Hindawi Publishing Corporation
ISRN Economics
Volume 2013, Article ID 158240, 16 pages
http://dx.doi.org/10.1155/2013/158240

Research Article

Choosing the Right Spatial Weighting Matrix in a Quantile Regression Model

Philip Kostov

Lancashire Business School, University of Central Lancashire, Greenbank Building, Preston, Lancashire PR1 2HE, UK

Correspondence should be addressed to Philip Kostov; pkostov@uclan.ac.uk

Received 4 December 2012; Accepted 27 December 2012

Academic Editors: D. M. Hanink and W. R. Reed

This paper proposes computationally tractable methods for selecting the appropriate spatial weighting matrix in the context of a spatial quantile regression model. This selection is a notoriously difficult problem even in linear spatial models and is even more difficult in a quantile regression setup. The proposal is illustrated by an empirical example and manages to produce tractable models. One important feature of the proposed methodology is that by allowing different degrees and forms of spatial dependence across quantiles it further relaxes the usual quantile restriction attributable to the linear quantile regression. In this way we can obtain a more robust, with regard to potential functional misspecification, model, but nevertheless preserve the parametric rate of convergence and the established inferential apparatus associated with the linear quantile regression approach.

1. The Spatial Quantile Regression Model

The spatial quantile regression model [1] is a straightforward quantile regression generalisation of the popular, in spatial econometrics, linear spatial lag model. More specifically it can be written as

$$y = \lambda(\tau) Wy + X\beta(\tau) + u, \tag{1}$$

where Wy is a spatially lagged dependent variable, specified via a predetermined spatial weighting matrix W, X is the design matrix containing the independent variables (covariates), and u is a residuals vector. Here we only have one spatially lagged dependent variable but this is not an essential assumption, and more than one spatial weighting matrix can be easily incorporated. This representation is similar to the linear spatial lag regression model, but here coefficients are allowed to vary with the quantile, rather than being assumed fixed.

This model has some attractive properties. First, the original motivation for Kostov's [1] proposal is to alleviate the potential bias arising from inappropriate functional form assumptions in a spatial model. In simple terms the underlying logic is as follows. Omitting spatial dependence typically introduces estimation bias in the presence of spatial lag dependence when the wrong functional form specification is employed. Hence a natural way to circumvent the problem is to estimate the underlying function nonparametrically. The sample sizes used in many empirical studies are however often too small for efficient application of nonparametric methods. Semiparametric methods could then be used to alleviate the problem. The linear quantile regression is such a semi-parametric method. Although it cannot be guaranteed to entirely eliminate the adverse effects of functional form assumptions, such methods can greatly reduce them. In particular Kostov [1] argues that for a typical hedonic model the (linear) quantile restriction is appropriate.

A major advantage of the quantile regression approach is the opportunity to estimate a flexible semiparametric model, which is nevertheless characterised by parametric rate of convergence, thus making it suitable for empirical analysis in small sample cases. Furthermore a well-developed set of tools for efficient inference is available (see [1] for details).

Spatial modelling has however been focused mostly on estimation issues. For example, Kostov [1] assumes that the exact form of the process generating the spatial dependence is given. This is a typical assumption of an "estimation focused" approach to spatial modelling in that the spatial weighting matrix used to specify the model is known. The spatial

weighting matrix is however a part of the specification process. It needs to be prespecified. There could be cases where the underlying theoretical model provides some guidance but more often than not this is not the case. Consequently in empirical applications of spatial models the selection of spatial weight matrices is characterised by a great deal of arbitrariness. This arbitrariness presents a serious problem to the inference in such models since estimation results have been shown to critically depend on the choice of spatial weighting matrix [2–4]. Even more importantly, there is an interplay between spatial weighting matrix and functional form choice. Using the wrong spatial weighting matrix has broadly speaking the same implications as ignoring existing spatial dependence. Therefore functional form and spatial weighting matrix specification have to be considered simultaneously. The problem is not as severe in nonparametric models, because most nonparametric estimation methods are typically consistent even in the presence of spatial dependence. The wrong spatial weighting matrix however would still introduce inefficiency in the non-parametric estimates, which with smaller samples can seriously impede inference. In a parametric setup, the wrong spatial weighting matrix introduces bias even when the right functional form is used.

2. Selection of Spatial Weighting Matrix

Owing to these considerations it would be advantageous to have methods to choose an appropriate spatial weighting matrix. Selecting the "right" spatial weighting matrix can serve twofold purpose. First, it will increase the efficiency of the model estimates, as discussed previously. Second, when the nature of the process generating spatial dependence is of particular interest (e.g., in social interaction models) the form of the spatial weighting matrices consistent with that data generation process becomes a major inferential problem. In such cases we need to find the appropriate spatial weighting matrix, since this is the explicit subject of the research problem. In this paper we consider the issue in a spatial quantile regression framework.

In the following we will briefly review some approaches designed to reduce the arbitrariness of spatial weighting matrix choice (mostly) in linear models. Then we will discuss the possible extensions to the spatial quantile regression. The approach taken in this paper falls in the framework of selecting the spatial weighting matrix either implicitly or explicitly from a pre-defined set of candidates.

Holloway and Lapar [5] used a Bayesian marginal likelihood approach to select a neighbourhood definition (cut-off points for the neighbourhood), but one can consider their approach as a general model selection approach, which could be applied to any other set of competing models. A particularly active strand of research is concerned with Bayesian model averaging (BMA) approaches. LeSage and Parent [6] proposed a BMA procedure for spatial model which incorporates the uncertainty about the correct spatial weighting matrix. LeSage and Fischer [7] extended the latter approach into an MC3 (Markov Chan Monte Carlo Model Composition) method to select an inverse distance nearest neighbour type of spatial weighting matrix for the linear spatial model. Crespo-Cuaresma and Feldkircher [8] further extend this procedure to deal with different types of spatial weighting matrices by introducing Bayesian model averaging inference conditional on a given spatial weighting matrix. Crespo-Cuaresma and Feldkircher [8] use spatial filtering to resolve the endogeneity issue and in this way focus on the regression part of the model rather than on the spatial dependence itself. The approach above implicitly assumes that the spatial dependence can be characterised by a single spatial weighting matrix. This assumption can be relaxed but at a considerable computational cost. Eicher et al. [9] proposed instrumental variables Bayesian model averaging procedure which is essentially a hierarchical Bayesian counterpart to the frequentist two-step estimation that accounts for model uncertainty in both steps. Although Eicher et al. [9] do not deal with spatial dependence, but only with the more general issue of endogeneity, since spatial lag dependence is a particular type of endogeneity, their approach can be readily applied to spatial lag models.

Finally from a non-Bayesian point of view Kostov [10] suggested a two-step procedure for selecting spatial weighting matrix that is applicable to a wide range of prespecified candidates. This procedure is motivated by considerations specific to spatial models (and the proposed computational algorithms are tuned for this purpose), but otherwise it deals with the endogeneity problem in the same way as Eicher et al. [9].

3. Proposal Outline

This paper proposes extending the methodology adopted in Kostov [10] to a quantile regression setting. In what follows we will first briefly explain the previously mentioned approach. We will then highlight the particularities of the extension of this procedure to quantile regression models. Furthermore we will briefly comment on the different alternative options and the reasons for the specific choices we adopt. Our contribution is twofold. First we adapt the approach of Kostov [10] to a (linear) quantile regression model. Second, since as we will explain later, the original approach has a prediction focus, we further expand it to focus on structure discovery (i.e., identifying the "true sparsity pattern").

Kostov's [10] approach is based on Kelejian and Prucha's [11] two-stage least squares method to estimate spatial models. In this method, spatially lagged independent variables are used as instruments for the spatially lagged dependent variable. The first step (instrumentation) is a least squares regression of the lagged dependent variable on the lagged independent variables. In the second step, the fitted values from the first stage regression replace the original endogenous variable in the estimation of the model's coefficients. Kostov [10] retains the first step of this procedure (which projects the spatially lagged dependent variable in the vector space of the instruments). He however suggests implementing this first step for a number of different spatial weighting matrices resulting in an augmented second stage model that includes a large number of transformed, in the first step, variables (instead of the original spatial weighting matrices) to be considered. In this way the problem of choice of spatial weighting matrix becomes a variable selection problem (amongst

the previously mentioned transformed variables). The other interesting feature of Kostov's [10] paper is the application of a component-wise boosting algorithm as a variable selection method in the second step. Any other variable selection method could be used but Kostov's [10] choice is mainly motivated by computational considerations in dealing with large number of potential alternatives.

In a nutshell the approach of Kostov [10] amounts to transforming the spatial weighting matrix selection problem into a high-dimensional (due to the potentially large number of alternatives) variable selection problem, for which "standard" methods could be applied. The crucial point is Kostov's [10] approach to establish equivalence between the two-stage spatial least squares method and the proposed component-wise boosting alternative. Therefore in order to extend the same logic to a spatial quantile regression model we need to find a variable selection equivalent to a quantile regression estimation method. We will deal with these two issues in turn.

The first issue is the estimation method for spatial quantile regression. We are aware of two main approaches able to consistently estimate such models. The first is the application of Zietz et al. [12] who use the results of Kim and Muller [13] for quantile regression estimation under endogeneity. The other approach is presented in Kostov [1] who builds upon the methods developed by Chernozhukov and Hansen [14, 15]. In Kostov's [1] application one minimises a matrix norm over a range of values for the spatial dependence parameter. This is convenient when there is a single spatial weighting matrix. With many candidates however this would involve such minimisation over a multidimensional grid, which makes such an approach prohibitively expensive in terms of computational requirements, particularly when the number of potential spatial weighting matrices is large. Alternatively the methods developed in Chernozhukov and Hong [16] could be used to estimate such a model, but this will still involve considerable computational costs, and we will not pursue this option here. Furthermore, the main appeal of this procedure over the two-stage quantile regression is the availability of robust inference tools, since it is computationally more demanding (see [1] for detailed comparison). Here we are interested in selecting the model specification, rather than estimating a prespecified model. With view to this simpler methods are preferable. Once the final model specification is established and inference is the main focus, any estimation method could be applied, depending on the purpose of the analysis.

The Zietz et al. [12] approach on the other hand represents a simple two-stage quantile regression. As such it is very similar to the spatial two-stage least squares approach of Kelejian and Prucha [11], which is being used in Kostov [10]. Therefore using the theoretical results of Kim and Muller [13] we can extend their two-stage quantile regression estimator to include variable selection, using essentially the same arguments as Kostov [10]. Such an extension however comes at a cost. The previous approach uses two consecutive quantile regression estimators defined at the same quantiles at both steps. In the context of selecting spatial weighting matrices the first step would carry considerable computational burden, mainly because of the large number of alternatives to be considered. This means that the computational burden will be increased since separate first step estimation would need to be carried over each quantile that is to be considered. It would therefore have been very useful if one could have replaced the first step with, for example, least squares estimation, because this would then only need to be carried once. There have been empirical applications of two-stage estimation where the estimators used in the first and the second stage are different. For example, Arias et al. [17] and Garcia et al. [18] used least squares in the first step followed by quantile regression in the second. Unfortunately in general settings such an approach could induce asymptotic bias in the overall estimator (see [13] for details). In simple terms the robustness of two-stage estimators could be lost when the first stage applies an estimator that is not robust. Owing to this we consider here only estimators that employ the same type of estimator for both steps. This means that we will have to use quantile regression in both steps. The use of quantile regression for each estimated quantile greatly increases the computational costs of the method compared to the linear model.

The proposal of Kostov [10] translates into using variable selection algorithm in the second stage estimation. As discussed previously this variable selection algorithm needs to be the same type as the one in the original two-stage estimator. Therefore we need a quantile regression variable selection method. There are several possibilities for the latter. First, the component-wise boosting approach used in Kostov [10] can be adapted to do variable selection in a quantile regression setting. At this end Fenske et al. [19] demonstrated that using the check function used to define the quantile regression as an empirical loss function leads to an alternative quantile regression estimator. Using this approach looks like a natural extension to the logic of Kostov [10], particularly since he does mention the potential use of alternative empirical risk functions.

Another option is to use regularised (i.e., penalised) quantile regression to select covariates. Two of the most popular regularisation approaches, namely the least absolute shrinkage and selection operator (lasso) of Tibshirani [20] and the smoothed clipped absolute deviations (SCAD) method of Fan and Li [21] have already been considered in quantile regression setting (see [22–24]. In general these papers have established the consistency of such regularised estimators for quantile regression problems, subject to appropriately chosen "optimal" penalty parameter(s).

So, a straightforward generalisation of the approach of Kostov [10] to quantile regression involves a similar two-step procedure. In the first step a number of quantile regressions are implemented (for each candidate spatial weighting matrix) regressing the spatially lagged dependent variable on the spatially lagged independent variables. The fitted values from the first step are then used as additional explanatory variables (thus augmenting the original set of covariates). This second step is estimated using variable selection methods to effectively select the appropriate spatial weighting matrix.

There are several important features of such implementation. First, since it is based on a consistent two-stage estimator (the two-stage quantile regression estimator of Kim and Muller [13]) it should retain the consistency properties

of the original estimator as long as the second step is also consistent. As already discussed the price we have to pay for maintaining such consistency is the need to estimate separate first step quantile regression for each quantile considered. Second, similarly to other the two-step procedures, standard errors, or indeed any inference based solely on the second step estimation would be invalid. One could consider asymptotic inference based on the results of Kim and Muller [13]. Alternatively the overall (two-step) estimator could be bootstrapped. Note however that due to the computational costs of the first step (details of which we present later on) such an implementation would be prohibitively expensive. The best option is to follow the suggestion of Kostov [10] and only use the proposed estimator to select the structure of the model, which can then be estimated using standard methods.

4. Variable Selection Step

From now on we will take the first (instrumentation) step as given and will focus entirely on the variable selection step. We will argue that in order to obtain efficient inference it is desirable that in the second step a variable selection procedure characterised by the so-called oracle property is implemented. In simple words if an estimator possesses the oracle property this means that the asymptotic distribution of the obtained estimates is the same as this of the "oracle estimator," that is, an estimator constructed from a priori knowledge of which coefficients should be zero. Therefore estimators possessing the oracle property can be used for both variable selection and inference. Here we deviate considerably from Kostov [10] who claimed that since the proposed procedure is only to be used for selecting the model structure, the oracle property in not essential. Actually the brief discussion provided in Kostov [10] implies (without explicitly mentioning it) that instead of consistency, the weaker condition of persistence [25] would be sufficient. While the oracle property aims at minimising prediction error, the persistence tries to avoid wrongly excluding significant variables.

Therefore using persistent estimator implicitly includes a measure of uncertainty very much in the spirit of Bayesian methods. The actual aim in many typical applications however would be to discover the "true sparsity pattern." For such purposes a combination of consistent and oracle estimators have been shown to be able to discover the underlying structure and retain the oracle property. This idea has been formalised and theoretically developed in Fan and Lv [26]. Their methodology consists of a screening step (using a consistent variable selection method) followed by an oracle method (estimation step) to produce the final model. Even if both methods used in such a combination do not possess the oracle property the overall procedure will gain from improved convergence rates and can still be consistent subject to some additional conditions (see e.g., [27] for detailed discussion and simulation evidence). Here however we prefer to avoid imposing such additional conditions and would prefer applying a method possessing the oracle property in the estimation step.

An additional advantage of combining screening and estimation steps is the reduction in computational requirements and improved convergence rates. The convergence rates of estimators possessing the oracle property depend on the relative (to the complexity of the employed model) sample size. Owing to this it would be desirable if the size of the initial model is reduced. Applying an estimator possessing the oracle property to such a reduced model will improve this estimator's efficiency (compared to the case when it is applied directly to the larger, unrestricted model). In addition to the theoretical efficiency gain, this could bring considerable practical gains in greatly reducing the computational requirements of the selection algorithm(s) involved. Such a reduced model can be produced by using any consistent estimator (i.e., an estimator that (asymptotically) retains the important variables (i.e., variables with nonzero coefficients)). In simple terms the combination of screening and estimation steps reduces the false positive discovery rate (i.e., falsely retaining unimportant variables) and hence is tuned to structure discovery. Retaining such unimportant variables often improves prediction accuracy or uncertainty measures and hence can result in larger models (see [27], for a detailed discussion).

So we propose applying a combination of screening and estimation steps to the already transformed model. Such a proposal can be viewed as unnecessary complication to an already involved procedure. Nevertheless it has significant advantages. First, as we will show, it nests within itself the straightforward implementation of the Kostov's [10] proposal. Second since the combination of screening and estimation steps is equivalent to a single step estimation, but has better convergence rates, one can potentially further reduce the set of potential spatial weighting matrices by maintaining the consistency of the overall estimation procedure. The previously mentioned equivalency means that the overall proposed spatial model estimator which comprises three distinct steps (instrumentation, screening, and estimation) is still equivalent to the two-step method used to motivate it (i.e., the two-step quantile regression).

As discussed previously using either a boosting or regularisation approach can be viewed as different implementations of the same idea, namely, implementing a variable selection step in a two-stage quantile regression estimator. In order to ascertain the relative merits of these two alternatives let us first consider their relative computational requirements. The boosting approach is considerably less intensive in terms of computation. It has another important, in the context of spatial weighting matrix selection, advantage over the regularisation approach. Since the component-wise boosting approach processes the candidate variables one by one (see the next section for description of the component-wise boosting algorithm), high degrees of correlation amongst variables (and therefore singularity issues due to a highly nonorthogonal design) do not present significant problem to effectively reduce the set of alternatives. The nature of the spatial weighting matrix selection problem could involve simultaneous consideration of numerically very similar alternatives, which could be infeasible in the regularisation approach. Furthermore although extensively studied and shown to be consistent it is unclear whether the boosting approach possesses the oracle property. It is therefore desirable to implement the component-wise boosting as a screening method.

Then an oracle property regularisation approach can be implemented in the estimation step. Note that if we stop after the screening step, we obtain a straightforward quantile regression generalisation of the approach of Kostov [10].

Due to the fact that component-wise boosting is much faster than direct implementation of any regularisation approach, the previous strategy achieves considerable reduction in the computational requirements and makes the overall approach computationally feasible. Note that in addition to the computational requirements, direct application of a regularisation estimator could be infeasible in many spatial problems, simply because of the nature of the spatial weighting matrices to be considered. When a large number of such matrices is considered (as in [10]), the resulting transformed variables could be quite similar numerically. This could result in singularities that would prevent direct application of a regularised quantile regression estimation of the transformed problem.

In addition to the approach outlined previously we will also consider adopting the stability selection approach of Meinshausen and Bühlmann [28] to the boosting estimation. Strictly speaking stability selection is not an estimator per se, but application of a combination of subsamplings (although other forms of bootstrap could be used) and a variable selection algorithm. It provides a measure of how often a variable is selected, and therefore by using a threshold only persistent variables can be selected.

5. Technical Implementation Details

The screening step will use component-wise boosting estimation of quantile regression, following Fenske et al. [19]. Consider the general linear quantile regression model:

$$y = \eta(X) + \xi = \beta_0(\tau) + \sum_{j=1}^{k} \beta_j(\tau) x_j + \xi, \tag{2}$$

where y and x_j are the dependent and independent variables (the latter collected in the matrix X) and τ is the quantile of interest.

Boosting can be viewed as a functional gradient descent method that minimises the constrained empirical risk function $(1/n)\sum_{i=1}^{n} L(y_i, \eta(X))$, where $L(\cdot)$ is some suitable loss function. The τth quantile regression is obtained when the so-called check function is used as empirical risk:

$$L_\tau(y_i, \eta(X)) = \rho_\tau(y_i - \eta(X)),$$
$$\rho_\tau(u) = \begin{cases} u\tau & u \geq 0 \\ u(\tau - 1) & u < 0. \end{cases} \tag{3}$$

In the a notation mentioned we intentionally use the general additive predictor $\eta(\cdot)$ since it allows for generalisation of the approach to nonlinear and indeed nonparametric versions of the quantile regression problem. Since the check function is used to define the conventional linear quantile regression estimator of Koenker and Basett [29], using it as an empirical risk function solves an equivalent optimisation problem.

The boosting algorithm is initialised by an initial value for η, for example, η_0. This implies an initial evaluation for the underlying function $\widehat{f}_0$. In this case all underlying functions will be linear. Typically one starts with an offset set to the unconditional mean of the response variable, but in the quantile regression the unconditional median is used instead (see [19] for details and justification of this choice).

Let $\widehat{g}_{j,m}$ and $\widehat{f}_{j,m}$ denote the evaluations of the corresponding learners (in this case linear functions) for component j at iteration m. $\widehat{g}_{j,m}$ represents the learner (i.e., linear function) fitted to the current "residuals" while $\widehat{f}_{j,m}$ is the "global" evaluation of the same function (see the following algorithm).

Then the component-wise boosting algorithm iteratively goes through the following steps.

(1) Compute the negative gradient of the empirical risk function evaluated at the current function estimate (η_m for every step from $m = 1, \ldots$):

$$u_i = -\left.\frac{\partial \rho_\tau(y_i - \eta(x_i))}{\partial \eta}\right|_{\eta = \widehat{\eta}_{m-1}(x_i)} \quad \text{for } i = 1, 2, \ldots, n. \tag{4}$$

(2) Use the previous calculated negative gradients to fit the underlying function $\widehat{g}_{j,m}(\cdot)$ for each dependent variable (component). Here $\widehat{g}_{j,m}(\cdot)$ is fitted to the current residuals value of the used function at iteration m.

Find the best fitting base learner $j^* = \operatorname{argmin}_{1 \leq j \leq k} \sum_{i=1}^{n} (u_i - \widehat{\beta}_{j,m} x_i)^2$.

(3) Update the best fitting base learner for a given step size ν:

$$\widehat{f}_{j^*,m} = \widehat{f}_{j^*,m-1} + \nu \widehat{g}_{j^*,m}(\cdot)$$
$$\widehat{f}_{j,m} = \widehat{f}_{j,m-1} \quad \text{for } \forall j \neq j^* \tag{5}$$

The algorithm iterates between steps (1 and 3) until a maximum number of iterations are reached. The algorithm described above needs an updating step ν. In this application we will use $\nu = 0.3$. See Kostov [10] and references therein for a discussion about this choice and demonstration that the final results are insensitive to a wide range of choices. The other element of interest is the criterion used to decide which is the "best fitting" component in step (2). Here we use $L2$ norm (see the aforementioned), but other choices are also possible. The greatest advantage of $L2$ norm is that the base learners can be updated by simple least squares fitting, which is computationally fast and convenient (see [19]). In this particular case, since we use linear quantile regression, updating the base learners amounts to applying univariate least squares.

A regularised linear quantile regression estimator can be formally defined as

$$\min_{\beta_\tau} \sum_{i=1}^{n} \rho_\tau\left(y_i - X^T \beta_\tau\right) + \lambda J\left(X^T \beta_\tau\right), \tag{6}$$

where β_τ is the vector of the linear coefficients pertaining to the covariates, that is, $\beta_\tau = (\beta_{1\tau}, \beta_{2\tau}, \ldots, \beta_{d\tau})^T$, and $J(\cdot)$ is a given penalty function.

The shrinkage effect is determined by the positive penalty parameter λ, that needs to be chosen according to some criterion (typically information criterion or cross-validation).

The SCAD penalty is symmetric around the origin (i.e., $\theta = 0$). It is defined as follows:

$$p_\lambda(\theta) = \begin{cases} \lambda|\theta| & \text{if } 0 \le |\theta| \le \lambda \\ -\dfrac{\theta^2 - 2a\lambda|\theta| + \lambda^2}{2a-1} & \text{if } \lambda < |\theta| < a\lambda \\ \dfrac{(a+1)\lambda^2}{2} & \text{if } a\lambda \le \theta, \end{cases} \tag{7}$$

where $a > 2$ and $\lambda > 0$ are tuning parameters. In this paper we will set $a = 3.7$, following Zou and Yuan [30], which would help us avoid searching for optimal tuning parameters over two-dimensional grid and for this reason suppress a in the notation previous.

The SCAD estimator can then be formally defined as

$$\min_{\beta_\tau} \sum_{i=1}^{n} \rho_\tau\left(y_i - X^T\beta_\tau\right) + \sum_{j=1}^{d} p_\lambda\left(\beta_{j\tau}\right). \tag{8}$$

Straightforward implementation of regularised estimators is however computationally demanding. The main issue is that expensive repeated optimisation calls are needed to select the regularisation parameter(s) typically via some form of cross-validation. Furthermore the nonconvex nature of the SCAD optimisation problem can lead to considerable increase of the computation time at some quantiles, particularly when larger number of spatial weighting matrices are retained by the screening step, which is consistent with the results of Wu and Liu [23]. In order to select the optimal amount of regularisation we need some criterion. Given the computational costs of SCAD estimation, information criteria would be preferable. Here we will employ the g-prior Minimum Description Length (gMDL) criterion used in Kostov's [10] boosting application. This choice is however dictated mostly by computational reasons, and up to the best of our knowledge there is no evidence (such as simulation studies) to ascertain the performance of this criterion in empirical studies of nonlinear models.

The adaptive lasso estimator for the linear quantile regression can be defined as a weighted lasso problem in the following way:

$$\min_{\beta_\tau} \sum_{i=1}^{n} \rho_\tau\left(y_i - X^T\beta_\tau\right) + \lambda \sum_{j=1}^{d} \widetilde{\omega}_j \left|\beta_{j\tau}\right|, \tag{9}$$

where $|\cdot|$ denotes the $L1$ norm, while the weights are given by $\widetilde{\omega}_j = 1/|\widetilde{\beta}_{j\tau}|^\gamma$ for some $\gamma > 0$, where $\widetilde{\beta}_{j\tau}$ are initial estimates for the parameters. In this case $\widetilde{\beta}_{j\tau}$ will be obtained by an unpenalised quantile regression. The conventional lasso estimator is a particular case when all weights are equal, rather than adaptively chosen.

The adaptive lasso when implemented in a quantile regression setting retains the oracle property [30] similarly to the mean regression case. Therefore the adaptive lasso estimator is a reasonable choice in this setting, particularly bearing in mind the computational cost associated with the transformation step. Furthermore $L1$ norm estimators are by far the most widely studied regularisation estimators for quantile regression (see, e.g., [23, 24, 30] for variable selection applications).

Li and Zhu [22] proposed an algorithm to estimate the whole regularisation path for lasso type of quantile regression problem. Their proposal is potentially valuable since it can be applied to non- (or semi-) parametric additive quantile regression models and therefore results in a much more general approach, intrinsically immune to functional form misspecification. The advantage to such algorithms is that since they exploit the piecewise linear property of the regularisation path, the latter can be obtained at a fraction of the computational cost of the overall regularised estimator. This facilitates implementation of cross-validation and/or information criteria.

The elastic net [31] penalty is a combination of $L1$ and $L2$ norms, and for the quantile regression the resulting estimator can be written as

$$\min_{\beta_\tau} \sum_{i=1}^{n} \rho_\tau\left(y_i - X^T\beta_\tau\right) + \lambda_1 \sum_{j=1}^{d} \left|\beta_{j\tau}\right| + \lambda_2 \sum_{j=1}^{d} \beta_{j\tau}^2. \tag{10}$$

An important property of the elastic net penalty is that the inclusion of the $L2$ norm induces a grouping effect in that correlated variables are grouped together. This would help avoid spuriously selecting only one variable from a group of highly correlated variables. Given that in many empirical problems the spatial weighting matrices considered can lead to highly correlated designs, it would be desirable to avoid such a pitfall. One should note however that elastic net penalisation could be expected to retain more variables compared to the other approaches.

The least squares approximation (LSA) estimator [32] is given by:

$$\min_{\beta} \left\{ \left(\beta - \widetilde{\beta}\right)^T \widetilde{\Sigma}^{-1} \left(\beta - \widetilde{\beta}\right) + \lambda \sum_{j=1}^{d} \widetilde{\omega}_j \left|\beta_j\right| \right\}, \tag{11}$$

where $\widetilde{\Sigma}^{-1} = n^{-1}(\partial^2 \ell(\widetilde{\beta})/\partial\widetilde{\beta})$ is the second derivative at the unpenalised loss function, evaluated at the unregularised estimates $\widetilde{\beta}$. It is technically obtained as an approximation based on first order Taylor series expansion (see [32]).

In the case of quantile regression, the respective loss function (i.e., the check function $\rho_\tau(\cdot)$) is not sufficiently smooth. Nevertheless, as long as $\widetilde{\Sigma}$, which is in principle any consistent covariance matrix estimate pertaining to the unpenalised problem, can be obtained, the corresponding LSA estimator, defined in (11) exists. Furthermore when regularisation parameters are chosen optimally it possesses the oracle property (see [32] for a formal proof). Since (11) is essentially a linear lasso type of problem, it can be estimated using standard methods. In particular the computationally efficient

least angle regression algorithm (LARS) of Efron et al. [33] can be used to compute the regularisation path. Here we will apply the BIC-type tuning parameter selector of Wang et al. [34] to select the optimal amount of shrinkage. Application of the LSA to a quantile regression requires a covariance matrix estimator for the latter. Any consistent estimator would be appropriate. In this paper we will use the kernel-based covariance estimator proposed in Newey and Powell [35].

6. Study Design and Implementation Details

For comparative purposes we follow closely the design outlined in Kostov [10]. This involves using the same dataset, model specification as well as a set of competing alternative spatial weighting matrices. Since all these are discussed in some detail in Kostov [10] we will only briefly sketch them here.

The corrected version of the popular Boston housing dataset [36] is used. It consists of 506 observations and incorporates some corrections and additional latitude and longitude information, due to Gilley and Pace [37]. This dataset contains one observation for each census tract in the Boston Standard Metropolitan Statistical Area. The variables comprise of proxies for pollution, crime, distance to employment centres, geographical features, accessibility, housing size, age, race, status, tax burden, educational quality, zoning, and industrial externalities. A detailed description of the variables, to be used in this study, is presented in Table 1.

The basic model as implemented in Kostov [10] is as follows:

$$\log(\text{MEDV}) = f\left\{\text{CRIM}, \text{ZN}, \text{INDUS}, \text{CHAS}, \text{NOX}^2, \text{RM}^2, \text{AGE}, \log(\text{DIS}), \log(\text{RAD}), \text{TAX}, \text{PTRATIO}, \text{B}, \log(\text{LSTAT})\right\}. \quad (12)$$

The basic specification mentioned previously is augmented with alternative candidate spatial weighting matrices, constructed using the longitude and latitude information. The set of alternative spatial weighting matrices is constructed using inverse distance raised on a power weights specification and nearest neighbours definition of the neighbourhood scheme. We will adopt the naming conventions used in Kostov [10] combining the codes for the neighbourhood definition and the weighting scheme to refer to the corresponding spatial weighting matrix and the resulting additional variables to be included in the boosting model. All these variables are named using the following convention: nxwy, where x is the number of neighbours and y is the weighting parameter (which is the inverse power of the weight decay). For example, the spatial weighting matrix with the nearest 50 observations as neighbours and inverse squared distance weights as well as the resulting transformed variable will be denoted as n50w2. We employ all values for number of neighbours from 1 to 50 and evaluate w in the interval [0.4, 4] using increments of 0.1. In simple words this means that we are combining 50 possible neighbourhood definitions with 37 alternatives for the weighting parameter resulting in 1,850 alternative spatial weighting matrices to be considered simultaneously.

Table 1: Description of variables.

Variable	Description
MEDV	Median values of owner-occupied housing in thousands of USD
LON	Tract point longitude in decimal degrees
LAT	Tract point latitude in decimal degrees
CRIM	Per capita crime
ZN	Proportion of residential land zoned for lots over 25,000 sq. ft per town
INDUS	Proportion of nonretail business acres per town
CHAS	An indicator: 1 if tract borders Charles River; 0 otherwise
NOX	Nitric oxides concentration (parts per 10 million) per town
RM	Average number of rooms per dwelling
AGE	Proportions of owner-occupied units built prior to 1940
DIS	Weighted distance to five Boston employment centres
RAD	Index of accessibility to radial highways per town
TAX	Property-tax rate per USD 10,000 per town
PTRATIO	Pupil-teacher ratio per town
B	Calculated as $1000*(\text{Bk}-0.63)^2$ where Bk is the proportion of blacks
LSTAT	Percentage of lower status population

Kostov [10] projects the spatially weighted dependent variable into the column vector space of the spatially weighted independent variables, by taking the fitted values from a least-squares regression to obtain the transformed variables, named according to the previous convention. As discussed before here we need to replace this first step with a quantile regression defined over a pre-determined quantile to obtain a model augmented with the alternative spatial weighting matrices. The second stage is then implemented in two consecutive steps. First we apply a component-wise boosting quantile regression, defined over the same quantile (as in the first stage) to the augmented model. This is the screening step that reduces the set of variables to be considered in the model. Then a regularized quantile regression (defined over the same quantile) is applied to the screened dataset. The previous three steps (transformation, screening and estimation) can be run over any prespecified quantile, and their consecutive implementation defines our estimator.

In the present setting some caution should be exercised in applying the estimation step. Note that in conditionally parametric models, there is a certain trade-off between variables and spatial dependence. The spatial dependence structure could approximate the effect of missing variables, provided these are spatially correlated. Therefore simultaneously shrinking the coefficients of both variables and spatial lags will be a manifestation of this trade-off. Whenever the model contains such related terms in both the spatial part (i.e., spatial weighting matrices) and in the regression part

(variables the effect of which could be approximated by these spatial weighting matrices) simultaneous shrinkage is undesirable. The danger here is that one can spuriously exclude important variables and approximate their effect by additional spatial terms. Note however that if we assume that the regression part is given, this trade-off will disappear. Ideally one would want to eliminate this trade-off. In order to avoid the impact of the approximation on this trade-off we suggest a two-step implementation of the estimation step. In the initial step only the spatial lag coefficients are penalised, while in the following final step all coefficients are penalised. In this way the initial step should select the appropriate spatial dependence structure, while the final step would perform final variable selection. Hence the initial step makes structural inference about the spatial part conditional on the regression part of the model. If the screening step has produced a model that is reasonably close to the true one, then the proposed approach should be able to discover the true underlying structure. Alternatively one may wish to implement an iterative estimation in which the estimator alternates between steps in which only the spatial structure is penalised and steps with only the regression part are penalised until convergence (defined in terms of obtaining a stable structure in that no more terms are eliminated). Such steps can be viewed as conditioning one part (spatial or regression) of the model on the other hence avoiding the trade-off. The latter approach would however be computationally more expensive.

Another issue is the highly correlated design of the spatial quantile regression model, when there are large number of potential spatial weighting matrices. Since in principle the variable selection methods rely upon marginal correlations, they could fail to perform in such highly correlated designs. For the mean regression model recent contributions by Wang [38] and Cho and Fryzlewicz [39] have suggested alternative methods that overcome such a reliance on marginal correlations and hence are applicable to highly correlated designs. It is however unclear how such methods can be extended to the quantile regression case. The two-step approach adopted in this paper conditions selection for the spatial and hedonic variables on the other part of the model and hence reduces this trade-off. Such an approach is justified if the regression part of the model is correctly specified, but could be suboptimal if this is not the case. This is of course an area that deserves further investigation.

7. Results

We implement the proposed estimator for the 0.1 to 0.9 quantiles with a step of 0.1 (i.e., 9 different quantile regressions). Table 2 presents comparative computational time details for the different procedures. All these are calculated from the first of the considered quantiles (i.e., the 0.1th one) and are given as a guidance only since the actual computational time could vary according to the nature of the optimisation problem that can change over different quantiles. All computations are undertaken using the statistical programming language R [40] on Intel Core2 2.13 GHz processor with 2 Gb RAM, not using any parallel computation. Parallelising some of the more computationally demanding tasks and/or using compiled code could considerably reduce the computational time. Furthermore it cannot be claimed that the actual implementation of these procedure is optimised in terms of computational time.

Table 2: Typical computational details for different procedures.

Procedure	Time (seconds)
Instrumentation step (1850 Ws)	1850.33
5000 boosting iterations on all data	5.41
gMDL criterion calculation	1108.42
5000 boosting iterations on reduced data	2.71
stability selection	106.35
LSA step 1	0.55
LSA step 2	0.05
Full path step 1 (with 5-fold CV)	9.08
Full path step 2 (with 5-fold CV)	79.11
SCAD step 1 (with gMDL, 50 penalty values)	2.18
SCAD step 2 (with gMDL, 50 penalty values)	2.88
Elastic net step 1	1.11
Elastic net step 2	7.00
Elastic net single step	7.82

The instrumentation step is the most time-consuming task. In our implementation it takes over 30 minutes for 1850 spatial weighting matrices. In many empirical problems one would probably consider much smaller number of alternative spatially weighting matrices. Furthermore most of the time in this step is spent on creating the spatially weighted dependent and independent variables, rather than fitting the actual quantile regressions.

The actual boosting procedure requires running the boosting algorithm for a large number of iterations and then calculating a stopping criterion to decide upon the estimated structure. The boosting algorithm is very efficient computationally. The stopping criterion calculation however takes considerable time. Efficient parallel implementations for the latter exist, and these can considerably reduce the computation time.

The time needed to calculate the stopping criterion is directly proportional to the number of boosting steps (which is effectively the number of alternative "models" for which it is calculated). Since in this case at all considered quantiles we need at least three times less iterations than the 5,000 used here, practical implementation would have taken 6-7 minutes rather than 18 as reported in Table 2.

We apply the stability selection to the already reduced (in the instrumentation step) dataset. Yet again this is relatively time-consuming procedure, but it can be parallelised for further computational gains.

One has to be careful in directly comparing these implementations of the estimation step, as the instrumentation step mentioned previously demonstrates; calculating the stopping criterion (i.e., the optimal penalty parameters) is by far the most computationally demanding part of these procedures and the reported implementations use different methods for this.

Table 3: Stability selection-derived inclusion probabilities for spatial weighting matrices.

	Quantiles								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
n1w0.4	0.56								
n2w0.5									0.80
n2w4							0.28		
n3w0.4			0.56	0.52			0.40		
n3w1.2			0.72						
n3w1.7			0.24						
n3w1.8		0.80							
n4w3						0.28			
n4w0.4			0.16						
n4w0.7		0.56							
n4w3.4						0.24			
n5w0.4			0.24	0.28	0.52				
n5w0.5		0.68							
n5w0.6	0.96								
n5w0.7							0.56		
n5w1.3				0.40					
n6w0.4		0.32	0.48	0.44	1.00	0.80	0.32		
n6w0.5		0.20		0.84			0.76		
n6w0.6						0.44			
n6w1.1								0.80	
n6w1.9							0.20		
n7w0.4		0.28	0.28						0.76
n7w1.2								0.72	
n7w1.3									0.56
n8w0.4						0.28			
n8w0.5			0.36			0.60			
n8w0.6			0.16						
n8w0.7							0.32		
n8w0.9							0.32		
n8w2.6									0.32
n9w3.9								0.28	
n10w2.5									0.20
n12w3.9								0.28	
n12w0.4	0.56	0.20							
n16w0.4				0.36					
n19w0.4				0.24					

With regard to the estimation methods we report separately the computation times for step one (where only the spatial weighting matrix coefficients are penalised) and the consecutive second step where all the coefficients are penalised. As it is to be expected the LSA is the fastest method. This is due to two underlying facts. The first is that it uses the efficient least angle regression algorithm [33] while the other refers to use of the BIC-type tuning parameter selector of Wang et al. [34] which is easy to compute.

The full path estimation for adaptive lasso, accompanied by cross-validation to choose the optimal amount or regularization, appears to be the most computationally demanding estimation method. Most of the computational costs however come from the use of cross-validation. Furthermore this is the most universally applicable method in the sense that many of the other methods can run into difficulties during the optimisation (at different quantiles) which can considerably inflate their computational costs.

We present computational details for implementing SCAD with gMDL over a predefined grid of 50 penalty values. Although the computational times appear acceptable, one has to take into account some caveats. The nonconvex nature of the SCAD optimisation problem means that in some cases the actual computation time can increase considerably (with a factor of over 100 in some cases). Furthermore we have opted to fix one of the regularisation parameters which artificially reduces the computational time. Another important point to make is that no set of penalisation parameters is ex ante guaranteed to span the whole regularisation path. In our implementation we run a preliminary SCAD estimation over a range of such values designed to identify a feasible set that does span most of the regularisation path and then manually select the grid of such values. In cases where the optimisation is difficult, this can lead to considerable increase of computational time. Therefore a path estimation algorithm for the SCAD estimator for quantile regression is essential if a reliable implementation of this method is to be designed. The use of the gMDL as an optimality criterion is also somehow ad hoc in that there is no firm evidence on its performance for this type of problems, and it is mostly dictated by computational reasons (since cross-validation, for example, would be very costly).

The elastic net implementation is reasonably efficient. Both the BIC and the generalised approximate cross-validation yield the same models. The reported computational costs refer to the routines that compute internally both of the above criteria, but this only marginally increases the computational costs. Most of the computational load comes from the double regularisation needed to solve for the two underlying penalties.

The component-wise boosting algorithm manages to achieve considerable reduction in the model space. It retains between three and eleven spatial weighting matrices across the different quantiles. We will not present these intermediate results here for brevity reasons, but details are available upon request. This intermediate step yields a reduced model space that can be explored for the underlying structure as discussed in the methodology section. Table 3 presents the results from the stability selection applied to the prescreened model (i.e., after the boosting application). Typically stability selection applies a prespecified probability threshold to select variables. Here instead of proper stability selection we present the corresponding inclusion probabilities for the spatial weighting matrices. We omit spatial weighting matrices with inclusion probability less than 10%. Full results are available upon request. Table 3 provides a background against which the actual estimation results can be evaluated. If one was to use a threshold of 0.6, most quantiles would have resulted in a single spatial weighting matrix being selected. Such a choice would however have been base solely on the component-wise boosting algorithm, which as already discussed may

Table 4: Estimation results at the 0.1th quantile.

	LSA		Full path		SCAD	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
Intercept	1.591	0.000	1.493	0.000	0.975	0.025
CRIM	−0.010	0.000	−0.011	0.000	−0.010	0.000
ZN			0.000	0.568	0.000	0.795
INDUS			−0.001	0.803	0.004	0.193
CHAS					0.033	0.342
NOX^2			−0.006	0.973		
RM^2	0.011	0.000	0.012	0.000	0.015	0.000
AGE	−0.001	0.032	−0.002	0.021	−0.002	0.013
log (DIS)	−0.091	0.001	−0.118	0.008	−0.039	0.405
log (RAD)	0.036	0.073	0.041	0.058	0.042	0.063
TAX	0.000	0.096	0.000	0.071	0.000	0.035
PTRATIO	−0.007	0.245	−0.006	0.389	−0.006	0.297
B	0.001	0.000	0.001	0.001	0.001	0.000
log (LSTAT)	−0.153	0.003	−0.123	0.028	−0.057	0.292
n1w0.4					0.066	0.458
n5w0.6	0.478	0.000	0.385	0.014	0.442	0.000
n12w0.4			0.094	0.481		

Table 5: Estimation results at the 0.2th quantile.

	LSA		Full path		SCAD	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
Intercept	2.013	0.000	1.663	0.000	1.390	0.000
CRIM	−0.013	0.000	−0.012	0.000	−0.012	0.000
ZN			0.000	0.683	−0.001	0.288
INDUS			−0.001	0.777	0.003	0.265
CHAS					0.041	0.196
NOX^2	−0.180	0.153				
RM^2	0.013	0.000	0.014	0.000	0.016	0.000
AGE	−0.001	0.019	−0.002	0.009	−0.001	0.015
log (DIS)	−0.144	0.000	−0.133	0.000	−0.025	0.544
log (RAD)	0.052	0.003	0.049	0.006	0.032	0.106
TAX	0.000	0.010	0.000	0.009	0.000	0.041
PTRATIO	−0.009	0.056	−0.004	0.423	−0.009	0.096
B	0.000	0.001	0.000	0.000	0.000	0.003
log (LSTAT)	−0.151	0.000	−0.138	0.001	−0.094	0.020
n3w1.8					0.439	0.000
n5w0.5			0.133	0.571		
n6w0.4	0.397	0.000	0.309	0.164		

not possess the oracle property and therefore may be inappropriate for structure discovery purposes. The estimation results from the least squares approximation (LSA), full-path adaptive lasso (full path), and smoothly clipped absolute deviations (SCAD) by quantile are presented in Tables 4–12. The elastic net results are not presented. In simple terms due to the implicit correlation penalty, the elastic net possesses a grouping property in that it groups together correlated variables. In this case due to the highly correlated nature of the spatial weighting matrices the elastic net groups them together and cannot exclude spatial weighting matrices whenever they are similar to the ones already included in the model.

Bearing in mind the relative computational cost for the different estimation approaches, it is useful to determine whether the LSA results differ substantially from the other approaches. Table 4 presents the results at the 0.1th quantile. With regard to the main (i.e., hedonic) variables the results are comparable across estimation methods. Where a variable is omitted by one of these methods, it is either dropped by the other methods or found to be statistically insignificant. With regard to the spatial weighting matrices the stability selection

Table 6: Estimation results at the 0.3th quantile.

	LSA		Full path		SCAD	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
Intercept	1.904	0.000	1.973	0.000	1.553	0.000
CRIM	−0.014	0.000	−0.014	0.000	−0.011	0.207
ZN			0.000	0.452	0.000	0.591
INDUS			0.003	0.158	0.004	0.041
CHAS			0.024	0.441		
NOX^2	−0.106	0.344	−0.187	0.132		
RM^2	0.013	0.000	0.014	0.000	0.015	0.000
AGE	−0.001	0.033	−0.001	0.052	−0.001	0.055
log (DIS)	−0.119	0.000	−0.110	0.002	−0.067	0.073
log (RAD)	0.053	0.003	0.061	0.002	0.045	0.037
TAX	0.000	0.034	0.000	0.014	0.000	0.013
PTRATIO	−0.009	0.039	−0.012	0.010	−0.011	0.011
B	0.000	0.000	0.000	0.000	0.000	0.000
Log (LSTAT)	−0.159	0.000	−0.163	0.000	−0.131	0.000
n3w0.4	0.221	0.048	0.141	0.202	0.192	0.384
n6w0.4	0.186	0.101	0.245	0.032		
n8w0.5					0.259	0.078

Table 7: Estimation results at the 0.4th quantile.

	LSA		Full path		SCAD	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
Intercept	2.020	0.000	2.151	0.000	1.697	0.000
CRIM	−0.007	0.000	−0.010	0.242	−0.007	0.000
ZN			0.000	0.518	0.000	0.827
INDUS			0.003	0.088	0.003	0.074
CHAS			0.014	0.653	0.021	0.478
NOX^2			−0.231	0.059		
RM^2	0.012	0.000	0.014	0.000	0.015	0.000
AGE			−0.001	0.186	−0.001	0.074
log (DIS)	−0.079	0.000	−0.112	0.002	−0.070	0.044
log (RAD)	0.036	0.031	0.053	0.011	0.044	0.015
TAX	0.000	0.004	0.000	0.012	0.000	0.001
PTRATIO	−0.010	0.013	−0.013	0.003	−0.010	0.010
B	0.000	0.001	0.000	0.000	0.000	0.000
log (LSTAT)	−0.206	0.000	−0.169	0.000	−0.150	0.000
n3w0.4	0.055	0.643	0.184	0.246	0.160	0.145
n5w1.3			0.157	0.362		
n6w0.4					0.263	0.016
n6w0.5	0.340	0.004				

approach (see Table 3) determined that three such matrices have inclusion probability of over 0.5, but one such matrix, namely, n5w0.6 has a very high inclusion probability of 0.96. The estimation results reflect this in that n5w0.6 is selected by all methods, while the full path and the SCAD estimators also include another spatial weighting matrix. It hence looks like the LSA chooses a slightly more parsimonious model. One should note however that the additional spatial weighting matrices chosen by the other two methods are statistically insignificant. Therefore all methods point to the same underlying model structure.

Table 5 presents the results for the 0.2th quantile. Again the LSA closely approximates the full path solution. The additional spatial weighting matrix selected by the full path estimation is insignificant, and both methods lead to the same conclusions. The SCAD estimation however produces slightly

Table 8: Estimation results at the 0.5th quantile.

	LSA		Full path		SCAD	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
Intercept	1.991	0.000	1.948	0.000	1.799	0.000
CRIM	−0.007	0.000	−0.007	0.000	−0.007	0.000
ZN			0.000	0.290	0.000	0.518
INDUS			0.002	0.216	0.003	0.178
CHAS					0.012	0.712
NOX^2						
RM^2	0.013	0.000	0.013	0.000	0.014	0.000
AGE	−0.001	0.102	−0.001	0.206	−0.001	0.096
log (DIS)	−0.111	0.000	−0.102	0.001	−0.082	0.013
log (RAD)	0.054	0.002	0.057	0.001	0.050	0.007
TAX	0.000	0.001	0.000	0.000	0.000	0.002
PTRATIO	−0.009	0.021	−0.009	0.027	−0.010	0.020
B	0.000	0.000	0.000	0.000	0.000	0.000
log (LSTAT)	−0.189	0.000	−0.187	0.000	−0.164	0.000
n5w0.4			0.256	0.296		
n6w0.4	0.406	0.000	0.146	0.547	0.416	0.000

Table 9: Estimation results at the 0.6th quantile.

	LSA		Full path		SCAD	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
Intercept	2.235	0.000	2.091	0.000	1.907	0.000
CRIM	−0.007	0.002	−0.006	0.013	−0.006	0.087
ZN			0.000	0.379	0.000	0.980
INDUS			0.001	0.418	0.002	0.343
CHAS					0.001	0.965
NOX^2						
RM^2	0.013	0.000	0.013	0.000	0.015	0.000
AGE	−0.001	0.047	−0.001	0.047	−0.001	0.066
log (DIS)	−0.120	0.000	−0.110	0.000	−0.072	0.035
log (RAD)	0.035	0.043	0.048	0.007	0.041	0.023
TAX	0.000	0.004	0.000	0.001	0.000	0.003
PTRATIO	−0.012	0.003	−0.009	0.029	−0.013	0.003
B	0.000	0.000	0.000	0.000	0.001	0.000
log (LSTAT)	−0.195	0.000	−0.190	0.000	−0.152	0.000
n4w3.4	−0.143	0.203				
n6w0.4					0.390	0.000
n6w0.6	0.510	0.000	0.388	0.000		

different results. Most notably it chooses a different spatial weighting matrix. In terms of implementation the SCAD optimisation problem at this quantile did take longer to solve. Nevertheless, the selected by SCAD spatial weighting matrix is the one characterised by highest inclusion probability, according to the stability selection. Due to the somewhat ad hoc implementation of the SCAD estimation in this paper (because it is evaluated over a grid, instead of computing the whole regularisation path and the use of the gMDL as stopping criterion) it is however difficult to evaluate the latter result. As far as the LSA is concerned however, it provides a reliable approximation of the adaptive lasso estimator at a fraction of its computational cost.

The results for the 0.3th quantile (Table 6) are broadly similar across different methods. LSA and the adaptive lasso select the same spatial weighting matrices and although these show some small differences in terms of statistical significance for the latter, the structural inference is essentially the same. Once again SCAD results yield a small difference with regard to the preferred spatial weighting matrices.

Table 10: Estimation results at the 0.7th quantile.

	LSA		Full path		SCAD	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
Intercept	2.134	0.000	2.408	0.000	2.205	0.000
CRIM	−0.004	0.016	−0.007	0.018	−0.007	0.003
ZN			0.000	0.967	0.000	0.790
INDUS			0.000	0.902		
CHAS						
NOX^2						
RM^2	0.011	0.000	0.013	0.000	0.014	0.000
AGE			−0.001	0.022	−0.001	0.041
log (DIS)	−0.109	0.000	−0.126	0.000	−0.096	0.003
log (RAD)			0.053	0.002	0.046	0.008
TAX	0.000	0.025	0.000	0.003	0.000	0.012
PTRATIO	−0.010	0.027	−0.014	0.001	−0.012	0.004
B	0.001	0.000	0.001	0.001	0.001	0.000
log (LSTAT)	−0.240	0.000	−0.185	0.000	−0.170	0.000
n3w0.4			0.308	0.000		
n5w0.7					0.320	0.003
n8w0.9	0.413	0.000				

Table 11: Estimation results at the 0.8th quantile.

	LSA		Full path		SCAD	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
Intercept	3.456	0.000	3.404	0.000	1.573	0.022
CRIM	−0.010	0.000	−0.010	0.000	−0.009	0.000
ZN			0.000	0.926	0.000	0.492
INDUS			−0.002	0.284	−0.003	0.208
CHAS	0.084	0.073	0.083	0.075	0.098	0.156
NOX^2	−0.506	0.000	−0.475	0.000	0.270	0.353
RM^2	0.010	0.000	0.010	0.000	0.012	0.000
AGE	−0.001	0.187	−0.001	0.167	−0.001	0.191
log (DIS)	−0.217	0.000	−0.232	0.000	−0.081	0.119
log (RAD)	0.042	0.007	0.044	0.006	0.008	0.713
TAX						
PTRATIO	−0.026	0.000	−0.024	0.000	−0.002	0.829
B						
log (LSTAT)	−0.257	0.000	−0.253	0.000	−0.189	0.000
n6w1.1	0.236	0.000	0.252	0.000	0.573	0.027
n9w3.9					−0.051	0.817

At the 0.4th quantile (Table 7) all methods choose n3w0.4 together with a slightly different second spatial weighting matrix. With regard to the second retained spatial weighting matrix LSA is very similar to the SCAD (n6w0.5 versus n6w0.4) while the adaptive lasso selection is slightly different. Hence although LSA results are comparable to the other methods, the approximation to the adaptive lasso is not as close as at the previously considered quantiles. One can note that in this case out of the 8 spatial weighting matrices originally retained by the boosting algorithm, 7 have a model inclusion probability exceeding 0.1 (see Table 3). Therefore one can tentatively conclude that LSA provides a reasonable approximation to the adaptive lasso for quantile regression whenever there are relatively few competing spatial weighting matrices. This property is to be expected given the trade-off between the spatial and the regression parts in a quantile regression model, that was discussed in the methodology section. It is possible that if one applies a different method to control for this trade-off, instead of the two-step implementation of the oracle estimator, implemented here, the quality of the LSA approximation may not deteriorate with increasing dimension of the spatial weighting matrices space.

The estimation results for the 0.5th quantile (see Table 8) are similar across the different methods. The adaptive lasso

Table 12: Estimation results at the 0.9th quantile.

	LSA		Full path		SCAD	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
Intercept	3.804	0.000	2.374	0.000	1.608	0.017
CRIM	−0.010	0.000	−0.002	0.789	−0.010	0.000
ZN			−0.001	0.048	0.000	0.975
INDUS			0.001	0.716	−0.002	0.300
CHAS					0.068	0.286
NOX^2	−0.818	0.000				
RM^2	0.007	0.000	0.008	0.001	0.012	0.000
AGE						
log (DIS)	−0.248	0.000			−0.069	0.164
log (RAD)	0.068	0.000			0.050	0.014
TAX						
PTRATIO	−0.029	0.000	−0.012	0.158	0.000	0.995
B	−0.279	0.000				
log (LSTAT)			−0.263	0.000	−0.168	0.000
n2w0.5			−0.098	0.549	0.152	0.140
n7w0.4			0.026	0.920	0.329	0.024
n8w2.6			0.517	0.069		
n10w2.5	0.260	0.000				

finds it difficult to discriminate between n5w0.4 and n6w0.4, but since these two are very difficult to distinguish between one can conclude that the methods agree. Interestingly although SCAD produces the same spatial weighting matrix as LSA, the technical difficulty in discriminating between two very similar spatial weighting matrices results in considerable increase in computational time.

Similarly at the 0.6th quantile (Table 9) results are consistent across methods. The LSA and adaptive lasso both select n6w0.6 with LSA also selecting another spatial weighting matrix, which is however statistically insignificant. The SCAD selects n6w0.4, which by the way is the highest inclusion probability spatial matrix, according to the stability selection results (see Table 3). This again raises the issue of the comparative performance of SCAD and adaptive lasso, but the results are nevertheless qualitatively similar.

Table 10 shows the estimation result for the 0.7th quantile. This is where there are considerable differences between the different methods. There is disagreement about the sign of TAX between SCAD and the other two methods. Furthermore all three methods choose different spatial weighting matrices. Once again the source for such difference is probably the size of the (screened) spatial weighting matrices space, which consists of 11 such matrices, 8 of which have stability selection-derived inclusion probability of at least 0.2.

At the 0.8th quantile all methods select n6w1.1 (see Table 11). SCAD selects another spatial weighting matrix, which leads to some differences in the regression part. However since this additional spatial weighting matrix is not statistically significant, omitting it would produce results consistent with the other methods.

Finally Table 12 presents the results for the 0.9th quantile which differ considerably amongst methods. The main reason for these differences is that different main hedonic variables are selected by different methods which results in differences for the spatial part. The latter raises the issue of the trade-off between the regression part and the spatial dependence.

8. Conclusions

This paper proposes methods for selecting spatial weighting matrix in a quantile regression context. We build upon previous work in this area. In discussing our proposal we outline the different alternatives and the potential different implementations of the same set of ideas. Our proposals are designed to reduce the arbitrariness of choice of spatial weighting matrix and the impact of the trade-off between functional form and spatial dependence specifications. The spatial quantile regression is already a quite flexible model allowing for different impacts, across different quantiles. Our procedure introduces additional flexibility in not only different degrees of spatial dependence but also different spatial weighting matrices for different quantiles, resulting in potentially interesting inferences about the nature of the underlying process.

The methodology proposed in this paper consists of three steps: instrumentation, screening, and estimation. Several different alternative methods for the estimation step are explored. This allows for conclusions to be drawn with regard to the performance of these estimators in different circumstances. In general terms, the proposed methods are tractable with small and moderate size samples, where the

advantages of the spatial quantile regression model are most pronounced.

References

[1] P. Kostov, "A spatial quantile regression hedonic model of agricultural land prices," *Spatial Economic Analysis*, vol. 4, no. 1, pp. 53–72, 2009.

[2] F. Stetzer, "Specifying weights in spatial forecasting models: the results of some experiments," *Environment & Planning A*, vol. 14, no. 5, pp. 571–584, 1982.

[3] L. Anselin, "Under the hood issues in the specification and interpretation of spatial regression models," *Agricultural Economics*, vol. 27, no. 3, pp. 247–267, 2002.

[4] B. Fingleton, "Externalities, economic geography, and spatial econometrics: conceptual and modeling developments," *International Regional Science Review*, vol. 26, no. 2, pp. 197–207, 2003.

[5] G. Holloway and M. L. A. Lapar, "How big is your neighbourhood? Spatial implications of market participation among filipino smallholders," *Journal of Agricultural Economics*, vol. 58, no. 1, pp. 37–60, 2007.

[6] J. P. LeSage and O. Parent, "Bayesian model averaging for spatial econometric models," *Geographical Analysis*, vol. 39, no. 3, pp. 241–267, 2007.

[7] J. P. LeSage and M. M. Fischer, "Spatial growth regressions: model specification, estimation and interpretation," *Spatial Economic Analysis*, vol. 3, no. 3, pp. 275–304, 2008.

[8] J. Crespo-Cuaresma and M. Feldkircher, "Spatial filtering, model uncertainty and the speed of income convergence in Europe," in *Annual Meeting of the Austrian Economic Association*, 2010, http://www.univie.ac.at/economics/noeg2010/papers/Crespo_Cuaresma_NOEG2010.pdf.

[9] T. S. Eicher, A. Lenkoski, and A. E. Raftery, "Bayesian model averaging and endogeneity under model uncertainty: an application to development determinants," UW Working Paper UWEC-2009-19, University of Washington, Department of Economics, 2009.

[10] P. Kostov, "Model boosting for spatial weighting matrix selection in spatial lag models," *Environment and Planning B*, vol. 37, no. 3, pp. 533–549, 2010.

[11] H. H. Kelejian and I. R. Prucha, "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances," *Journal of Real Estate Finance and Economics*, vol. 17, no. 1, pp. 99–121, 1998.

[12] J. Zietz, E. N. Zietz, and G. S. Sirmans, "Determinants of house prices: a quantile regression approach," *Journal of Real Estate Finance and Economics*, vol. 37, no. 4, pp. 317–333, 2008.

[13] T. H. Kim and C. Muller, "Two-stage quantile regression when the first stage is based on quantile regression," *Econometrics Journal*, vol. 7, pp. 218–231, 2004.

[14] V. Chernozhukov and C. Hansen, "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, vol. 132, no. 2, pp. 491–525, 2006.

[15] V. Chernozhukov and C. Hansen, "Instrumental variable quantile regression: a robust inference approach," *Journal of Econometrics*, vol. 142, no. 1, pp. 379–398, 2008.

[16] V. Chernozhukov and H. Hong, "An MCMC approach to classical estimation," *Journal of Econometrics*, vol. 115, no. 2, pp. 293–346, 2003.

[17] O. Arias, K. F. Hallock, and W. Sosa-Escudero, "Individual heterogeneity in the returns to schooling: instrumental variables quantile regression using twins data," *Empirical Economics*, vol. 26, no. 1, pp. 7–40, 2001.

[18] J. García, P. J. Hernández, and A. López-Nicolás, "How wide is the gap? An investigation of gender wage differences using quantile regression," *Empirical Economics*, vol. 26, no. 1, pp. 149–167, 2001.

[19] N. Fenske, T. Kneib, and T. Hothorn, "Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression," Technical Report No. 52, Department of Statistics, LMU München, 2009.

[20] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society B*, vol. 58, pp. 267–288, 1996.

[21] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.

[22] Y. Li and J. Zhu, "L1-norm quantile regression," *Journal of Computational and Graphical Statistics*, vol. 17, no. 1, pp. 163–185, 2008.

[23] Y. Wu and Y. Liu, "Variable selection in quantile regression," *Statistica Sinica*, vol. 19, no. 2, pp. 801–817, 2009.

[24] A. Belloni and V. Chernozhukov, "L1-penalized quantile regression in high-dimensional sparse models," WP10/09, Centre For Microdata Methods and Practice, Institute for Fiscal Studies, 2009.

[25] E. Greenshtein and Y. Ritov, "Persistence in high-dimensional linear predictor selection and the virtue of overparametrization," *Bernoulli*, vol. 10, no. 6, pp. 971–988, 2004.

[26] J. Fan and J. Lv, "Sure independence screening for ultrahigh dimensional feature space," *Journal of the Royal Statistical Society B*, vol. 70, no. 5, pp. 849–911, 2008.

[27] L. Wasserman and K. Roeder, "High-dimensional variable selection," *Annals of Statistics A*, vol. 37, no. 5, pp. 2178–2201, 2009.

[28] N. Meinshausen and P. Bühlmann, "Stability selection," *Journal of the Royal Statistical Society B*, vol. 72, no. 4, pp. 417–473, 2010.

[29] R. Koenker and G. Bassett, "Regression quantiles," *Econometrica*, vol. 46, pp. 33–50, 1978.

[30] H. Zou and M. Yuan, "Regularized simultaneous model selection in multiple quantiles regression," *Computational Statistics and Data Analysis*, vol. 52, no. 12, pp. 5296–5304, 2008.

[31] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society B*, vol. 67, no. 2, pp. 301–320, 2005.

[32] H. Wang and C. Leng, "Unified LASSO estimation by least squares approximation," *Journal of the American Statistical Association*, vol. 102, no. 479, pp. 1039–1048, 2007.

[33] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[34] H. Wang, R. Li, and C. L. Tsai, "Regression coefficient and autoregressive order shrinkage and selection via the lasso," *Journal of the Royal Statistical Society B*, vol. 69, no. 1, pp. 63–78, 2007.

[35] W. K. Newey and J. L. Powell, "Efficient estimation of linear and type I censored regression models under conditional quantile restrictions," *Econometric Theory*, vol. 6, pp. 295–317, 1990.

[36] D. Harrison and D. L. Rubinfeld, "Hedonic Housing Prices and the Demand for Clean Air," *Journal of Environmental Economics and Management*, vol. 5, pp. 81–102, 1978.

[37] O.W. Gilley and R. K. Pace, "On the Harrison and Rubinfeld Data," *Journal of Environmental Economics and Management*, vol. 3, pp. 403–405, 1996.

[38] H. Wang, "Factor profiling for ultra high dimensional variable selection," Working Paper, 2011, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1613452.

[39] H. Cho and P. Fryzlewicz, "High-dimensional variable selection via tilting," 2010, High-dimensional variable selection, http://stats.lse.ac.uk/fryzlewicz/tilt/tilt.pdf.

[40] R: Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2009, http://www.R-project.org.