

Jens Kolbe, Rainer Schulz, Martin Wersing, Axel Werwatz

Identifying Berlin's land value map using adaptive weights smoothing

Article, Postprint version

This version is available at <https://depositonce.tu-berlin.de//handle/11303/7380>.



Suggested Citation

Kolbe, Jens; Schulz, Rainer; Wersing, Martin; Werwatz, Axel: Identifying Berlin's land value map using adaptive weights smoothing. In: Computational Statistics. - ISSN: 1613-9658. - 30 (2015). - pp. 767-790. DOI: 10.1007/s00180-015-0559-9.

Terms of Use

German copyright applies. This is a post-peer-review, pre-copyedit version of an article published in Computational statistics. The final authenticated version is available online at: <https://doi.org/10.1007/s00180-015-0559-9>.

Identifying Berlin's land value map using Adaptive Weights Smoothing

Jens Kolbe, Rainer Schulz, Martin Wersing, and Axel Werwatz*

*Kolbe and Werwatz: Technische Universität Berlin, Institut für Volkswirtschaftslehre und Wirtschaftsrecht, Straße des 17. Juni 135, 10623 Berlin, Germany, and Collaborative Research Center 649 *Economic Risk*, Humboldt-Universität zu Berlin. Emails: j.kolbe@tu-berlin.de and axel.werwatz@tu-berlin.de. Schulz and Wersing: University of Aberdeen Business School, Edward Wright Building, Dunbar Street, Aberdeen AB24 3QY, United Kingdom. Emails: r.schulz@abdn.ac.uk and martin.wersing@abdn.ac.uk.

Abstract

We use Adaptive Weights Smoothing (AWS) of Polzehl and Spokoiny (2000, 2003, 2006) to estimate a map of land values for Berlin, Germany. Our data are prices of undeveloped land that was transacted between 1996-2009. Even though the observed land price is an indicator of the respective land value, it is influenced by transaction noise. The iterative AWS applies piecewise constant regression to reduce this noise and tests at each location for constancy at the margin. If not rejected, further observations are included in the local regression. The estimated land value map conforms overall well with expert-based land values. Our application suggests that the transparent AWS could prove a useful tool for researchers and real estate practitioners alike.

Keywords: land value, adaptive weight smoothing, spatial modeling

JEL Classification: C14, R14, R15

1 Introduction

The scope of flexible, local regression modelling has been greatly expanded in the past decades, both by computational advances and increased data availability. Nonparametric kernel regression, in particular, has become a standard tool of applied statistics and is nowadays implemented in most statistical software packages. Kernel regression only assumes that the regression function is smooth. Smoothness, however, is an unattractive assumption in applications where the regression function is likely to possess jumps or sharp edges. Moreover, smoothness is at the heart of the well-known ‘curse of dimensionality’: the implied continuously changing regression function makes kernel regression very data-hungry, resulting in the low precision of kernel estimates in multivariate settings.

Semiparametric regression models, the topic of this special issue of Computational Statistics, have been developed to overcome the problems associated with the ‘over-flexibility’ of nonparametric regression. Additive regression models, for instance, impose an additively separable structure on the regression relation in order to overcome the curse of dimensionality (Stone, 1985). They maintain, however, that each component function is smooth. Additive models thus are also not suitable for situations with ‘edgy’ regression functions.

A well-known example of edgy regression is image denoising. From a statistical perspective, image data can be regarded as a noisy representation of the image of interest. The underlying image is regarded as the regression function to be recovered by a suitable estimation method. A challenge for regression modelling of image data are the specific structural features of images: they are typically composed of several regions (e.g. organs or tumors in medical images) with pronounced edges. Moreover, within each region of the image (i.e. the regression function) image values are rather homogenous. Smooth nonparametric regression is unsuitable in this setting as it can neither cope with the edges of the regression function nor does it exploit its local homogeneity. A suitable regression method for this situation was proposed by Polzehl and Spokoiny (2000, 2003). They developed Adaptive Weights Smoothing (AWS) originally in the context of image denoising. In this paper, we demonstrate that AWS can also be applied to land value estimation, a problem of considerable interest in economics that shares the main structural features of the image denoising problem.

AWS, which has been extended into the propagation-separation approach¹, replaces the

¹See Polzehl and Spokoiny (2006). For simplicity, we nonetheless refer to the approach as AWS throughout

smoothness assumption of nonparametric kernel regression with the assumption that the unknown regression function is constant or can be at least approximated by a constant for a set of observations $U(\mathbf{x}_i)$ around a location \mathbf{x}_i . AWS uses an iterative algorithm to determine this set. For each observation i , the algorithm starts with a small neighborhood $U^0(\mathbf{x}_i)$ of locally close observations and uses these to estimate the local constant. In each step k , the algorithm enlarges the set of observations for each i and includes all those observations \mathbf{x}_j into $U^k(\mathbf{x}_i)$ for which the hypothesis of local constancy at i and j cannot be rejected. In this situation, propagation of the set of observations takes place. If, however, local constancy for observations i and j is rejected, separation takes place. \mathbf{x}_j does not become member of $U^k(\mathbf{x}_i)$. The algorithm stops in step k^* and the observations in $U^{k^*}(\mathbf{x}_i)$ are used for the final estimate of the constant at \mathbf{x}_i .

AWS has several remarkable properties. First, AWS does not suffer from the ‘curse of dimensionality’, because the dimension of \mathbf{x}_i plays no role for the iterative procedure. Second, unlike nonparametric kernel regression, AWS can handle ‘edges’ in the regression function. Third, by successively increasing the bandwidth in the direction of \mathbf{x}_i , AWS allows more distant observations to be included in an estimate at any location as long as this is justified by approximate constancy of the estimate. The resulting areas of local constancy are completely determined by the data. Areas with identical values need not be of the same shape, say rectangular or radial (as in kernel regression), and also do not have to be adjacent.

We show in this paper that AWS and its properties are well-suited not only for image denoising but also for our economic application of land value estimation. Instead of a smudged image, we work with a set of land prices from a city (Berlin, Germany); instead of pixels, our design points \mathbf{x}_i are location coordinates; instead of finding areas with the same contrast, we find areas with approximately constant land values. There are strong reasons for using AWS in our application. First, piecewise constancy agrees with the block-wise layout of a city, formed by streets and transport infrastructure and shaped by planning and zoning regulations. It also fits with the prediction of the monocentric city model that land should have the same value at locations with equal travel distance to the central business district. Second, in most practical situations, fine graduation of land values will not be economically relevant, making the assumption of approximate local constancy reasonable. Third, a smooth map of land values

the paper

could only be estimated imprecisely, as it requires many local observations, which might not be available.

Land value estimation is an important problem in economics. Knowledge of a city's land values is of high interest to both real estate market participants and urban economists. Market participants need information on land values for purchasing decisions, development decisions, property taxation, and compulsory purchases. Urban economists have used (expert-based) land values to decompose house price dynamics (Bostic et al., 2007) and to assess the effect of local (dis)amenities on land values (Diamond, 1980; Ahlfeldt and Maenning, 2010).

Indeed, in Germany, information on the real estate market is seen as so important that expert-based land values (Bodenrichtwerte, BRW) must be published at least every two years. BRW are assessed by independent surveyors, following detailed guidelines. They serve as the benchmark for our AWS-based land values in this paper. Despite being based on detailed guidelines, it is fair to say that BRW rely heavily on local surveyors' knowledge and expertise. AWS, on the other hand, is a transparent statistical approach for the problem of jointly determining areas and land values. We apply AWS to our data of undeveloped land transacted between 1996-2009. AWS splits each price into the land value (the expected price) and transaction noise, which could have occurred during the business dealings. Our application of AWS to Berlin land prices reveals that overall the estimated land values are close to BRW. For specific parts of Berlin, however, land values estimated with AWS and BRW can show different local behavior, particularly if AWS is applied with a low degree of smoothing or if local circumstances (such as a lake-side location) demand (favor) expert knowledge. In summary, our paper demonstrates how land values can be estimated using a transparent statistical procedure. In addition to academic research, the procedure should be useful for practitioners as it provides a statistical method to which they can bring their expert knowledge.

The remainder of the paper is organized as follows. In Section 2, we explain the BRW, which serve as benchmark in our study. In Section 3, we explain AWS in the context of our application and detail in Section 4 the computational aspects of our implementation. Section 5 describes the transaction data. Section 6 presents the results of our empirical analysis. The final section concludes.

2 Expert-based land values

Local surveyor commissions are obliged by the German Building Law (Baugesetzbuch) to assess land values (Bodenrichtwerte, BRW) at least every two years. The members of a surveyor commission act on an honorary basis and are independent from the local administration. The local administration provides support, however, by collating information on the market and individual transactions.

The surveyors assess BRW with the sales comparison approach using land prices of comparable lots that were sold recently. Land transactions dating further back can be considered, too, once adjusted for the general price trend. Effects of unusual site conditions on the transaction price should be corrected for. Besides land prices, other real estate market information should be used too whenever deemed as being relevant. Such information can include transaction prices of developed land, zoning restrictions, and rent levels. In addition to the land value for a location, the surveyors have also to decide on areas for which BRW are equal. Guidelines provide further details on how such areas should be found and on how BRW should be derived (Richtlinie zur Ermittlung von Bodenrichtwerten).

Figure 1 shows the BRW map for Berlin, with 1 January 2010 as the reference date. The expert-based land values are therefore effectively for the year 2009, as the map is based on transaction information up and including this year. Land values within the first percentile of the value distribution are shown in light grey (bright yellow). The intensity increases to dark grey (bright red) for land values that fall within the tenth percentile of the distribution. From this distant view, the dominating feature of the map is the declining (color) intensity in the outward direction away from the city center. At the aggregate level, the land value map seems to confirm the prediction of the monocentric city model that the land value gradient falls with distance from the city center.² At close view, however, there is indication that the land value gradient is not strictly monotone in all directions. Figure 2 shows a detail of the BRW map for the center of Berlin. BRW areas are indicated by the dashed (red) lines and the solid (red) lines in case that the areas share borders with special development areas (Entwicklungsbereiche). Area 1132 in the lower left quadrant of the map detail, for instance, has a BRW of 2900 EUR per square meter. M₁ indicates that this area is developed densely for commercial use mainly.

²Fitting a third degree polynomial for the distance to the CBD gives such a decreasing function with an $R^2 = 0.2345$.

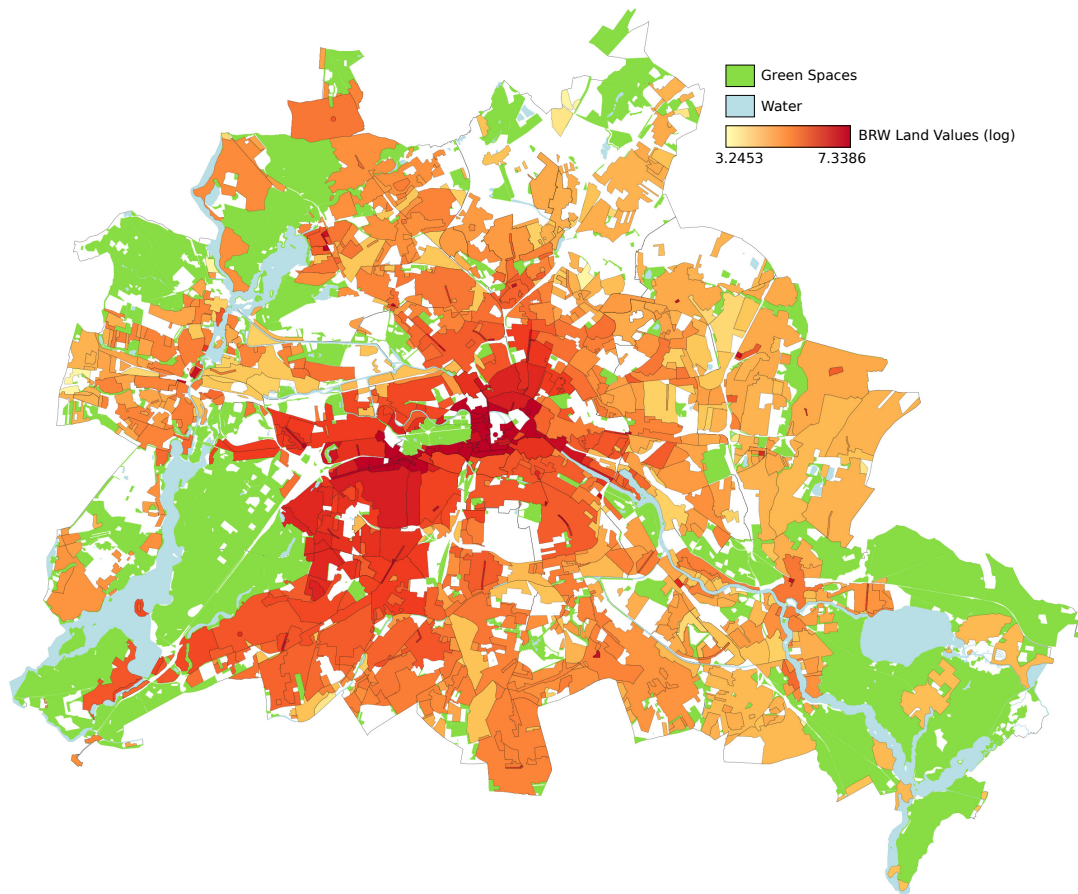


Figure 1: Expert-based land value (BRW) map for Berlin. Shows map of expert-based land values (BRW, in logs) for Berlin. Reference date for map is 1 January 2010. Source: Geoportal Berlin/Bodenrichtwerte 01.01.2010.

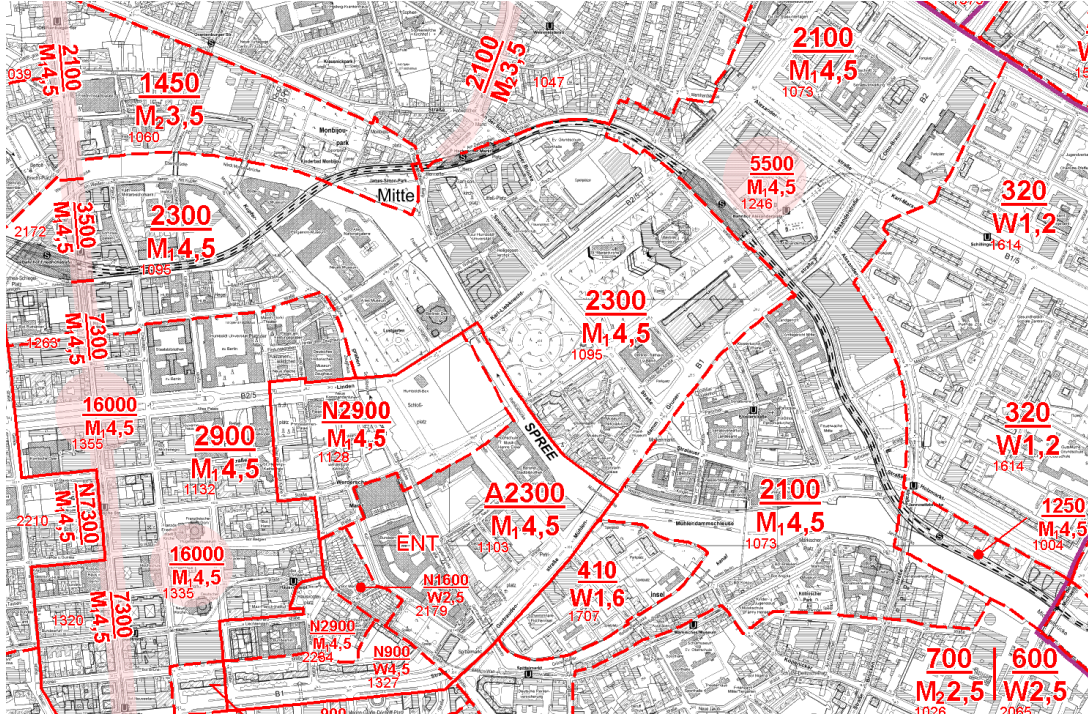


Figure 2: Detail of expert-based land value (BRW) map. Shows central business district including the boulevard Unter den Linden, the Museumsinsel, and the Alexanderplatz. Reference date for map is 1 January 2010. Source: Geoportal Berlin/Bodenrichtwerte 01.01.2010.

Within area 1132, there are also areas with a higher BRW, such as the Gendarmenmarkt square in area 1335. The detail reveals even better than Figure 1 that the BRW map consists of areas with identical land values and that the areas have sharp boundaries. Even between adjacent areas, BRW can vary substantially. At the same time, BRW can be the same in areas that are disconnected. Any statistical estimator of land values should be able to mimic this behavior. We now explain that AWS is such an estimator.

3 Adaptive Weights Smoothing

We consider the nonparametric regression model

$$y_i = \theta(\mathbf{x}_i) + \epsilon_i \tag{1}$$

for our application of land value estimation. y_i is the log land price per sqm of a site with location coordinates $\mathbf{x}_i = [x_{1i}, x_{2i}]$. $\theta(\mathbf{x}_i)$ is the log land value at \mathbf{x}_i and $\epsilon_i \sim N(0, \sigma^2)$ is independent of \mathbf{x}_i .³ As discussed in Section 1, it is reasonable to assume that land values can be the same at different locations, whereby the locations do not have to be adjacent. We denote with $U(\mathbf{x}_i)$ the set of locations where the log land value is $\theta(\mathbf{x}_i)$. If we knew the set $U(\mathbf{x}_i)$, it would be straightforward to estimate $\theta(\mathbf{x}_i)$. On the other hand, if we knew $\theta(\mathbf{x}_i)$ for all observations, then it would be straightforward to determine the set $U(\mathbf{x}_i)$.

In an iterative procedure AWS identifies, for each location \mathbf{x}_i , such sets of observations and uses the kernel estimator

$$\hat{\theta}_i \equiv \hat{\theta}(\mathbf{x}_i) = \frac{\sum_{j=1}^n w_{ij} y_j}{\sum_{j=1}^n w_{ij}} . \quad (2)$$

to estimate $\theta(\mathbf{x}_i)$. The weights are

$$w_{ij} = K_1(\text{dist}_{ij}) \times K_2(\text{lev}_{ij}) . \quad (3)$$

The kernel functions $K_1(u)$ and $K_2(u)$ work on a positive argument u , are strictly positive and non-increasing on the support $[0, 1]$ and zero otherwise. Examples are the uniform kernel function $K(u) = I(u \in [0, 1])$, which is constant on the support and the triangular kernel function $K(u) = (1 - u)I(u \in [0, 1])$, which is decreasing on the support. The arguments in the kernel functions in Eq. 3 are the distance measure for the locations i and j scaled by bandwidth h

$$\text{dist}_{ij} \equiv \frac{\rho(\mathbf{x}_i, \mathbf{x}_j)}{h} \quad (4)$$

and the test statistic for a constant local level of land values scaled by its critical value

$$\text{lev}_{ij} \equiv \frac{T_{ij}}{\lambda} . \quad (5)$$

While the role of the first kernel in Eq. 3 should be clear to readers familiar with nonparametric regression, the role of the second kernel needs motivation. This is best done by going through the first two iterative steps of AWS, where the second kernel becomes relevant after the initial step and plays a role at the extensive and the intensive margin of weighting.

³Scatterplots of the estimated AWS land values, $\hat{\theta}(\mathbf{x}_i)$, against the land prices, y_i , (not reported) as well as kernel density estimates of the estimated AWS residuals, $\hat{\epsilon}_i$, (not reported) indicate that the assumption of normally distributed and homoscedastic (log) land prices is approximately satisfied in our application below.

In the initial estimation step, $k = 0$, $lev_{ij}^0 = 0$ for all i, j , so that $K(lev_{ij}^0) = 1$. The bandwidth h^0 is set very small, leading to a small set $U^0(\mathbf{x}_i)$. The observations in this set are spatially close to \mathbf{x}_i and we can be confident that their land values are approximately the same. Using Eq. 2, we estimate

$$\hat{\theta}_i^0 = \frac{\sum_{j=1}^n w_{ij}^0 y_j}{\sum_{j=1}^n w_{ij}^0} \quad (6)$$

with weights

$$w_{ij}^0 = K_1 \left(dist_{ij}^0 \right) . \quad (7)$$

Eq. 6 is the conventional Nadaraya-Watson kernel estimator of $\theta(\mathbf{x}_i)$. Because only few spatially close observations are used, $\hat{\theta}_i^0$ will be a very local estimate of $\theta(\mathbf{x}_i)$. It might be that the initial bandwidth is too small and that the estimate could be improved by increasing the bandwidth to $h^1 > h^0$, thereby increasing the number of observations used to estimate $\theta(\mathbf{x}_i)$. Increasing the bandwidth has the well-known effect of reducing the variability of the estimates, but may introduce bias. This will not occur if those observations in the now enlarged neighborhood can be identified that have the same land value as observation i . In this case, the larger bandwidth leads to variance reduction, because more observations are used to estimate the land value without introducing bias.

It is therefore crucial to test if $\theta(\mathbf{x}_j)$ is the same as $\theta(\mathbf{x}_i)$. To understand how this is implemented in AWS, suppose we used initially the most local version of the estimator in Eq. 6 by setting h^0 so small that only observation i would be used. The land value estimate would be $\hat{\theta}_i^0 = y_i$. Increasing the bandwidth to $h^1 > h^0$ would imply that more observations become available that could be used for the estimation of $\theta(\mathbf{x}_i)$. For each such observation j , we had to decide if we should include the observation into $U^1(\mathbf{x}_i)$. Obviously, we should do this only if the land values are the same at both locations and if

$$H_0 : \theta(\mathbf{x}_i) = \theta(\mathbf{x}_j) \quad (8)$$

is true. Given the model from Eq. 1, the null hypothesis implies $\hat{\theta}_i^0 - \hat{\theta}_j^0 = y_i - y_j = \epsilon_i - \epsilon_j$ and $\hat{\theta}_i^0 - \hat{\theta}_j^0 \stackrel{H_0}{\sim} N(0, 2\sigma^2)$. We can therefore construct the test statistic

$$\frac{\hat{\theta}_i^0 - \hat{\theta}_j^0}{\sqrt{2\sigma^2}} \stackrel{H_0}{\sim} N(0, 1) \quad \Rightarrow \quad T_{ij}^1 \equiv \left(\frac{\hat{\theta}_i^0 - \hat{\theta}_j^0}{\sqrt{2\sigma^2}} \right)^2 \stackrel{H_0}{\sim} \chi_1^2 . \quad (9)$$

If we knew σ^2 , we could compute T_{ij}^1 and compare it with the critical value λ of the χ_1^2 distribution at some significance level α . Whenever $T_{ij}^1 \lambda^{-1} = lev_{ij}^1 < 1$, we could not reject the null hypothesis in Eq. 8 at significance level α and $K_2(lev_{ij}^1) > 0$. The observation with location \mathbf{x}_j should become member of $U^1(\mathbf{x}_i)$ and y_j should be used for the estimation of $\theta(\mathbf{x}_i)$. If, on the other hand, $lev_{ij}^1 > 1$, then observation j should not become member of $U^1(\mathbf{x}_i)$ and y_j should not be used to estimate $\theta(\mathbf{x}_i)$ and $K_2(lev_{ij}^1) = 0$. The treatment of the borderline case $lev_{ij}^1 = 1$ depends on the chosen kernel function and could lead to the in- or exclusion of observation j .

We see that $K_2(lev_{ij}^1)$ works at the extensive margin by indicating whether observation j should be included in $U^1(\mathbf{x}_i)$ or not. But $K_2(lev_{ij}^1)$ plays also a role at the intensive margin as long as $K_2'(u) < 0$. For observations in $U^1(\mathbf{x}_i)$, the magnitude of $K_2(lev_{ij}^1)$ is then inversely related to the magnitude of lev_{ij}^1 . If we now consider the weights from Eq. 3 for step $k = 1$

$$w_{ij}^1 = K_1 \left(dist_{ij}^1 \right) \times K_2 \left(lev_{ij}^1 \right) , \quad (10)$$

we see that the weight of observation y_j in the average formed at i is determined by *both* kernel functions. The first kernel plays the same role as in conventional nonparametric regression. It will allocate more weight to observation j the closer \mathbf{x}_j is to \mathbf{x}_i . In step $k = 1$, the distance penalty is relaxed because a larger bandwidth h^1 is employed than in the initial step. Hence, more observations are used for $\hat{\theta}_i^1$. The second kernel controls that this does not introduce bias. At the extensive margin, only those observations receive a nonzero weight for which the assumption of local constancy is not rejected. At the intensive margin, the included observations receive a higher weight the more likely it is that they fit with the null hypothesis.

Comparing the weights from Eqs. 7 and 10 reveals how AWS differs from and improves upon conventional kernel smoothing. Introducing a second kernel function, which kicks in at iteration $k = 1$, makes AWS ‘structurally adaptive’. It incorporates information about the local properties of the regression function from the previous iterative step. If there is evidence for local constancy, then increasing the bandwidth h is beneficial (‘propagation’) as the constant land value will be estimated with more land price observations. If, however, the previous step estimates $\hat{\theta}_i^0$ and $\hat{\theta}_j^0$ differ considerably, then T_{ij}^1 will be large relative to λ and local homogeneity will be rejected and $K_2(lev_{ij}^1)$ will assign a zero weight to observation j (‘separation’). Successive estimation and testing are thus intertwined to improve estimation precision and to identify the set of locations where land values are the same.

In the motivation just given, we made several simplifying assumptions that need qualification. First, we assumed that the initial bandwidth is set so small that only observation y_i is considered. This will lead to a very imprecise initial estimate of $\theta(\mathbf{x}_i)$. As a consequence, AWS starts with a small bandwidth, but one that is typically large enough to allow computation of $\hat{\theta}_i^0$ at most locations with more than one observation. The test statistic becomes

$$T_{ij}^1 = A_i^0 \left(\frac{\hat{\theta}_i^0 - \hat{\theta}_j^0}{\sqrt{2\hat{\sigma}^2}} \right)^2, \quad (11)$$

where $A_i^0 = \sum_{j=1}^n w_{ij}^0$, i.e., the sum of the weights at i from step 0, which can be regarded as the ‘local sample size’.⁴ Second, AWS is an iterative algorithm and the distributional assumption regarding T_{ij}^k will only hold in the first iteration, $k = 1$.

We summarize the iterative AWS algorithm. In step 0, $\hat{\theta}_i^0$ is estimated for each location with an observation using a Nadaraya-Watson kernel estimator using a small bandwidth h^0 . The estimated land values are used to compute the pairwise terms lev_{ij}^0 . In step 1, the bandwidth is increased to h^1 . More observations are now available for the estimation of $\theta(\mathbf{x}_i)$. However, some observations will have markedly different land values from $\theta(\mathbf{x}_i)$ and should not be used. The estimator from Eq. 2 controls for this via the second kernel. Observations with land values similar to $\theta(\mathbf{x}_i)$ will receive positive weight in the estimator $\hat{\theta}_i^1$, those with markedly different values will receive no weight at all. The estimated land values $\hat{\theta}_i^1$ are used to compute the pairwise terms lev_{ij}^1 . In step 2, the bandwidth is increased to h^2 and so on. The procedure terminates in step k^* , when the bandwidth reaches the threshold h^* .

4 Implementation

For our application, we use the *R-package: Adaptive weights smoothing* (Polzehl, 2014). We work on a grid of bins instead of the individual site location to increase computational speed. In our preferred implementation, bins have the size of 152×152 meters, which is smaller than the average BRW area. Combinations of adjacent bins therefore have the potential to recombine into BRW areas. Given that Berlin covers an area of 891 km^2 , with 45 km distance from west to east and 38 km from south to north, this leads to a grid of $300 \times 300 = 90,000$ equally-sized bins.

⁴Obviously, if only the observation itself was used in step 0, $A_i^0 = 1$, and Eq. 11 becomes Eq. 9.

Within each bin, we average the log transaction prices $p_{j,i}$. These bin-averages become the dependent variable y_i . Binning smoothes the transaction data and increases computational speed with only small cost to accuracy (Fan and Marron, 1994). For instance, given our choice of k^* , the algorithm takes about 14 minutes of CPU time on the preferred grid. On a 500×500 grid, the algorithm takes already about 29 minutes to run. The algorithm would take much longer if the individual transaction data were used instead. Visual inspection of the land values estimated on the 500×500 and the 300×300 grids show no discernable differences. A grid coarser than 300×300 speeds up the computation further, but differences in the estimated land value maps become discernable.

Given the grid structure of our data, the location coordinates are integers and indicate the row and column position of bin i in the grid. Therefore, $\mathbf{x}_i = [r_i, c_i]$ with $r, c \in \{1, \dots, 300\}$. Correspondingly, the bandwidth h is an integer. To measure the distance between observations in Eq. 4, we use the Manhattan distance

$$\rho(\mathbf{x}_i, \mathbf{x}_j) = |r_i - r_j| + |c_i - c_j|. \quad (12)$$

We use the triangular kernel function on the positive semi-axis

$$K(u) = (1 - u)I(u \in [0, 1]) \quad (13)$$

for both $K_1(u)$ and $K_2(u)$. We have experimented also with other kernel functions, but the choice was negligible for the results.

To motivate what these choices imply, we use Figure 3 and focus on bin $i = 25$ in the center. For $h = 2$, we have $\rho(\mathbf{x}_{25}, \mathbf{x}_{25}) = 0$, $\rho(\mathbf{x}_{25}, \mathbf{x}_{24}) = \rho(\mathbf{x}_{25}, \mathbf{x}_{26}) = 1$ and $\rho(\mathbf{x}_{25}, \mathbf{x}_{27}) = 2$. The scaled distance measures are $dist_{25,25} = 0$, $dist_{25,24} = dist_{25,26} = 0.5$ and bins further to the left or right have a distance measure of at least one. This implies that the distance kernel gives y_{25} the most weight, y_{24} and y_{26} lesser weight and, for instance, y_{27} no weight at all. We estimate the variance for the level penalty term in Eq. 11 with

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N \left\{ \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (p_{j,i} - y_i)^2 \right\}. \quad (14)$$

Only bins i with at least two transaction prices $p_{j,i}$ are relevant for this average over estimated

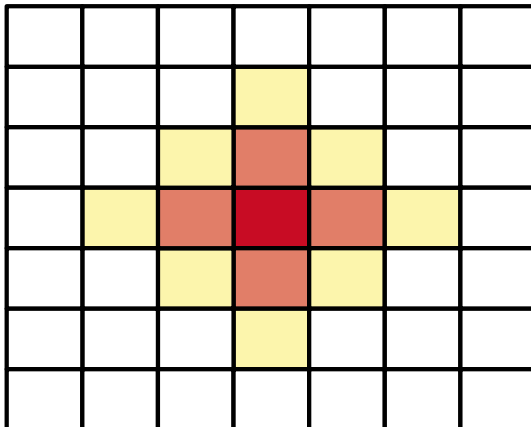


Figure 3: Distance weighting of observations in AWS estimator. Illustrates scaled Manhattan distance and Kernel weight for bandwidth $h^0 = 2$. We want to compute $\theta(\mathbf{x}_{25})$ for the dark red shaded bin 25 in the center. Scaled distance for this bin is zero and y_{25} receives the highest distance weight in $\hat{\theta}_{25}^0$. The light red shaded bins have a scaled distance of 0.5 to the reference bin. Prices y_j from these bins receive less weight in $\hat{\theta}_{25}^0$. The scaled distance of the light yellow shaded bins is 1. These bins (as those further away) are not considered in $\hat{\theta}_{25}^0$.

within-bin variances. N is the total number of such bins.

The crucial smoothing parameter of AWS is λ , which can be viewed as the ‘bandwidth’ in the level direction. Choosing λ too high will lead to a loss of sensitivity to changes in the land value map. Choosing λ too low will result in forgoing the benefits of extending the number of observations that can be used to estimate the local land value. This trade-off is akin to the familiar bias-variance trade-off of non-adaptive, conventional nonparametric regression. In the latter case, the desirable properties of data-driven smoothing parameter selectors such as cross-validation have been established. However, working out such a theory for the adaptive, iterative AWS procedure is very challenging and has not yet been accomplished.

While there is no established data-driven method for choosing λ , Polzehl and Spokoiny (2006, 2008) have proposed an approach to arrive at a suitable, objective value. This approach is based on the observation that the level penalty of AWS can be viewed as a test statistic, T_{ij}^k , with λ acting as the critical value (see Eqs. 9 and 11 above). For $T_{ij}^k > \lambda$, the null hypothesis of local homogeneity is rejected and y_j should not be used to estimate $\theta(\mathbf{x}_i)$. If the null hypothesis is true, then this amounts to a type-I error: an artificial ‘edge’ has been detected when there should have been propagation. Hence, one may choose λ such that the probability of a type-I error is sufficiently small. This suggests to choose λ as the $1-\alpha$ percentile of the

χ_1^2 -distribution.⁵ Due to the multiple testing of the null hypothesis in each iteration step, at every design point, however, α does not equal the probability of a type-I error. Thus, no unique, well-defined choice for λ can be made within this approach.

To arrive at a unique value of λ , Polzehl and Spokoiny (2006, 2008) have suggested to consider the following extreme situation: the case where $\theta(\mathbf{x}_i) = \theta$, i.e. the regression function is globally homogenous. In this hypothetical situation, the null hypothesis of homogeneity holds everywhere and the AWS algorithm should continue to propagate in each iteration step such that it arrives at the global model in the final iteration step k^* at all design points. Obviously, this can be achieved by choosing $\lambda \rightarrow \infty$. To arrive at a sensible value of λ , Polzehl and Spokoiny thus propose to choose the smallest value of λ such that the algorithm will, with a high probability, continue to propagate everywhere in the case of the globally constant model. We will refer to this smallest value of λ satisfying the propagation condition as λ^* .

Since the hypothetical situation of a globally constant regression model is considered, λ^* must be found by Monte Carlo simulation. We follow the simulation design proposed by Becker (2014) and Becker and Mathé (2013) and adapt it to our context. We maintain the assumption that the errors of the regression model are normally distributed and homoscedastic. The variance used to simulate the data equals the conditional variance estimate obtained from our transactions data using Eq. 14. Becker (2014) and Becker and Mathé (2013) have shown that in our case of Gaussian regression, λ^* does not depend on the value of θ . We thus set $\theta(\mathbf{x}_i) = \theta = 0$, for simplicity. We simulate data for this normal regression with constant mean and fixed variance on a quadratic grid of 100×100 points. This is close to the number of bins for which we observe land transactions and is exactly a third of the grid used in our application. The bandwidth for the location penalty is set to 45. This is a third of the bandwidth $h^* = 135$ in our actual application, see Section 6 below.

The definition of λ^* requires to specify the desired level at which one safeguards against not propagating everywhere at the end of the iteration process in the case of the globally constant model. In the formulation of Becker (2014) and Becker and Mathé (2013), this ‘propagation level’ ϵ can be interpreted as the expected probability that AWS (erroneously) separates in the homogenous situation $\theta(\mathbf{x}_i) = \theta = 0$. We set $\epsilon = 0.0005$. Given ϵ , Becker (2014) and Becker

⁵While the exact sampling distribution of T_{ij}^k can only be derived in iteration step $k = 1$ if the bandwidth h^0 is very small, the χ_1^2 -distribution may still be a good approximation in every iteration step k under the assumptions of normally distributed, homoscedastic errors.

and Mathé (2013) provide a sufficient condition for checking whether a candidate value of λ is meeting the desired propagation level.⁶ λ^* is then found as the smallest value of λ meeting this requirement for the simulated data.

5 Data

The transaction data has been provided by Berlin’s surveyor commission (Gutachterausschuss für Grundstückswerte, GAA). The GAA is entitled by law to request and collect information on all real estate transactions occurring in Berlin. The data has 24,519 observations and covers arms-length transactions of undeveloped land during the years 1996-2009. For each observation, information is provided on the transaction price per square meter (sqm), geo-coordinates, unusual features of the site, and information on aspects of the business dealings. The BRW map shown in Figure 1 is based on these land transactions, but the GAA surveyors will have considered also other real estate market information. We have therefore only a subset of the information that was available to the surveyors. Figure 4 shows the locations of the transacted sites. Most transactions of undeveloped land took place in the residential areas at the outskirts of the city. Less transactions of undeveloped land took place in the densely developed city center.

Table 1 presents summary statistics for the transaction data. The land prices are in real terms and adjusted for unusual features of the site and unusual aspects of the business dealing. Unusual features of the site include physical aspects such as structural damage or flooding risk and legal aspects such as rights of way or use for pipes or cables. Such easements are rather common. Unusual aspects of business transaction include price rebates, installment payment, and investment obligations. The Appendix explains the adjustments. We could match 23,950 transactions with their BRW using the geo-coordinates. For the remaining 569 transactions, the GAA did not compute a BRW, perhaps because of restrictions on land use or land disposal.

⁶The condition involves the probability that the Kullback Leibler divergence between the adaptive AWS estimate and the globally constant θ is bounded and that the bound does not increase in the iteration process for a given propagation level ϵ . Becker and Mathé (2013) also propose a method for estimating this probability from the simulated data.

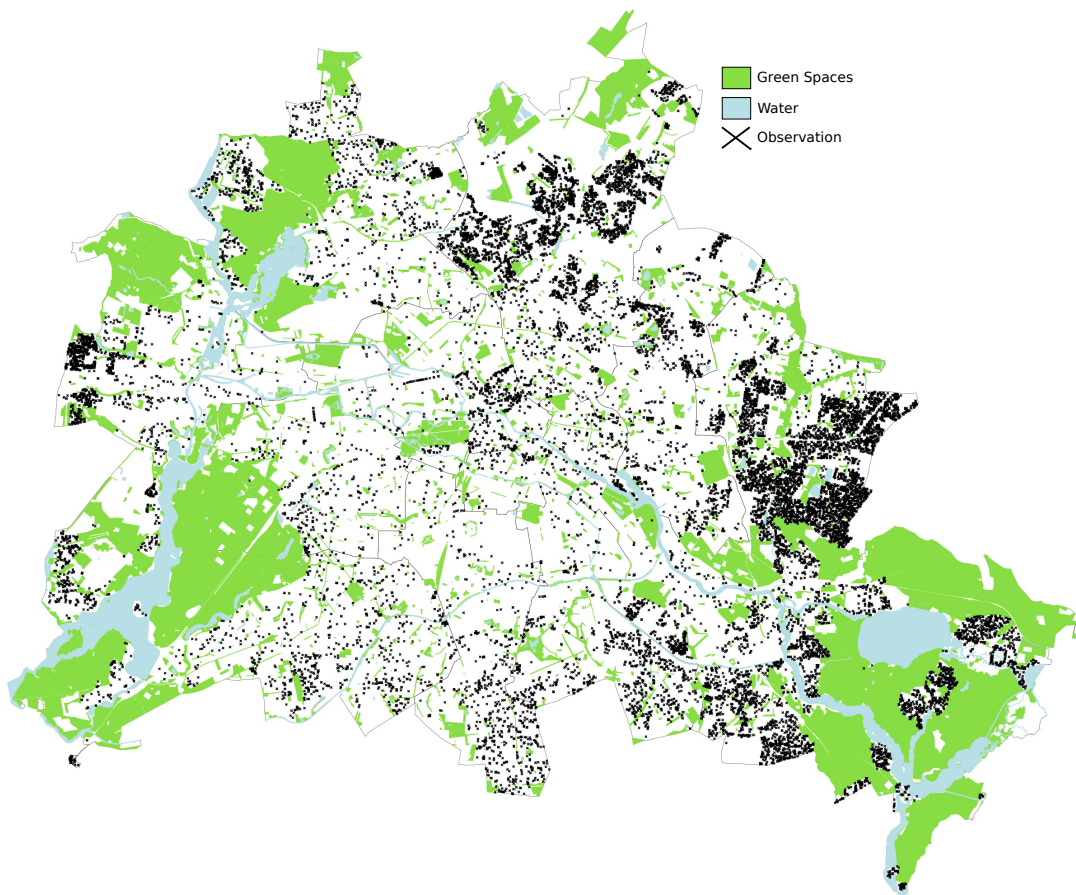


Figure 4: Location of transacted sites within Berlin. Shows the location of 24,519 undeveloped sites that have been transacted between 1996 to 2009. Solid lines represent the borders of Berlin's 12 administrative districts (as of the year 2000).

Table 1: Summary statistics for transacted land.

	Mean	Median	Std. Dev.
<i>Land price per sqm</i>			
Natural scale	182.65	117.24	416.45
Log scale	4.774	4.667	0.647
<i>Expert-based land value per sqm (BRW)</i>			
Natural scale	190.65	110.00	407.69
Log scale	4.943	4.701	0.590
<i>Unusual features of the lot</i>			
Ground monument	0.009		
Contaminated soil	0.030		
Demolished structure	0.350		
Land easement	0.211		
<i>Aspects of business dealing</i>			
Non-private seller	0.352		
Non-private buyer	0.179		
Infrastructure charge	0.418		
Other unusual aspects	0.098		
<i>Amenities and expert-based location rating</i>			
Lake side	0.025		
Low quality	0.417		
Medium quality	0.458		
High quality	0.114		
Very high quality	0.009		

Number of observations is 24,519. BRW information is available for 23,950 of these observations. Land prices are adjusted for unusual features of the site and unusual aspects of the business dealing. Land price and BRW are in real (year 2009) Euro. Discrete characteristics are in percent. Other unusual aspects of business transaction include price rebates, installment payment, and investment obligations. Expert-based location rating comes from Berlin's Senate Department for Urban Development.

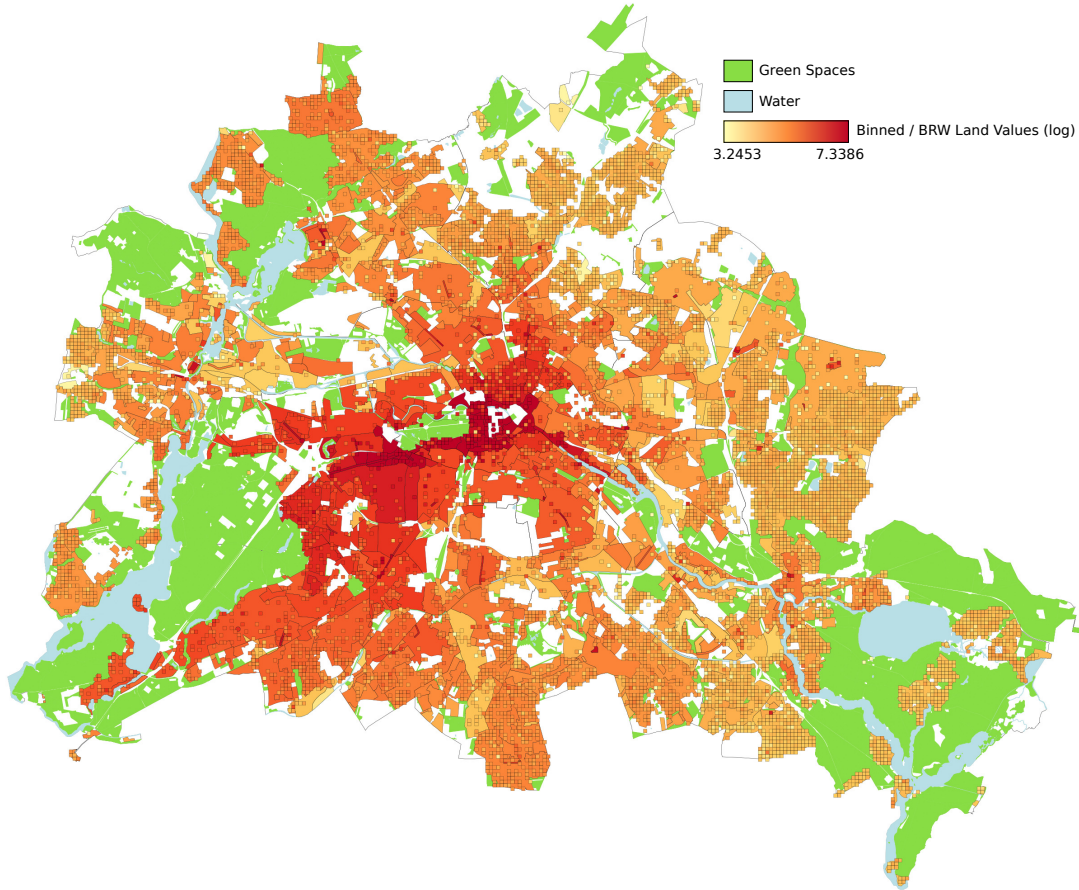


Figure 5: Land price bins for Berlin. Shows the 7,448 non-empty bins with y_i , i.e., the log land prices averaged by bin. Backdrop is expert-based land value map.

Land prices and the corresponding BRW show substantial variation, both at the natural and the log scale, as indicated by their standard deviations. The average levels of land prices and BRW are quite similar, indicating that our adjustments for the general price trend and unusual aspects are in line with adjustments done by the GAA surveyors. Applying the 300×300 grid to the transaction data, 7,448 bins end up with transactions, of which 2,924 bins contain exactly one transaction. On average, the filled bins contain 3.29 land transactions. Figure 5 shows a map with the non-empty land price bins and as backdrop the map of expert-based log land values. The more satiated a bin, the higher is y_i . Inspection reveals that in some areas land prices y_i are similar to surrounding expert-based land values, but in others they are not. As to be expected, the expert-based land values are smoother than the raw land prices, see Figure 6. If we apply the same grid to the expert-based land values, we obtain 21,842 non-empty bins. Denoting with BRW_i the bin-average of the log expert-based land values, we observe this



Figure 6: Detail of raw land prices and BRW land values. Upper panel shows bins with raw land prices. Lower panel shows bin average of BRW land values.

variable for 7,222 of the 7,448 bins with land transactions. Table 2 presents summary statistics for the binned data conditional on the bin containing at least one transaction.

Table 2: Summary statistics for binned data.

	Mean	Median	Std. Dev.
<i>Land price per sqm</i>			
Natural scale	225.66	126.63	545.97
Log scale	4.880	4.727	0.736
<i>Expert-based land value per sqm (BRW)</i>			
Natural scale	240.03	140.00	562.55
Log scale	5.085	4.942	0.683

Reports summary statistics for land prices and BRW at the bin level, conditional on the bin covering at least one transaction. Of the 7,448 bins fulfilling this criterion, 7,222 have also BRW information.

Binning corrects for the fact that we observe relatively more transactions in the lower-priced residential than the higher-priced central areas of the city. This explains why binned land prices and binned BRW are higher on average and show larger variation than the observations at the level of individual transactions, see the relevant statistics in Tables 1 and 2.

6 Results

We start by motivating the numerical values of the two smoothing parameters, h and λ , employed in our AWS application. We set the initial bandwidth $h^0 = 2$. A smaller bandwidth corresponds to y_i and therefore an imprecise estimate of $\hat{\theta}^0$. Even our initial bandwidth choice leads to a jaggy land value map. The iterative algorithm stops when the location bandwidth reaches $h^* = 135$.⁷ This corresponds to a radius of 20,520 meters.

To assess if AWS can reproduce the BRW values, we estimate the land values for an array of different λ values. We choose critical values of the χ_1^2 -distribution for this array. The

⁷We use the sequence $\mathbf{h} = \{1, 2, 3, 4, 5, 7, 9, 11, 14, 18, 22, 28, 35, 44, 55, 69, 86, 108, 135\}$ to increase h^k in the $k^* = 19$ iteration steps.

critical values correspond to the $1-\alpha$ percentiles of the χ_1^2 -distribution with $\alpha^0 = 0.05$ and $1-\alpha^p = 0.05^{1.125^k}$ for $p = 1 \dots 6$. As we discussed in Section 3, the test statistic in iteration step $k = 1$ will follow the χ_1^2 -distribution under the null hypothesis if h^0 is very small. Even though our initial bandwidth is slightly larger, Figure 7 shows that T_{ij}^1 follows the χ_1^2 -distribution fairly. Although this distributional result is unlikely to hold in later steps of the algorithm, we feel that our choice of λ is less arbitrary than others. To further safeguard against separation in homogenous regions, we estimate also the land values using $\lambda^* = 19.9$ which has been chosen with the simulation explained in Section 4.

We begin our discussion of the AWS results by showing in Figure 8 the estimated Berlin-wide land value map for λ^* . As our AWS procedure works on a bin grid, the estimated land values are visualized by framed bins. For coloring the bins in Figure 8 we employ the same scheme as in the BRW map of Figure 1. Comparing both maps shows that the AWS bins in Figure 8 only cover a part of the continuously shaded BRW areas of Figure 1. In view of the map of transactions in Figure 4, it becomes clear that the expert-based BRW map involves a substantial amount of extrapolation as there were no transactions in the areas not covered by one of the AWS bins. Moreover, the comparison of the BRW and AWS maps shows that both approaches agree fairly well in terms of their spatial color patterns. Hence, at least for λ^* , the estimated log-land values based on AWS appear to correspond quite closely to their expert-based benchmark.

Both the differences of AWS and BRW in terms of their spatial coverage as well as their general agreement in log-land values is summarized in Figure 9 where both maps have been superimposed (rectangular AWS bins are in the foreground, continuously shaded BRW areas in the background). In order to investigate the dependency of the AWS land values on λ , particularly the (dis-)agreement between AWS and BRW land values, we run a series of bivariate least-squares regressions. The dependent variable in each regression is the AWS land value obtained for a given value of λ .

We start by regressing the AWS land values on the land prices y_i . The resulting coefficients of determination, R^2 , is reported in the ‘Land price’ row of Table 3. We find that the in-sample fit is fairly good for all values of λ . Moreover, the R^2 s from these regressions show the expected pattern: more smoothing (i.e. a larger value of λ) leads to a worse in-sample fit. While these results suggest that AWS delivers sensible results for a range of values of λ , they can not give

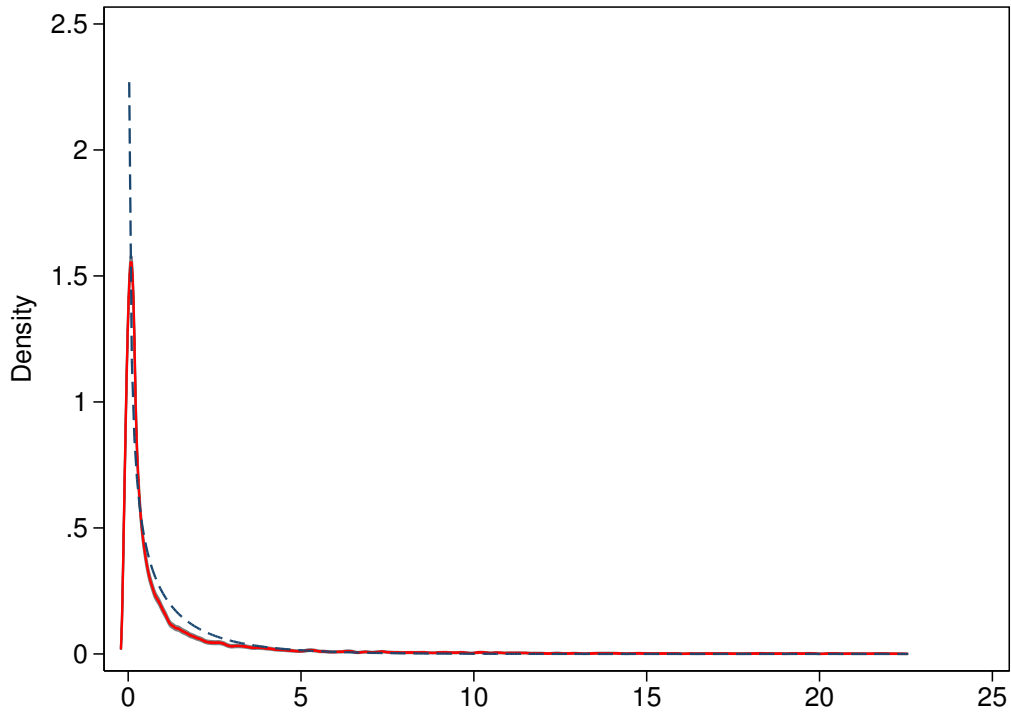


Figure 7: Kernel density estimate of realizations of the test statistic. Shows kernel density estimate of realizations of test statistic in iteration $k = 1$. Bandwidth is computed according to Silverman's rule of thumb. Shaded area is 95% pointwise confidence interval. Dashed curve is χ^2 -density with 1 degree of freedoms. Number of observations is 16,022.

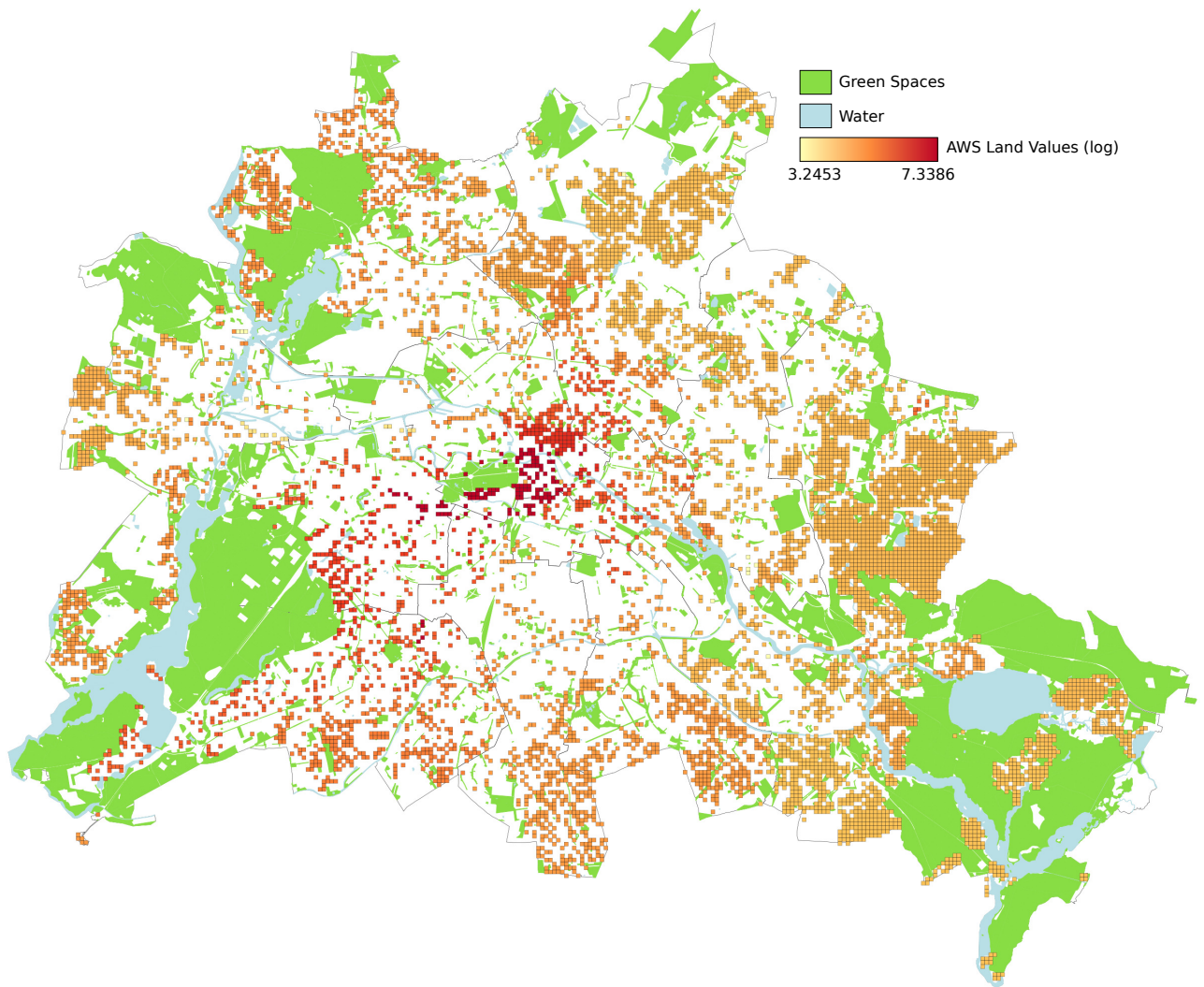


Figure 8: AWS land value map for Berlin. Shows contour map of estimated AWS land values. Level bandwidth is set to $\lambda^* = 19.9$.

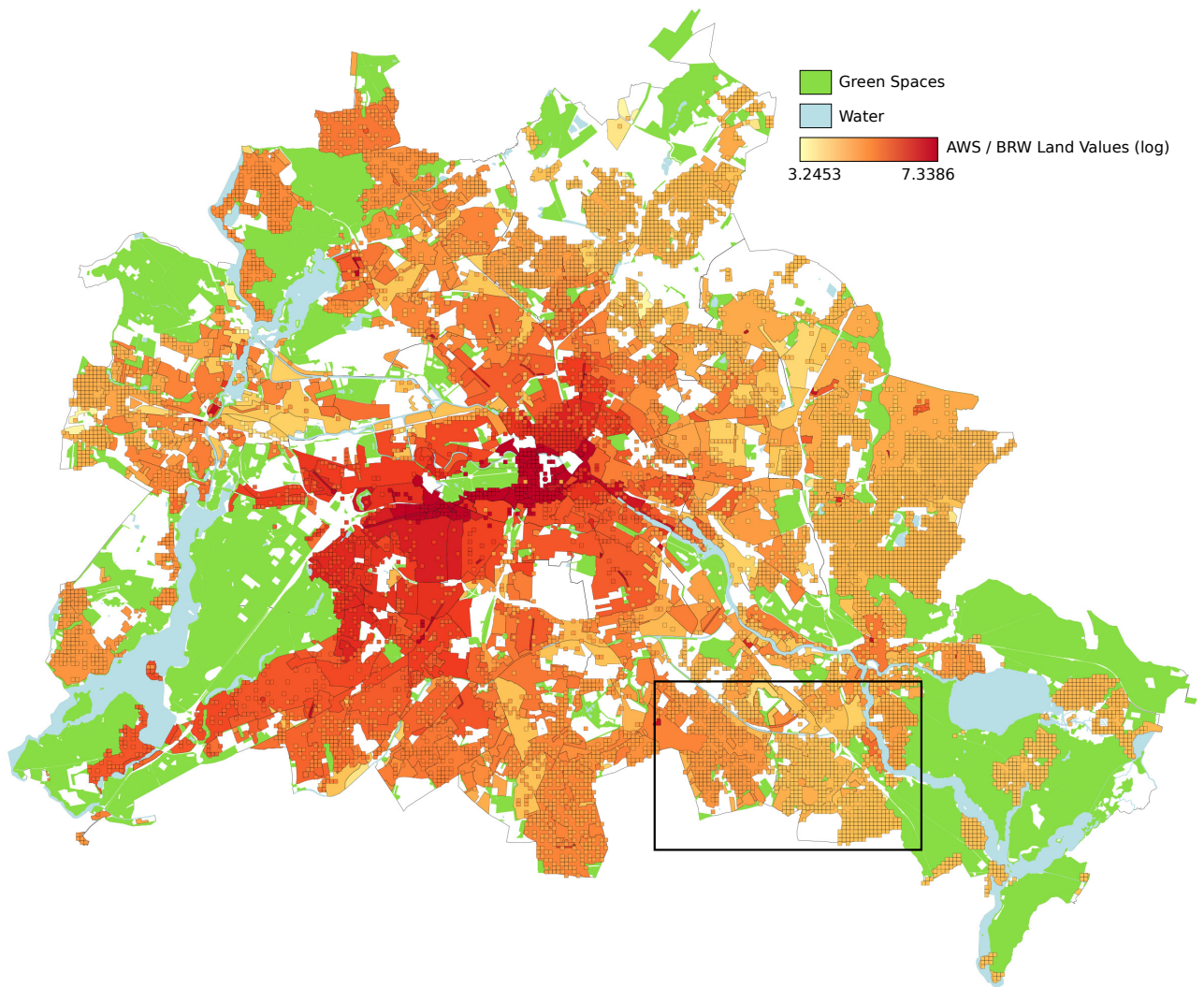


Figure 9: AWS land value map for Berlin. Shows contour map of estimated AWS land values. Level bandwidth is set to $\lambda^* = 19.9$. Coloured areas in background are BRW.

an answer about which value of λ to prefer.

Table 3: Explanatory power.

	λ^*	λ						
	19.9	3.8415	4.4756	10.5180	16.8410	23.2840	29.7938	36.346
BRW	0.7747	0.7274	0.7390	0.7640	0.7720	0.7828	0.7733	0.7690
Land price	0.6992	0.8661	0.8526	0.7734	0.7195	0.6764	0.6525	0.6418

Reports coefficient of determination R^2 for bivariate regressions of BRW_i and land prices y_i on AWS land values $\hat{\theta}_i$. Regressions include a constant. Number of observations used for regressions in first row is 7,222 and 7,448 for regressions in second row.

We therefore run a second set of regressions, where we regress the AWS land values on the corresponding bin averages of the expert-based land value, i.e. the BRW_i . The R^2 s of these regressions are reported in the row ‘BRW’ of Table 3. There is generally good agreement between the AWS and BRW land values for all considered values of λ . Though, the R^2 values show an inverse U-shape pattern: they initially increase but eventually decrease for successively higher values of λ . The strongest agreement occurs at $\lambda = 23.284$ ($R^2 = 0.783$). For our favoured λ^* , the coefficient of determination is only slightly smaller. While this exercise is not the equivalent of independent out-of sample evidence, the observed pattern still suggests that the simulation based method delivers a reasonable choice of λ . That is, if we regard the expert-based BRW as a benchmark, λ^* offers a good-compromise between over- and under-smoothing for our data.

While Table 3 collects some evidence on the impact of λ at the global level, its influence on the AWS procedure is also noticeable at a more local level. This is illustrated in Figure 10 which shows AWS results for a particular area of the city that roughly corresponds to the Rudow and Alt-Glienicke localities of the boroughs of Neukölln and Treptow (to locate this area within the Berlin-map it is marked by a rectangle in Figure 9). The estimates in the upper panel are based on λ^* while the lower panel shows results for the considerably smaller $\lambda = 3.8415$.⁸ In the upper panel, it can be seen that with λ^* AWS identifies three distinct areas of homogenous land values marked by dashed ellipsoids. While roughly these three areas are also visible in the lower panel, the AWS estimates for the small value of λ show considerably more variation

⁸ $\lambda = 3.8415$ corresponds to the 5% percentile of the χ^2_1 -distribution.

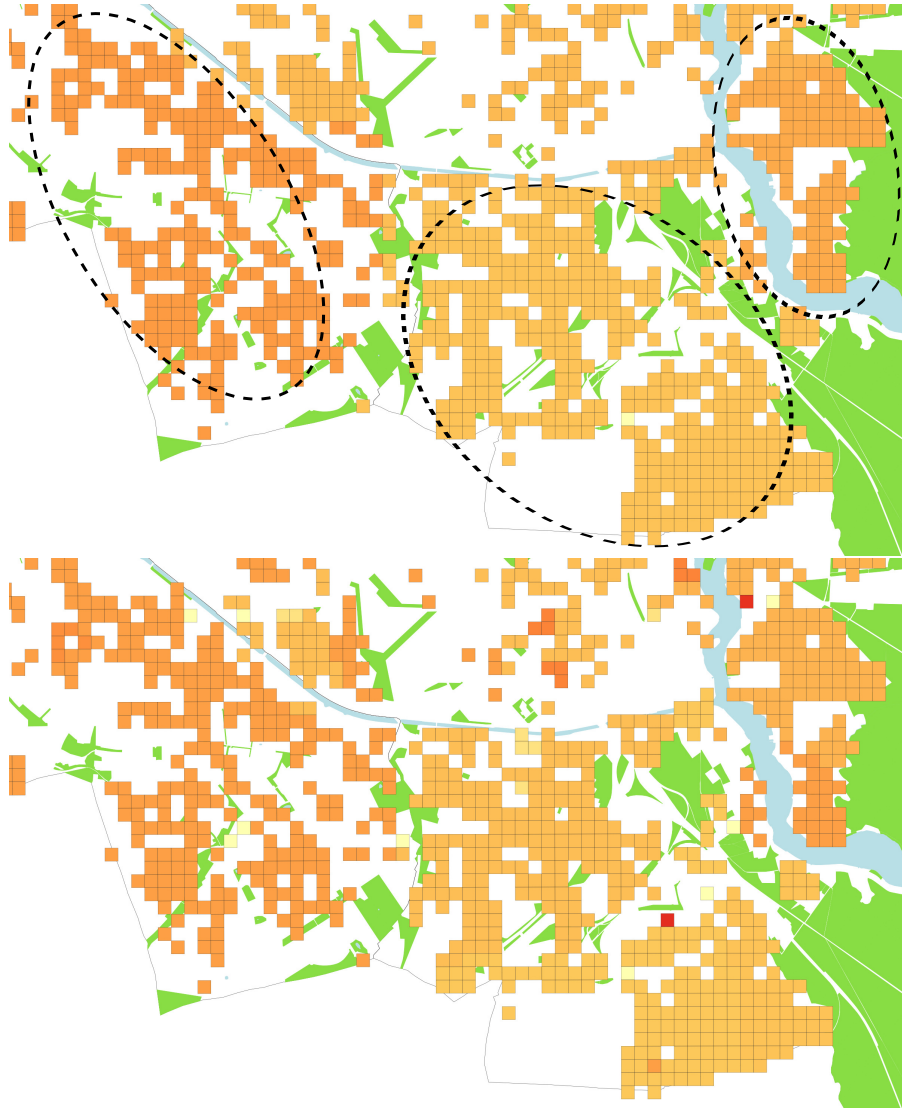


Figure 10: Detail of AWS land values for different λ values. Upper panel sets level bandwidth to $\lambda^* = 19.9$. Lower panel sets level bandwidth to $\lambda = 3.8415$.

in the levels of land values and in the shapes of locally homogenous areas. In the light of the discussion in the previous paragraph, where λ^* was found to overall deliver a good agreement between AWS and BRW land values, Figure 10 suggests that the surveyors appear to implicitly apply a relatively high degree of smoothing when judging differences in land values at different locales rather than aiming at a very detailed map that AWS delivers for small λ .

Finally, in Figure 11 we consider the local degree of (dis-)agreement between the estimated AWS land values (shown in the upper panel) and the BRW_i land values (shown in the lower panel). The depicted area is again the locality defined by the black rectangular in Figure 9. The AWS land values are estimated using our favored λ^* . It is apparent that the AWS land values divide the locality into roughly the same three areas of local homogeneity (indicated by the encircled areas) as the BRW_i . At the very local level, however, the BRW_i bring the surveyors' expert-knowledge to the fore. For instance, the BRW_i close to the river 'Dahme' (indicated by the arrows) are significantly higher than the corresponding AWS land values which implies an area of local homogeneity across the river banks. It is clear, that at this micro-level, the surveyors have a knowledge advantage, as compared to the data-driven AWS algorithm which must rely on the globally set λ^* .

7 Conclusions

In this paper, we applied adaptive weights smoothing (AWS) to a problem of substantial economic relevance, the estimation of the contour map of land values of a city. Using AWS has several advantages over standard nonparametric regression. It allows the size and shape of areas with a common land value to be completely determined by the data. As illustrated by our application, these areas need not be symmetric or adhere to a particular shape. Moreover, unlike kernel regression, AWS does not require the land value map to be smooth. AWS identifies these areas by relaxing the distance penalty in successive iterations and implicitly testing for local constancy of land values. As long as the land values are sufficiently similar, relaxing the distance penalty is justified and adjacent areas are subsumed into one.

Our application to the Berlin market revealed that estimated land value maps based on AWS generally agree fairly well with the benchmark of expert-based land values (BRW). This is particularly true if AWS' crucial smoothing parameter is chosen by the simulation based method based on the propagation condition. We also find, though, that land values estimated

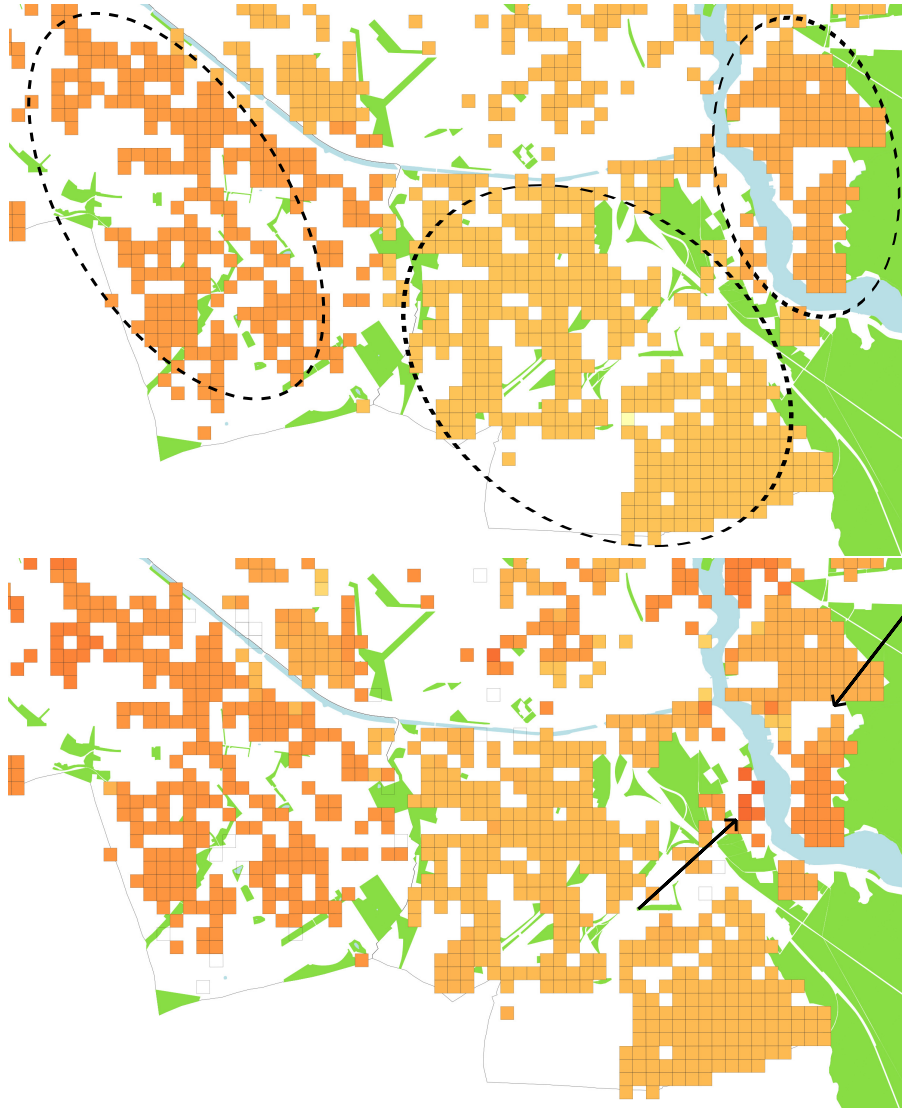


Figure 11: Detail of AWS land values and BRW land values. Upper panel shows estimated AWS land values. Level bandwidth is set to $\lambda^* = 19.9$. Lower panel shows bin average of BRW land values.

with AWS and BRW can show different local behavior, for instance if local circumstances (such as a lake-side location) favor expert knowledge. However, such expert-knowledge may contain subjective judgement or may be unavailable altogether. Our paper demonstrates that AWS, on the other hand, is a transparent statistical procedure. It delivers in our application land values that should at least provide a sound basis for academic research. AWS land values may also be useful for practitioners as a statistical basis to which they can bring their expert knowledge.

In future research, we will enlarge our land data by including house and condominium prices purged of the building component as in Bryan and Satre (2009). This will increase our data set to about 250,000 observations. The larger data set will allow us to fill gaps in the current AWS land value map. Remaining gaps for residential land (developed or undeveloped) will then be interpolated.

8 Appendix

We adjust the land prices as follows. First, we run the regression

$$p_{j,t} = \mathbf{d}_j \boldsymbol{\alpha} + \mathbf{x}_j \boldsymbol{\beta} + \mathbf{z}_j \boldsymbol{\gamma} + \epsilon_{j,t} , \quad (15)$$

where $p_{j,t}$ is the log price of site j transacted in quarter t . The column vector \mathbf{d}_j has T elements: the first element for the overall constant is one, the period t element is one if site j was transacted in this period, zero otherwise. 1996Q1 is the omitted reference period. The vector \mathbf{x}_j contains binary indicators for unusual features of the site and for unusual aspects of the business dealings. The vector \mathbf{z}_j contains binary indicators for Berlin's 96 administrative sub-districts (Ortsteile). The site will be located in one of these sub-districts. The vector \mathbf{z}_j also contains a binary indicator for site location adjacent to a lake or the bank of a river. Finally, the vector contains binary indicators for site's location rating. This rating comes from Berlin's Senate Department for Urban Development and rates natural amenities, the quality of existing buildings, access to public transport and shopping facilities with in the neighborhood. The rating for a site takes one of four values: low, medium, high, very high.

Table 4: Parameter estimates for land price adjustment model.

Ground monument	-0.104	[0.056]
Contaminated soil	-0.120**	[0.033]
Demolished structure	0.030**	[0.006]
Land easement	-0.113**	[0.008]
Recoupment charge	0.082**	[0.007]
Non-private buyer	0.098**	[0.011]
Non-private seller	-0.075**	[0.008]
Other unusual circumstances	0.063**	[0.016]
Lake side	0.212**	[0.024]
Medium location quality	0.067**	[0.009]
High location quality	0.191**	[0.015]
Very high location quality	0.246**	[0.095]
$\hat{\sigma}_\epsilon^2$	0.194	
R^2	0.581	

Reports OLS estimates of the parameters in Eq. 15. Dependent variable is log price per sqm of undeveloped land. Regression uses 24,519 observations. Expert-based location rating comes from Berlin's Senate Department for Urban Development. Reference category is low location quality. Sub-district dummies, quarterly time dummies and overall constant are included, but estimated coefficients are not reported. Heteroscedasticity robust standard errors are reported in brackets. ** significant at 1%-level * significant at 5%-level.

The variables in \mathbf{z}_j control crudely for location effects. Without the inclusion, the estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ may suffer from omitted variable bias. Table 4 presents least squares estimates of the model in Eq. 15. The in-sample fit, as measured by the R^2 , is reasonably good. Except for the coefficient for ground monument, all coefficients are statistically significant at the 5% significance level. The signs of the point estimates, as well as their magnitude, are plausible.

Second, given the coefficient estimates, we compute the adjusted log real land price as

$$p_j = p_{j,t} - (\mathbf{d}_j - \mathbf{d}_b)\hat{\boldsymbol{\alpha}} - \mathbf{x}_{j,t}\hat{\boldsymbol{\beta}}. \quad (16)$$

The first entry of \mathbf{d}_b is one, the entries for the four quarters of the year 2009 are 0.25 each, the remaining entries are zero. The term in brackets in Eq. 16 converts prices to the base year 2009. The estimated value of $\mathbf{z}_j\boldsymbol{\gamma}$ is not considered for p_j , because it enters Eq. 15 only to prevent bias. The resulting p_j is in real terms and adjusted for unusual circumstances of the site. Using it in our analysis puts us on an equal footing with the land price information used by local surveyors to produce the BRW. The summary statistics for prices in natural scale in Table 1 are computed using $P_j = \exp\{p_j + 0.5\hat{\sigma}_\epsilon^2\}$, where $\hat{\sigma}_\epsilon^2$ is the estimated variance of the error term in Eq. 15 (Kennedy, 1983).

References

- Ahlfeldt, G. and Maenning (2010). Impact of sport arenas on land values: Evidence from Berlin. *Annals of Regional Science* 44, 205–227.
- Becker, S. (2014). *The Propagation-Separation approach. Theoretical study and application to magnetic resonance imaging*. Ph. D. thesis, Humboldt-Universität zu Berlin.
- Becker, S. and P. Mathé (2013). A different perspective on the propagation-separation approach. *Electronic Journal of Statistics* 7, 2702–2736.
- Bostic, R. W., S. D. Longhofer, and C. L. Redfearn (2007). Land leverage: decomposing home price dynamics. *Real Estate Economics* 35, 183–208.
- Bryan and Satre (2009). Semiparametric estimation of land price gradients using large data sets. *Federal Reserve Bank of Richmond Economic Quarterly* 95, 53–74.
- Diamond, D. B. (1980). The relationship between amenities and urban land prices. *Land Economics* 56.
- Fan, J. and J. S. Marron (1994). Fast implementations of nonparametric curve estimators. *Journal of Computational and Graphical Statistics* 3, 35–56.

- Kennedy, P. (1983). Logarithmic dependent variables and prediction bias. *Oxford Bulletin of Economics and Statistics* 45, 389–92.
- Polzehl, J. (2014). Adaptive weights smoothing, r package version 1.9-4. Technical report, CRAN R-Project. <http://CRAN.R-project.org/package=aws>.
- Polzehl, J. and V. Spokoiny (2000). Adaptive weights smoothing with applications to image restoration. *Journal of the Royal Statistical Society Series B* 62, 335–354.
- Polzehl, J. and V. Spokoiny (2003). Image denoising: pointwise adaptive approach. *Annals of Statistics* 31, 30–57.
- Polzehl, J. and V. Spokoiny (2006). Propagation-separation approach for local likelihood estimation. *Probability Theory and Related Fields* 135, 335–362.
- Polzehl, J. and V. Spokoiny (2008). Structural adaptive smoothing by propagation-separation-methods. In C. Chen, W. Härdle, and A. Unwin (Eds.), *Handbook of Data Visualization*, Berlin and Heidelberg, pp. 471–492. Springer.
- Stone, C. J. (1985). Additive regression and other nonparametric models. *Annals of Statistics* 13, 689–705.