neighbourhood structure can be given by an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ that consists of vertices $\mathcal{V} = (1, 2, \ldots, q)$ and a set of edges $\mathcal{E}$. A typical edge of the graph is (m, t), $t, m \in \mathcal{V}$. With respect to the graph, a random vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^T$ is called a GRMF with mean $\boldsymbol{\mu}$ and symmetric precision matrix $\lambda\mathbf{G}$, if and only if its density is given by

$$\pi(\boldsymbol{\gamma}) \propto exp\left[-\frac{1}{2}\lambda(\boldsymbol{\gamma} - \boldsymbol{\mu})^T \mathbf{G}(\boldsymbol{\gamma} - \boldsymbol{\mu})\right] \tag{2}$$

and

$$G_{mt} \neq 0 \Leftrightarrow (m, t) \in \mathcal{E} \text{ for } m \neq t, \tag{3}$$

where G_{mt} is the element of matrix $\mathbf{G}$ for row m and column t (Rue and Held, 2005). $\mathbf{G}$ contains the information about adjacent regions. When $\mathbf{G}$ is a non-singular matrix, the GMRF model is called conditional autoregressive (CAR) model (Besag, 1974) and can be defined by

$$\gamma_i | \boldsymbol{\gamma}_{-i} \sim N\left(\sum_j \alpha_{ij}\gamma_j, k_i\right), \tag{4}$$

where $\gamma_{-i} = (\gamma_1, \gamma_2, \ldots, \gamma_{i-1}, \gamma_{i+1}, \ldots, \gamma_q)$, $\alpha_{ii} = 0, \alpha_{ij} = -G_{ij}/G_{ii}\,(i \neq j)$ and $k_i = 1/(\lambda G_{ii})$, for $i = 1, 2, \ldots, q$. If $\mathbf{G}$ is symmetric, then $\boldsymbol{\mu} = 0$ (Besag and Kooperberg, 1995). For GAMs, the intrinsic autoregressive model (IAR), which is a limiting case of the CAR, is typically used to model spatially structured random effects (De Bastiani *et al.*, 2018). In order to incorporate an IAR model in the GAMLSS, its respective $\mathbf{Z}$ is set to be an index matrix indicating which observation belongs to which region. Then $\boldsymbol{\gamma}$ is a vector of q spatial random effects and $\boldsymbol{\gamma} \sim N(0, \lambda^{-1}\mathbf{G}^{-1})$. The intuitive interpretation is that such an effect follows Tobler's law and that estimated values in $\boldsymbol{\gamma}$ for neighbouring regions are closer to each other than for non-neighbouring regions. For more details, please refer to De Bastiani *et al.* (2018).

The inferential framework for the estimation of a GAMLSS is derived from an empirical Bayesian argument. Assuming independent normal priors for γ_{jk}, it can be shown that the maximum-a-posteriori is equivalent to the penalised likelihood estimation for fixed smoothing (or hyper-) parameters. For this, algorithms relying on backfitting methods are used. These can be nested into methods for the estimation of the hyperparameters, which allows for an automated determination of the model's smoothing parameters (Rigby and Stasinopoulos, 2005).

The designated dependent variable (RPR) in the present study is logically restricted to the $(0,1)$ interval.[1] There are multiple distributions which could be used to model such variables. A common choice is the beta distribution. The beta-distribution is defined by two parameters and allows for a lot of flexibility (see Rigby *et al.*, 2019). One way to parameterise the PDF of the beta-distribution is:

$$f_Y(y|\mu,\sigma) = \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1}. \tag{5}$$

The parameters μ and σ are the location and scale parameters, respectively. They refer to the mean and the standard deviation of the variable Y. In this parameterisation, $\alpha = \mu(1-\sigma^2)/\sigma^2$ and $\beta = (1-\mu)(1-\sigma^2)\sigma^2$, while $0 < \mu < 1$ and $0 < \sigma < 1$. $B(\alpha,\beta)$ represents the evaluation of the beta-function for α and β. The mean of Y is given by $E(Y) = \mu$, the variance by $Var(Y) = \sigma^2\mu(1-\mu)$ (Rigby *et al.*, 2019).

Other, less common, alternatives are the logistic-normal (also called logit-normal) and the simplex distribution (Rigby *et al.*, 2019). The PDF of the logistic-normal distribution is given by

$$f_Y(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{y(1-y)} \exp\left(-\frac{\left\{\log\left[\frac{y}{1-y}\right] - \log\left[\frac{\mu}{1-\mu}\right]\right\}^2}{2\sigma^2}\right), \tag{6}$$

The PDF of the Simplex-distribution by

$$f_Y(y|\mu,\sigma) = \frac{1}{\left[2\pi\sigma^2\mu^3(1-y)^3\right]^{1/2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2 y(1-y)\mu^2(1-\mu)^2}\right). \tag{7}$$

Both PDFs are defined for the same range as the beta-distribution, with $0 < \mu < 1$ and $\sigma > 0$. μ represents the median and the mean of the logistic-normal distribution and the simplex distribution, respectively. The variance is undefined in both cases; for more details, see Rigby *et al.* (2019) and the references therein. All model specifications discussed in the following are considered for all three distributions (see Appendix A1 and Appendix A2).

In the present study, two parameter vectors, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, are estimated. The respective predictor vectors are $\boldsymbol{\eta}_\mu$ and $\boldsymbol{\eta}_\sigma$. For estimations assuming the beta distribution, the logit-link function is used for both parameters. For the other estimations, the logit-link function is used for the μ parameter and the log-link function for the σ parameter.

1 Under the plausible assumption that the rental rate will always be smaller than the sale price.

3. Data description and variable selection

3.1. Data description and processing

Calculating the RPR for a given plot of land would ideally be based on information regarding both its rent and sale price. In reality, a plot is usually either sold or rented out at a given point in time, but this kind of information is not available. Instead, local averages of land rents and prices are used in the present study. Therefore, a data set was compiled from three sources. The first source is the data from the German agricultural census 2010 (*Landwirtschaftszählung* 2010), which is latest available data set providing comprehensive information on all farms in Germany. It contains information on all farms, including data on the rent paid by the farmers. The second source of data is a data set provided by the OGA Lower Saxony, which consists of plot-level data for all agricultural land sales in Lower Saxony during the time period between the latest large scale statistical farm survey (*Agrarstrukturerhebung* 2007) and the agricultural census (second quarter of 2007 until the first quarter of 2010). These data were originally collected by the OGA Lower Saxony for the purposes of the German Federal Building Code. The third data source is the regional statistical database of the Federal Statistical Office and Statistical Offices of the Länder, which provides general economic indicators for the analysis (Statistische Ämter des Bundes und der Länder, 2020a, 2020b, 2020c, 2020d).

The local average land rent and land purchase prices per hectare (ha) were calculated using a standardised spatial grid which was the smallest grid used for official agricultural statistical purposes in Germany (e.g. Destatis, 2021). This grid has a cell size of 5×5 km and was used to merge the two data sources. In both data sets, observations were assigned to the grid based on their geo-referenced locations. Then the RPR of the respective cell was calculated.[2] It is worth noting that within Germany there is the major distinction between 'arable land' and 'grassland' (or 'pasture'). The difference is that grassland is considered to be permanently used for forage production and that there are additional legal restrictions for its usage, most importantly the ban on ploughing. Still, this does not imply that arable land is not used for forage production at all. Nevertheless, the profitability of the investment in land may vary between the two types. As the data differentiate between arable and grassland, the RPR was calculated separately on the cell level. This led to a total of 2,794 local observations of the RPR. In order to allow for a differentiation between the land type of each observation of the RPR, a cell level dummy variable *d_grassl* (1 if the observation is based on grassland land, 0 otherwise) was included in the final data set. All other explanatory variables considered were calculated at the district level, using the farm-level data from the agricultural census and regional statistics database. Apart from our own

2 As the rental rates are only available at the farm level, an unobserved measurement error for the cell level averages is potentially present, as farms may rent plots located in other grid cells compared to the ones they are located in.

considerations, these variables are motivated by previous research on the determinants of rental rates for land (Habermann and Breustedt, 2011; Habermann and Ernst, 2010; März *et al.*, 2016).

With respect to the farming structure, the share of farms, which are legal entities (*farm_type_share*), the share of part time farms (*parttime_share*), the share of rented land on the agricultural land (*rent_share1*), the average share of rented land on agricultural land per farm (*rent_share2*) and the share of organic farms (*organic_share*) in the district are calculated. Furthermore, the labour intensity per ha (*labour*) and the average farm size (*size*) were considered. In order to account for the district-level competition, a concentration measure was included in the analysis; it is defined in the form of a Herfindahl–Hirschmann Index (*hhi*) and calculated as:

$$hhi_j = \sum_{i=1}^{N_j} \left(\frac{Area\_of\_farm_i}{Total\_area_j} \right)^2, \quad (8)$$

for the N_j farms in a district j and measured for the concentration of farmland in that district. If each of the farms in the district were of the same size, the index would be equal to $\frac{1}{N}$ (thus approaching 0 with an increasing number of equal sized farms) and would be equal to 1 if there were only one farm in the district (thus the land is fully concentrated).

In addition to the farming structure, there are two additional dimensions, which should be accounted for: the average production programme and the general economic situation in the district. With respect to the average production programme, the average density of cattle (*cattle*), as well as hogs and poultry (*hog_poultry*) animal units (AU) per ha were calculated. In terms of crop production, shares of *potato, rye, sugarbeet* and *winterwheat* in the cropping pattern were considered for the analysis. As discussed above, the potential effects of biogas production on land markets have gained interest in recent years (Habermann and Breustedt, 2011; Hennig and Latacz-Lohmann, 2017). To account for such potential effects, two variables reflecting biogas production in terms of the agricultural production direction as well as the farming structure are considered for the analysis. For the former case, this is the district's average biogas capacity (in kWh) per ha (*biogas_cap*), for the latter it is the share of farms with biogas plants in the district (*biogas_share*). Lastly, the share of pasture on the total agricultural land in the district (*grass_share*) is considered. This variable is linked to the district's average production system. In order to account for the general economic situation, the district-level unemployment rate (*unemployment*), population density (*pop_density*) and average income (*income*) are also considered for the analysis. Finally, the average *landprice* in each district is also included. The variables are summarised in Table 1. All variables are considered for both the predictor of the mean and the scale parameter of the RPR. As discussed earlier, potential remaining spatial heterogeneity could be addressed by including structured spatial effects (f_{str}) and unstructured spatial effects $\left(f_{unstr}\right)$ in the predictors.[3]

3.2. Variable selection

One of the major advantages of GAMLSS is the ease to study complex models with various potential effect specifications. This also represents a potential drawback, as overly complex models are prone to overfitting, leading to inadequate model specifications. One way to select the terms to be included in the model is model comparisons based on the generalised Akaike information criterion (GAIC; Stasinopoulos *et al.*, 2017). Therefore, a modification of the procedure outlined by De Bastiani *et al.* (2018) was applied. In a first step, an appropriate set of variables was selected. In a second step, it was evaluated whether remaining heterogeneity could be explained by spatial effects. The procedure was as follows:

1. Estimate a 'Null model', containing only a constant in η_μ and η_σ.
2. Select variables to be included in the model, based on the GAIC, by:
   - 2.1. Applying a forward-stepwise selection procedure on η_μ.
   - 2.2. Applying a forward-stepwise selection procedure on η_σ, given the model obtained by step 2.1.
   - 2.3. Applying a backward-stepwise elimination procedure on the variables in η_μ, given the model obtained by step 2.2.
3. Consider the inclusion of structured and unstructured spatial effects in the model obtained in step 2.3. by:
   - 3.1. Applying a forward-stepwise selection procedure on η_μ.
   - 3.2. Applying a forward-stepwise selection procedure on η_σ, given the model obtained by step 3.1.
   - 3.3. Applying a backward-stepwise elimination procedure on the spatial effects in η_μ, given the model obtained by step 3.2.

The model obtained in step 3.3 will be used as the final model for the analysis. Note that the degree of smoothing for structured and unstructured spatial effects is not fixed, but rather determined during the estimation process (cf. Rigby and Stasinopoulos, 2014). In the case that the variable *d_grassl* is selected in step 1, it is reasonable to also control for potential interaction effects between the variable *d_grassl* and other variables included in the model. These potential interactions are considered by the selection algorithm applied in the steps 2.1–2.3. The final model selected by the procedure is presented and discussed in the following section.

3 The reader may ask why no alternative specification with a set of dummy variables indicating the district were considered. As the other explanatory variables are also considered at the district level, doing so would lead to multicollinearity issues.

Table 1. Descriptive statistics of the RPR and the predictor variables considered for the analysis

Variable	Description	Mean	SD	*n*
biogas_cap	Average biogas capacity (in kWh) per ha	0.0995	0.1036	46
biogas_share	Share of farms with biogas plants	0.0143	0.0141	46
cattle	Average cattle density in animal units (AU) per ha	0.6341	0.5072	46
d_grassl	Dummy variable, 1 if the observation refers to grassland, 0 otherwise	0.4528	–	2794
farm_type_share	Share of farms solely in proprietorship	0.9051	0.0377	46
grassl_share	Share of grassland on total agricultural land in production	0.3030	0.2394	46
hhi	Herfindahl–Hirschman index, based on the farm size in ha and the total amount of land under production in the district	0.0084	0.0126	46
hog_poultry	Average density of hogs and poultry in AU per ha	0.2705	0.4024	46
income	Average taxable income (in 10,000 EUR per taxable person)	3.0365	0.2800	46
labour	Average labour force per farm	1.5572	0.2128	46
landprice	Average land price (10,000 EUR per ha)	1.6710	0.7494	46
organic_share	Share of organic farms	0.0328	0.0221	46
parttime_share	Share of part-time farms	0.3524	0.0844	46
pop_density	Population density (per 1000 inhabitants per km²)	2.9058	3.6301	46
potato	Share of potato in the cropping pattern	0.0146	0.0193	46
rent_share1	Share of rented land on total agricultural land	0.4617	0.0594	46
rent_share2	Average share of rented land on agricultural land per farm	0.3819	0.0543	46
RPR	Rent price ratio	0.0198	0.0121	2794
rye	Share of rye in the cropping pattern	0.0519	0.0470	46
size	Average farm size in 100 ha	0.6729	0.1817	46
sugarbeet	Share of sugar beet in the cropping pattern	0.0537	0.0701	46
unemployment	Unemployment rate	7.9413	1.9375	46
winterwheat	Share of winter wheat in the cropping pattern	0.2687	0.1758	46

Note: all variables except *RPR* and *d_grassl* are given at the district level, *RPR* and *d_grassl* are given on the grid level.
Sources: RDC of the Federal Statistical Office and Statistical Offices of the Länder, agricultural census 2010 and data from the OGA Lower Saxony and the Federal Statistical Office and Statistical Offices of the Länder, own calculations.

4. Analysis of the RPR

4.1. Results of the model selection and overview of the studied model

In this section, the results of the model selection procedure are presented in Tables 2–4. Estimations were done using the 'R'-software-package[4] (R Core Team, 2019). Additionally, alternative, fixed model specifications as well as all models under different distributional assumptions (logistic normal distribution and simplex distribution instead of the beta distribution) were estimated. The models vary in their complexity, sequentially including all variables considered in the variable selection, their interactions with *d_grassl*, and the spatial effects in the predictors; first only for the μ parameter and then for both the μ and σ parameters. Thus, the comparison covers regular linear as well as additive regression models and different distributional regression specifications in both the conventional setting (beta distribution) and non-standard distributions. The GAIC and the (effective) degrees of freedom (DFs) for the fits of all models are presented in Appendix A1.[5] Summarising, the variable selection (under the assumption of a beta distribution) yields a model which outperforms all alternative model specifications in terms of the GAIC, while only using a moderate number of DF. Note that the DF can be interpreted as the effective number of parameters in the model, which becomes continuous in presence of smoothing terms (Stasinopoulos *et al.*, 2017). It is also important to note that regular standard errors obtained by the GAMLSS implementation may not be accurate when the model includes additive smoothing terms. Additionally, the standard errors do not account for the variable selection procedure, which further renders the interpretation of these effects unreliable (Hastie, Tibshirani and Friedman, 2009). In order to be able to assess the statistical significance, an additional non-parametric bootstrap procedure with 1,000 samples was carried out (Stasinopoulos *et al.*, 2017).

The selected variables include the local farming structure (e.g. *farm_type_share* and *labour*), the production programme (e.g. *winterwheat* and *potato*) and also the overall economic structure within the district (e.g. *unemployment*). In particular, *d_grassl* is selected for the mean and scale predictor, while the district level *landprice* is only selected for the mean predictor. In total, more variables are selected for the scale predictor than for the mean predictor. Further, a series of interaction effects between *d_grassl* and other variables are selected. Also here more effects are selected for the scale predictor. These effects are discussed in section 4.2. In both predictors, structured and unstructured spatial effects are included in the model. This is an indication that after

4 For the estimations of the models, the 'gamlss' package (Rigby and Stasinopoulos, 2005; Stasinopoulos and Rigby, 2007) was used. The structured spatial effects rely on the implementation of the 'gamlss.spatial' package (De Bastiani, Stasinopoulos and Rigby, 2018). The bootstrap procedure was implemented using functions of the 'boot' package (Canty and Ripley, 2019; Davison and Hinkley, 1997).

5 The variables included by the variable selection procedure under the three distributional assumptions are summarised in Appendix A2. The complete estimation results are available as supplementary material.

Table 2. Parameter estimates for the mean parameter of the RPR ($n = 2{,}794$)

Variable	β	95 % CI Lower bound	95 % CI Upper bound
Intercept	−0.7891	−3.6424	0.1047
biogas_cap	0.5973	−0.4569	1.0268
biogas_share	−6.3747	−9.1365	3.4462
d_grassland	−0.3059	−0.6355	0.2132
farm_type_share	−1.5637	−2.6208	1.1547
grassl_share	−0.4315	−0.6195	−0.1742
income	−0.2111	−0.3118	−0.0678
landprice	−0.2186	−0.3270	−0.1485
parttime_share	−0.9171	−1.2980	−0.3110
pop_density	−0.0597	−0.0757	−0.0112
rent_share2	−0.1614	−1.2750	0.5918
unemployment	−0.0122	−0.0328	0.0160
Interaction effects			
d_grassl × *landprice*	−0.2072	−0.2738	−0.1425
d_grassl × *pop_density*	0.0220	−0.0090	0.0548
d_grassl × *rent_share2*	2.2916	0.9292	3.3734
d_grassl × *unemployment*	−0.0367	−0.0666	−0.0123

Note: CI: confidence interval, based on 1,000 bootstrap samples.
Sources: RDC of the Federal Statistical Office and Statistical Offices of the Länder, agricultural census 2010 and data from the OGA Lower Saxony and the Federal Statistical Office and Statistical Offices of the Länder, own calculations.

controlling for the selected variables, the remaining heterogeneity has spatial components. These effects are discussed in section 4.3.

4.2. Results of the variables selected in the model

The effects of the selected variables on both the mean and the scale of the RPR's distribution are presented and discussed. For the interpretation of the effects of the explanatory variables, it is helpful to recall that the RPR is a crude measure for the profitability of an investment. Thus, taking the perspective of a potential investor (and less of a farmer) as a potential buyer into account allows for an intuitive interpretation of the results. Also, while the analysis shares some similarities with hedonic pricing studies, the results are not directly comparable, as the present study considers a relative and not an absolute measure. Table 2 shows the effects of the variables selected in the mean predictor and thus the effects on the average profitability of an investment in land (at the district level). Additionally, the respective 95 per cent confidence intervals (CIs) are presented, which are used to assess the statistical significance of the individual effects.

The statistically significant effects indicate that investments in land are less profitable in districts with a higher share of grassland (*grassl_share*) as well as with a higher share of part-time farmers (*parttime_share*). The average *landprice* in a district has a statistically significant negative effect on the RPR, indicating that in districts with higher average land prices, investments are less profitable. This indicates that rental rate on average does not increase at the same rate as land prices. Still, this interpretation has to be considered with caution, as rental rates used in this study represent all paid rents during the considered period (regardless of the contract date), whereas sale prices are limited to this period. The statistically non-significant effect of *d_grassland* indicates that when considering only the main effect, the profitability of investments in grassland does not significantly differ from investments in arable land. Still, the statistically significant interaction effects of *d_grassland* indicate that the RPR for grasslands is lower in districts with higher average land prices (*d_grassl* × *landprice*) and a higher unemployment rate (*d_grassl* × *unemployment*), but higher in districts with a higher average share of rented land per farm (*d_grassl* × *rent_share2*). Arguable, the latter two effects lack a direct economic interpretation. Still, an improved general economic situation may lead to increases of leisure-related pastures usages (e.g. for sports or private equine husbandry). Thus, the unemployment rate may capture such effects as a proxy variable. The results also indicate that investments in land are less profitable in districts with a higher population density (*pop_density*) and a higher average *income.* Assuming a higher competition for land with other economic sectors in these districts allows for the interpretation that this increases efficiency on the agricultural land market. An analogue argument for the effect of *income* can be made as for the interaction effect *d_grassl* × *unemployment.*

In principle, the RPR should be homogeneous in efficient markets. Thus, the results for the scale predictor indicate which variables are potentially associated with the district-level market efficiency. The heterogeneity of the RPR (Table 3) increases with the average farm size (*size*) but decreases with the degree of land concentration (*hhi*). If larger farms are active on a larger spatial scale, their activity potentially helps to equalise the RPR in different locations, explaining these effects. The RPR is also lower in districts with a higher share of farms in sole proprietorship (*farm_type_share).* Here, a potential explanation is that farms of this type are usually family farms, where the family members have a good knowledge of the local market conditions. This higher level of knowledge thus leads to sales prices and rental rates closer to those at the theoretical market equilibrium. This would be reflected in a lower variability of the RPR. The results also indicate that the RPR heterogeneity decreases with the share of *winterwheat.* This appears reasonable, as wheat requires less specific production conditions compared to other considered crops (e.g. potatoes). Hence, a higher share in the average production programme could be seen as an indicator for more homogenous natural conditions and indicate more homogenous returns in the district. The negative effect of *labour* could readily be explained by the fact that labour is a costly production factor which has to be compensated. Interestingly, the two variables relating to biogas production have statistically significant effects of opposite sign. The share of farms with biogas plants (*biogas_share*) reduces the district-level heterogeneity of the RPR, while the average capacity per ha (*biogas_cap*) increases it. Here, one interpretation is that if the biogas capacity is spatially unevenly distributed,

Table 3. Parameter estimates for the scale parameter of the RPR ($n = 2{,}794$)

Variable	β	95 % CI Lower bound	Upper bound
Intercept	3.7028	−1.0366	7.2977
biogas_cap	1.6886	0.4229	3.7242
biogas_share	−20.0327	−36.1906	−9.5104
d_grassland	−4.4121	−9.6543	1.7782
farm_type_share	−5.4918	−9.3334	−1.0785
hhi	−27.3018	−67.8583	−13.5538
hog_poultry	−0.2004	−0.4140	0.0091
income	−0.2938	−0.5628	−0.0376
labour	−0.8331	−1.3477	−0.2701
organic_share	3.6574	−1.7769	8.2914
parttime_share	0.5304	−0.6645	1.9825
pop_density	0.0692	0.0165	0.1324
potato	−0.0600	−4.3113	2.9674
size	1.3542	0.7222	2.4660
winterwheat	−1.7000	−2.5154	−0.9413
Interaction effects			
d_grassl × *farm_type_share*	4.8465	−0.7248	9.7712
d_grassl × *hhi*	23.2742	−15.0156	58.4324
d_grassl × *hog_poultry*	0.2004	−0.0453	0.4513
d_grassl × *income*	−0.2464	−0.6355	0.1860
d_grassl × *labour*	0.3033	−0.6526	1.0437
d_grassl × *organic_share*	−8.6041	−13.7022	−1.1562
d_grassl × *parttime_share*	1.0753	−1.3130	3.0351
d_grassl × *pop_density*	−0.0906	−0.1818	−0.0055
d_grassl × *potato*	5.9402	0.5318	10.8336
d_grassl × *winterwheat*	1.4770	0.3319	2.5049

Note: CI: confidence interval, based on 1,000 bootstrap samples.
Sources: RDC of the Federal Statistical Office and Statistical Offices of the Länder, agricultural census 2010 and data from the OGA Lower Saxony and the Federal Statistical Office and Statistical Offices of the Länder, own calculations.

it can lead to local differences in competition on the land markets. Population density (*pop_density*) increases the heterogeneity, which appears reasonable, as the population within a district is usually unevenly distributed in space, for example, between cities and rural areas, leading to locally varying competition of the agricultural land market with other land usage forms. *Income* has a negative effect on the RPR's scale. Again, this can be linked with the economic situation within the district.

As with the mean predictor, the main effect of a plot being grassland or not (*d_grassl*) is not statistically significant, an indication that the local profitability heterogeneity did not significantly vary between pasture and arable land (*d_grassl*) when solely considering this difference. Still, a number of relevant interaction effects can be found. These include the share of organic farms (*d_grassl* × *organic_share*), the population density (*d_grassl* × *pop_density*) and the average production programme (*d_grassl* × *winterwheat* and *d_grassl* × *potato*). Here, it is interesting that the results indicate that heterogeneity of the RPR for grassland is higher in districts with a higher share of wheat and potato production, which can be seen as competition for forage production on arable land.

The remaining included variables have wide CIs covering zero; thus, it cannot be stated that they significantly contribute to the explanation of the distribution of the RPR and thus they do not influence the profitability of investments in agricultural land and the market efficiency. The same holds for the variables not selected in the model (cf. Appendix A2). Comparing the type of variables with statistically significant effects in the mean and the scale predictor, differences can be found. For the mean predictor, these variables can be linked to more general structural differences between the districts, e.g. general economic indicators (e.g. population density) or the share of part-time farms. In contrast, for the scale predictor, variables relating to the average production programme can also be found (e.g. the biogas capacity or the share of winter wheat). This serves as an indication that the average profitability in land is not affected by the regional production differences, while its heterogeneity is.

While these results are readily interpreted from an investor's perspective, an interpretation from the farmer's perspective is more challenging. Still, the results could be used to give some decision support. The RPR in a district on average decreases with higher average land prices, this indicates that it is relatively 'cheaper' to rent land in these districts. The same holds for districts with higher shares of grassland and part-time farmers. The final decision whether to buy or rent land would still have to be made individually, as the actual profitability would depend on the realisable returns from farming activities. Additionally, this result may only hold in the short run.

It also has to be taken into account that the results (analogously to hedonic price studies) may include expectations of actors on the land market. As the studied time period covers the land market before assumed increased interest of non-agricultural investors, it can be argued that if such expectations were present, they would most likely be the expectations of the farming sector and the expectations of non-agricultural investors. Furthermore, the data on land rents average information on all rental contracts, including potential older and long-running contracts. This is a fundamental issue of data availability, as land sales and rental contracts are commonly not observed at the same point in time.[6] Lastly, many of the results discussed above have to be considered exploratory, as they lack direct interpretation based on theoretical considerations. Rather, they can be seen as the starting point for more theory driven investigations of these effects.

6 The only hypothetical exception would be a case where a plot of land (without a current rental contract) is sold to an investor, who immediately rents it out to a farmer.

Table 4. Parameter estimates of the spatial effects

Predictor	η_μ		η_σ	
Spatial effect	Structured	Unstructured	Structured	Unstructured
σ_b	0.0037	0.0398	0.0033	Not included
Effective DF used in the fit	1.8052	15.6454	1.2587	

Sources: RDC of the Federal Statistical Office and Statistical Offices of the Länder, agricultural census 2010 and data from the OGA Lower Saxony and the Federal Statistical Office and Statistical Offices of the Länder, own calculations.

4.3. Spatial effects

As shown in Table 4, structured and unstructured spatial effect terms for the mean parameter are selected in the final model. For the scale parameter, only the structured term is included. This indicates that the spatial effects lead to an overall improvement of the model and that spatial heterogeneity is present with respect both to the mean and to the scale of the RPR's distribution. For a first assessment of the spatial effects, Table 4 shows their standard deviations (σ_b), as well as the DF used for their fit. For the RPR's mean, the standard deviation of the unstructured effect is larger by one order of magnitude and more DF are used for the fit as for the structured one (recall that effective DF are a crude measure for the complexity of the respective fit). For the spatial effects of the scale parameter, the standard deviation of the structured effect is similar to the one of unstructured spatial effect in the mean predictor.

While the inclusion of the spatial effects leads to an overall improvement of the model, the question arises whether this spatial variation also indicates a significant difference at the district level (cf. März *et al.*, 2016). As the mean of both the structured and unstructured effects are zero by construction, a district-level effect significantly different from zero would indicate a positive or negative deviation from the overall intercept. To assess their statistical significance, CIs of the district level effects were also obtained by the bootstrap procedure.

Overall, statistically significant district-level effects were only found for the spatial effects terms of the mean parameter. Following März *et al.* (2016), the statistical significance of the structured and unstructured spatial effects for the RPR's mean is depicted in Figure 1. District-level effects whose CIs lie below zero are in light grey, while black ones indicate CIs above zero. Grey indicate districts where the effects' CI overlap zero. Ceteris paribus, this can be interpreted as such that the respective parameter values of the RPR's distribution in a given district are larger or smaller.

The difference of the structured and unstructured effect term standard deviations indicates that the district-level effect sizes of the unstructured term are larger than the ones of the structured term. Most of the statistically significantly positive effects are found for districts located between the federal city state of Bremen and the intensive livestock production districts (cf. NMELV, 2017).

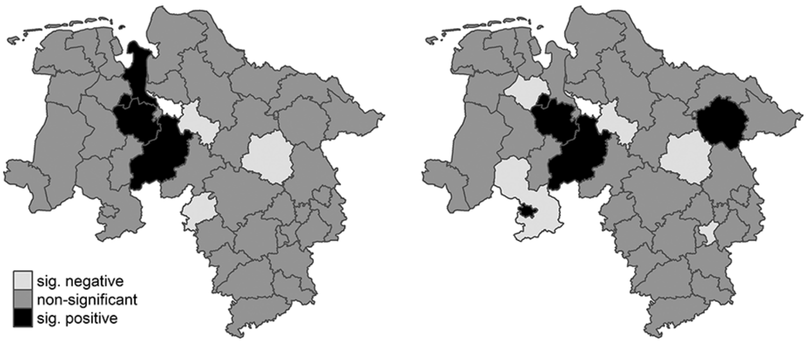

Fig. 1. Statistical significance of the structured (left hand) and unstructured (right hand) district-level effects for the mean of the rent–price ratio distribution based on bootstrapped 95 per cent confidence intervals.
Note: the empty polygon represents the federal state Bremen.
Sources: RDC of the Federal Statistical Office and Statistical Offices of the Länder, agricultural census 2010 and data from the OGA Lower Saxony and the Federal Statistical Office and Statistical Offices of the Länder, own calculations.

This is an indication that the profitability of investments in agricultural land is, ceteris paribus, higher in this region. Identifying the underlying reasons for these apparent spillovers from regions with either a high density of animal or humans (respectively their potential interations) would require further investigation. Further, two districts have both statistically significant negative structured and unstructured effects, indicating that investments in land in these districts are less profitable than Lower Saxony's average. This shows that unobserved district-level heterogeneity remains after adjusting for the explanatory variables for a limited number of districts.

With respect to the spatial effects for the scale parameter, only the structured effect term was included in the model selection procedure. Still, none of its values have a confidence interval excluding zero (not depicted). This implies that all meaningful heterogeneity of the RPR's scale is captured by the other explanatory variables.

5. Conclusions

Recent price increases for agricultural land in many European countries have reinforced the need for the understanding of the related land markets. By focusing on the cross-sectional profitability of investments in farmland, the results presented in this paper complement and extend prior research. The district-level analysis of the RPR of agricultural land revealed that the average RPR as well as its heterogeneity are influenced by the district average production programme, the farming structure and the general economic condition within the district. Remaining district-level spatial heterogeneity was identified by including structured and unstructured spatial effects in the regression model.

The results have some implications for policymakers, farmers and other actors on markets for agricultural land. They provide general insights on distribution of the profitability of agricultural land, which can be utilised by decision-makers at different spatial scales. There are differences between investments in grassland and arable land. This is important insofar, as the status of arable land can change to grassland but, under current legislation, not vice versa. The results indicate that the market for agricultural land cannot be seen being independent from the non-agricultural economic situation, as the average RPR varies with general economic indicators like the average income. Also, the RPR in districts with a higher land concentration varied less, which is an indication that (at current levels) the local land market was potentially more efficient. In the light of political and societal discussions about the effects of structural change and land ownership in the agricultural sector, this would be particularly relevant. Still, present results should only be seen as an indication; they cannot be interpreted as a causal relationship. Here, more research is needed.

In future research, GAMLSS variants (e.g. for high-dimensional settings relying on boosting methods, Mayr *et al.*, 2012) could be considered. Given the recent developments on the land markets, further insights could be gained by expanding the data set, in either the spatial, temporal or even jointly in the spatiotemporal domain. Such extensions would be straightforward from a methodological perspective. More interestingly though, future research could also aim to extent the present work to allow for causal interpretations. By nature, this research would need to have a narrower focus in terms of the effects of interest. Based on the results presented here, this could, for example, be the difference between both arable and pasture lands or the influence of the general economic structure on the profitability of investments in agricultural land. In this context, it could also be fruitful to consider the market for agricultural land in integration with the market for land in general. In any case, such work would crucially depend on the availability of larger amounts of more comprehensive data.

Acknowledgements

The authors thank the members of the FDZ in Kiel, especially Dr. Steffi Dierks and Janne Timmermann for their outstanding assistance and the OGA Lower Saxony, especially Peter Ache, for providing data access. Three anonymous reviewers and the editor provided comments and suggestions which helped to substantially improve the manuscript.

Funding

The authors gratefully acknowledge financial support from the German Research Foundation (DFG) through Research Unit 2569 ‘Agricultural Land Markets – Efficiency and Regulation’.

Supplementary data

Supplementary data are available at *ERAE* online.

References

Alston, J. M. (1986). An analysis of growth of U.S. farmland prices, 1963–82. *American Journal of Agricultural Economics* 68(1): 1–9.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)* 36(2): 192–225.

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregression. *Biometrika* 82(4): 733.

Burt, O. R. (1986). Econometric modeling of the capitalization formula for farmland prices. *American Journal of Agricultural Economics* 68(1): 10–26.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies* 1(3): 195–228.

Canty, A. and Ripley, B. D. (2019). *Boot: Bootstrap R (S-plus) Functions*. (R package version 1.3-24) [Computer software]. https://CRAN.R-project.org/package=boot. Accessed 11 December 2019.

Clark, J. S., Fulton, M. and Scott, J. T. (1993). The inconsistency of land values, land rents, and capitalization formulas. *American Journal of Agricultural Economics* 75(1): 147–155.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Applications*. Cambridge: Cambridge University Press.

De Bastiani, F., Rigby, R. A., Stasinopoulous, D. M., Cysneiros, A. H. M. A. and Uribe-Opazo, M. A. (2018). Gaussian Markov random field spatial models in GAMLSS. *Journal of Applied Statistics* 45(1): 168–186.

De Bastiani, F., Stasinopoulos, D. M. and Rigby, R. A. (2018). Gamlss.spatial: spatial terms in generalized additive models for location scale and shape models (R package version 2.0.0) [Computer software]. https://CRAN.R-project.org/package=gamlss.spatial. Accessed 11 December 2019.

Destatis. (2021). Atlas Agrarstatistik Deutschland [Map]. https://agraratlas.statistikportal.de/. Accessed 30 November 2021.

Eurostat. (2019, May 29). *Agricultural land prices by region* [Database]. Agricultural Land Prices by Region. https://ec.europa.eu/eurostat/web/products-datasets/product?code=apri_lprc. Accessed 11 December 2019.

Fahrmeir, L. and Kneib, T. (2011). *Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data*. Oxford, New York: Oxford University Press.

Falk, B. (1991). Formally testing the present value model of farmland prices. *American Journal of Agricultural Economics* 73(1): 1–10.

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(2): 1–15.

Gutierrez, L., Westerlund, J. and Erickson, K. (2007). Farmland prices, structural breaks and panel data. *European Review of Agricultural Economics* 34(2): 161–179.

Habermann, H. and Breustedt, G. (2011). Einfluss der Biogaserzeugung auf landwirtschaftliche Pachtpreise in Deutschland. *German Journal of Agricultural Economics* 51(2): 85–100.

Habermann, H. and Ernst, C. (2010). Entwicklungen und Bestimmungsgründe der Landpachtpreise in Deutschland. *Berichte Uber Landwirtschaft* 88(1): 57–85.

Hallam, D., Machado, F. and Rapsomanikis, G. (1992). Co-integration analysis and the determinants of land prices. *Journal of Agricultural Economics* 43(1): 28–37.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. New York: Springer-Verlag.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Boca Raton: Chapman & Hall/CRC.

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(4): 219–239.

Hennig, S. and Latacz-Lohmann, U. (2017). The incidence of biogas feed-in tariffs on farmland rental rates – evidence from northern Germany. *European Review of Agricultural Economics* 44(2): 231–254.

Hüttel, S., Wildermann, L. and Croonenbroeck, C. (2016). How do institutional market players matter in farmland pricing? *Land Use Policy* 59: 154–167.

Hyder, K. and Maunder, A. H. (1974). The price of farms. *Oxford Agrarian Studies* 3(1): 3–14.

Ibendahl, G. and Griffin, T. (2013). The connection between cash rents and land values. *Journal of the American Society of Farm Managers and Rural Appraisers (ASFMRA)* 2013: 239–247.

Latruffe, L. and Le Mouël, C. (2009). Capitalization of government support in agricultural land prices: what do we know? *Journal of Economic Surveys* 23(4): 659–691.

März, A., Klein, N., Kneib, T. and Musshoff, O. (2016). Analysing farmland rental rates using Bayesian geoadditive quantile regression. *European Review of Agricultural Economics* 43(4): 663–698.

Mayr, A., Fenske, N., Hofner, B., Kneib, T. and Schmid, M. (2012). Generalized additive models for location, scale and shape for high dimensional data-a flexible approach based on boosting: boosting generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society. Series C, Applied Statistics* 61(3): 403–427.

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.

NMELV (Niedersächsisches Ministerium für Ernährung, Landwirtschaft und Verbraucherschutz). (2017). Die niedersächsische Landwirtschaft in Zahlen 2017. Niedersächsisches Ministerium für Ernährung, Landwirtschaft und Verbraucherschutz (p. 100). https://www.ml.niedersachsen.de/download/124920/Die_niedersaeschsische_Landwirtschaft_in_Zahlen_2017.pdf. Accessed 13 December 2019.

Phipps, T. T. (1984). Land prices and farm-based returns. *American Journal of Agricultural Economics* 66(4): 422–429.

Plogmann, J., Mußhoff, O., Odening, M. and Ritter, M. (2020). What moves the german land market? A decomposition of the land rent-price ratio. *German Journal of Agricultural Economics* 69: 1–18.

R Core Team. (2019). R: A language and environment for statistical computing. https://www.R-project.org/. Accessed 11 December 2019.

Rigby, R. A., Stasinopoulos, D. M., Heller, G. Z. and De Bastiani, F. (2019). Distributions for modelling location, scale, and shape using GAMLSS in R. Boca Raton: CRC Press LLC.

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape (with discussion). *Journal of the Royal Statistical Society. Series C, Applied Statistics* 54(3): 507–554.

Rigby, R. A. and Stasinopoulos, D. M. (2014). Automatic smoothing parameter selection in GAMLSS with an application to centile estimation. *Statistical Methods in Medical Research* 23(4): 318–332.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton: Chapman & Hall/CRC.

Saguatti, A., Erickson, K. and Gutierrez, L. (2014). Spatial panel models for the analysis of land prices In: 2014 International Congress, 26–29 August 2014. No. 182757. Ljubljana, Slovenia: European Association of Agricultural Economists. https://econpapers.repec.org/paper/agseaae14/182757.htm.

Stasinopoulos, D. M., Rigby, R. A., Heller, G. Z., Voudouris, V. and De Bastiani, F. (2017). *Flexible Regression and Smoothing: Using GAMLSS in R*. Boca Raton: CRC Press/Taylor & Francis Group.

Stasinopoulos, D. M. and Rigby, R. A. (2007). Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software* 23(7): 1–46.

Statistische Ämter des Bundes und der Länder. (2020a). Regionalstatistik Tabelle 11111-01-01-4. www.regionalstatistik.de, Data licence Germany – attribution – Version 2.0 (https://www.govdata.de/dl-de/by-2-0). Accessed 13 December 2020.

Statistische Ämter des Bundes und der Länder. (2020b). Regionalstatistik Tabelle 12411-01-01-4. www.regionalstatistik.de, Data licence Germany – attribution – Version 2.0 (https://www.govdata.de/dl-de/by-2-0). Accessed 13 December 2020.

Statistische Ämter des Bundes und der Länder. (2020c). Regionalstatistik Tabelle 13211-02-05-4. www.regionalstatistik.de, Data licence Germany – attribution – Version 2.0 (https://www.govdata.de/dl-de/by-2-0). Accessed 13 December 2020.

Statistische Ämter des Bundes und der Länder. (2020d). Regionalstatistik Tabelle 73111-02-02-4. www.regionalstatistik.de, Data licence Germany – attribution – Version 2.0 (https://www.govdata.de/dl-de/by-2-0). Accessed 3 December 2020.

Tietz, A., Forstner, B. and Weingarten, P. (2013). Non-agricultural and supra-regional investors on the German agricultural land market: an empirical analysis of their significance and impacts. *German Journal of Agricultural Economics* 62(2): 86–98.

Traill, B. (1979). An empirical model of the U.K. land market and the impact of price policy on land values and rents. *European Review of Agricultural Economics* 6(2): 209–232.

Turvey, C. (2003). Hysteresis and the value of farmland: a real-options approach to farmland valuation. In C. B. Moss and A. Schmitz (eds), *Government Policy and Farmland Markets*. Ames, Iowa: Iowa State Press, 179–207.

Umlauf, N. and Kneib, T. (2018). A primer on Bayesian distributional regression. *Statistical Modelling* 18(3–4): 219–247.

Yang, X., Ritter, M. and Odening, M. (2017). Testing for regional convergence of agricultural land prices. *Land Use Policy* 64: 64–75.

Appendix A.

Appendix A1. Comparison of different model specifications with different distributional assumptions for the RPR ($n = 2794$)

Included effects	Model specifications								
	1	2	3	4	5	6	7	8	9
Effects in η_μ	Variable selection, see Appendix A2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Effects in η_σ		No	No	No	No	Yes	Yes	Yes	Yes
Interaction effects		No	Yes	No	Yes	No	Yes	No	Yes
Spatial effects		No	No	Yes	Yes	No	No	Yes	Yes
Beta distribution									
DF for used the fit	55.72	24.00	45.00	34.27	54.41	46.00	88.00	57.12	104.54
GAIC	−19,255.69	−18,819.15	−18,870.49	−18,823.21	−18,873.50	−19,070.46	−19,230.77	−19,077.79	−19,242.07
Rank (within/all)	1/1	9/21	7/17	8/20	6/16	5/8	3/4	4/7	2/2
Logistic normal distribution									
DF for used the fit	51.22	24.00	45.00	36.52	59.01	46.00	88.00	50.90	90.32
GAIC	−19,236.25	−18,840.09	−18,910.19	−18,847.49	−18,918.79	−19,039.03	−19,188.53	−19,043.25	−19,191.91
Rank (within/all)	1/3	9/19	7/15	8/18	6/14	5/10	3/6	4/9	2/5
Simplex distribution									
DF for used the fit	48.53	24.00	45.00	35.64	61.85	46.00	88.00	45.90	87.91
GAIC	−18,934.48	−18,287.79	−18,342.37	−18,298.71	−18,357.90	−18,732.35	−18,930.32	−18,733.90	−18,932.19
Rank (within/all)	1/11	9/27	7/25	8/26	6/24	5/23	3/13	4/22	2/12

Note: the row 'Rank' indicates the rank of the respective model when comparing all models with the same distributional assumption ('within') and all considered models ('all') based on the GAIC.
Sources: RDC of the Federal Statistical Office and Statistical Offices of the Länder, agricultural census 2010 and data from the OGA Lower Saxony and the Federal Statistical Office and Statistical Offices of the Länder, own calculations.

Appendix A2. Variables selected in model specification 1 under the different distributional assumptions ($n = 2{,}794$)

Distribution	Beta		Logistic normal		Simplex	
Variable	μ	σ	μ	σ	μ	σ
biogas_cap	✓	✓	✓			✓
biogas_share	✓	✓	✓	✓		
cattle						
d_grassl	✓	✓	✓	✓	✓	✓
farm_type_share	✓	✓	✓	✓	✓	
grass_share	✓		✓			✓
hhi		✓		✓		
hog_poultry		✓				✓
income	✓	✓	✓		✓	
labour		✓		✓		✓
landprice	✓		✓		✓	✓
organic_share		✓		✓		✓
parttime_share	✓	✓	✓			
pop_density	✓	✓	✓	✓	✓	
potato		✓		✓		✓
rent_share1				✓		✓
rent_share2	✓		✓	✓		✓
rye				✓		
size		✓		✓	✓	✓
sugarbeet				✓		
unemployment	✓		✓	✓	✓	✓
winterwheat		✓		✓		✓
f_{str}	✓	✓	✓	✓	✓	✓
f_{unstr}	✓		✓	✓	✓	
Interaction effects[a]						
d_grassl × *biogas_share*						✓
d_grassl × *farm_type_share*		✓		✓		
d_grassl × *hhi*		✓				
d_grassl × *hog_poultry*		✓				
d_grassl × *income*		✓				
d_grassl × *labour*		✓		✓		✓
d_grassl × *landprice*	✓		✓		✓	✓
d_grassl × *organic_share*		✓		✓		✓
d_grassl × *parttime_share*		✓				✓
d_grassl × *pop_density*	✓	✓	✓	✓		
d_grassl × *potato*		✓		✓		✓
d_grassl × *rent_share1*				✓		
d_grassl × *rent_share2*	✓		✓	✓		✓
d_grassl × *rye*				✓		
d_grassl × *size*					✓	
d_grassl × *sugarbeet*				✓		
d_grassl × *unemployment*	✓		✓	✓		✓
d_grassl × *winterwheat*		✓				

Note: [a]all other potential interaction effects were not selected in one of the models.