ESTIMATING LOCATION VALUES OF
AGRICULTURAL LAND

Matthias Ritter, Georg Helbing, Zhiwei Shen, Martin Odening

Matthias.Ritter@agrar.hu-berlin.de

Department for Agricultural Economics, Humboldt-Universität zu Berlin,
Philippstr. 13, 10115 Berlin, Germany

Financial support from the German Research Foundation (DFG) and the German Academic Exchange Service (DAAD) is gratefully acknowledged. We thank P. Kutschke and T. Gehrke of the *Oberer Gutachterausschuss für Grundstückswerte im Land Mecklenburg-Vorpommern* for providing the data.

2017

Vortrag anlässlich der 57. Jahrestagung der GEWISOLA
(Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaues e.V.)
und der 27. Jahrestagung der ÖGA
(Österreichische Gesellschaft für Agrarökonomie)
„Agrar- und Ernährungswirtschaft zwischen Ressourceneffizienz und gesellschaftlichen Erwartungen"
Weihenstephan, 13. bis 15. September 2017

ESTIMATING LOCATION VALUES OF AGRICULTURAL LAND

Abstract

"Bodenrichtwerte" reflect the average location value of land plots within a specific area. They constitute an important source of information that contributes to price transparency on land markets. In Germany, "Bodenrichtwerte" are provided by publicly appointed expert groups (Gutachterausschüsse). Using empirical data from Mecklenburg-Western Pomerania between 2013 and 2015, this article examines the relation between "Bodenrichtwerten" and statistically determined location values. It turns out that "Bodenrichtwerte" tend to underestimate location values of arable land by 11.5 percent on average. This underestimation can be traced back to the pronounced increase of land prices in the observation period. As an alternative to the expert-based determination of location values, we suggest a nonparametric smoothing procedure that rests on the Propagation-Separation Approach. The application of this data-driven procedure achieves an accuracy comparable to that of official "Bodenrichtwerte" at the one-year ahead prediction of location values without the requirement of expert knowledge.

Keywords

Land value; adaptive weight smoothing; agricultural land markets; propagation-separation approach; PSA.

1 Introduction

Information about realized prices is crucial for the price formation process on land markets. An important source of information that contributes to price transparency on land markets are location values, estimates for which (referred to as *Bodenrichtwerte*, BRW) are provided by publicly appointed expert groups (*Gutachterausschüsse*) in Germany. According to the Federal Building Code (BAUGESETZBUCH), BRW are intended to reflect the average location value (per square meter) of pieces of land. The purpose of these values is to reduce transaction costs related to real estate transactions by offering reliable benchmarks for purchases and taxation.

Unfortunately, three features of land markets impede the accurate estimation of location values. First, land markets are characterised by a relatively low liquidity. For example, in Germany on average only less than one percent of the agricultural area is sold each year (STATISTISCHES BUNDESAMT 2015). Actually, it may happen that only a few or even no land transactions take place within a particular sub-district (*Gemarkung*) during one or two years. As a consequence, estimating location values typically warrants pooling observations from sub-districts for which one can assume a similar location value. In practice, this entails a bias-variance trade-off: by including weighted observations from other sub-districts, one can reduce variance, but if the assumption of equal location value is violated, considerable bias may be incurred. The second feature that impedes estimation of location values is that land is an extremely heterogeneous asset: Its value depends on a variety of attributes and conditioning variables, such as soil quality, plot size, land use systems, or distance to cities[1]. This heterogeneity complicates a direct comparison of observed prices. The third characteristic that complicates the determination of BRW is the dynamics inherent to land markets. Changes in the location value of land may arise from changes in interest rates or agricultural product prices, technological change, or changes in legislation. To capture these dynamics, BRW are updated every two years at the latest. The

[1] See HÜTTEL ET AL. (2013) and the literature cited therein for an overview on land price determinants.

method to be applied in this task is comparative analysis, i.e., pooling prices of similar plots and adjusting prices for deviations of the underlying plot to make them comparable. For this purpose, homogeneous sub-districts showing similar price determining attributes, so-called location value zones (*Bodenrichtwertzonen*), are defined.

In view of the aforementioned characteristics of land markets, it is quite obvious that expert groups face a challenging statistical estimation problem. Observed transactions have to be filtered to reflect market conditions, i.e., purchases between family members, forced sales, or seizure should be ruled out. Moreover, prices that are untypical need to be identified as outliers and either adjusted or dropped. Finally, observed transactions need to be 'translated' to reflect typical land characteristics of the sub-district, which implies that observed prices have to be weighted or otherwise adjusted. While there are some clear procedures for filtering, much intuition is required for adjusting and weighting observed land prices when updating the location value estimates. In practice, expert knowledge comes into play at this point. In the case that no sufficient amount of transactions for pooling is available, 'deductive methods' may be applied (BUNDESMINISTERIUM FÜR VERKEHR, BAU UND STADTENTWICKLUNG 2011). These include the consideration of past location values and general market trends.

From a scientific point of view, the question arises if BRW actually reflect location values and how the procedure applied by the experts can be assessed. In particular, it would be interesting to analyse if BRW show systematic biases and if so, where and why these biases occur[2]. Any answer to these questions has to cope with the problem that location values are hypothetical values and thus unobservable. Nonetheless, given their definition, one would expect that BRW do not systematically deviate from realized prices in a location value zone.

This paper contributes to the evaluation of BRW as indicators of location values of agricultural land. However, we do not confine our analysis to a comparison of BRW and sample statistics of observed land prices. Rather, we propose a statistical smoothing procedure as a data driven alternative to the expert-based approach. More specifically, we make use of an adaptive smoothing procedure that has been introduced as the "Propagation-Separation Approach" (PSA) by POLZEHL and SPOKOINY (2006) into the literature. This method was originally developed as "Adaptive Weights Smoothing" in the context of image denoising (POLZEHL and SPOKOINY 2000). Recently, it has also been used in geology for the estimation of seismic parameter fields (GITIS ET AL. 2015) and in econometrics for the estimation of land values in an urban context (KOLBE ET AL. 2015). PSA is a nonparametric regression method that allows separating the underlying structure in the data from distorting noise by means of an iterative locally adaptive smoothing algorithm. Unlike conventional smoothing algorithms, such as fixed-bandwidth kernel regression, PSA does not only consider the distance between two locations when determining the weight of observations; rather, it adds a second component that takes into account the difference in resulting regression estimates. The attractiveness of PSA is based on an appealing statistical property: The estimator obeys a "small modelling bias condition" meaning that it shows the smallest variance given a predetermined bias which can be controlled by the econometrician (POLZEHL and SPOKOINY 2006). Thus, PSA addresses the variance-bias trade-off in pooling observations from different sub-districts. Previous applications have documented that PSA performs well, if data show large homogeneous zones that are separated by sharp discontinuities (BECKER and MATHÉ 2013). In contrast, SHEN ET AL. (2016) report that PSA has difficulties to identify outliers in otherwise homogeneous data. Thus, it is not clear whether PSA constitutes a superior alternative to the expert-based determination of location values. The application and the evaluation of this rather new statistical method constitutes the second contribution of our study.

[2] Biases of BRW could be rooted in the underlying methodological procedure. Apart from that, expert groups might have a tendency to update BRW conservatively in phases of booming land prices to dampen further price increases.

The remainder of the article is organized as follows: Section 2 describes the land transaction data from Mecklenburg-Western Pomerania that we use as the empirical basis of our analysis. Afterwards, we derive a benchmark for assessing the performance of location value estimators. In Section 3, we analyse whether BRW show a significant bias and what factors this hinges on. In particular, we are interested in whether there are any significant differences in bias between different expert groups. In Section 4, we introduce the PSA method in general and demonstrate how it can be applied to our data. Section 5 presents the results of an out-of-sample forecast application, which compares the performance of BRW and PSA at the one-year ahead prediction of location value. The paper ends with an assessment of the current practice of calculating BRW and answers the question if the use of formal statistical procedures can improve the informational content of BRW.

2 Empirical Data and Derivation of a Benchmark

In this study, we use a data set of purchases of arable land in Mecklenburg-Western Pomerania through the years 2013–2015.[3] We drop some transactions that are labelled as 'unsuited to analysis', since they took place between family members or show other irregularities that mark them as not being representative. We also cut off the lowest and highest percentiles of the prices for each year from 2013-15. This serves to remove extreme prices, which are unrealistic for agricultural land and are therefore most likely affected by some sort of error, e.g., a misplaced decimal point, or are a very untypical sale. Altogether, we obtain 4,374 observations over three years. The summary statistics in Table 1 depict an almost linear increase in mean land prices of about 0.23 EUR/m² from 2013 to 2015. The spatial unit of analysis that is used for location value estimation is the sub-district (*Gemarkung*), a historic administrative unit that is usually situated at a sub-municipality level. In Mecklenburg-Western Pomerania, there are 3,557 sub-districts altogether, which implies that in most years there is not even one observation per sub-district available. This gives rise to the necessity of using observations from several years for location value prediction. Experts may use deductive methods and their experience for this purpose. For PSA, we will pool time-adjusted prices from 2013 and 2014 as the basis for predicting the location values of 2015.

Table 1: Summary statistics of observed purchase prices, plot size and soil quality of sold pieces of land

Summary statistics	Plot size (ha)	Soil quality	Prices (EUR/m²)	Prices 2013 (EUR/m²)	Prices 2014 (EUR/m²)	Prices 2015 (EUR/m²)
Mean	8.89	38.18	1.64	1.43	1.64	1.92
Standard Deviation	19.94	8.14	0.76	0.66	0.72	0.82
Observations	4,374	4,278	4,374	1,651	1,479	1,244

Note: Soil quality is measured on a scale from 0 to 120 in ascending order. Different total counts result from missing soil quality values in the data set. In the subsequent analyses, the largest possible datasets are used.

In order to assess the predictive performance of BRW and PSA, we need to establish a benchmark, given that the true location values are not observable. We call this benchmark empirical location values (ELV). An important property of ELV is that by design they are an unbiased estimator of location value. Briefly, they are obtained by calculating the average price of a sub-district in a given year. However, we first perform an adjustment of the observed purchase prices. This step serves to reduce the variance of ELV by shifting observed prices towards the expected value, which is particularly useful to mitigate extreme prices and to some

[3] Data source: *Landesweite Datensammlung des Oberen Gutachterausschusses für Grundstückswerte im Land Mecklenburg-Vorpommern (OGAA M-V)*

extent should compensate the fact that in many sub-districts only few transactions are observed per year. Adjustment consists in subtracting from the observed prices the effects of certain individual plot characteristics, e.g., an above-average fertility, so that we obtain the price that would have been realised had the transacted plot been ‘typical’ for its sub-district.

To calculate the effects of conditioning variables, we set up a linear regression model for (log) land-prices. We consider soil quality and plot size as covariates. Soil quality is known to have a considerable influence on land prices (e.g., HENNIG ET AL. 2014). Plot size on the other hand is included because we hypothesize that large plot sizes tend to be sold by the federal trust (BVVG) that is in charge of administrating formerly state owned land. It is not unlikely that the prices from these sales differ from sales among private parties (HÜTTEL ET AL. 2016). Given the observed linear trend in our data, we also account for temporal effects by including time dummy variables. Hence, we fit the following log-linear regression model to our data (see next paragraph for details):

$$\log(p_{i,j,t}) = \alpha s_{i,j,t} + \beta q_{i,j,t} + \gamma I_{i,j,t,2014} + \delta I_{i,j,t,2015} + b + \varepsilon_{i,j,t} \qquad (1)$$

where, $s_{i,j,t}$ denotes plot size of transaction i in sub-district j in year t, $q_{i,j,t}$ denotes the corresponding soil quality, $I_{i,j,t,2014}$ and $I_{i,j,t,2015}$ are time dummy variables indicating the year the transaction took place in, and b is a constant. The subsequent adjustment step corrects actual prices for effects of above-average or below-average values of soil quality and plot size:

$$\log(\tilde{p}_{i,j,t}) = \log(p_{i,j,t}) - \hat{\alpha}(s_{i,j,t} - \overline{s}_j) - \hat{\beta}(q_{i,j,t} - \overline{q}_j) \qquad (2)$$

where $\tilde{p}_{i,j,t}$ denotes adjusted prices. We determine average soil quality $\overline{q}_j$ and average plot size $\overline{s}_j$ of sub-district j by taking the mean soil quality and plot size of all sold plots in that sub-district from 2013–15 (see Table 1 for summary statistics). Note that we do not adjust for temporal effects, because we want to estimate time-varying location values. In a final step, the ELV $\tilde{\theta}_{j,t}$ of sub-district j in year t is derived by retransforming the adjusted log-price with the exponential and taking the sub-district- and year-wise mean of the adjusted prices:

$$\tilde{\theta}_{j,t} = \frac{1}{n_{j,t}} \sum_{i=1}^{n_{j,t}} \tilde{p}_{i,j,t} \qquad (3)$$

where $n_{j,t}$ denotes the number of observations in sub-district j in year t.

The model in Eq. (1) is estimated with OLS yielding highly significant effects for all covariates, as displayed in Table 2. The effects of the years 2014 and 2015 reflect the upward trend of land prices observed in our data. Soil quality has a positive effect on land prices as expected. Plot size, too, shows a positive effect. We performed a Beusch-Pagan test confirming that the residuals are homoscedastic (p-value 0.0494). We are aware that the rather simple model in Eq. (1) may not capture heterogeneity of land prices completely, but the moderate model fit suggests that ELV constitute a fair approximation of the true location value.

Table 2: Regression model for price adjustment

Covariate	Effect (EUR/m²)	Standard error
Intercept	-0.7142***	0.0326
Year 2014	0.1327***	0.0156
Year 2015	0.2881***	0.0163
Soil Quality	0.0243***	0.0008
Plot Size (ha)	0.0044***	0.0000

Note: The effects refer to log-prices. $R^2 = 0.26$. *** denotes significance at the 1 percent level.

To measure the performance of a location value predictor, we use the mean squared error (MSE) and the bias. The calculation basis for these measures is the so-termed 'observed deviation', which denotes the deviation $\hat{\theta}_{j,t} - \tilde{\theta}_{j,t}$ of a predicted value $\hat{\theta}_{j,t}$ from the ELV $\tilde{\theta}_{j,t}$, that we observe for each sub-district j and year t. To compare different predictors, we use the MSE with respect to ELV as a measure of performance in this study. Finally, we are interested in the bias of a predictor, which we estimate with the mean observed deviation.

3 Analysing BRW Bias and Deviations from Empirical Location Values

In this section, we will have a closer look at BRW as one-year ahead predictor with the goal of assessing bias and identifying the factors that explain the observed deviation. It is important to note that our data set does not contain BRW for all sub-districts, so we have to perform this analysis on the subset ('BRW test set') of sub-districts and years for which we have a BRW and at least one suitable transaction. This leaves us with 900 (in 2013), 664 (in 2014) and 808 (in 2015) sub-districts, respectively.

Figure 1. Distribution of ELV and BRW and observed deviation in the BRW test set

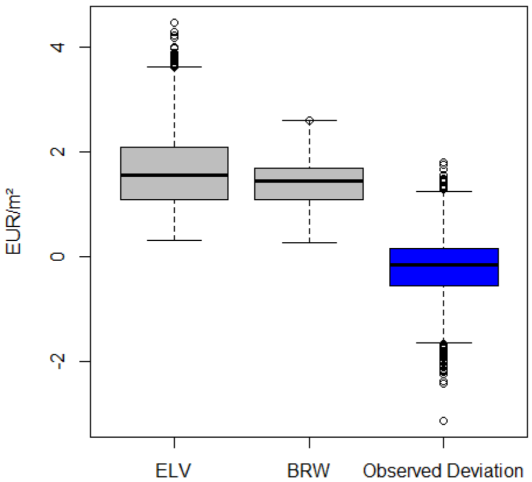
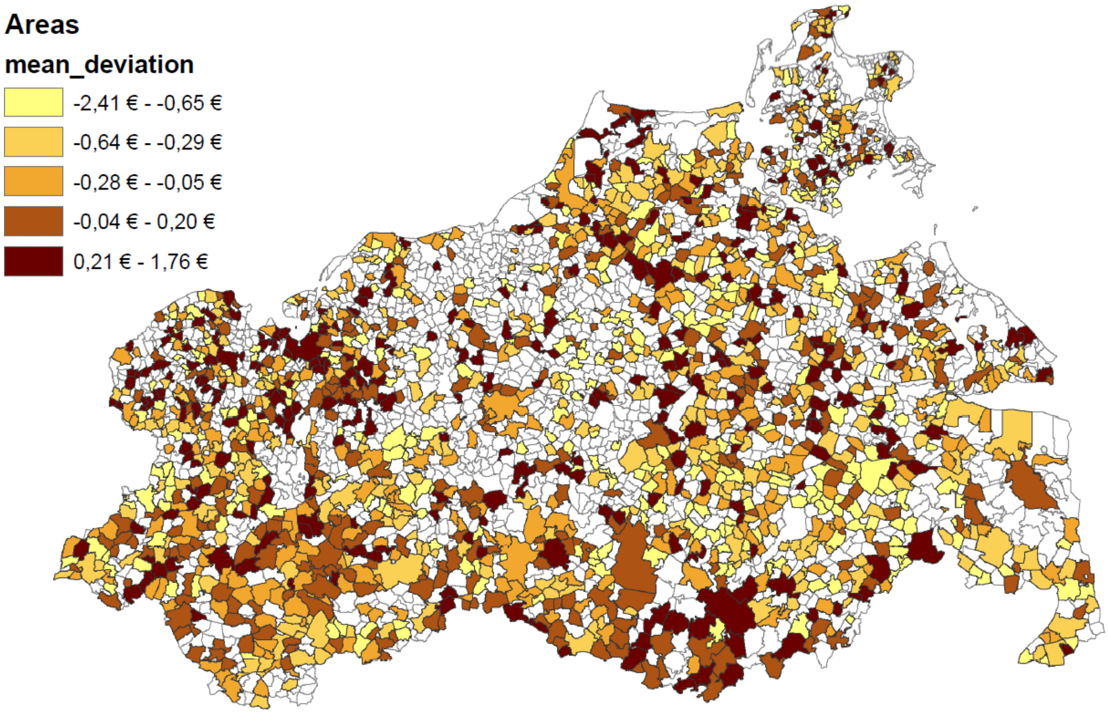


Figure 1 displays boxplots of ELV, BRW, and the observed deviation of BRW. We find that compared to ELV, BRW show a smaller variability as well as a lower price level, which points at an underestimation of location values. We therefore expect to find a significant bias for BRW. For the test set, we obtain for BRW a bias of -0.22 EUR/m². This means an underestimation by 11.5 percent in relation to average land prices in 2015. In order to infer whether this figure is statistically significant, we perform a one-sample t-test for the null hypothesis of zero bias. From the resulting p-value $< 10^{-15}$ we conclude that BRW has a significant, negative bias. To make our result more robust to violations of the normality assumption underlying the t-test, we also perform Wilcoxon's signed-rank test for the null-hypothesis of the median being equal to zero. This test only requires the weaker assumption that the distribution of the observed deviation is symmetric, which is approximately given, as illustrated by Figure 1. Here again we obtain a p-value $< 10^{-15}$, which corroborates the previous result. This shows that BRW actually tend to underestimate location values.

Figure 2 depicts the spatial distribution of observed differences between BRW and ELV. Apparently, there are some regional clusters, in particular in the central South and the North-West. This observation suggests that systematic factors exist that explain the bias of BRW. To analyse the observed deviation of BRW from ELV further, we develop a linear regression model for the absolute value of the observed deviation – this does not cover the direction of the deviation, but only its magnitude.

Figure 2. Mean deviation of BRW from ELV per sub-district from 2013 to 2015



Note: Categories are quintiles. The blank sub-districts denote a lack of BRW values and/or transactions

To determine what factors lead to an over- or underestimation, we furthermore perform a logistic regression of the sign of observed deviation against the same factors. As explanatory variables, we again consider the indicators of soil quality and plot size, a time dummy and a categorical variable indicating which expert group determined the BRW. The rationale of choosing these covariates is as follows: One might conjecture that experts tend to oversmooth location values in areas with high soil quality, i.e., high land prices. Likewise, experts may have difficulties to smooth prices for small plots, which are often sold at high prices (per square meter). Moreover, since BRW are not continuously updated, they may lag behind the actual development of location values, particularly during a period of booming prices. Finally, the expert groups themselves may have an impact on the bias, because BRW are not calculated with a clear algorithm but involve personal judgements that may differ among expert groups. However, the effect of this variable has to be interpreted with caution, because it is difficult to separate the impact of experts from unobserved regional effects. As both expert group and year are categorical variables and we use a model without a constant, we have to exclude one dummy variable from the model. We chose the time dummy for 2013, which is then the reference year. All expert group dummies are included so that they can be interpreted as regional fixed effects. To better quantify the regional effects, we use centred versions of the variables 'average plot size' and 'average soil quality' by subtracting their individual means.

Table 3 summarises the results of the regression model estimated with OLS. Both years as well as average plot size and average soil quality are significant at least at the 5 percent level. Note that we have already adjusted ELV for the effects of soil quality and plot size of individual transactions. The effects of this regression model therefore refer to properties of a sub-district, not of transactions. The effect of average plot size is significant at the 5 percent level, yet – at less than 0.01 EUR/ha and 0.05 EUR for a sub-district with mean average plot size $\overline{s}_j$ of 9.77 ha – rather small in magnitude. Average soil quality has an effect of 0.17 EUR/m² for a sub-district of mean average soil quality $\overline{q}_j$ of 38.94. Temporal effects are in the same order of magnitude as average soil quality. The magnitude of bias in BRW increases with every year, which we attribute to the linear increase in mean land prices that we have observed between 2013 and

2015. It seems as though BRW do not sufficiently take market trends into account. As for expert effects, we find that all expert groups show effects significant at the 1 percent level, ranging from 0.36 EUR/m² to 0.48 EUR/m². This means that there is a significant deviation in 2013 for all expert groups, which even increases in the following years. To determine, however, if a systematic over- or underestimation is present, we perform a logistic regression.

Table 4 summarises the effects of our covariates on the probability of BRW overestimating (positive sign) or underestimating (negative sign) location value. We find that average plot size shows a significant negative effect, meaning that the larger transacted plots in a sub-district on average, the more does BRW tend to underestimate its location value. The results for average soil quality do not show any significant effect for the direction of the bias. Finally, we see that all expert groups tend to underestimate location values in 2013, even though this effect is comparatively weak and not significant for Group 4. In the years 2014 and 2015, no significant change occurs in this regard.

Table 3: Coefficients of BRW deviation OLS linear regression.

Covariate	Coeff.	Std. error
Year 2014	0.0467**	0.0213
Year 2015	0.1179***	0.0197
Avg. Plot Size (ha)	0.0051***	0.0000
Avg. Soil Quality	0.0043***	0.0015
Expert Group 1	0.4502***	0.0213
Expert Group 2	0.4583***	0.0300
Expert Group 3	0.4779***	0.0224
Expert Group 4	0.4535***	0.0268
Expert Group 5	0.4051***	0.0281
Expert Group 6	0.3600***	0.0231

Note: Dependent variable is the absolute value of deviation. $R^2 = 0.603$. ** and *** denote significance at the 5 and 1 percent levels, respectively.

Table 4: Coefficients of BRW deviation logistic regression.

Covariate	Coeff.	Std. error
Year 2014	-0.0101	0.1103
Year 2015	0.0598	0.1030
Avg. Plot Size (ha)	-0.0349***	0.0000
Avg. Soil Quality	-0.0049	0.0077
Expert Group 1	-0.6448***	0.1119
Expert Group 2	-0.8678***	0.1634
Expert Group 3	-0.4858***	0.1162
Expert Group 4	-0.0799	0.1374
Expert Group 5	-0.9822***	0.1535
Expert Group 6	-0.6232***	0.1204

Note: Dependent variable is the sign of deviation (1 = non-negative, 0 = negative). The model has been fit with ML. Nagelkerke pseudo-$R^2 = 0.054$. *** denotes significance at the 1 percent level.

To summarise the findings of this section, our analysis shows that there is a significant negative bias in BRW – meaning that experts systematically underestimate location value in our BRW test set. This underestimation may be linked to the fact that in the years covered by our study, we observe a nearly linear increase in land prices, suggesting that experts do not sufficiently take the trend into consideration, which is corroborated by our regression analysis of BRW deviation from ELV. This analysis has further shown that high average soil quality in a sub-district likewise increases deviation, but in both directions; market trend therefore does not appear to be the only source of erroneous assessment, but it accounts more than other factors for the observed bias. Finally, we have found some heterogeneity between expert groups, which can also be interpreted as regional heterogeneity.

4 A Propagation-Separation Approach for Estimating Location Value

In the introduction to this paper, we have pointed out that data scarcity requires to pool observations from different sub-districts to estimate location values. Depending on how the pooling is carried out, it trades a reduced variance for an increased bias. In the previous section, we have seen that BRW show a relatively low variance compared to the benchmark, but at the same time are afflicted by a significant bias. In the present section, we introduce a statistical procedure, which unlike BRW selects the sub-districts used for pooling in a purely data-driven way for every sub-district.

The "Propagation-Separation Approach" (PSA; POLZEHL and SPOKOINY 2006) is an iterative, adaptive procedure based on local constant regression. The underlying idea of this approach is to find for every point x_i a maximal local neighbourhood in which the local constant parametric assumption is not violated – in other words, in which we can assume equal location value. At the beginning of the procedure, a small neighbourhood $U^0(x_i)$ of every point x_i is considered to estimate the location value $\theta(x_i)$.[4] Afterwards, in each step k, we update the initial location value estimate by including new points x_j from an extended neighbourhood $U^k(x_i)$; but those candidates x_j are tested for homogeneous location value and only used for re-estimation of location value if the hypothesis of local homogeneity $\theta(x_i) = \theta(x_j)$ is not rejected. This iterative procedure is continued until we reach a pre-defined maximal radius of the neighbourhood.

The underlying local regression model for estimating the location values can be described as

$$y_i = \theta(x_i) + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma_i^2) \tag{4}$$

where y_i denotes the observed log price of agricultural land, x_i is a vector of explanatory variables which determine the distribution of observation y_i. Since we are interested in finding sub-districts with homogeneous location values, x_i simply refers to location coordinates $[x_{1i}, x_{2i}]$ in our case.[5] In a local regression model, the local parameter $\theta(x_i)$ can be estimated by the weighted maximum likelihood estimation where a nonnegative weight $w_{ij} = w_{ij}(x_i) \leq 1$ is given to each observation y_j, $i, j = 1, \dots, n$. The corresponding local maximum likelihood estimator for a fixed x_i is given by:

$$\tilde{\theta}(x_i) = \underset{\theta}{\operatorname{argmax}} \sum_{j=1}^{n} w_{ij}(x_i) \log p(y_j, \theta), \tag{5}$$

where $p(\cdot, \theta)$ denotes the density function. In the case of the density function $p(\cdot, \theta)$ from the exponential family functions, for instance Gaussian distribution, POLZEHL and SPOKOINY (2006) have shown that the explicit solution of (5) is in fact a Nadaraya-Watson estimator:

$$\tilde{\theta}(x_i) = \frac{\sum_{j=1}^{n} w_{ij}(x_i) y_j}{\sum_{j=1}^{n} w_{ij}(x_i)}. \tag{6}$$

As above mentioned, the PS approach is an iterative procedure, and in each iteration step, the local estimator is defined as a weighted mean of observations. Therefore, in iteration step k (i.e., within the neighbourhood $U^k(x_i)$), the adaptive local estimator $\tilde{\theta}^k(x_i)$ is

$$\tilde{\theta}^k(x_i) = \frac{\sum_{j=1}^{n} w_{ij}^k(x_i) y_j}{\sum_{j=1}^{n} w_{ij}^k(x_i)}. \tag{7}$$

The main advantage of the PS approach arises from the construction of the weights $w_{ij}^k(x_i)$. The determination of weights in the PS approach not only considers the likeness of the data with the sub-district of interest, but also controls the bias possibly introduced from the extension of data samples. To be specific, the weights depend on the product of two components: the location component K_{loc} and the homogeneity component K_{hom}:

$$w_{ij}^k = K_{\text{loc}}(l_{ij}^k) K_{\text{hom}}(s_{ij}^k), \tag{8}$$

[4] For our analysis, the initial neighbourhood includes only x_i itself.

[5] We use the coordinates of a sub-district's centre point as coordinates of the sub-district.

where $K_{\mathrm{loc}}(\cdot)$ and $K_{\mathrm{hom}}(\cdot)$ are two kernel functions that are non-negative and strictly monotonically decreasing on the support [0, 1], for example the triangular kernel function. Similar to the standard nonparametric regression, the argument in the location component K_{loc} is the Euclidean distance measure between the locations i and j divided by the bandwidth h^k:

$$l_{ij}^k = \frac{\rho(x_i, x_j)}{h^k} \qquad \textbf{(9)}$$

On the other hand, $s_{ij}^k = \frac{T_{ij}^k}{\lambda}$ in the homogeneity component is a statistical penalty where T_{ij}^k is the test statistic for a constant local parametric estimate and λ is the critical value of the test statistic T_{ij}^k. The homogeneity component $K_{\mathrm{hom}}\left(s_{ij}^k\right)$ becomes relevant for controlling the bias when extending the size of neighbourhood $U^k(x_i)$. To test the hypothesis of local homogeneity $\theta(x_i) = \theta(x_j)$ at each step k, the estimates $\tilde{\theta}_i^{k-1}(x_i)$ and $\tilde{\theta}_j^{k-1}\left(x_j\right)$ obtained from the previous iteration is compared. Following POLZEHL and SPOLOINY (2006) and BECKER and MATHÉ (2013), the test statistic T_{ij}^k is constructed based on the Kullback-Leibler divergence between the pointwise parameter estimates of the previous iteration step at two different points.

The crucial parameter of PS approach is the critical value λ that determines the number of observations to be used in the estimation of each location value. Greater values of λ allow the inclusion of more points into a homogeneous region, leading to a smoother parameter surface and potentially a higher bias at reduced variance. In fact, for $\lambda \to \infty$, we obtain a non-adaptive kernel smoother. On the other hand, smaller values of λ will lead to a stricter selection of homogeneous regions and less points being included into the estimation. As a result, less available information is used and the variance of the estimate is generally higher. Due to the multiple testing procedure in this adaptive algorithm, there is no well-defined choice of λ (KOLBE ET AL., 2015). POLZEHL and SPOKOINY (2006) suggest performing Monte Carlo simulations of the relevant likelihood function with globally constant parameters on the design space. λ can then be chosen as the smallest value that ensures the homogeneity assumption holds everywhere with a high probability. For computing PSA estimates, we use the package 'aws' for the statistical software R (POLZEHL 2016). Here, an adequate simulation-based choice of λ is provided automatically for a given set-up.

5 Comparing BRW and PSA

In this section, we compare the performance of BRW and PSA at the one-year ahead predictions of location values. For this purpose, we use a training set for PSA based on adjusted prices from 2013 and 2014, and a test set of ELV from 2015 for validation purposes. As explained in Section 2, the low number of observations in 2014 requires that we pool data from 2013 and 2014. Moreover, it is convenient that for obtaining our training set, we use the same procedure that we previously applied to compute ELV, but only taking into account observations from 2013 and 2014 since we cannot include information from the test set. In particular, we use Eq. (2) for price adjustment where we furthermore add the estimated temporal effect $\hat{\gamma}$ to observations from the year 2014. This approach to temporal pooling is very similar to the deductive methods available to land price experts. The resulting prices reflect the 2014 price level of typical plots. As with ELV, we compute the mean per sub-district and obtain a training set of 1,556 average prices that represent the initial location value estimates for PSA. There are, however, 3,557 sub-districts in Mecklenburg-Western Pomerania, so we do not have PSA estimates of the 2015 location values for all sub-districts; moreover, we do not have corresponding BRW for all sub-districts, either. Consequently, we have to filter the 2015 ELV data by selecting only those sub-districts, for which we have a value in the PSA training set and a BRW to enable a fair comparison. This shrinks the number of sub-districts in the test set to 502.

As explained above, PSA has two parameters λ and h_{max} that control the threshold of the homogeneity test and the maximum distance of observations that are included in local estimation, respectively. For our PSA baseline predictor, λ is set to 9.72 by suggestion of POLZEHL AND SPOKOINY (2006) (cf. Section 4). h_{max} can be selected such that for any cell on the grid, all other cells lie within the maximum distance. As we use a 100x100 grid, we set h_{max} to 150, which is slightly greater than the length of the grid's diagonal. This is the configuration of our default PSA predictor 'PSA1'.
To demonstrate the sensitivity of the results to parameter choice, we also perform PSA with a reduced value of h_{max} ('PSA2') as well as with smaller ('PSA3') and greater ('PSA4') values of λ. Furthermore, we seek to account for the fact that expert-based estimates can leverage trends observed during the past years for prediction, whereas PSA is limited to synchronous data. To reflect this possibility, we combine PSA with a linear trend, based on the effect γ from the regression model in Eq. (1) fitted to the 2013/14 data. We compute this trend-adjusted predictor ('PSA5') as $\frac{\hat{\theta}}{e^{\hat{\gamma}}}$, where $\hat{\theta}$ is the PSA baseline predictor. If no abrupt change in trend is expected for the next year, this should improve the PSA estimate significantly. An overview of the predictors used in our analysis and of their characteristics is provided in Table 5.

Table 5: Characteristics of the used predictors

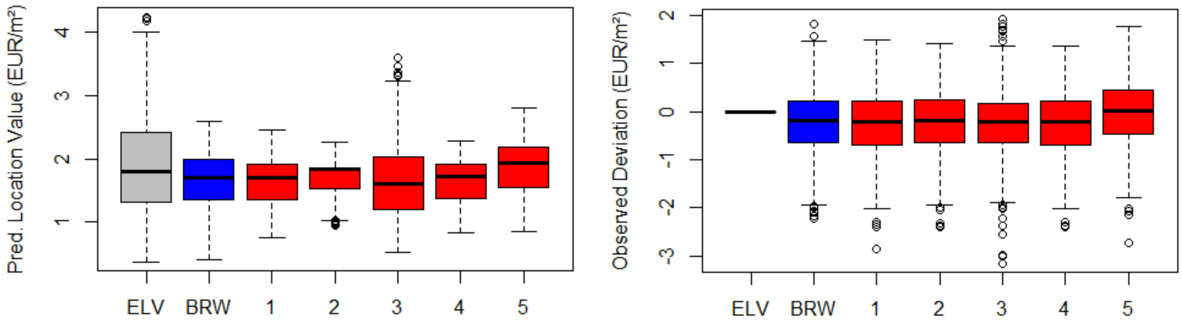
Predictor	Description
BRW	Expert based location value
PSA1	$\lambda = 9.72$; $h_{max} = 150$
PSA2	$\lambda = 9.72$; $h_{max} = 10$
PSA3	$\lambda = 0.972$; $h_{max} = 150$
PSA4	$\lambda = 97.2$; $h_{max} = 150$
PSA5	PSA1 trend-adjusted

Figure 3 contains in its left panel boxplots of the distributions of empirical location values in the test set and the predicted values. The right panel displays boxplots of the differences between predicted and empirical location values. The more a predictor's deviations from ELV are centred around zero, the less bias it has. A first impression is that BRW as well as PSA predictors have a significant bias, with the single exception to the trend-adjusted PSA5. Altogether, the distributions of observed deviation are quite similar. To formally compare the predictors, we compute the MSE and test whether the predictors (i) have a bias significantly different from zero and (ii) have a significantly smaller bias than BRW. For (i), we perform one-sample t-tests assuming a non-homogeneous variance, and additional non-parametric Wilcoxon's signed-rank tests as in Section 3. For (ii), we carry out two-sample t-tests assuming a non-homogeneous variance (Welch test) and, to make the results robust against a violation of the t-test's normality assumption, Wilcoxon's rank-sum test.

Table 6 lists the results of these tests as well as the mean squared error (MSE) for every predictor. We find that all predictors except for PSA5 have a significant, negative bias in the same order of magnitude. The MSE, too, indicates a similar performance of all PSA predictors and BRW, with the MSE of PSA5 of course being lower due to its reduced bias.

In summary, our results show that PSA in various configurations can reach the same level of accuracy in terms of MSE and bias as BRW. Since, apart from PSA5, none of the PSA estimators have shown less bias than BRW, we find that its data-driven approach to pooling does not show any apparent advantage over the fixed BRW zones. The substantial improvement of PSA5 achieved by considering linear trend on top of PSA indicates how strongly the general market trend from 2013–15 impacts on the performance of predictors. Indeed, the fact that, like PSA, BRW does not seem to take trend into consideration would explain the negative bias, especially seeing as the increase in mean land prices from 2014 to 2015 (0.28 EUR/m²) lies in the same order of magnitude.

Figure 3. Left: Empirical and predicted location values for 2015. Right: Observed deviation in 2015.



Note: Observed deviation is the differences between predicted and empirical location value. 1–5 denote PSA1–PSA5.

Table 6: Estimated bias, test statistics of the applied tests and the MSE

Predictor	Bias (EUR/m²)	Test (i): One-sample t-test	Test (i): Wilcoxon signed rank test	Test (ii): Welch test	Test (ii): Wilcoxon rank sum test	MSE (EUR/m²)²
BRW	-0.25	-8.1858***	39368***	-	-	0.5194
PSA1	-0.26	-8.9151***	38215***	-0.3835	125210	0.5081
PSA2	-0.26	-8.8788***	38513***	-0.3699	125480	0.5095
PSA3	-0.25	-7.9708***	38067***	-0.1433	124400	0.5725
PSA4	-0.26	-8.8116***	38818***	-0.3056	125700	0.5055
PSA5	-0.03	-0.94044	62974	5.21***	150250***	0.4347

Note: ELV for all predictors; *** denotes significance at the 1 percent significance level.

6 Discussion and Conclusion

In our analysis based on purchase prices of arable land in Mecklenburg-Western Pomerania over the years 2013–15, we have found that BRW significantly underestimates location values of the following year. A regression analysis of the observed deviation has pointed towards regional heterogeneity, soil quality, and temporal effects as explanatory factors of this deviation. Indeed, we observe a strong linear increase in mean land prices for every year from 2013–15, which suggests that the time trend is not sufficiently taken into account in BRW estimation. However, soil quality also shows a strong effect, suggesting that experts have difficulties in correctly considering soil quality for location value estimation. Secondly, we find that on our 2015 test data, PSA predicts location values with an accuracy comparable to that of BRW, both in terms of bias and MSE. These findings are in line with KOLBE ET AL. (2015), who find that PSA is able to replicate BRW in an urban context. The performance depends to a limited degree on the choice of the algorithm's parameters, but neither bias nor MSE have proven too sensitive in this regard. Since PSA does not achieve a reduction of bias, it appears as though its adaptive approach does not hold any advantage over fixed BRW zones in the estimation of location values of agricultural land. On the other hand, the substantial performance improvement when linear trend is taken into consideration hints at a great potential for improving BRW as location value predictor by complementing this approach with conventional forecasting techniques.

A practical issue with PSA is that outliers are usually not smoothed by PSA – the reason being that the homogeneity test, which is performed at every iteration when smoothing the sub-district with the outlier, will most certainly result in zero weights for most values other than the outlier itself. On the one hand, this is precisely the sort of behaviour that we wish, because it keeps the bias low when pooling values. On the other hand, it does not allow us to reach a reasonable estimate for the outlier itself. The reasons for the occurrence of such singular values may be manifold, and it is impractical to derive a general rule of treating them – in this analysis, we have opted for an *a priori* removal of the highest and lowest percentiles of prices. Our original concern that the results might be too sensitive to the choice of parameters has proven unjustified after this outlier removal. It seems that results for different parameters diverge more strongly in the presence of extreme values.

One limitation to our results is that our data set is of rather limited size. Carrying out similar calculations for other regions with a longer time series of land prices and BRW could improve the reliability of our findings. Moreover, our observation period is characterised by a strong linear upward trend of mean land prices. Further assessment of BRW and PSA on data without such a trend might elucidate if the performance of PSA holds under different market conditions, too. This caveat notwithstanding, we have found PSA to be a convenient tool for the automatic estimation of location value of agricultural land in a transparent way since no expert knowledge is required for the procedure. Such a tool can complement the expert-based approach and serve as a benchmark.

References

BAUGESETZBUCH as published on 23rd September 2004 (BGBI. I S. 2414), last changed by article 6 of the law of 20th October 2015 (BGBI. I p. 1722).

BECKER, S. (2014): The Propagation-Separation Approach - Theoretical Study and Application to Magnetic Resonance Imaging. PhD Thesis, Humboldt University at Berlin, Mathematisch-Naturwissenschaftliche Fakultät II.

BECKER, S., and MATHÉ, P. (2013): A different perspective on the Propagation-Separation Approach. Electronic Journal of Statistics, Vol. 7 (2013) 2702–2736, DOI: 10.1214/13-EJS860.

BUNDESMINISTERIUM FÜR VERKEHR, BAU UND STADTENTWICKLUNG (2011): Richtlinie zur Ermittlung von Bodenrichtwerten (Bodenrichtwertrichtlinie – BRW-RL). Bundesanzeiger of 11th February 2011 (no. 24, p. 597).

GITIS, V. G. A., DERENDYAEV, B., PIROGOV, S. A., SPOKOINY, V. G., YURKOV, E. F. (2015): Adaptive Estimation of Seismic Parameter Fields from Earthquake Catalogs. Journal of Communications Technology and Electronics, 2015, Vol. 60, No. 12, pp. 1459–1465.

HENNIG, S., BREUSTEDT, G., LATACZ-LOHMANN, U. (2014): Zum Einfluss mitgehandelter Zahlungsansprüche auf die Kauf- und Pachtpreise von Ackerland in Schleswig-Holstein. German Journal of Agricultural Economics 63: 219–239.

HÜTTEL, S., WILDERMANN, L., CROONENBROEK, C. (2016): How do institutional market players matter in farmland pricing? Land Use Policy 59: 154–167.

HÜTTEL, S., ODENING, M., BALMANN, A. (eds.) (2013): Agricultural Land Markets – Recent Developments and Determinants. German Journal of Agricultural Economics – Special Issue Vol. 62(2).

KOLBE, J., SCHULZ, R., WERSING, M., WERWATZ, A. (2015): Identifying Berlin's land value map using adaptive weights smoothing. Computational Statistics (2015) 30:767–790, DOI 10.1007/s00180-015-0559-9.

POLZEHL, J., and SPOKOINY, V. G. (2000): Adaptive Weights Smoothing with Applications to Image Restoration. Journal of the Royal Statistical Society. Series B (Statistical Methodology), Vol. 62, No. 2 (2000), pp. 335–354.

POLZEHL, J., and SPOKOINY, V. G. (2006): Propagation-Separation Approach for Local Likelihood Estimation. Probability Theory and Related Fields 135:335–362.

POLZEHL J. (2016): aws: Adaptive Weights Smoothing. http://cran.r-project.org/package=aws. R-package version 1.9-6.

SHEN, Z., ODENING, M., OKHRIN, O. (2016): Adaptive local parametric estimation of crop yields: Implication for crop insurance ratemaking. Paper prepared for presentation at the 156th EAAE Seminar 'Prospects for agricultural insurance in Europe', Wageningen, Netherlands, October 3–4, 2016.

STATISTISCHES BUNDESAMT (2015): Kaufwerte für landwirtschaftliche Grundstücke 2015. Fachserie 3, Reihe 2.4. Wiesbaden.