

Spatial smoothing revisited: An application to rental data in Munich

Ludwig Fahrmeir¹, Göran Kauermann¹, Gerhard Tutz¹ and Michael Windmann¹

¹Institut für Statistik, München, Germany

Abstract: Spatial smoothing makes use of spatial information to obtain better estimates in regression models. In particular flexible smoothing with B-splines and penalties, which has been propagated by Eilers and Marx (1996), provides strong tools that can be used to include available spatial information. We consider alternative smoothing methods in spatial additive regression and employ them for analysing rental data in Munich. The first method applies tensor product P-splines to the geolocation of apartments, measured on a continuous scale through the centroid of the quarter where an apartment is. The alternative approach exploits the neighbourhood structure of districts on a discrete scale, where districts consist of a set of neighbouring quarters. The discrete modelling approach yields smooth estimates when using ridge-type penalties but can also enforce spatial clustering of districts with a homogeneous structure when using Lasso-type penalties.

Key words: P-splines smoothing, lattice data, lasso, spatial clustering

Received September 2022; **revised** January 2023; **accepted** January 2023

1 Introduction

According to German Law, increases in rents for apartments can be justified on ‘average rents’ for apartments that are comparable in size, location, equipment and quality. Such average rents are thereby published in official rental guides (Mietspiegel). Munich and most other larger cities publish rental guides, usually based on regression models with net rent or net rent per square meter as the dependent variable and characteristics of the apartment as explanatory variables. The models are based on data from surveys and are an official instrument in the German apartment rental market (see e.g., Fahrmeir et al., 1995 or Fitzenberger and Fuchs, 2017). Resulting rental guides appear in form of tables that are easy to use for both tenants and landlords. Therefore, suitable regression models should provide good predictive performance but should not be unnecessarily complex. A general discussion of statistical aspects in rental guides in Germany can be found in Kauermann and Windmann (2016).

Statistical consulting and analyses for rental guides for the city of Munich are carried out by the Department of Statistics, LMU Munich, since 1992. Coincidentally, this is the year when the first

Address for correspondence: Göran Kauermann, Institut für Statistik, Ludwigstraße 33, München 80539, Germany.

E-mail: goeran.kauermann@lmu.de



version of P-splines was presented by Eilers and Marx (1992) at the GLIM and Statistical Modelling Meeting in Munich. The use of P-spline for rental guides however occurred much later, after the seminal publication of Eilers and Marx (1996), for modelling the distinctly non-linear effect of size (in square meters) of an apartment—and possibly also of the year of construction—on its net rent per square meter. The inclusion of categorical variables characterizing equipment and quality then leads to additive regression models. Given that the data are available for research, they have been used as example in many further papers, including for instance Fahrmeir et al. (1998), Stasinopoulos et al. (2000), Kneib (2013) or De Bastiani et al. (2018).

It is well known that the location of an apartment has high predictive value, but suitable inclusion and modelling of this important spatial variable is non-trivial. The current Munich rental guide contains two types of location variables. First, as a categorical variable obtained from expert assessment in combination with exploratory statistical analysis. This variable describes the local neighbourhood, which is categorized into the average, high, and top residential areas. Additionally, the location of the apartment is also included in the rental guide, categorized into central and non-central locations. We refer to the webpage mietspiegel-muenchen.de for an exact definition. While the first variable describes the quality of the local residential area, the second variable refers to the spatial location within the city borders of Munich. The combination of the two discrete variables leads to 6 ($= 2 \times 3$) categories referring to location.

In this article we focus on the influence of the location of the apartment in a rigorous data-based manner, utilizing more detailed information from the data. That is we aim at extending the currently used rental guide. For reasons of data protection, the exact address of an apartment is not provided, but the city of Munich is divided into 475 quarters. For each apartment, we know the quarter the apartment is located in. Calculating the centroid for each quarter will allow for spatial smoothing by substituting the exact location of the apartment through the corresponding quarter centroid. The quarters themselves are grouped into 25 districts as a coarser categorization. We will also propose to use the districts and the resulting neighbourhood structure of the districts for smoothing as well as spatial clustering. Both, the centroids of quarters and the 25 districts are visualized in Figure 1.

We focus on spatial additive regression models, also called geoadditive models (Kammann and Wand, 2003), evaluating and comparing different but related forms of spatial smoothing. Bivariate and spatial smoothing for continuous and discrete spatial variables is described, for example, in Fahrmeir et al. (2021), covering tensor product P-splines, kriging, thin plate splines, radial basis function and Markov random field approaches. For P-spline smoothing in one or more dimensions we refer to the very readable surveys of Eilers et al. (2015) and Eilers and Marx (1992). Smoothing on lattice data, in particular, its contrasts to spatial econometrics, is discussed in Kauermann et al. (2012).

The article is organized as follows. In Section 3, we use centroids as continuous spatial variable and apply tensor product P-splines for estimating a smooth spatial effect. In Section 4, we use the districts as discrete spatial variable and smoothing methods based on lattice data. Section 4.2 considers lattice smoothing using a Ridge-type penalty, derived from Gaussian Markov random fields for the effects of neighbouring districts. In Section 4.3, we replace this smoothing penalty with a Lasso-type penalty, enforcing spatial clustering through a selection of neighbouring districts with differences of effects close to or equal to zero. We also consider the identification of clusters of neighbouring districts with (approximately) the same spatial effects, which is a sensible goal to formulate rental guides that are easier to interpret and communicate. Though our comparison of the three versions of spatial smoothing is illustrated and motivated by application to the Munich rental guide

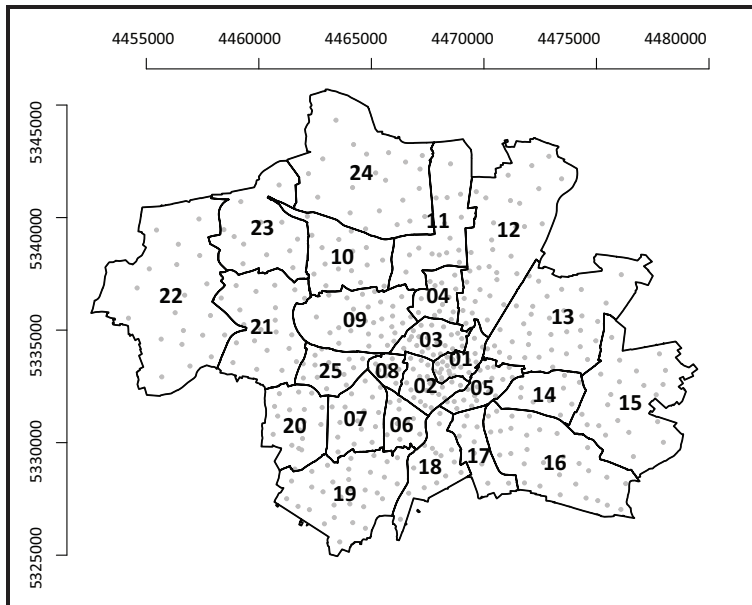


Figure 1 Map of districts 1 to 25 and centroids of quarters in Munich with coordinates.

data, we are convinced that this comparison will be rather useful for many other fields of application, for example, in epidemiology or labour market research.

2 Munich rental guide data

We analyse data from the 2019 Munich rental guide. The data were collected through a survey where the sample was drawn from the residential registration office. The survey was carried out through personal or video-based interviews. Only apartments were included that fulfilled particular legal criteria, which are of no particular interest to the application given here. All in all, we have data on 3 255 apartments. As response variable y , we use the net rent per square meter. The only metrical covariate is the size of the apartment in square meters, while the year of construction will be considered a categorical covariate. Spatial location will be denoted by r and is given either as a discrete variable with district numbers ranging from 1 (the city centre) to 25 or is considered a continuous variable representing the centroid of the quarter in which the apartment is. For the purpose of this application, categorical covariates are chosen as a subset of the covariates used for the official Munich rental guide. All variables included in our analyses are listed in Table 1. The assessment of the residential area is thereby carried out by a standing expert panel and provided for data analysis by the city council. The official denotation is ‘normal’, ‘good’ and ‘best’, which we rephrased here for better understanding. More detailed information is given in the documentation to this guide (see <https://2019.mietspiegel-muenchen.de/dokumentation.php>).

Table 1 Explanatory variables included in the models with rent as response variable.

Variable Name	Description	Range/Categories
Size $r = (r_{lon}, r_{lat})$	Size in square meter Coordinates of the centroid of quarter i	[20 m ² , 160 m ²] Longitude and latitude
Year	Year of construction	(1900–1948], (1948–1966], (1966–1977], (1977–1998] (> 1998)
Residential area	Quality of residential area	Standard, good, upscale
Modern/new floor	Binary	0: no modern/new floor, 1: modern/new floor
No balcony/terrace	Binary	0: with balcony/terrace, 1: no balcony/terrace

3 Penalized spline smoothing

3.1 Smooth spatial additive regression

We assume the net rent per square meter to follow a spatial, additive regression model of the form

$$y_i = s(r_i) + h(size_i) + x_i^T \beta + \varepsilon_i, \tag{3.1}$$

where $s(\cdot)$ is a smooth surface with r_i as the centroid of the quarter where apartment i is located and $h(\cdot)$ is a smooth function with $size_i$ as floor space of the apartment. For identifiability reasons it is assumed that $s(\cdot)$ and $h(\cdot)$ are centred around 0. Finally, we assume homoskedastic normality for the residual ε_i . The vector x_i includes 1 for the intercept and categorical covariates in binary coding. To be specific we set

$$x^T = \left(x_{year}^T, x_{residential\ area}^T, x_{modern/new\ floor}^T, x_{no\ balcony/terrace}^T \right)$$

where x_{year} is a four-dimensional dummy coded indicator vector with reference category (> 1998) and accordingly $x_{residential\ area}$ is three dimensional and also dummy coded with ‘standard’ as the reference category. Finally $x_{modern/new\ floor}$ and $x_{no\ balcony/terrace}$ are 0/1 variables as defined in Table 1.

The smooth functions can be estimated by P-splines, as originally introduced by Eilers and Marx (1992) in a first version and in the well-known article of Eilers and Marx (1996). The method gained massive interest in the last 25 years and we refer to Eilers et al. (2015) or Eilers and Marx (2021) for survey work. We do not give many technical details here but refer to the comprehensive survey work cited above. Instead, we want to focus on the idea of penalization and do this in the spatial context for the estimation of function $s(\cdot)$.

3.2 Spatial smoothing with P splines

To estimate $s(r)$ with centroids $r = (r_{lon}, r_{lat})$ measured on a continuous scale, several methods for spatial smoothing with continuous location variables are available, such as radial basis functions,

thin plate splines, kriging, and tensor product P-splines, see, for example, Ch.8 in Fahrmeir et al. (2021). We use the latter, having computational advantages in our application.

First, a two-dimensional B-spline basis is constructed on the convex hull of the spatial locations r_i , which are given by longitude and latitude values. To do so we construct a J dimensional B-spline basis on the longitude, denoted as B_{lon} and accordingly a K dimensional B-spline basis on the latitude, labelled as B_{lat} . The tensor product B-spline basis functions for estimation of the two-dimensional surface $s(r)$ are then defined through

$$W_{jk}(r) = B_{lon,j}(r_{lon}) \cdot B_{lat,k}(r_{lat}) \text{ for } j = 1, \dots, J, k = 1, \dots, K, \tag{3.2}$$

consisting of all pairs of univariate B-splines in the longitudinal and latitudinal directions. Here, $B_{lon,j}(\cdot)$ is the j -th column of B_{lon} and $B_{lat,k}$ the k -th column of B_{lat} . For simplicity, we assume equidistant knots in each direction, and we suppress the dependence on knots and the order of the B-splines notationally. Let $s(r)$ be approximated through

$$s(r_i) = W(r_i)\boldsymbol{\gamma} \tag{3.3}$$

where $W()$ has columns W_{jk} for $j = 1, \dots, J, k = 1, \dots, K$. Evaluation of the functions at the corresponding observations leads to the representation

$$y = W\boldsymbol{\gamma} + B\boldsymbol{\alpha} + X\boldsymbol{\beta} + \epsilon, \tag{3.4}$$

where B is the design matrix of univariate quadratic B-spline basis function evaluations referring to the modelling of apartment-size effects and W is the matrix with rows $W(r_i)$. The vector $\boldsymbol{\alpha}$ of coefficients is penalized through the well-known penalty

$$P_s(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^\top L\boldsymbol{\alpha}, \tag{3.5}$$

with D as the penalty matrix $L = D^\top D$ and D the (second-order) difference matrix. The coefficients γ_{jk} in (3.3) are penalized through row- and column-wise sums of squared univariate (second-order) differences

$$\begin{aligned} & \sum_j \sum_k (\gamma_{jk} - 2\gamma_{j-1,k} + \gamma_{j-2,k})^2, \\ & \sum_j \sum_k (\gamma_{jk} - 2\gamma_{j,k-1} + \gamma_{j,k-2})^2. \end{aligned} \tag{3.6}$$

These ridge-type penalties (3.6) for the location effect can be expressed as

$$P_l(\boldsymbol{\gamma}) = \boldsymbol{\gamma}^\top M\boldsymbol{\gamma}, \tag{3.7}$$

where the penalty matrix M is constructed from Kronecker products of univariate penalty matrices for both directions, see Fahrmeir et al. (2021, Section 8.2). Parameter estimates are obtained as

minimizers of the penalized L2-criterion

$$\|y - W\boldsymbol{\gamma} - B\boldsymbol{\alpha} - X\boldsymbol{\beta}\|^2 + \lambda_1 P_l(\boldsymbol{\gamma}) + \lambda_2 P_s(\boldsymbol{\alpha}). \tag{3.8}$$

Smoothing parameters can be estimated via the `mgcv` package in R (Wood, 2017) or mixed model approaches after parametrizing $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$. We emphasize that using (3.7) in (3.8) is very memory hungry due to the construction and high dimension of the penalty matrix M . A numerically more efficient version, utilizing (3.7) has been proposed in Currie et al. (2006), see also Xiao et al. (2013).

3.3 Results

We show the fitted spatial effect in Figure 2, where one can clearly see that the centre part of Munich has higher apartment rents which get smaller with more distance from the city centre. It is also seen that the decrease in rent in the north/south direction is less visible than in the east/west direction. This can be explained by the topology of Munich, with the river Isar running from south to north through Munich, and proximity to the river is mirrored in higher apartment rents.

The effect of apartment size is visualized in Figure 3. We see a decreasing effect but generally not a complicated structure of the function. In the remainder of the article, we will therefore model this effect with a six-dimensional quadratic B-spline basis and omit the penalty on the coefficients $\boldsymbol{\alpha}$ for simplicity. Finally, the effects of the categorical variables are listed in Table 2. The standard errors rely on the mixed model framework used for fitting and the standard output of the `gam()` procedure, see Wood (2017). It is seen that newer apartments are more expensive and that the quality of the

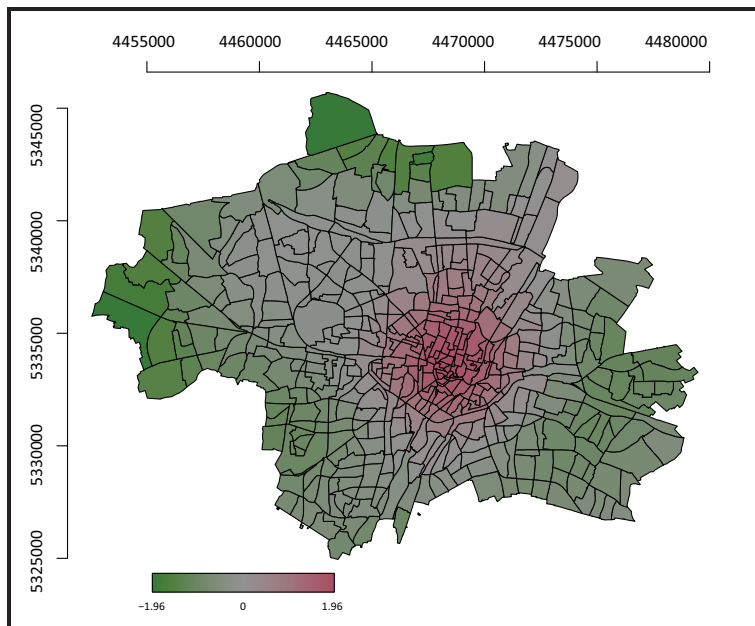
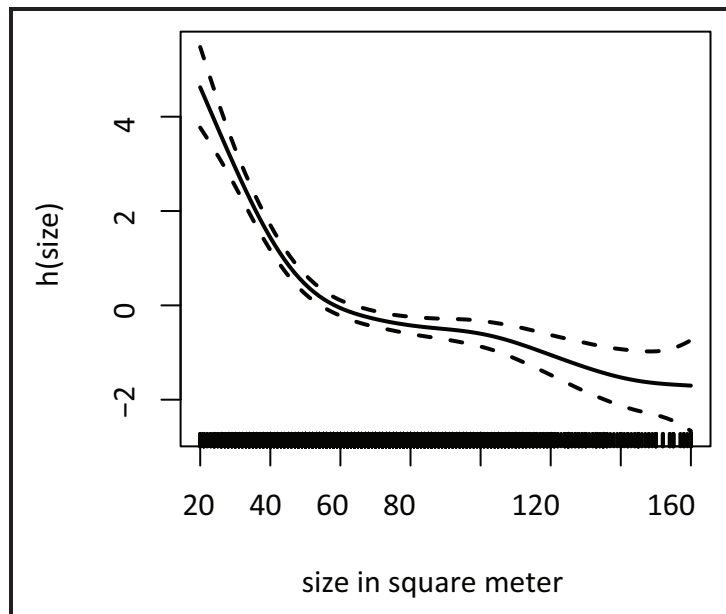


Figure 2 Estimated effect $\hat{s}(r)$ for every quarter.

Table 2 Estimates, standard error and resulting standardized values resulting from the P-Spline model.

Variable	Estimate	Standard Error	<i>t</i> -values
Year(1900–1948]	−3.3248	0.2106	−15.788
Year(1949–1966]	−3.2201	0.1980	−16.260
Year(1967–1977]	−2.5723	0.2150	−11.964
Year(1978–1998]	−2.1521	0.2122	−10.141
Good residential area	0.8477	0.1468	5.773
Upscale residential area	2.2882	0.2878	7.952
Modern/new floor	2.1137	0.1928	10.961
No balcony/terrace	−0.6543	0.1476	−4.434

**Figure 3** Effect of apartment size in P-Spline smoothing.

residential area has a significant positive effect. Moreover, if the apartment has a new floor and a balcony or terrace, the rent per square meter increases.

4 Lattice smoothing

4.1 Lattice data

As already discussed, the geolocation used for each apartment in the previous section is not its exact longitude and latitude coordinates, but the centroid of the corresponding quarter so that spatial smoothing as proposed above is carried out over a finite set of distinct centroids. For practical purposes, this is still burdensome when applying the rental guide, due to a large number of quarters

in Munich. It is therefore easier to coarsen the spatial variable and work with districts instead of quarters. We visualized the step from quarters to districts already in Figure 1. In this case, centroids of districts are less useful to work with, since there are only 25 distinct values. We now consider the district to correspond to a lattice with a neighbour structure, as visualized in Figure 1. This opens a new avenue of spatial smoothing by taking the lattice structure as spatial information instead of the Euclidean distance between the district centroids. We, therefore, use the following neighbourhood structure and numerate the districts from 1 to K . Hence, we can relate the location r_i to one of the districts and with a slight twist in the notation, we define with $r_i \in \{1, \dots, K\}$ the district in which apartment i is located. For each district, k we define with \mathcal{N}_k the neighbouring districts, hence those districts that have a common border with district k . Apparently, for $j \in \mathcal{N}_k$ we have the symmetric relation $k \in \mathcal{N}_j$. Smoothing can now be carried out by assuming that neighbouring districts have a similar rent level.

Note that smoothing on lattice data ignores the Euclidean distance. In spatial smoothing, as carried out in the previous section, we imposed through the penalty (3.5), that apartments lying close together have a similar rent level. In lattice smoothing, instead, it is only postulated that neighbouring districts have similar rent levels, regardless of their Euclidean distance. While in general, it sounds less plausible for smoothing to rely on lattice data if coordinates are available, it does make sense for rental guide data. First, districts themselves have some homogeneity regardless of their size. This is ignored in spatial smoothing but implicitly taken into account in lattice smoothing. Second, in terms of applicability, it is far easier to give a general rent level per district instead of per geolocation or quarter.

4.2 Lattice smoothing with ridge-type penalties

We consider again model (3.1), but now $s(r)$ is a discrete valued function, that is,

$$s : \{1, \dots, K\} \rightarrow \mathbb{R}. \tag{4.1}$$

We define $\gamma_k := s(k)$ so that estimation of $s(\cdot)$ corresponds to estimation of the parameter vector $\boldsymbol{\gamma}^T = (\gamma_1, \dots, \gamma_K)$. The aim is to achieve a smooth fit, such that neighboring districts do not differ strongly. This aim is reflected in the penalty

$$P(\boldsymbol{\gamma}) = \sum_{k=1}^K \sum_{j>k; j \in \mathcal{N}_k} (\gamma_k - \gamma_j)^2. \tag{4.2}$$

The penalty consists of squared differences of all possible combinations of neighbouring districts, where each combination is considered only once. This yields a penalty that discourages large deviations of effects associated with neighbouring regions. The penalty can also be derived from the Gaussian Markov random field approach, see Fahrmeir et al. (2021, Section 8.2.4).

Let the $K \times K$ dimensional difference matrix D have entries

$$D_{kl} = \begin{cases} 1 & \text{for } k = l \\ -1 & \text{for } l \in \mathcal{N}_k \text{ and } l > k \\ 0 & \text{otherwise,} \end{cases} \tag{4.3}$$

where $R = N(N - 1)/2$ with $N = \sum_{k=1}^K |\mathcal{N}_k|$ is the number of pairwise neighbor relations. Then the penalty component in the penalized L2-criterion is defined through the quadratic term

$$P(\boldsymbol{\gamma}) = \boldsymbol{\gamma}^\top L \boldsymbol{\gamma}, \tag{4.4}$$

where $L = D^\top D$. Since D does not have full rank also L does not have full rank. Generally, however, that does not matter and λ can be fixed by standard smoothing parameter selection tools.

By setting $\gamma_k = s(k)$, model (3.1) is overparameterized due to the intercept which is still included. We circumvent this problem by setting $\gamma_1 \equiv 0$ leading to parameter $\tilde{\boldsymbol{\gamma}} = (\gamma_2, \dots, \gamma_K)^\top$. We also reduce D to the $(K - 1) \times R$ dimensional difference matrix \tilde{D} which is obtained from D by deleting the first column in D . The penalty (4.4) for subvector $\tilde{\boldsymbol{\gamma}}$ then results to

$$P(\tilde{\boldsymbol{\gamma}}) = \tilde{\boldsymbol{\gamma}}^\top \tilde{L} \tilde{\boldsymbol{\gamma}} \tag{4.5}$$

with $\tilde{L} = \tilde{D}^\top \tilde{D}$, which now has full rank. Using standard linear algebra we can reparameterize model (3.1) to incorporate the penalty structure (4.5) in the covariates, leading to simple squared penalties. Let therefore $\mathbf{z}_i = (1(r_i = 1), \dots, 1(r_i = K))^\top$ be the indicator vector referring to the corresponding district of apartment i and accordingly, $\tilde{\mathbf{z}}_i$ be the vector with the first element dropped. Function $s(\cdot)$ can then be expressed as $s(r_i) = \tilde{\mathbf{z}}_i^\top \tilde{\boldsymbol{\gamma}}$, assuming $s(1) = 0$. This can be rewritten as

$$\begin{aligned} \tilde{\mathbf{z}}_i^\top \tilde{\boldsymbol{\gamma}} &= \underbrace{\tilde{\mathbf{z}}_i^\top}_{=: \tilde{\mathbf{v}}_i^\top} \underbrace{\tilde{L}^{1/2} \tilde{L}^{-1/2} \tilde{\boldsymbol{\gamma}}}_{=: \tilde{\boldsymbol{\delta}}} = \tilde{\mathbf{v}}_i^\top \tilde{\boldsymbol{\delta}}. \end{aligned} \tag{4.6}$$

The penalty (4.5) then has the form

$$P(\tilde{\boldsymbol{\gamma}}) = P(\tilde{\boldsymbol{\delta}}) = \tilde{\boldsymbol{\delta}}^\top \tilde{\boldsymbol{\delta}} \tag{4.7}$$

yielding as simple squared regularization. As remarked in Section 3, the non-linear effect of the apartment size is fitted by a simple six-dimensional quadratic B-spline and the remaining covariates are included in the model as before. This leads to the penalized L2 loss which we aim to minimize

$$\|y - \tilde{V} \tilde{\boldsymbol{\delta}} - B\boldsymbol{\alpha} - X\boldsymbol{\beta}\|^2 + \lambda P(\tilde{\boldsymbol{\delta}}). \tag{4.8}$$

where matrix \tilde{V} has rows $\tilde{\mathbf{v}}_i^\top$.

The corresponding smooth district effect is shown in Figure 4. The resulting effects for the covariates are given in Table 3. The effect of size is not visualized as this looks very much like the penalized fit shown in Figure 3. Generally, Table 3 shows comparable results to Table 2, hence changing the spatial model has only a small impact on the categorical covariates. Looking at Figure 4 we see little variation in the eastern and western suburbs of Munich, which already appeared from Figure 2. We may therefore question, whether some neighbouring districts in fact have the same rent level. This is further examined in the next modelling step.

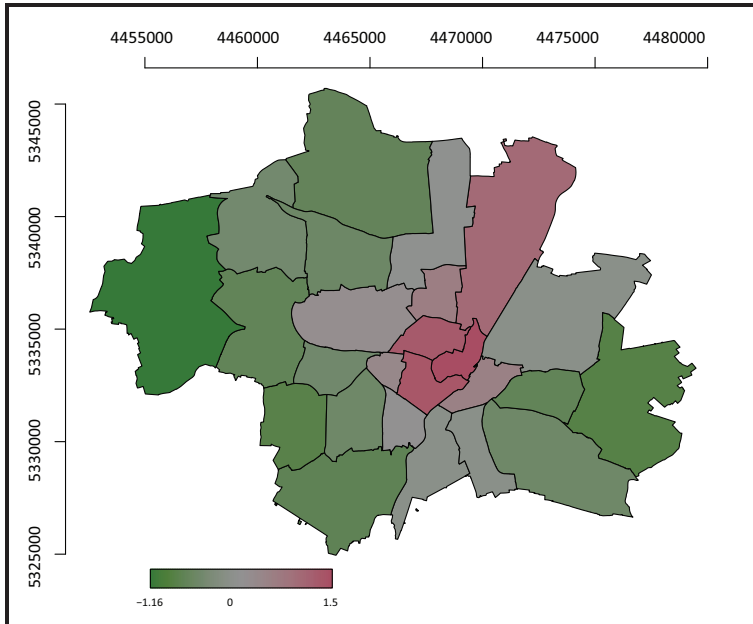


Figure 4 Estimated effect $\hat{\zeta}(r_i)$ for every quarter i .

Table 3 Estimates, standard error and resulting standardized values in the smooth lattice model.

Variable	Estimate	Standard Error	t-values
Year(1900–1948]	−3.08653	0.2080285	−14.83704
Year(1949–1966]	−3.01229	0.1969827	−15.29215
Year(1967–1977]	−2.48845	0.2153085	−11.55758
Year(1978–1998]	−2.10879	0.2124739	−9.92492
Good residential area	0.91586	0.1438097	6.36855
Upscale residential area	2.25760	0.2837630	7.95592
Modern/new floor	2.12058	0.1933850	10.96558
No balcony/terrace	−0.60723	0.1480415	−4.1017

4.3 Lattice smoothing with LASSO-type penalties

The penalty (4.7) corresponds to simple ridge regression, that is, we impose an L2 component on the coefficients. An alternative is to proceed with LASSO estimation, that is, replacing the L2 with L1. We, therefore, postulate that the difference between neighboring districts should be small in absolute terms, that is,

$$\sum_{k=1}^K \sum_{\substack{j \in \mathcal{N}_k \\ j > k}} |\gamma_k - \gamma_j| \tag{4.9}$$

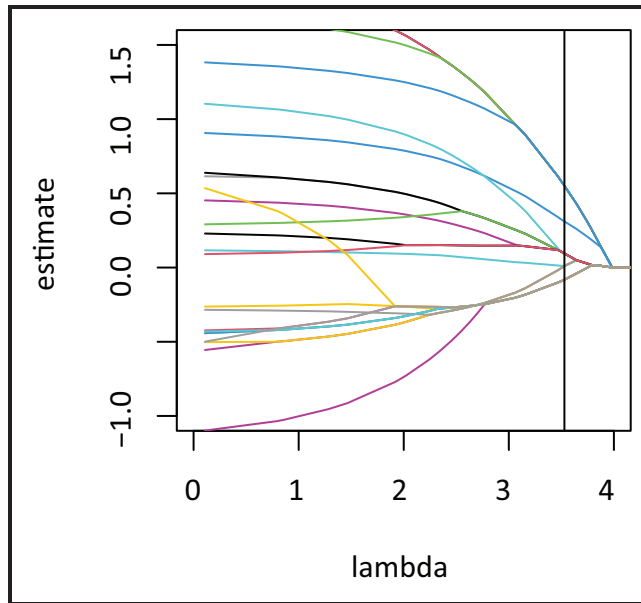


Figure 5 Generalized Lasso estimates for different penalization parameters. The vertical line shows the model with the smallest BIC.

which can be rewritten as $|D\boldsymbol{\gamma}|$ where $|\cdot|$ refers to the L1 norm, that is, the sum of absolute terms. Setting $\gamma_1 = 0$ replaces (4.9) accordingly which can be written as $|\tilde{D}\tilde{\boldsymbol{\gamma}}|$. The penalty has the form of a spatial-fused Lasso penalty. It enforces spatial clustering of adjacent districts with only small differences in spatial effects, which is very helpful for rental guides, but also in other areas of application, see Choi et al. (2018), Li and Sang (2019), Rahardiantoro and Sakamoto (2022), Masuda and Inoue (2022), and Ohishi et al. (2019) for similar fused Lasso penalties. A closely related Lasso-type penalty is the all-pairs penalty for clustering categories in categorical covariates, see, for example, the survey in Gertheiss and Tutz (2023). To solve the resulting optimization problem, the generalized Lasso (Tibshirani and Taylor, 2011) and its implementation in the R package `genlasso` (Arnold and Tibshirani, 2019) can be used. This minimizes the penalized L2 loss

$$\| \mathbf{y} - \tilde{E}\tilde{\boldsymbol{\gamma}} - B\boldsymbol{\alpha} - X\boldsymbol{\beta} \|^2 + \lambda |\tilde{D}\tilde{\boldsymbol{\gamma}}| \tag{4.10}$$

for different values of the penalty parameter λ , where \tilde{E} is the $n \times (K - 1)$ dimensional indicator matrix with entries $E_{i(k-1)} = 1$ if apartment i lies in district k and zero entries otherwise.

In Figure 5 we show the estimates for parameters γ_k for $k = 1, \dots, 25$ for different values of the penalty λ . For decreasing λ we obtain more differences in the rent levels for neighbouring districts. For each value of λ where estimation traces in Figure 5 split, we refit the model in unpenalized form with the Lasso constraints fulfilled. In other words, if the lasso estimates $\hat{\gamma}_{k,\lambda}$ for a specific value of λ fulfill $\hat{\gamma}_{k,\lambda} = \hat{\gamma}_{l,\lambda}$ we set impose the hard constraints $\gamma_k \equiv \gamma_l$ while for $\hat{\gamma}_{k,\lambda} \neq \hat{\gamma}_{l,\lambda}$ we allow γ_k and γ_l to differ, for $k, l = 1, \dots, 25$ and $k \neq l$. With these hard constraints, we refit the model without further penalty and calculate the resulting BIC value. This is shown in Figure 7. Instead of plotting

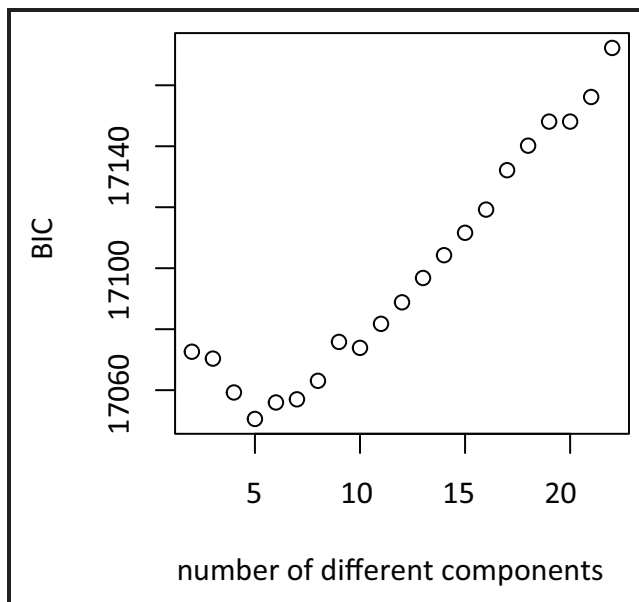


Figure 6 BIC for different models. The number of components corresponds to the number of splits resulting for different values of λ in Figure 5.

Table 4 Estimates, standard error and resulting standardized values in selected model based on BIC and unpenalized fit.

Variable	Estimate	Standard Error	<i>t</i> -values
Year(1900–1948]	−3.0195	0.2054	−14.699
Year(1949–1966]	−2.9757	0.1951	−15.254
Year(1967–1977]	−2.4416	0.2138	−11.420
Year(1978–1998]	−2.0833	0.2116	−9.846
Good residential area	0.9739	0.1329	7.330
Upscale residential area	2.2632	0.2669	8.479
Modern/new floor	2.1612	0.1931	11.191
No balcony/terrace	−0.6130	0.1477	−4.150

the BIC value against λ we plot this against the number of components, that is the number of splits shown in Figure 5. We see a clear minimum which results in the final model. We show the resulting smooth estimate in Figure 5. While there are five different levels, visually we can only recognize three distinct areas. The centre part includes a stripe towards the north with high rent levels. This area is the centre of Munich and the area along the river Isar towards the north. This is neighbored towards the east and west by some intermediate rent levels and lower rent levels in the eastern and western suburbs. The corresponding parameter effects are shown in Table 4, which are comparable to the estimates obtained before. Hence, changing the modelling of the spatial component does not have a big impact on the parametric effects.

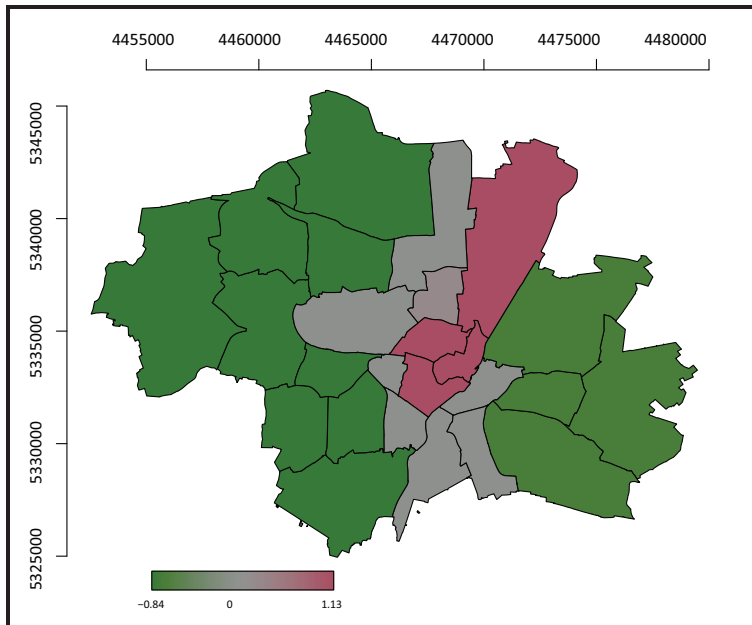


Figure 7 Estimated effects $\hat{s}(r_i)$ after Lasso model selection.

5 Discussion

In this article, we considered three different versions of smoothing applied to the Munich rental data. We used smoothing based on geolocations of apartments, grouped to centroids of quarters. We also grouped the quarters into districts and considered these as lattice, penalized by both, an L2 as well as an L1 loss.

At least in our application, the estimates of other covariate effects are quite robust with respect to the different versions of smoothing. Therefore, the choice of a specific spatial model will be guided by the specific goals of smoothing. While P-spline smoothing for quarters is well-suited for providing a refined global picture of spatial effects, Lasso-type smoothing for districts with automated spatial clustering is very useful for developing applicable rental guides.

We chose to run this comparison with data from Munich to remember Brian Marx's many visits to our home town. Beyond working on statistical modelling and giving lectures at the Department of Statistics, LMU Munich, Brian enjoyed good things in life together with us, such as hiking or biking along the banks of the river Isar, visiting beergardens in and around Munich, and having fun with friends. Although this article focuses on spatial smoothing for rental guides, spatial clustering as in Section 4.3. can be of interest in many other applications. This motivates further methodological research: In its current version, the generalized Lasso is restricted to linear models with an additional spatial component. It might be useful or even necessary, to allow for Ridge-type penalties, as for P-splines, in combination with generalized Lasso penalties. The resulting optimization problem is challenging. However, We see three possible approaches: First, an alternating optimization algorithm, switching between optimizing the Ridge component for fixed parameters of the Lasso

component, and vice versa, see Ohishi et al. (2019) for a grouped Lasso part and a generalized Lasso part. Second, approximation of the Lasso penalty through a differentiable function, see Oelker and Tutz (2017). Third, a Bayesian approach based on the Bayesian Lasso, with a conditionally Gaussian prior, along the lines of Masuda and Inoue (2022).

Acknowledgment

The second author acknowledges the support of the Munich Center for Machine Learning (MCML).

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

Supplementary Material

Supplemental material for this article is available online.

References

- Arnold T and Tibshirani R (2019). genlasso: Path Algorithm for Generalized Lasso Problems. R package version 1.4. <https://github.com/glmgen/genlasso>
- Choi H, Song E, Hwang S and Lee W (2018). A modified generalized lasso algorithm to detect local spatial clusters for count data. *AStA – Advances in Statistical Analysis*, **102**, 537–563.
- Currie ID, Durban M and Eilers PH (2006). Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**, 259–280.
- De Bastiani F, Rigby RA, Stasinopoulous DM, Cysneiros AH and Uribe-Opazo MA (2018). Gaussian Markov random field spatial models in GAMLSS. *Journal of Applied Statistics*, **45**, 168–186.
- Eilers PH and Marx BD (1992). Generalized linear models with p-splines. In *Advances in GLIM and Statistical Modelling*, edited by Ludwig Fahrmeir, Brian Francis, Robert Gilchrist, Gerhard Tutz, pages 72–77. New York: Springer.
- Eilers PH and Marx BD (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 89–121.

- Eilers PH and Marx BD (2021). *Practical Smoothing: The Joys of P-splines*. Cambridge University Press.
- Eilers PH, Marx BD and Durbán M (2015). Twenty years of P-splines. *SORT: Statistics and Operations Research Transactions*, **39**, 0149–186.
- Fahrmeir L, Gieger C and Klinger A (1995). Additive, dynamic and multiplicative regression. In *Applied statistics—recent developments. Proceedings. Pfingsttagung 1994 der Deutschen Statistischen Gesellschaft, Festkolloquium der 20-Jahrfeier des Fachbereichs Statistik*, Universität Dortmund, Germany, Mai 1994, pages 95–130. Göttingen: Vandenhoeck & Ruprecht.
- Fahrmeir L, Gieger C and Klinger A (1998). Regression approaches to rental guides. In *Econometrics in Theory and Practice*, pages 241–254. Springer.
- Fahrmeir L, Kneib T, Lang S and Marx BD (2021). *Regression* (2nd ed.). Springer.
- Fitzenberger B and Fuchs B (2017). The residency discount for rents in Germany and the tenancy law reform act 2001: Evidence from quantile regressions. *German Economic Review*, **18**, 212–236.
- Gertheiss J and Tutz G (2023). Regularization and model selection for categorical/ordinal data. In *Trends and Challenges for Categorical Data Analysis*, edited by M Kateri and I Moustaki. Springer.
- Kammann E and Wand MP (2003). Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **52**, 1–18.
- Kauermann G, Haupt H and Kaufmann N (2012). A hitchhiker's view on spatial statistics and spatial econometrics for lattice data. *Statistical Modelling*, **12**, 419–440.
- Kauermann G and Windmann M (2016). Mietpiegel heute. *ASta Wirtschafts-und Sozialstatistisches Archiv*, **10**, 205–223.
- Kneib T (2013). Beyond mean regression. *Statistical Modelling*, **13**, 275–303.
- Li F and Sang H (2019). Spatial homogeneity pursuit of regression coefficients for large datasets. *Journal of the American Statistical Association*, **114**, 1050–1062.
- Masuda R and Inoue R (2022). Point event cluster detection via the bayesian generalized fused lasso. *ISPRS International Journal of Geo-Information*, **11**.
- Oelker M-R and Tutz G (2017). A uniform framework for the combination of penalties in generalized structured models. *Advances in Data Analysis and Classification*, **11**, 97–120.
- Ohishi M, Fukui K, Okamura K, Itoh Y and Yanagihara H (2019). *Estimation for spatial effects using the fused lasso* (Technical Report 19-07). Hiroshima Statistical Research Group.
- Rahardiantoro S and Sakamoto W (2022). Optimum tuning parameter selection in generalized lasso for clustering with spatially varying coefficient models. *IOP Conference Series: Earth and Environmental Science*, **950**, 012093.
- Stasinopoulos D, Rigby R and Fahrmeir L (2000). Modelling rental guide data using mean and dispersion additive models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **49**, 479–493.
- Tibshirani RJ and Taylor J (2011). The solution path of the generalized lasso. *The Annals of Statistics*, **39**, 1335–1371.
- Wood, S. N. (2017). *Generalized additive models: An introduction with R*. Boca Raton: CRC press.
- Xiao L, Li Y and Ruppert D (2013). Fast bivariate p-splines: The sandwich smoother. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **75**, 577–599.