

López, Fernando A.; Kholodilin, Konstantin A.

Working Paper

The MARS algorithm in the spatial framework: Non-linearities and spatial effects in hedonic models

DIW Discussion Papers, No. 1842

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: López, Fernando A.; Kholodilin, Konstantin A. (2020) : The MARS algorithm in the spatial framework: Non-linearities and spatial effects in hedonic models, DIW Discussion Papers, No. 1842, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/213313>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Discussion Papers

Deutsches Institut für Wirtschaftsforschung

2020

The MARS Algorithm in the Spatial Framework: Non-Linearities and Spatial Effects in Hedonic Models

Fernando A. López and Konstantin A. Kholodilin

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

IMPRESSUM

© DIW Berlin, 2020

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
<http://www.diw.de>

ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
<http://www.diw.de/discussionpapers>

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
<http://ideas.repec.org/s/diw/diwwpp.html>
<http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html>

The MARS algorithm in the spatial framework: Non-linearities and spatial effects in hedonic models[☆]

Fernando A. López^a, Konstantin A. Kholodilin^{b,c}

^a*Universidad Politécnica de Cartagena, Spain*

^b*DIW Berlin, Mohrenstraße 58, 10117, Berlin, Germany*

^c*NRU HSE, Kantemirovskaya ul., 3, 194100, St. Petersburg, Russia*

Abstract

Multivariate Adaptive Regression Spline (MARS) is a simple and powerful non-parametric technique that automatizes the selection of non-linear terms in regression models. Non-linearities and spatial effects are natural characteristics in numerous spatial hedonic pricing models. In this paper, we propose using the MARS data-driven methodology combined with the Instrumental Variables method in order to account for potential non-linearities and spatial effects in hedonic models. Using a large data set of more than 6,000 dwellings in Hamburg and about 17,000 in St. Petersburg, we confirm the presence of both effects (non-linearities and spatial autocorrelation). The results also show that there is a non-linear effect of the prices of neighboring houses on the price of each house. High prices for neighboring houses have a lower impact on the house price than low prices of neighboring houses. Finally, an extensive Monte Carlo exercise evaluates the ability of MARS to incorporate the correct spatial spillover terms in spatial regression models simultaneously including at same time non-linear effects.

Keywords: Multivariate Adaptive Regression Spline; spatial regression models; hedonic models; housing prices; Hamburg; St. Petersburg.

JEL codes: C4, C5, R1.

[☆]The authors express their gratitude to Sebastian Brandt and Wolfgang Maennig for kindly sharing their data. The usual disclaimer applies.

Contents

1	Introduction	1
2	Methodology and data	4
2.1	Methodology	4
2.1.1	The MARS algorithm	4
2.1.2	The learning phase or forward pass	5
2.1.3	The pruning process	6
2.2	Parameter options	7
2.2.1	Illustration	8
3	Non-linear spatial spillovers: empirical illustrations	9
4	Monte Carlo evidence	19
4.1	Results: NL-SLX	22
4.2	Results: NL-SAR	23
4.3	Results: NL-SDM	25
4.4	The impact of the number of neighborhoods and the level of autocorrelation	27
5	Conclusions	30
	Literature	31
	Supplementary material. Tables	36
	Results: NL-SEM	37

List of Tables

1	Example Forward and Backward of MARS algorithm	8
2	Regression results for Hamburg	11
3	Regression results for St. Petersburg	15
4	Characteristic of MARS and RMSE for NL-SLX	22
5	Characteristics of MARS and RMSE for NL-SAR	24
6	Characteristic of MARS and RMSE for NL-SDM	27
7	Sensibility of MARS algorithm to knn and ρ ($n=3600$). Irregular lattice	29
S-1	Characteristic of MARS and RMSE for NL-SLX ($\sigma = 1$)	36
S-2	Characteristic of MARS and RMSE for NL-SAR ($\sigma = 1$)	36
S-3	Characteristic of MARS and RMSE for NL-SDM-MARS ($\sigma = 1$)	36
S-4	Characteristic of MARS and RMSE for NL-SEM-MARS ($\sigma = 0.5$)	37
S-5	Characteristic of MARS and RMSE for NL-SEM ($\sigma = 1$)	37

List of Figures

1	Example MARS algorithm	9
2	Effects of distance to railway stations in Hamburg	14
3	Non-linear apartment price effects for St. Petersburg	17
4	Boxplot for c and c' knots for NL-SLX. Regular lattice	23
5	Boxplot c knots and ρ values in NL-SAR. Hexagonal lattice	26
6	Boxplot c and c' knots for different ρ values in NL-SDM. Regular lattice	28
7	Boxplot [‡] c and c' knots for different ρ values in NL-SDM-MARS. Regular lattice	29

1. Introduction

The advent of the Internet paved the way for using the big data in economic investigations. New techniques have been developed to work with large volumes of data ([Gandomi and Haider, 2015](#)) and, in other cases, methodologies introduced long time ago, developed when the data were scarce, have become more useful in the big data era ([Yadav et al., 2013](#)). Machine learning algorithms are a powerful techniques applied to big data sets in order to discover patterns using only the information contained in the data. Regional science and the urban economy are not oblivious to these changes and some contributions greatly benefit the statistical analysis of data when the spatial reference of data is present. Some of more popular of machine learning methods are adapted to include the spatial component in them with the objective of incorporating spatial references in the analysis. For example, a modification of the regression trees algorithm to identify converge clubs using filtered spatial regression models is proposed by [Postiglione et al. \(2010\)](#). In the same vein, [Paez et al. \(2019\)](#) include the coordinates of observations and interactions to improve the classification algorithm of regression trees. [Sommervoll and Sommervoll \(2019\)](#) use machine learning methods, in form of genetic algorithms, to aggregate spatial units showing that outperform the conventional econometric models with spatial fixed effects; [Dong et al. \(2015\)](#) show that Support Vector Machine (SVM) models accounting for spatial proximity outperform the non-spatial models in terms of model fitting and predictive performance. [Andris et al. \(2013\)](#) use SVM to investigate geographic patterns. [Sinha et al. \(2019\)](#) show that the random forest model is sensitive to spatial autocorrelation, which leads to an increase in the variance of the residuals; [Cajias and Ertl \(2018\)](#) test explanatory power by means of out-of-sample validation approaches using Generalized Additive Model (GAM) for spatial hedonic model. In general, this research shows how regional science and urban economics can benefit from the use of machine learning.

In this context, the Multivariate Adaptive Regression Splines (MARS) algorithm, as suggested by [Friedman \(1991\)](#), is not yet used in regional science, to the best of our knowledge, even as it is frequently used in other fields, such as finance, geography, medicine, industry, engineering, etc. For example, in finance, [De Andrés et al. \(2011\)](#) use MARS for the forecasting of bankruptcy; in business in mining the customer credit ([Lee et al., 2006](#)); in geography, [Haleem et al. \(2010\)](#) apply MARS to predict motor vehicle crashes; in medicine, [Lin et al. \(2008\)](#) detect

interaction between genetic factors in prostate cancer; [Austin \(2007\)](#) take advantage of MARS to predict acute myocardial infarction mortality; [York et al. \(2006\)](#) use it for detecting disease–risk relationship; and [Zhang et al. \(2015\)](#) in geotechnical earthquake engineering to analyze the soil liquefaction. These are only some examples showing the multidisciplinary and popularity of the MARS methodology.

MARS is a simple and powerful non-parametric regression algorithm based on a “divide and conquer” (*divide et impera*) strategy, partitioning the data sets into separate regions; each getting its own regression line. The MARS algorithm has two main objectives: the first is to make automatic selection of relevant variables in a classical regression model, while the second allows for capturing potential non-linearities of data using a set of specific functions named base functions. Both objectives have had an unequal treatment in the field of spatial econometrics ([Anselin, 1988](#)). With respect to the technique of automatic selection of variables as a method for the selection of spatial-terms ([Florax et al., 2003](#)), as far as we know, no research has been developed for automated specification strategies looking for the correct specification of the spatial regression model. Perhaps, it is due to the fact that the main estimation method for such models is maximum likelihood (ML), although an instrumental variable approach has also been developed ([Kelejian and Prucha, 1998](#)). By contrast, with respect to capturing potential non-linearities, there is a considerable literature accounting for non-linearities in spatial regression models ([Basile and De Benedictis, 2008](#); [Basile et al., 2014](#)). For example, [Sangalli et al. \(2013\)](#) consider Spatial Autoregressive Semiparametric Geoadditive Models, combining penalized regression spline (PS) methods ([Eilers et al., 2015](#)) with standard spatial autoregressive models (such as SAR, SEM, and SDM). With a similar objective, [Basile et al. \(2014\)](#) combine penalized regression spline (PS) methods with standard cross-section spatial autoregressive models. The authors develop new models named PS-SAR, PS-SEM, PS-SDM, and PS-SLX with the objective of including within the same specification non-linearities and spatial autoregressive terms, the latter capturing spatial interactions. Moreover, there are numerous applied papers that include splines in classical spatial regression models. Two nice examples are [Anselin and Le Gallo \(2006\)](#), who include splines to model the impact of independent terms on the price of houses, and [Brunauer et al. \(2010\)](#), who apply additive mixed regression models (AMM) to estimate hedonic price equations for rents in Vienna.

Despite its power, the MARS algorithm is still seldom used in spatial economics. To the best of our knowledge, there are only two regional economics studies using MARS: [Martinetti and Geniaux \(2017\)](#) and [De la Llave et al. \(2019\)](#). However, both cases only apply the MARS algorithm to make an automatic pre-selection of non-spatial variables and not to select spatial term in a model with spatial effects. In part, this lack of applications of MARS can be explained by the limitation of standard estimation techniques.

This paper fills this gap in the literature, showing how the MARS algorithm works in the framework of spatial regression models. We claim that the MARS algorithm combined with the instrumental variables (IV) regression method ([Kelejian and Prucha, 1998](#)) is very useful in the spatial econometric context, both for the selection of the correct specification of the model and for the identification of non-linear structures. The combination of both methods (MARS and IV) will be especially useful when handling large data sets with spatial autocorrelation. Note that, as shown in [Arbia et al. \(2019\)](#), the standard ML methods already become prohibitive and unreliable already with large sample sizes. Thus, the fast growth of the big data sets with spatial reference make it necessary to explore new tools to be successfully discover unseen patterns in large and high-dimensional data sets.

In this study, we present two examples of MARS applications using big data sets with spatial autocorrelation. The first represents two empirical evidences using real-life hedonic models illustrating the application of spatial MARS to the housing markets of Hamburg (Germany) and St. Petersburg (Russia). We select the real estate markets because they are typically characterized by non-linearity and spatial autocorrelation ([Anselin and Le Gallo, 2006](#); [Brunauer et al., 2010](#); [Cajias and Ertl, 2018](#)). The second example is an extensive Monte Carlo exercise used to evaluate the performance of MARS in presence of data with spatial autocorrelation and non-linearities.

Therefore, the present study makes a twofold contribution. On the one hand, it extends MARS to allow for spatial dependency (both of the dependent and explanatory variables). On the other hand, it proposes the use of the IV method, which makes feasible the application of MARS to very large data sets.

The paper is organized as follows. A short description of the MARS algorithm is present in Section 2. Section 3 contains two empirical applications to spatial regression hedonic model

using large data sets of dwellings in Hamburg and St. Petersburg. Section 4 presents the Monte Carlo evidence about the power of MARS to select the correct specification of spatial regression model in presence of non-linear terms. Finally, main conclusions appear in Section 5.

2. Methodology and data

2.1. Methodology

MARS is a flexible non-parametric piecewise regression technique introduced by [Friedman \(1991\)](#). This data-driven technique is specifically useful for identifying non-linearities in regression models without previous assumptions about the functional form or what are the explanatory variables or the number of them. The main characteristic of this methodology is that the econometric model considers different regression slopes in distinct intervals for each predictor. Unlike better known linear regression techniques, MARS does not assume that coefficients are stable across the entire range of each variable and instead uses splines in order to fit piecewise continuous functions to model responses. In general, MARS constructs a piecewise linear function to capture non-linear relationship in an adaptive manner ([Hoang et al., 2017](#)). The principal advantage of this methodology compared to similar algorithms (e.g., polynomial models) is the simplicity of the resulting econometric model and its easy interpretability. Moreover, the MARS models are reported to work satisfactorily in terms of computational cost irrespective of dimension of the data set. This is very useful when it is suspected that model inputs have varying optima across different levels of the model inputs ([Crino and Brown, 2007](#)).

2.1.1. The MARS algorithm

Like in any regression model, the objective of this methodology is to build an econometric model in order to explain the variation of a dependent variable $Y = (y_1, \dots, y_n)'$ with a set of potential independent variables $X = (X_1, \dots, X_p)$, where $X_i = (x_{1i}, x_{2i}, \dots, x_{ni})'$. In order to achieve this classical objective, the MARS uses the named basic functions (BF) of the form $(x - c)_+ = \max\{0, x - c\}$ and $(c - x)_+ = \max\{0, c - x\}$, where the subscript “+” means that the function takes only the positive value or zero in case of negative difference. Such pairs of linear functions are called “hinge functions” (or two-sided truncated functions) and the constant c denotes a knot, where the slope changes. The collection of all possible BFs, \mathcal{C} , is used to

construct the following econometric model:

$$\mathcal{C} = \{(x - c)_+, (c - x)_+\} \quad \text{with } c \in \{x_{1i}, x_{2i}, \dots, x_{ni}\} \quad \text{and } i = 1, \dots, p \quad (1)$$

Each function is piecewise linear with a knot c at every x_{ij} , and, in case all input values are distinct, there are np hinge functions, or equivalently $2np$ basic functions. Using those BFs, the model-building strategy is similar to a classical forward stepwise regression, which uses as inputs the functions from the set \mathcal{C} and their products. The final expression of the model is as follows:

$$Y = \beta_0 + \sum_{m=1}^M \beta_m h_m(X) + \epsilon \quad (2)$$

where $h_m(X)$ is a BF, or a product of two or more such functions, if interactions between variables are permitted, or perhaps the original predictor, if it has a linear impact on the dependent variable. The coefficients β_m are estimated by minimizing the sum of squared (residual) errors (SSE) similarly to a standard linear regression model.

2.1.2. The learning phase or forward pass

The model-training process iteratively selects and adds some of hinge functions into the model (or the original predictor). During the training process at each step, MARS selects new terms that minimize the SSE using ordinary least squares (OLS). In this forward pass, the MARS algorithm starts with a single model including only the intercept term β_0 . At each subsequent step, a reflected pair of hinge functions (or an original predictor) are selected and added to the model. The selected pair of hinge functions (or original predictor) can enter the model directly. Alternatively, they can be multiplied by an existing basic function that is already in the model, thus becoming new basis functions. The second case allows modeling the interaction between and/or among different predictors. Note that a reflected pair of hinge functions always enter the model together (but may be removed separately in the pruning process; see subsection 2.1.3). The forward pass continues until it meets one of many conditions such as: (i) maximum number of model terms (chosen by the user) before pruning is reached; or (ii) the change in the SSE is too small to continue. The search of hinge functions at each step can be done using brute force, but this search can be sped up with a heuristic that reduces

the number of parent terms to consider (see [Friedman 1993](#)).

In general, at the end of this process, we have a large model of the form of equation (1). The MARS model obtained in this forward pass is adaptive and can exhibit a high degree of flexibility that can ultimately result in overfitting, if no measures are taken to counteract it. To solve the overfitting problem and build a model with better generalization ability a pruning procedure must be applied.

2.1.3. The pruning process

Although there are other methods, MARS typically applies a backward deletion procedure to prune the model. Using this procedure, the second phase of this algorithm is the pruning step, where a “one-at-a-time” backward deletion procedure is applied during which the basis function with the least contribution to the model is eliminated, repeatedly, until the best submodel is found. This pruning is based on a generalized cross-validation (GCV) criterion originally proposed by [Craven and Wahba \(1979\)](#) and adapted by [Friedman and Silverman \(1989\)](#) known as Lack-of-Fit (LOF). Note that the raw SSE on the training data is inadequate for comparing models, because the SSE always increases as MARS terms are dropped and, therefore, the application of this criterion in the backward pass always results in selecting the largest model. Therefore, the GCV criterion is used to find the overall best model from a sequence of fitted models, where a larger GCV value tends to produce a smaller model and *vice versa*. The GCV criterion is estimated as the LOF criterion ([Friedman et al., 2001](#)). At each step, the algorithm removes a term in the model that results in the smallest increase in the sum of squared errors, obtaining an optimal model. This criterion is defined as,

$$GCV = \frac{1}{n} \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\left[1 - \frac{\hat{C}(M)}{n}\right]^2} \quad (3)$$

where $\hat{C}(M) = C(M) + dK$ and $C(M)$ are the number of parameters being estimated (the number of linearly independent basis functions without the intercept term); K is the number of knots selected in the forward process, and d represents a cost for each basic function optimization. Usually, $d = 2$, if the model does not involve interaction terms, and $d = 3$, otherwise ([Friedman, 1991](#)). Thus, the GCV formula adjusts the SSE to take into account the flexibility of the model. Larger values of d result in fewer knots and smoother function estimates. The

best MARS approximation is the one with the highest GCV value. In order to get a measure similar to R^2 , the GCV coefficient can be standardized and a new coefficient is defined as:

$$GRSq = 1 - \frac{GCV}{GCV.tot} \quad (4)$$

where $GCV.tot$ is the GCV of a model with only the intercept term. In this pruning phase, the researcher can determine the maximum number of knots considered, the minimum number of observations between knots, and the highest order of interaction terms. In addition, a critical strength of MARS is that it can train models that are very easily interpretable. However, increasing the flexibility generally reduces the interpretability. If interpretability is not a key consideration, then perhaps a more flexible algorithm, such as a random forest, should be used instead. This is the reason why the degree of interaction is usually limited to one or two, rarely more.

A useful option in the MARS procedure is to set an upper limit on the order of interaction. For example, one can set a limit of two, allowing pairwise products of piecewise linear functions, but not third or higher order products. This can aid in the interpretation of the final model. An upper limit of one results in an additive model.

2.2. Parameter options

The basic inputs to the MARS algorithm are the predictors and the response variables. A set of parameter options can be prespecified by the researcher: The order of interaction; the inclusion of linear variables; the maximum or minimum number of nodes; and the number of observations between nodes. [Friedman \(1991\)](#) suggests default values for these parameters, which we select in the Monte-Carlo section. This doubles the number of predictors, forces it into the range of 20 to 200, and, finally, adds 1 for the intercept $nk = \min(200, \max(20, 2 \times ncol(x))) + 1$.

There is no single regression modeling technique that is best for all situations. The algorithm has its pros and cons. For example, MARS models are simple to understand and interpret. MARS is suitable for handling fairly large data sets. As an example, the function *earth* in the R package *earth* ([Milborrow, 2018](#)) suite has additive models with 8 million cases and 100 variables, and 80 million cases and 2 variables. Larger models are also possible.

Table 1: Example Forward and Backward of MARS algorithm

	Forward pass					Backward	
	Model1	Model2	Model3	Model4	Model5	Model4	Model3
(Intercept)	1.917*** (0.061)	2.668*** (0.063)	3.766*** (0.371)	3.794*** (0.370)	3.757*** (0.370)	3.794*** (0.370)	3.766*** (0.371)
$(X - 0.5857)_+$		0.685* (0.281)	4.096*** (1.169)	4.410*** (1.180)	3.966** (1.208)	4.410*** (1.180)	4.096*** (1.169)
$(0.5857 - X)_+$		-5.159*** (0.202)	-7.531*** (0.814)	-7.584*** (0.812)	-7.513*** (0.812)	-7.584*** (0.812)	-7.531*** (0.814)
$(X - 0.2042)_+$			-3.107** (1.034)	-3.231** (1.034)	-3.067** (1.036)	-3.231** (1.034)	-3.107** (1.034)
$(X - 0.9755)_+$				-20.731 (11.907)	-44.955* (18.951)	-20.731 (11.907)	
$(X - 0.9337)_+$					8.651 (5.273)		
RSq	0.0000	0.7513	0.7569	0.7587	0.7604	0.759	0.757
ΔRSq		0.7513	0.0055	0.0019	0.0016		
GRSq	0.0000	0.7463	0.7494	0.7488	0.7479	0.7462	
$\Delta GRSq$		0.7463	0.0031	-0.0006	-0.0009		

*** p - value < 0.001, ** p - value < 0.01, * p - value < 0.05

The algorithm has a problem of overfitting and a propensity to identify knots in case of linear relationship. Statistically insignificant changes in the slope coefficients can be identified using some standard tests (for example, LR test).

2.2.1. Illustration

To show how the algorithm works, a data set is generated,

$$y_i = 1 + 4(x_i - 0.2)_+ + 8(0.2 - x_i)_+ - 3(0.6 - x_i) + \epsilon_i \quad (5)$$

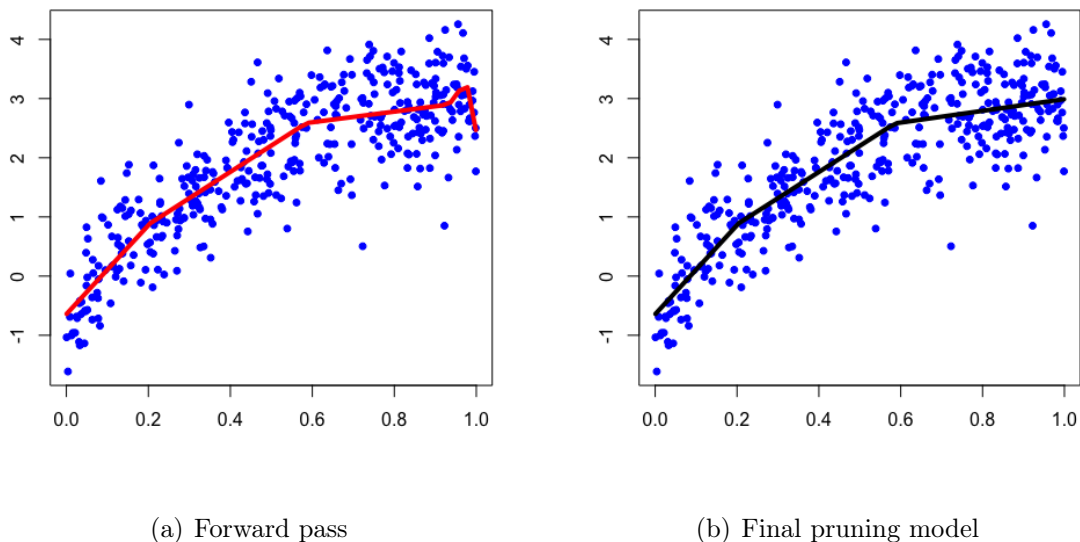
The DGP consider two knots $c = 0.2, 0.6$, with $\epsilon_i = N(0, 1)$ and $i = 1, \dots, n = 400$.

Table 1 shows the models estimated by OLS with the terms selected by MARS algorithm in each step. In the first phase, the algorithm incorporates, in an iterative process, a total of 6 terms in the forward pass (Models 1–5). In each step, select the term that maximizes the $RSq = R^2$. Model 1 is the baseline model and only includes the intercept. Model 2 selects the knot 0.5857 and two terms are incorporated into the model: $(X - 0.5857)_+$ and $(0.5857 - X)_+$ with an increase of RSq of $\Delta RSq = 0.7513$. Model 3 selects the knot 0.2042 and includes the term $(X - 0.2042)_+$. In this case, $\Delta RSq = 0.0055$. Model 4 selects the term $(X - 0.9755)_+$ with an increase of $\Delta RSq = 0.0019$ and Model 5 selects $(X - 0.9337)_+$ with an

increase of $\Delta RSq = 0.0016$. The process stops because the RSq increase with the next term $(X - 0.73776)_+$ is less than 0.001, a value fixed by the researcher, $\Delta RSq=0.00086$ (this model is not included in the Table to save space). In the second phase, the MARS algorithm prunes Model 5 using the $GRSq$ as criterion (with $d = 2$). *A priori*, any of the terms selected in the forward pass can be removed. In our case, note that for Model 4 and Model 5 the $\Delta GRSq$ values are negative. Finally, Model 3 is the one selected by the MARS algorithm.

Figure 1 shows the algorithm steps. First, the red line in Figure 2(a) shows the model before the pruning. Second, the black line in Figure 2(b) shows the final model after the pruning. The knots identified in the final model by the algorithm (0.5857 and 0.2042) are very close to the true knots (0.6 and 0.2).

Figure 1: Example MARS algorithm



3. Non-linear spatial spillovers: empirical illustrations

Here, we illustrate the application of the spatial MARS using two examples. The first consists of applying MARS to the data of [Brandt and Maennig \(2012\)](#), who assess the house price effects of the proximity to city rail transit in Hamburg using a linear model. In the second example, we will estimate a hedonic model of housing prices in St. Petersburg.

Hamburg. Brandt and Maennig (2012) estimate a linear hedonic model of the asking prices for dwellings in condominiums in Hamburg. They use asking prices announced in Internet ads between April 1, 2002, and March 31, 2008. The authors find non-linear effects of the proximity to the railway (including subway) stations: in the immediate neighborhood of the stations, the negative effects (e.g., noise) outweigh the positive effects of better transport accessibility. The highest benefit of the proximity is observed at the distance between 250 and 750 meters from the nearest station. We use their data set and estimate several models, see Table 2. Model (H-1) is a linear model, including both structural characteristics (construction year, total area, number of rooms, availability of balcony, terrace, fireplace, fitted kitchen, swimming pool, garage, and the state of dwelling) and locational features, such as the neighborhood characteristics (share of population over 65 years old, average income, share of foreigners, and share of social dwellings in the total housing stock) and distances from major points (like city hall, subcenters B and C, water, parks, industrial areas, and railway stations, the last being the variable of interest). It is kept as close as possible to the model of Brandt and Maennig (2012) and uses data kindly provided by authors.¹ Models (H-2) and (H-3) are the models produced by MARS and estimated by the OLS: without and with spatial lag of the dependent variable. In both cases, we restrict all variables, save the distance to the nearest station, to be linear. Model (H-2) includes spatial lags of all structural explanatory variables. In model (H-3), we use the IV approach (Kelejian and Prucha, 1998) in order to take care of the spatial lag denoted as Wy . The weighting matrix, W , is constructed using the k -neighbors approach. Different numbers of neighbors are tried: between 20 and 100. Starting from 20 neighbors, models produce similar results. Here, we report the models with 60 neighbors.² Not all variables included in the initial model specification survived the MARS selection procedure due to their lack of statistical significance.

The lowest row of Table 2 reports the Moran’s I calculated for residuals.³ It is statistically

¹However, we could not exactly replicate the data set, as some transformations undertaken by Brandt and Maennig (2012) cannot be traced and some variables could not be found.

²In addition, we estimated the models with W based on the inverse distances following the formulation used in Brandt and Maennig (2012). No non-linear effects are found using this definition of W . To save space the results are not reported here but are available to all interested upon request.

³To do this we use the function *moran.test* of the **R** package *spdep*. It computes the significance of test statistic using asymptotic distribution.

significant for all three models, but its value decreases substantially when spatial lags are introduced.

Table 2: Regression results for Hamburg

Parameter	Model (H-1)	Model (H-2)	Model (H-3)
Constant	11.375*** (0.036)	10.843*** (0.033)	5.144*** (0.230)
Year_2002	-0.036 (0.025)		
Year_2003	0.008 (0.021)		
Year_2004	-0.023* (0.012)		
Year_2005	-0.022** (0.011)		
Year_2006	-0.033*** (0.011)		
Year_2007	-0.016 (0.011)		
Size	0.014*** (0.0001)	0.013*** (0.0001)	0.013*** (0.0001)
SizeSq	-0.00001*** (0.00000)	-0.00001*** (0.00000)	-0.00001*** (0.00000)
$W \times \text{Size}$		0.004*** (0.0003)	-0.003*** (0.0005)
Age	-0.015*** (0.0003)	-0.013*** (0.0004)	-0.013*** (0.0003)
AgeSq	0.0001*** (0.00000)	0.0001*** (0.00000)	0.0001*** (0.00000)
$W \times \text{Age}$		-0.007*** (0.001)	0.003*** (0.0002)
$W \times \text{AgeSq}$		0.0001*** (0.00001)	
Rooms	0.030*** (0.003)	0.031*** (0.003)	0.031*** (0.003)
Garage	0.019** (0.008)		
Balcony	0.061*** (0.011)	0.087*** (0.008)	0.089*** (0.008)
Terrace	0.038*** (0.010)		
Kitchen	0.027*** (0.007)		
Pool	0.013 (0.020)		
Fireplace	0.017		

Parameter	Model (H-1)	Model (H-2)	Model (H-3)
	(0.016)		
Bad condition	-0.104***	-0.117***	-0.122***
	(0.014)	(0.014)	(0.014)
Good condition	0.070***	0.070***	0.070***
	(0.009)	(0.009)	(0.008)
Elderly population	-0.015***		
	(0.001)		
Income	0.004***		
	(0.0002)		
$W \times \text{Income}$		0.003***	
		(0.0002)	
Foreign population	-0.009***	-0.008***	
	(0.001)	(0.001)	
Social housing	-0.003***		-0.002***
	(0.0004)		(0.0003)
Total noise	-0.002***	-0.002***	-0.002***
	(0.0002)	(0.0002)	(0.0002)
Distance to city hall	-0.024***	-0.021***	
	(0.002)	(0.001)	
Distance to center B	0.005**		
	(0.003)		
Distance to center C	-0.004		
	(0.004)		
Distance to water	0.007		
	(0.005)		
Distance to park	-0.045***		
	(0.007)		
Distance to production sites	0.082***		
	(0.008)		
Distance to station	-0.033***		
	(0.007)		
(Distance to station-0.279) ₊		-0.050***	
		(0.007)	
(0.279-Distance to station) ₊		-0.397***	
		(0.080)	
(Distance to station-0.365) ₊			-0.058***
			(0.007)

Parameter	Model (H-1)	Model (H-2)	Model (H-3)
(0.365-Distance station) ₊	to		-0.241***
<i>Wy</i>			(0.051) 0.509*** (0.022)
Observations	6,332	6,332	6,332
Adjusted R ²	0.866	0.870	0.870
Moran's <i>I</i>	0.105	0.073	0.075

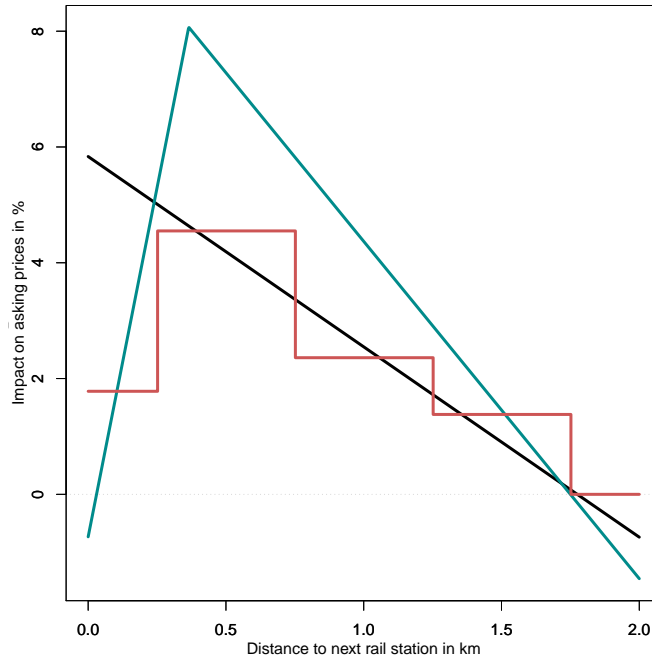
Note: *p<0.1; **p<0.05; ***p<0.01

The signs of estimated coefficients appear to be plausible. Spatial lags of surface, age of building, and income of neighborhood are also statistically significant, highlighting that the size and quality of closely located dwellings as well as the income levels in the area affect condominium prices. The spatial lag of dependent variable, *Wy*, turns out to be significant and positive, indicating that the prices of dwellings are affected by those of their neighbors.

Figure 2 depicts the effects of proximity to a railway station measured with respect to a dwelling located at 1750 m from the nearest station. The black line reflects the linear specification, where the effect is measured by a single coefficient of the distance variable. The brown stepwise line corresponds to the formulation of [Brandt and Maennig \(2012\)](#), who model the distance non-linearly by introducing several dummies for the distances between 0 and 250 m, between 250 and 750 m, between 750 and 1250 m, and between 1260 and 1750 m. The green line is obtained by our MARS model.

In the [Brandt and Maennig \(2012\)](#) model, the highest effects are obtained for the dwellings located between 250 and 750 m from the nearest railway station, while for those situated closer or farther to the station the effects are smaller, if not statistically insignificant. When dwellings are too close to the station, the negative impact of noise and visibility outweighs the positive accessibility effects. When they are too far away from the station, the accessibility factor no longer works. The MARS produces very similar effects. The effect is almost zero at very short distances and increases until 279 m (model (2)) or 365 m (model (3)) distance, where the largest positive effect is attained. After this threshold, the effects become negative: for dwellings located farther away from the nearest station, prices fall. Thus, MARS predicts the highest benefits at about the same distance from the railway station as [Brandt and Maennig](#)

Figure 2: Effects of distance to railway stations in Hamburg



(2012) do. Whereas in the original paper, the results were obtained a bit arbitrarily by trying various dummy variables of distance, the MARS provides the data-driven threshold.

St. Petersburg. A similar exercise is conducted for St. Petersburg, which is the second largest city of Russia and its former capital. We consider dwellings offered for sale in the multi-apartment buildings, since these represent the bulk of St. Petersburg’s housing stock.⁴ The data are collected from the website EMLS.⁵ The data set comprises over 17,000 announcements published in 2017. They contain some basic information about the structural characteristics of dwellings and the buildings in which they are located. Additional locational features are derived using the coordinates of dwellings and various urban amenities (city center, subway stations, and restaurants).⁶ We compute the distances between each dwelling and both the city center and the nearest subway station. For restaurants, we took instead the number of eating establishments within 1 km (walking distance) from each dwelling. We argue that it is

⁴According to the results of the All-Russia 2010 census, in St. Petersburg only 18,584 households out of 1,882,236 lived in single-family houses, which makes up less than 1% (https://gks.ru/free_doc/new_site/perepis2010/croc/perepis_itogi1612.htm).

⁵<https://emls.ru/>.

⁶The data on amenities are thoroughly described and the coordinates of consumer center of St. Petersburg used here as those of the city center are determined in Kholodilin et al. (2017).

more important to have a richer selection of restaurants in the neighborhood than just have one eating establishment at a short distance.

The estimation results for the logarithm of asking price per square meter are reported in Table 3.

Table 3: Regression results for St. Petersburg

Parameter	Dependent variable: logarithm of asking price		
	Model (SP-1)	Model (SP-2)	Model (SP-3)
Constant	7.986*** (0.009)	8.895*** (0.007)	3.838*** (0.297)
Area_total	0.017*** (0.0001)		
(Area_total-67.1) ₊		0.014*** (0.0001)	0.013*** (0.0001)
(67.1-Area_total) ₊		-0.022*** (0.0003)	-0.022*** (0.0003)
(WArea_total-68.1114) ₊		0.010*** (0.0002)	-0.002*** (0.001)
(68.1114-WArea_total) ₊		-0.001*** (0.0003)	0.010*** (0.001)
Rooms	-0.078*** (0.003)		
(Rooms-2) ₊		-0.132*** (0.004)	-0.127*** (0.004)
(2-Rooms) ₊		0.053*** (0.006)	0.050*** (0.006)
Floor	0.001** (0.0005)		
(4-Floor) ₊		-0.025*** (0.002)	-0.025*** (0.002)
(WFloor-7.85) ₊		-0.018*** (0.001)	
(7.85-WFloor) ₊		-0.005*** (0.002)	
NFloor	0.005*** (0.0004)		
(NFloor-9) ₊		0.003*** (0.0004)	0.002*** (0.0004)
(9-NFloor) ₊		-0.026*** (0.001)	-0.026*** (0.001)
Balcony	0.024*** (0.004)		
(WBalcony-0.42) ₊		-0.232*** (0.013)	-0.163*** (0.014)

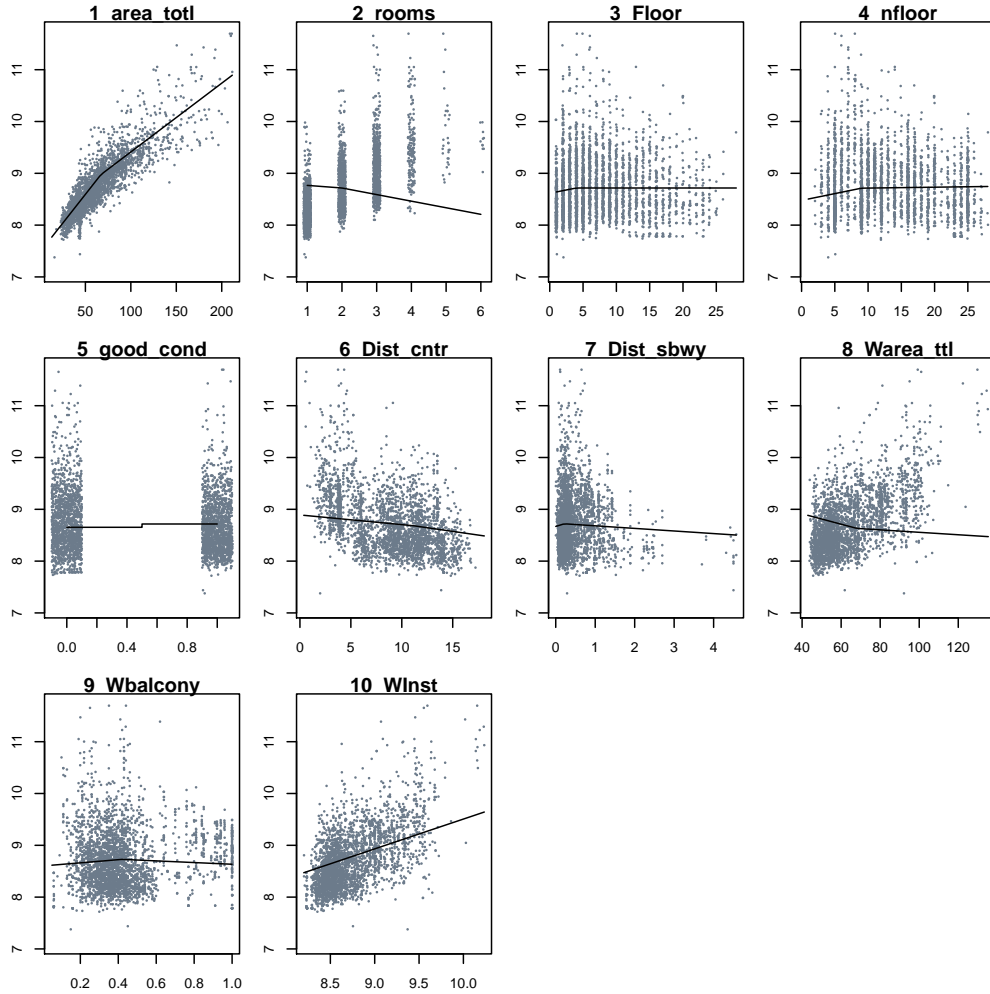
Parameter	Model (SP-1)	Model (SP-2)	Model (SP-3)
(0.42-WBalcony) ₊		-0.296*** (0.024)	-0.304*** (0.023)
Good condition	0.071*** (0.004)	0.061*** (0.004)	0.064*** (0.004)
Distance to center	-0.032*** (0.001)		
(Distance to center-11.17) ₊		-0.027*** (0.002)	-0.028*** (0.002)
(11.17-Distance to center) ₊		0.022*** (0.001)	0.019*** (0.001)
Distance to subway	-0.037*** (0.004)		
(Distance to subway- 0.217) ₊			-0.050*** (0.004)
(0.217-Distance to subway) ₊			-0.236*** (0.038)
Walk_restaurant	0.004*** (0.001)		
<i>Wy</i>			0.576*** (0.034)
Observations	16,959	16,959	16,959
Adjusted R ²	0.814	0.861	0.862
Moran's <i>I</i>	0.225	0.194	0.184

Note: *p<0.1; **p<0.05; ***p<0.01

Model SP-1 corresponds to the specification including only linear effects and no spatial lags. The signs of the coefficients are as expected. Model SP-2 includes spatial lags of all explanatory variables, *WX*, and is obtained using the MARS algorithm. Note that some variables (e.g., *Walk_restaurant*) are automatically dropped due to the lack of their significance, while several explanatory variables and their spatial lags are now included in non-linear form. Finally, model SP-3 extends model SP-2 by considering the spatial lag, *Wy*. This model is estimated using IV and “forces” MARS to include spatial lag of dependent variable as linear, provided that it is significant. In case of St. Petersburg, unlike for Hamburg, we allow for potential non-linearity of all variables. Seven variables (the total area, the number of rooms, the floor, the number of floors, the availability of balcony, the distance to the city center, and the distance to the subway stations) are found to have non-linear effects. These are graphically represented

in Figure 3. Its horizontal axis depicts the values of the corresponding explanatory variable, while the vertical axes reflects the log of asking square-meter price. The measure of spatial autocorrelation (Moran's I) is shown in the last row of Table 3. In models with spatial effects, it declines substantially compared to model SP-1.

Figure 3: Non-linear apartment price effects for St. Petersburg



The floor in which a dwelling is located exerts positive effect up to the 4th floor⁷ and then has a negative, but very small, impact. Living on lower floors can be unpleasant due to the noise and pollution, which are most intensive up to some height from the ground. It is known from anecdotal evidence that ground floor flats are not popular among Russians. The number

⁷In Russia, floor numbering starts from the ground floor, which is the first floor. Thus, the 4th floor in Russia is the 3rd floor, say, in Germany and much of Western Europe.

of floors in the house affects the price positively, however, the effect becomes negligible beyond the 9th floor. This can be the result of two factors. First, central parts of the city have relatively low-rise buildings due to governmental regulations, while in peripheral areas high-rise (sometimes more than 20 floors) buildings abound. Second, high-rise buildings constructed after 1960 are typically considered as low-quality housing: the construction materials are bad and the dwelling sizes small.

Housing prices seem to be most sensitive to the proximity to the city center within 11.2 km (the price decreases by 2.7–2.8% for each km), while after that distance, the effect declines (minus 1.9–2.2%). In other words, after about 11 km, proximity to city center becomes less relevant, possibly because the dwellings fall into attraction sphere of some subcenters. Interestingly, the effect, according to the linear model, is -3.2%, meaning that it is not an average of non-linear effects. For the proximity of subway stations, we obtain very similar results to those for Hamburg: up to 218 meters from the nearest station, the effects are positive, while beyond this distance, the prices decrease the further dwellings are located from the station. In fact, the distances are very similar: 218 meters for St. Petersburg vs. 279–365 meters for Hamburg. Apparently, the negative effects of station proximity (noise and crowds of people and vehicles) are most noticeable within about 300 meters around stations. The magnitude of proximity-to-station effects is also remarkably similar for Hamburg and St. Petersburg: 24.1–39.7% vs. 23.6% and -5.0 – -5.8% vs. 5%, respectively.

The spatial lag of dependent variable is statistically significant. In addition, non-linear spatial lags of three explanatory variables—the total area, the floor, and the availability of a balcony— exert a significant non-linear impact on the dwelling prices. For example, up to the surface of 67.1 m², the additional square meter adds 2.2% to the price, while beyond this threshold, the asking price increases by 1.4%. In case of spatial lag of total area, the effects are opposite. If the surface of neighbor apartments is below 68 m², each additional square meter leads to a price decrease of about 1%. When the nearby dwellings have a surface above 68 m², there is still a negative effect, but to a much smaller extent than in the case that such a dwellings is surrounded by small apartments. One interpretation is that the proximity of small dwellings indicates that the neighborhood consists of cheaper apartments, which are possibly inhabited by low-income persons.

4. Monte Carlo evidence

In this section, we present evidence about the performance of the MARS algorithm in presence of spatial terms. The objective is to evaluate the ability of MARS to select the correct model in presence of non-linear X term, non-linear spatial term (Wx), and the spatial lag term (Wy) for the more popular spatial regression models: SLX, SAR, and SDM.⁸ The ability of MARS to identify correctly the knot(s) value and the values of parameters of spatial dependence estimated by instrumental variables methodology is also evaluated.

The following DGPs are considered:

DGP: NL-SLX

$$y = \beta_0 + \beta_1(x - c)_+ + \beta_2(c - x)_+ + \beta_3(Wx - c')_+ + \beta_4(c' - Wx)_+ + \epsilon \quad (6)$$

DGP: NL-SAR

$$y = (I_n - \rho W)^{-1}(\beta_0 + \beta_1(x - c)_+ + \beta_2(c - x)_+ + \epsilon) \quad (7)$$

DGP: NL-SDM

$$y = (I_n - \rho W)^{-1}(\beta_0 + \beta_1(x - c)_+ + \beta_2(c - x)_+ + \beta_3(Wx - c')_+ + \beta_4(c' - Wx)_+ + \epsilon) \quad (8)$$

where x is a $n \times 1$ vector of x_i elements obtained from a $U(0, 1)$ distribution, c and c' are the knots of the hinge functions, and W is the classical weight matrix use in spatial econometrics to codify the relationship between observations.

The setup of the Monte-Carlo study is as follows:

label=MC-0: In DGP we select the parameters $\beta_0 = 4$; $\beta_1 = 4$; $\beta_2 = -2$; $\beta_3 = 4$; $\beta_4 = -2$ and $c = 0.5$; $c' = 0.5$.

lbbel=MC-0: We consider homoscedastic error term $\epsilon = N(0, \sigma^2 I_n)$ with $\sigma^2 = 0.5$, with this parameters we obtain coefficients R^2 around 0.75 (Supplementary material provides

⁸We select only spatial regression models that can be estimated by IV. For SEM models, Appendix 2 provides information about the performance of the MARS algorithm in presence of spatially autocorrelated errors

complementary results for $\sigma^2 = 1$).

lcbel=MC-0: Two types of lattices: Regular and irregular. Regular lattices with hexagonal forms of orders (20×20) ; (30×30) ; (60×60) ; (100×100) ; and (150×150) are used as the spatial support for the data. Irregular lattice with observations randomly distributed on a square (coordinates x and y obtained from a $U(0, 1)$ distribution). In both cases, sample sizes, n , of 400, 900, 3600, 10,000, and 22,500 observations, respectively, are used.

ldbel=MC-0: We consider W based on the rook contiguity criterion, where $w_{ij} = 1$, if x_i and x_j share a border for regular lattice. The criterion for irregular lattice is based on the $knn = 20$ near neighborhood. The W matrices are row-standardized as usual.

lebel=MC-0: Three levels of spatial dependence (low, medium, and high) are considered ($\rho = 0.25, 0.50, 0.75$).

lfbel=MC-0: 1000 replicas of DGP are considered.

The models are estimated by two-stage least squares (2SLS) using the MARS algorithm with the specifications listed below. The 2SLS approach is introduced by [Kelejian and Prucha \(1998\)](#) and [Kelejian and Prucha \(1999\)](#) in the spatial context as an alternative method to ML in order to avoid the need to compute the Jacobian term of the log-likelihood function. Kelejian and Prucha suggest a set of instruments obtained from a sequence of lags of the regressors, $H = [x, Wx, W^2x, W^3x, \dots]$. The sequence should be short due to potential problems of multicollinearity.

A large number of parameters can be customized when the MARS algorithm is launched. In this study, the selection of the MARS parameters is as follows:

lAbel=MARS-0 The input of the MARS algorithm includes five explanatory variables, x , Wx , z , Wz , and Wy^* . The variable z is obtained from a $U(0, 1)$ distribution independent of y and x . The variables z and Wz are included as input in the MARS algorithm to determine the ability of MARS to exclude non-related

terms. The Wy^* variable is the instrument used to the spatial lag, Wy , and obtained as the fitted values of the regression $Wy = x + Wx + W^2x + \epsilon$.

lBbel=MBRS-0 We force the term Wy^* to be linear in the MARS algorithm (argument *linpreds* in **R** package *earth*). Note that the MARS algorithm does not specify that a predictor must enter the model, only that, if it enters, it enters linearly.

lCbhel=MCRS-0 The values for the minimum number of observations between knots and the minimum number of observations before the first and after the final knot are considered by default with $\alpha = 0.05$. See equation (43) and (45) of [Friedman \(1991\)](#). In the pruning process, the penalty $d = 2$.

Two aspects are relevant as output of this Monte-Carlo. The first one is the ability of the MARS algorithm to select the “correct” specification (see below). The second one — in case the MARS algorithm selects the correct specification — is the error in the estimation of the knot (c and c').

For the final pruning model (FPM) estimate using the MARS algorithm, we are interested only in some characteristics of the model output:

- (i) mnc = Mean of the number of coefficients considered in the FPM.
- (ii) cs = Percentage of models with the correct specification. We understand by correct specification, if the FPM includes the intercept, two functions of type $(x - r)_+$; $(r - x)_+$, where the knot r is not necessarily equal to c , and the Wy^* term in case of NL-SAR and NL-SDM. In case of NL-SLX, we consider that MARS selects the correct model, if it includes the intercept and two functions of type $(x - r)_+$; $(r - x)_+$, where the knot r is not necessarily equal to c . Finally, in case of DGP NL-SAR and NL-SDM, the “correct specification” is a model with six terms (intercept, $(x - r)_+$; $(r - x)_+$; $(Wx - r')_+$; and $(r' - Wx)_+$).
- (iii) Wyt = Percentage of models that include the Wy^* term.
- (iv) The RMSE of the coefficients β_s , coefficients in case if the final pruning model selected the correct specification.

- (v) The RMSE coefficient ρ of the term Wy^* (in case if this coefficient is included in the FPM).

The root-mean-square error (RMSE) for the β_s and ρ coefficients for all models.

In this section, we report only the results for $\sigma = 0.5$ in order to save space. The results for $\sigma = 1$ are available in the Supplementary Material.

4.1. Results: NL-SLX

Table 4 shows the characteristics of FPM and values of RMSE of β_s coefficients in case of NL-SLX DGP. The more relevant results are:

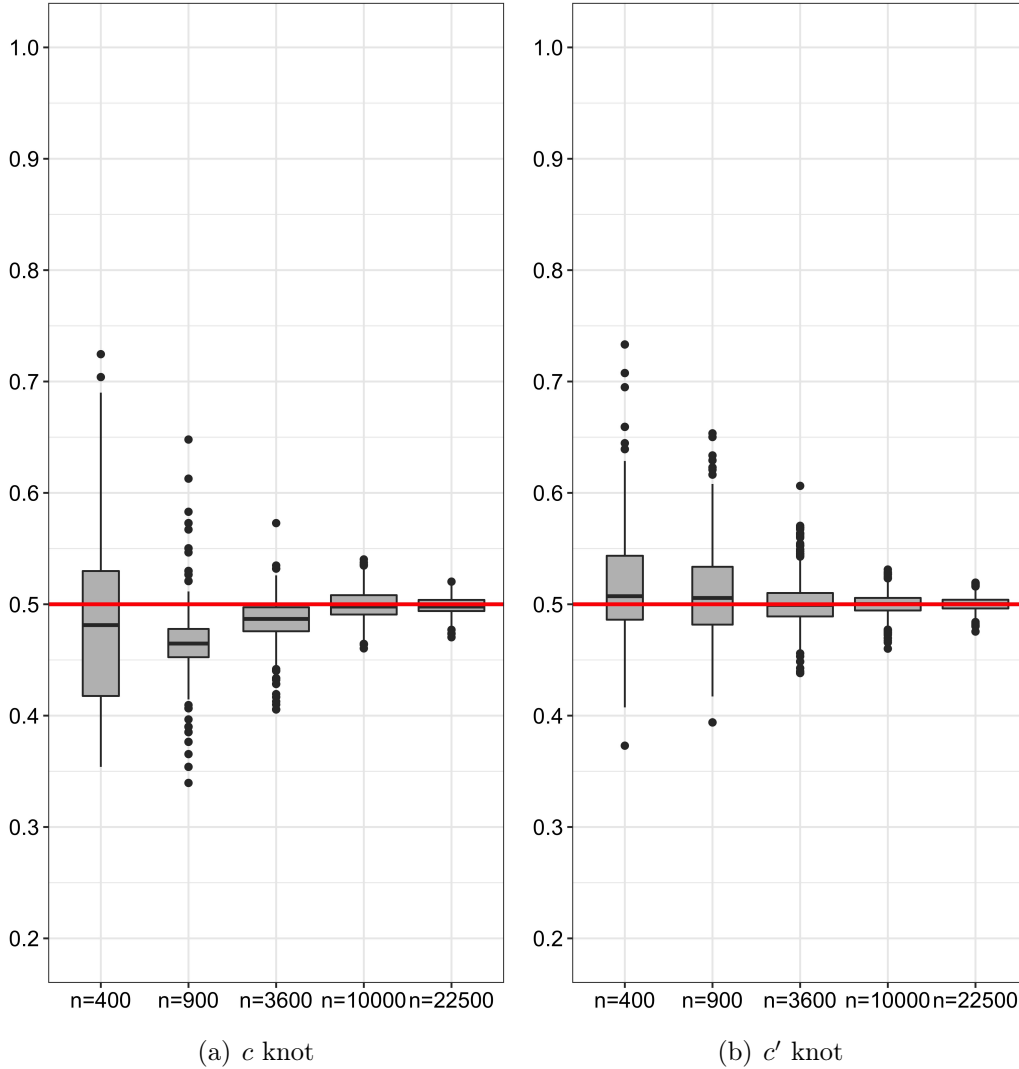
Table 4: Characteristic of MARS and RMSE for NL-SLX

		Characteristic FPM					RMSE			
lattice	n	mnc	cs	Wyt	β_0	β_1	β_2	β_3	β_4	
regular	400	5.629	0.024	0.116	0.834	18.216	0.525	9.553	6.046	
	900	6.026	0.086	0.103	0.463	0.912	0.412	10.237	3.150	
	3600	5.125	0.442	0.015	0.227	0.195	0.198	8.469	0.977	
	10000	4.944	0.942	0.000	0.107	0.103	0.098	1.925	0.444	
	22500	4.996	0.996	0.000	0.068	0.072	0.066	0.365	0.281	
irregular	400	5.791	0.117	0.105	0.287	0.314	0.281	4.557	2.292	
	900	5.668	0.324	0.043	0.248	0.211	0.184	3.095	0.856	
	3600	4.985	0.963	0.000	0.100	0.093	0.089	1.446	0.387	
	10000	5.000	1.000	0.000	0.045	0.051	0.048	0.204	0.200	
	22500	5.000	1.000	0.000	0.029	0.035	0.032	0.135	0.127	

- i. The sample size (n) is the more relevant parameter. For high values of n , the MARS algorithm always selects the correct specification.
- ii. The values of RMSE decrease when n increases.
- iii. The RMSE of coefficients β_3 and β_4 of spatial terms (WX) are higher than β_1 and β_2 .
- iv. The MARS algorithm does not confuse the spatial terms (WX vs. Wy). The term Wy^* is only retained in the model a very limited number of times by the algorithm and only when the sample size is small ($n \leq 900$).

With respect to the identification of knots, Figure 4 shows the boxplots of c and c' knots obtained using the MARS algorithm. Note that when the sample size increases, the MARS algorithm estimates both knots unbiased and with small variability.

Figure 4: Boxplot for c and c' knots for NL-SLX. Regular lattice



4.2. Results: NL-SAR

Table 5 shows the characteristics of FPM and values of RMSE in case of DGP NL-SAR.

- i. The sample size (n) is the more relevant parameter. For small sample size, the percentage of correct specifications cs obtained in the FPM is lower ($cs = 0.052$ for $n = 400$; $\rho = 0.25$; regular lattice). On the contrary, for a big sample size ($n \geq 10,000$; $\rho \geq 0.50$), 100% of models estimated using the MARS algorithm select the correct specification.
- ii. The degree of spatial autocorrelation $\rho = 0.5$ slightly improves the selection of correct model, if we compare it with cases of low or high ρ values ($\rho=0.25$; 0.75).

Table 5: Characteristics of MARS and RMSE for NL-SAR

lattice	n	ρ	Characteristic FPM			RMSE			
			mnc	cs	Wyt	β_0	β_1	β_2	ρ
regular	400	0.25	5.721	0.052	0.190	0.394	0.242	0.274	0.062
		0.50	5.871	0.175	0.674	0.469	0.336	0.346	0.050
		0.75	6.119	0.172	0.999	0.668	0.716	0.449	0.036
	900	0.25	5.287	0.065	0.184	0.314	0.187	0.221	0.047
		0.50	5.402	0.250	0.802	0.379	0.203	0.258	0.040
		0.75	5.667	0.228	1.000	0.508	0.238	0.442	0.028
	3600	0.25	4.397	0.288	0.293	0.151	0.088	0.097	0.023
		0.50	4.056	0.951	0.985	0.175	0.089	0.129	0.018
		0.75	4.198	0.851	1.000	0.249	0.131	0.250	0.013
	10000	0.25	4.323	0.541	0.541	0.080	0.057	0.052	0.013
		0.50	4.000	1.000	1.000	0.098	0.091	0.080	0.010
		0.75	4.005	0.997	1.000	0.133	0.219	0.179	0.007
	22500	0.25	4.239	0.697	0.697	0.056	0.042	0.037	0.009
		0.50	4.000	1.000	1.000	0.068	0.086	0.055	0.007
		0.75	4.000	1.000	1.000	0.090	0.222	0.146	0.005
irregular	400	0.25	5.198	0.017	0.105	0.907	0.271	0.324	0.143
		0.59	5.929	0.078	0.396	0.802	0.272	0.289	0.092
		0.75	6.543	0.083	0.834	1.356	0.276	0.296	0.078
	900	0.25	5.056	0.024	0.096	0.362	0.164	0.164	0.061
		0.50	5.637	0.122	0.496	0.442	0.200	0.182	0.051
		0.75	6.224	0.105	0.987	0.878	0.259	0.204	0.049
	3600	0.25	4.325	0.170	0.171	0.220	0.078	0.087	0.037
		0.50	4.146	0.821	0.865	0.251	0.094	0.092	0.029
		0.75	4.408	0.721	1.000	0.334	0.116	0.128	0.019
	10000	0.25	4.402	0.277	0.277	0.138	0.055	0.050	0.023
		0.50	4.010	0.986	0.986	0.148	0.055	0.054	0.017
		0.75	4.012	0.992	1.000	0.186	0.072	0.081	0.011
	22500	0.25	4.366	0.427	0.427	0.100	0.040	0.036	0.017
		0.50	4.000	1.000	1.000	0.110	0.047	0.040	0.013
		0.75	4.000	1.000	1.000	0.142	0.091	0.062	0.008

- iii. The percentage of correct specifications, cs , is higher for regular lattices than for irregular lattice.
- iv. The percentage of times that the term Wy^* is included in the 2SLS-MARS models depends on the level of spatial autocorrelation as expected. This percentage grows fast and for most cases is 100% ($Wyt = 1$).
- v. The mnc parameter converges to the number of terms. In case of $n \geq 10,000$, it is always equal to 4. There is a slight increase of mnc for high values of ρ .
- vi. The σ^2 parameter is another of the key elements of the correct model selection. In case of $\sigma^2 = 0.5$, the MARS algorithm obtains the best results for all parameters.
- vii. With respect to the RMSE of β_s parameters:

- The RMSE increases, when the level of spatial dependence increases.
 - For big sample sizes ($n \geq 10,000$), the RMSE of β_s parameters are very similar to those of the model NL-SAR-IV.
- viii. With respect to the parameter of spatial dependence, ρ , the three models show similar results. In case of $\sigma^2 = 0.5$, the RSMEs are lower than $\sigma^2 = 1$ (see Supplementary material) as expected.

Figures 6(a), 6(b), and 6(c) show the values obtained for knot c for different levels of spatial autocorrelation for regular lattices. The boxplots represent knots values only in cases when the model selection is correct. The width of the box is related to cs . The main result is that the level of spatial autocorrelation affects the selection of the knots. In case of high spatial autocorrelation, knot c is biased.

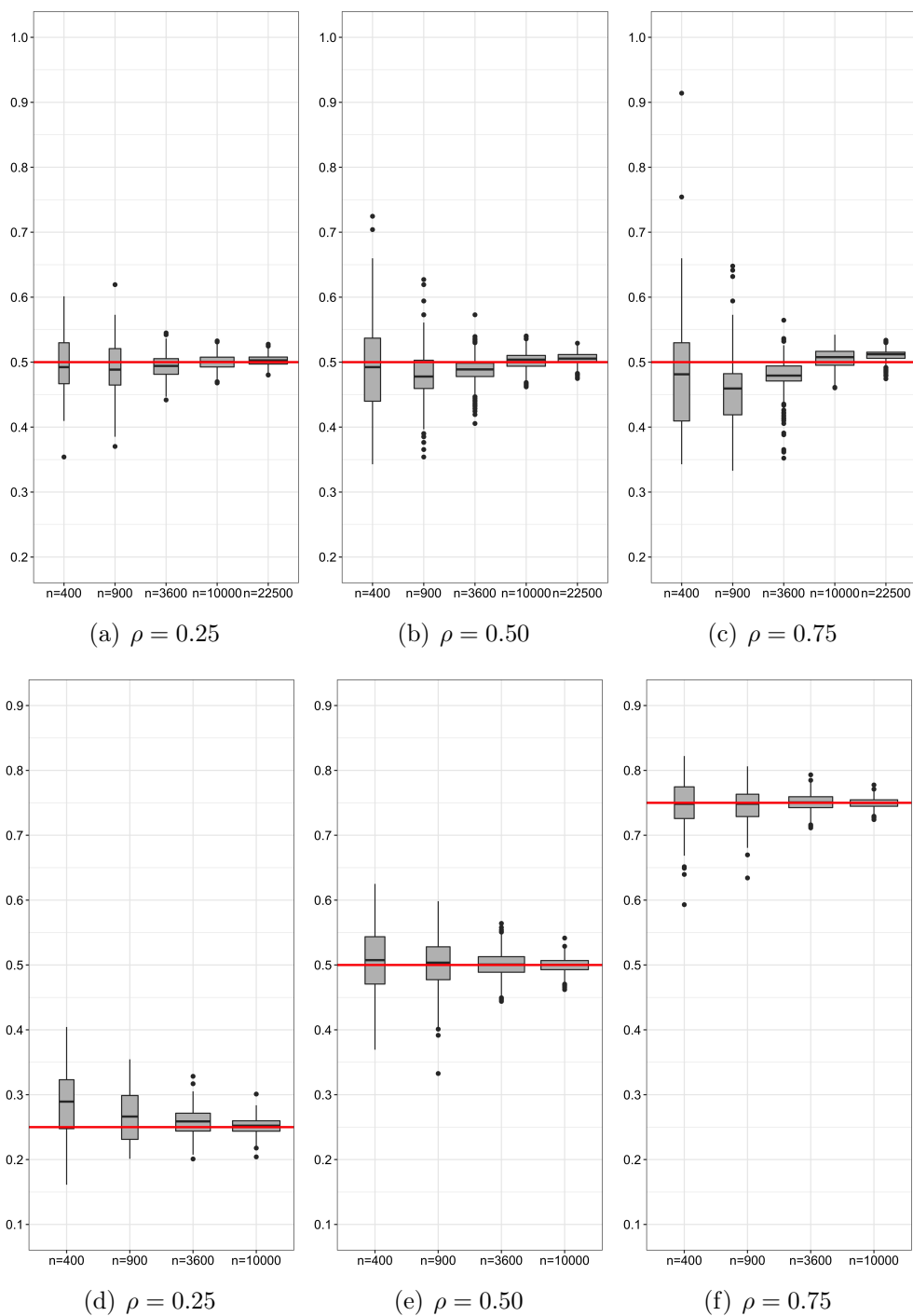
Finally, Figures 6(d), 6(e), and 6(f) show the boxplots for the estimated values of ρ parameter in SAR model for regular lattices. The estimation of this parameter of spatial dependence using IV is unbiased for big sample sizes.

4.3. Results: NL-SDM

Table 6 shows the main outputs of MARS algorithm for NL-SDM model. The main results are:

- i. For small sample size ($n = 400$) and low levels of spatial dependence, the number of cs is extremely low. In case of irregular lattice, no model with the correct specification is selected for $\rho = 0.25$ and only 1.8% ($cs = 0.018$) in case of $\rho = 0.50$. The increase of sample size leads to better results: For $n = 22,500$, the MARS algorithm selected 100% of correct specifications for medium and high levels of spatial autocorrelation.
- ii. Like in the NL-SLX model, the terms β_3 and β_4 associated with the spatial terms in X has higher RSME than the term not associated with spatial terms β_1 and β_2 .
- iii. The NL-SDM model is a bit harder to identify than NL-SAR model. The percentage of times that the MARS algorithm selects the correct model cs is lower for NL-SDM than for NL-SAR.

Figure 5: Boxplot c knots and ρ values in NL-SAR. Hexagonal lattice



iv. Like in case of NL-SAR model, the percentage of cs is higher for $\rho = 0.5$. Lower values are obtained if $\rho = 0.25$ or 0.75 .

Figure 6 shows the knots values obtained in the estimation of NL-SDM using the MARS

Table 6: Characteristic of MARS and RMSE for NL-SDM

lattice	n	ρ	Characteristic FPM			RMSE					
			mnc	cs	Wyt	β_0	β_1	β_2	β_3	β_4	ρ
regular	400	0.250	6.359	0.068	0.511	0.493	0.296	0.343	0.644	0.356	0.080
		0.500	6.723	0.176	0.999	0.711	0.418	0.358	1.278	0.949	0.081
		0.750	6.416	0.063	1.000	0.841	0.424	0.462	2.121	2.643	0.039
	900	0.250	6.577	0.272	0.719	0.482	0.219	0.283	0.576	0.438	0.070
		0.500	6.859	0.370	1.000	0.524	0.206	0.358	0.788	0.460	0.058
		0.750	6.630	0.414	1.000	0.560	0.320	0.351	1.444	0.464	0.031
	3600	0.250	5.932	0.921	0.926	0.249	0.094	0.127	0.278	0.282	0.037
		0.500	6.013	0.988	1.000	0.271	0.096	0.192	0.374	0.358	0.027
		0.750	5.985	0.981	1.000	0.289	0.187	0.266	0.585	0.661	0.014
	10000	0.250	5.994	0.994	0.994	0.154	0.073	0.067	0.188	0.185	0.024
		0.500	6.000	1.000	1.000	0.153	0.128	0.108	0.302	0.291	0.016
		0.750	6.000	1.000	1.000	0.185	0.236	0.279	0.661	0.587	0.008
	22500	0.250	6.000	1.000	1.000	0.103	0.042	0.046	0.151	0.142	0.017
		0.500	6.000	1.000	1.000	0.101	0.094	0.095	0.301	0.280	0.010
		0.750	6.000	1.000	1.000	0.133	0.246	0.241	0.708	0.659	0.006
irregular	400	0.25	6.099	0.000	0.248	—	—	—	—	—	—
		0.50	6.585	0.018	0.755	0.862	0.251	0.362	4.277	2.665	0.091
		0.75	7.291	0.022	0.998	1.915	0.288	0.342	5.707	6.403	0.110
	900	0.25	5.983	0.015	0.309	0.562	0.220	0.226	2.580	3.139	0.123
		0.50	6.763	0.137	0.989	1.033	0.280	0.211	2.871	1.238	0.104
		0.75	7.184	0.070	1.000	1.028	0.326	0.256	3.607	1.062	0.047
	3600	0.25	5.280	0.272	0.340	0.427	0.096	0.098	1.929	0.541	0.076
		0.50	5.919	0.866	1.000	0.370	0.111	0.105	1.351	0.557	0.041
		0.75	5.882	0.684	1.000	0.379	0.141	0.175	2.032	0.702	0.021
	10000	0.25	5.193	0.193	0.193	0.375	0.054	0.055	0.305	0.524	0.064
		0.50	5.999	0.999	1.000	0.228	0.055	0.058	0.505	0.466	0.026
		0.75	5.981	0.981	1.000	0.216	0.081	0.086	0.962	0.809	0.012
	22500	0.25	5.074	0.074	0.074	0.362	0.037	0.044	0.248	0.479	0.062
		0.50	6.000	1.000	1.000	0.172	0.058	0.042	0.412	0.388	0.019
		0.75	5.998	0.998	1.000	0.176	0.123	0.059	0.827	0.750	0.010

algorithm. For c and c' as well as for the three levels of spatial autocorrelation, the results show a clear improvement of the algorithm when the sample size increases.

Figure 7 displays the boxplots of the ρ coefficients estimated by IV. According to the empirical distribution, the parameter estimates appear to be unbiased.

4.4. The impact of the number of neighborhoods and the level of autocorrelation

The results shown in subsections 4.2 and 4.3 reveal that the more relevant parameter to obtain high cs percentage is the sample size (n). Other two factors impacting the performance of the MARS algorithm that are related to the spatial framework, where the MARS algorithm is developed: (i) the level of spatial autocorrelation; and (ii) the degree of connectivity of W matrix. In order to determine the impact of these factors in this subsection, we explore how do they affect the selection of the correct model by the MARS algorithm.

Figure 6: Boxplot c and c' knots for different ρ values in NL-SDM. Regular lattice

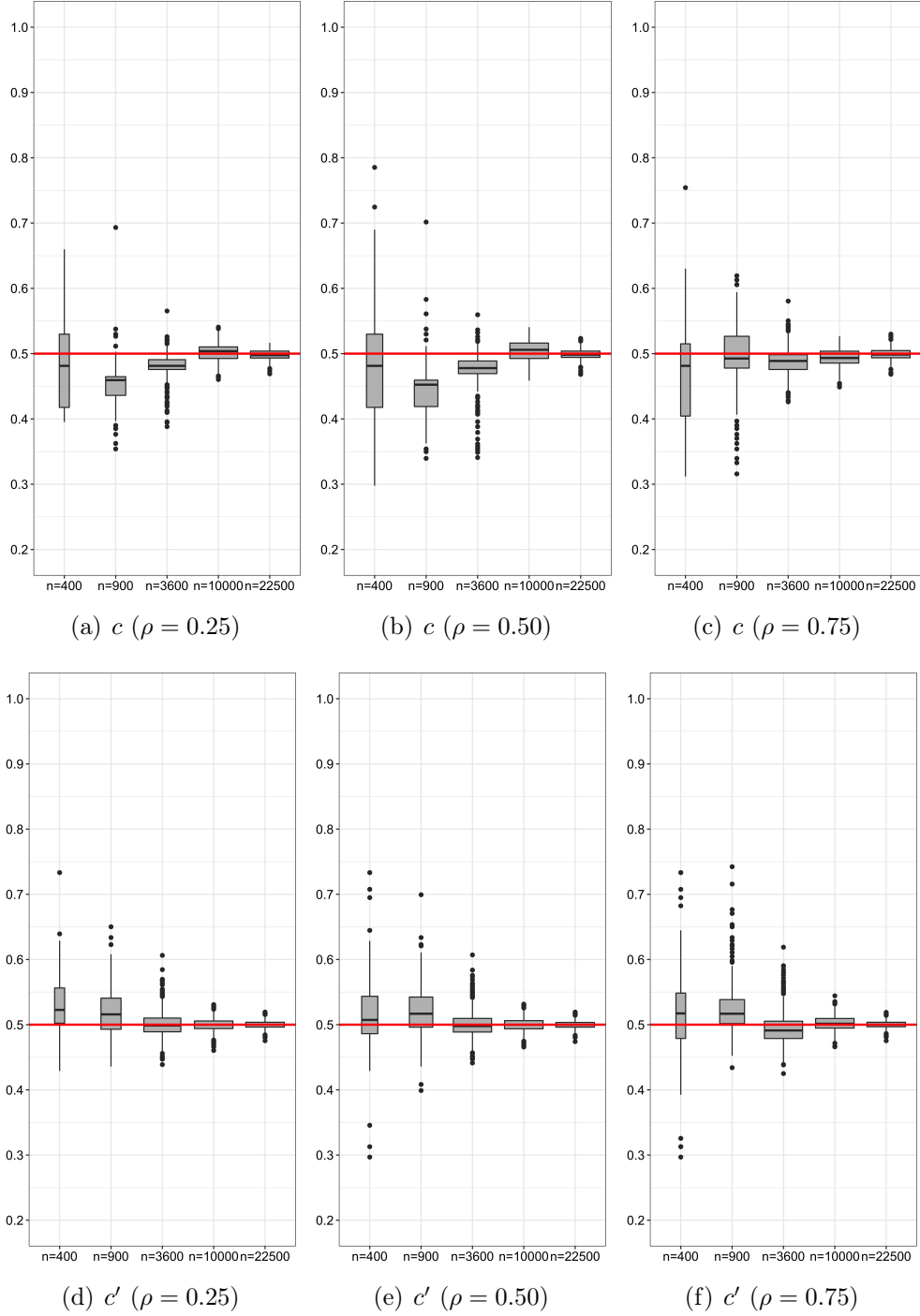
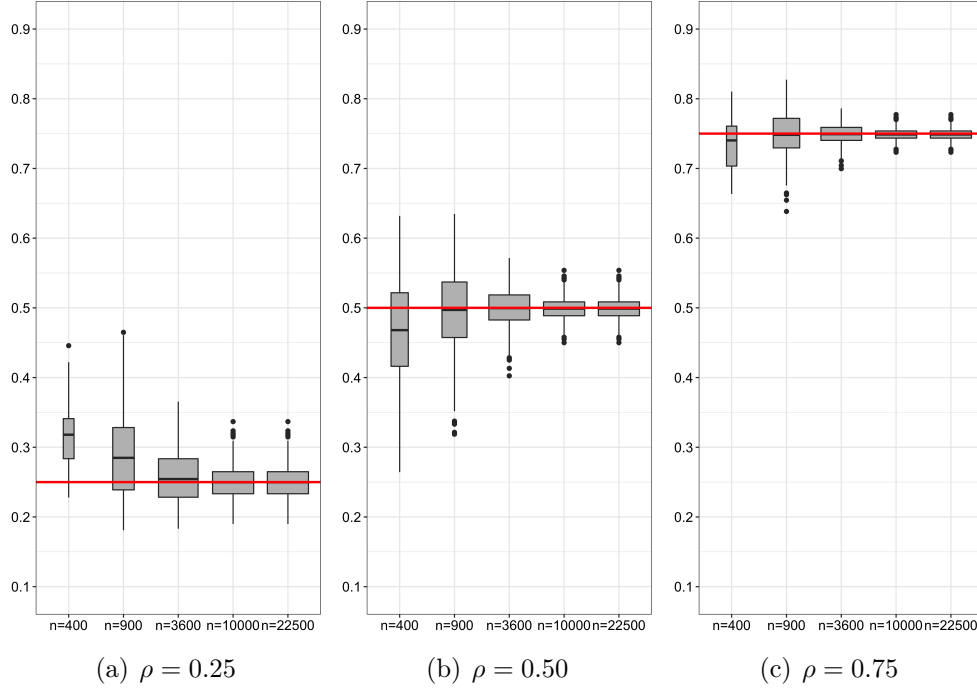


Table 7 shows the sensibility of MARS algorithm for different degrees of connectivity ($knn = 5, 10, 15, 20,$ and 30) of the W matrix and different levels of spatial autocorrelation ($\rho = 0.1, 0.25, 0.33, 0.50, 0.66,$ and 0.90). Two models $NL - SAR$ and $NL - SDM$ with

Figure 7: Boxplot[‡] c and c' knots for different ρ values in NL-SDM-MARS. Regular lattice



sample size $n = 3600$ over an irregular lattice are considered.

Table 7: Sensibility of MARS algorithm to knn and ρ ($n=3600$). Irregular lattice

Model	ρ	knn=5			knn=10			knn=15			knn=20			knn=30		
		mnc	cs	Wyt	mnc	cs	Wyt	mnc	cs	Wyt	mnc	cs	Wyt	mnc	cs	Wyt
NL-SAR	0.10	4.24	0.01	0.01	3.74	0.01	0.01	3.53	0.01	0.01	3.34	0.00	0.01	3.19	0.00	0.00
	0.25	4.38	0.40	0.40	4.44	0.27	0.28	4.43	0.17	0.18	4.40	0.14	0.14	4.35	0.11	0.11
	0.33	4.20	0.73	0.75	4.34	0.54	0.55	4.39	0.39	0.41	4.40	0.33	0.34	4.43	0.27	0.28
	0.50	4.12	0.90	1.00	4.12	0.90	0.99	4.15	0.87	0.95	4.17	0.81	0.88	4.28	0.71	0.78
	0.66	4.25	0.80	1.00	4.30	0.79	1.00	4.30	0.79	1.00	4.26	0.81	1.00	4.35	0.75	0.99
	0.75	4.38	0.73	1.00	4.45	0.70	1.00	4.50	0.67	1.00	4.47	0.67	1.00	4.58	0.61	1.00
	0.90	4.54	0.66	1.00	4.98	0.43	1.00	5.42	0.28	1.00	5.35	0.30	1.00	5.67	0.19	1.00
NL-SDM	0.10	5.04	0.04	0.04	5.02	0.01	0.01	5.01	0.00	0.01	5.00	0.00	0.01	4.98	0.00	0.00
	0.25	6.00	0.99	1.00	5.78	0.76	0.78	5.45	0.43	0.48	5.21	0.20	0.27	5.02	0.04	0.13
	0.33	6.01	0.99	1.00	6.00	0.99	1.00	5.90	0.88	0.94	5.77	0.74	0.83	5.37	0.33	0.49
	0.50	6.00	1.00	1.00	6.02	0.98	1.00	5.98	0.91	1.00	5.94	0.89	1.00	5.90	0.78	1.00
	0.66	5.98	0.97	1.00	5.91	0.87	1.00	5.82	0.73	1.00	5.80	0.67	1.00	5.71	0.54	1.00
	0.75	5.79	0.77	1.00	5.74	0.69	1.00	5.62	0.48	1.00	5.74	0.51	1.00	5.75	0.47	1.00
	0.90	5.39	0.34	1.00	5.31	0.10	1.00	5.69	0.28	1.00	6.02	0.25	1.00	6.41	0.37	1.00

The results of Table 7 are very informative:

- i. A strong “degradation” of the MARS algorithm is observed for high/low levels of spatial autocorrelation and high degree of connectivity, knn . This “degradation” is present for both models.
- ii. The percentage of times that the term Wyt is included in the model is higher for NL-SDM than for NL-SAR.

- iii. cs is high for NL-SDM with medium/low levels of spatial autocorrelation $\rho \leq 0.5$. For high levels of spatial autocorrelation, cs for NL-SAR is higher than for NL-SDM.

5. Conclusions

With the rapid increases in processing speed and memory of lower-cost computers, it is not surprising that various advanced computational learning tools are increasingly used for analyzing or modeling highly non-linear multivariate problems. These algorithms are useful for analyzing many problems, especially those that lack a precise analytical theory or understanding of the phenomena involved.

The field of spatial econometrics needs new methodologies to adapt to the era of big data. The problem of the maximum likelihood estimations for big data sets (Arbia et al., 2019) must be resolved using alternative estimation methods. The IV methodologies (Kelejian and Prucha, 1998) appear to be a real alternative for big data sets. The main advantages of MARS are its capacity to produce simple, easy-to-interpret models, its ability to estimate the contributions of the input variables, and its computational efficiency. The problem of selecting the correct specification for spatial regression models Mur and Angulo (2009); Mur et al. (2010) has similar problems. On the other hand, when working with big data sets, the hypothesis of linearity in the relationship between variables is improbable.

The MARS algorithm, a technique developed by Friedman (1991) in conjunction with the IV method developed by Kelejian and Prucha (1998), can be a solution to this problem. The results shown in the Monte Carlo section confirm the good performance of this methodology when working with a big data set. The algorithm helps to select the correct model and the IV inference framework work well in the case of big sample size.

Finally, we must admit that the approach used in the paper may have at least three fundamental weaknesses that merit some further comments. First, the IV method cannot be used to estimate SEM models. These models are frequent in spatial econometrics and no IV methods can be used in such cases. In this case, note that the MARS algorithm does not select spatial terms. Therefore, if the test of spatial dependence rejects the null hypothesis in an OLS model and the MARS does not select SLX/SAR/SDM, the researcher must suspect that he is dealing with an SEM model. Second, the MARS algorithm can cause overfitting problems. In this case,

we recommend using validation techniques to assess the results of the statistical analysis. The **R** package *earth* includes the option *nfold* to select the number of cross-validation folds. Third, a large number of options can be selected to launch the MARS algorithm (maximum degree of interactions between variables; maximum number of knots, maximum number of terms (including intercept) in the pruned model; minimum number of observations between knots and before the first and after the final knot; forward stepping threshold; etc.). Differing selection of this parameters could give rise to different models. It is the task of the researcher to choose the correct selection of the parameters based on his knowledge of the problem.

Literature

- Andris, C., Cowen, D., and Wittenbach, J. (2013). Support vector machine for spatial variation. *Transactions in GIS*, 17(1):41–61.
- Anselin, L. (1988). Spatial econometrics: Models and methods. *Dordrecht: Kluwer Academic Publishers*.
- Anselin, L. and Le Gallo, J. (2006). Interpolation of air quality measures in hedonic house price models: spatial aspects. *Spatial Economic Analysis*, 1(1):31–52.
- Arbia, G., Ghiringhelli, C., and Mira, A. (2019). Estimation of spatial econometric linear models with large datasets: How big can spatial big data be? *Regional Science and Urban Economics*.
- Austin, P. C. (2007). A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting ami mortality. *Statistics in Medicine*, 26(15):2937–2957.
- Basile, R. and De Benedictis, L. (2008). Regional unemployment and productivity in Europe. *Papers in Regional Science*, 87(2):173–192.
- Basile, R., Durbán, M., Mínguez, R., Montero, J. M., and Mur, J. (2014). Modeling regional economic dynamics: Spatial dependence, spatial heterogeneity and nonlinearities. *Journal of Economic Dynamics and Control*, 48:229–245.

- Brandt, S. and Maennig, W. (2012). The impact of rail access on condominium prices in Hamburg. *Transportation*, 39(5):997–1017.
- Brunauer, W. A., Lang, S., Wechselberger, P., and Bienert, S. (2010). Additive hedonic regression models with spatial scaling factors: An application for rents in Vienna. *The Journal of Real Estate Finance and Economics*, 41(4):390–411.
- Cajias, M. and Ertl, S. (2018). Spatial effects and non-linearity in hedonic modeling: Will large data sets change our assumptions? *Journal of Property Investment & Finance*, 36(1):32–49.
- Craven, P. and Wahba, G. (1979). Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 31:377–403.
- Crino, S. and Brown, D. E. (2007). Global optimization with multivariate adaptive regression splines. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(2):333–340.
- De Andrés, J., Lorca, P., de Cos Juez, F. J., and Sánchez-Lasheras, F. (2011). Bankruptcy forecasting: A hybrid approach using fuzzy c-means clustering and multivariate adaptive regression splines (mars). *Expert Systems with Applications*, 38(3):1866–1875.
- De la Llave, M. A., López, F. A., and Angulo, A. (2019). The impact of geographical factors on churn prediction: an application to an insurance company in Madrid’s urban area. *Scandinavian Actuarial Journal*, 2019(3):188–203.
- Dong, N., Huang, H., and Zheng, L. (2015). Support vector machine in crash prediction at the level of traffic analysis zones: assessing the spatial proximity effects. *Accident Analysis & Prevention*, 82:192–198.
- Eilers, P. H., Marx, B. D., and Durbán, M. (2015). Twenty years of P-splines. *SORT: statistics and operations research transactions*, 39(2):0149–186.
- Florax, R. J., Folmer, H., and Rey, S. J. (2003). Specification searches in spatial econometrics: the relevance of Hendry’s methodology. *Regional Science and Urban Economics*, 33(5):557–579.

- Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York.
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1):1–67.
- Friedman, J. H. (1993). *Fast MARS*. Stanford University. Laboratory for Computational Statistics.
- Friedman, J. H. and Silverman, B. W. (1989). Flexible parsimonious smoothing and additive modeling. *Technometrics*, 31(1):3–21.
- Gandomi, A. and Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2):137–144.
- Haleem, K., Abdel-Aty, M., and Santos, J. (2010). Multiple applications of multivariate adaptive regression splines technique to predict rear-end crashes at unsignalized intersections. *Transportation Research Record*, 2165(1):33–41.
- Hoang, N. D., Chen, C. T., and Liao, K. W. (2017). Prediction of chloride diffusion in cement mortar using multi-gene genetic programming and multivariate adaptive regression splines. *Measurement*, 112:141–149.
- Kelejian, H. H. and Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1):99–121.
- Kelejian, H. H. and Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40:509–533.
- Kholodilin, K. A., Krylova, I., and Kryutchenko, D. (2017). Where is the consumer center of St. Petersburg? DIW Berlin Discussion Paper No 1666.
- Lee, T.-S., Chiu, C.-C., Chou, Y.-C., and Lu, C.-J. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis*, 50(4):1113–1130.

- Lin, H.-Y., Wang, W., Liu, Y.-H., Soong, S.-J., York, T. P., Myers, L., and Hu, J. J. (2008). Comparison of multivariate adaptive regression splines and logistic regression in detecting snp-snp interactions and their application in prostate cancer. *Journal of Human Genetics*, 53(9):802.
- Martinetti, D. and Geniaux, G. (2017). Approximate likelihood estimation of spatial probit models. *Regional Science and Urban Economics*, 64:30–45.
- Milborrow, S. (2018). earth: Multivariate adaptive regression splines. R package.
- Mur, J. and Angulo, A. (2009). Model selection strategies in a spatial setting: Some additional results. *Regional Science and Urban Economics*, 39(2):200–213.
- Mur, J., López, F., and Herrera, M. (2010). Testing for spatial effects in seemingly unrelated regressions. *Spatial Economic Analysis*, 5(4):399–440.
- Paez, A., López, F., Ruiz, M., and Camacho, M. (2019). Inducing non-orthogonal and non-linear decision boundaries in decision trees via interactive basis functions. *Expert Systems with Applications*, 122:183–206.
- Postiglione, P., Benedetti, R., and Lafratta, G. (2010). A regression tree algorithm for the identification of convergence clubs. *Computational Statistics & Data Analysis*, 54(11):2776–2785.
- Sangalli, L. M., Ramsay, J. O., and Ramsay, T. O. (2013). Spatial spline regression models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):681–703.
- Sinha, P., Gaughan, A. E., Stevens, F. R., Nieves, J. J., Sorichetta, A., and Tatem, A. J. (2019). Assessing the spatial sensitivity of a random forest model: Application in gridded population modeling. *Computers, Environment and Urban Systems*, 75:132–145.
- Sommervoll, Å. and Sommervoll, D. E. (2019). Learning from man or machine: Spatial fixed effects in urban econometrics. *Regional Science and Urban Economics*, 77:239–252.
- Yadav, C., Wang, S., and Kumar, M. (2013). Algorithm and approaches to handle large data-a survey. *arXiv preprint arXiv:1307.5437*.

York, T. P., Eaves, L. J., and van den Oord, E. J. (2006). Multivariate adaptive regression splines: a powerful method for detecting disease–risk relationship differences among subgroups. *Statistics in Medicine*, 25(8):1355–1367.

Zhang, W., Goh, A. T., Zhang, Y., Chen, Y., and Xiao, Y. (2015). Assessment of soil liquefaction based on capacity energy concept and multivariate adaptive regression splines. *Engineering Geology*, 188:29–37.

Supplementary material. Tables

Table S-1: Characteristic of MARS and RMSE for NL-SLX ($\sigma = 1$)

n	Regular lattice								Irregular lattice								
	Charact. FPM			NL-SLX-MARS					Charact. FPM			NL-SLX-MARS					
	mnc	cs	W_{yt}	β_0	β_1	β_2	β_3	β_4	mnc	cs	W_{yt}	β_0	β_1	β_2	β_3	β_4	
$\sigma = 1$																	
400	5.946	0.078	0.108	0.483	1.697	0.441	3.725	1.469	6.150	0.303	0.087	0.261	0.301	0.294	0.859	0.460	
900	6.309	0.166	0.098	0.383	1.482	0.404	3.372	0.954	5.973	0.378	0.050	0.177	0.187	0.215	0.502	0.374	
3600	5.369	0.659	0.009	0.197	0.180	0.210	1.357	0.385	5.008	0.995	0.000	0.089	0.09	0.096	0.198	0.191	
10000	5.001	0.999	0.000	0.107	0.111	0.106	0.239	0.230	5.000	1.000	0.000	0.054	0.059	0.056	0.105	0.105	
22500	5.000	1.000	0.000	0.060	0.070	0.065	0.150	0.141	5.000	1.000	0.000	0.029	0.035	0.032	0.071	0.067	

Table S-2: Characteristic of MARS and RMSE for NL-SAR ($\sigma = 1$)

n	ρ	Regular lattice							Irregular lattice						
		Charact. FPM				RMSE			Charact. FPM			RMSE			
		mnc	cs	W_{yt}	β_0	β_1	β_2	ρ	mnc	cs	W_{yt}	β_0	β_1	β_2	ρ
400	0.25	5.245	0.021	0.136	0.806	2.364	0.459	0.134	5.366	0.000	0.081	-	-	-	-
	0.50	5.799	0.081	0.399	0.923	8.118	0.489	0.090	6.602	0.000	0.145	-	-	-	-
	0.75	5.896	0.106	0.841	1.253	11.746	0.579	0.065	6.815	0.000	0.244	-	-	-	-
900	0.25	5.808	0.034	0.153	0.630	0.511	0.426	0.095	5.469	0.012	0.103	0.852	2.217	0.475	0.137
	0.50	5.735	0.119	0.473	0.700	1.259	0.406	0.074	6.190	0.055	0.334	0.814	0.941	0.416	0.091
	0.75	6.196	0.154	0.947	1.072	1.229	0.482	0.058	6.922	0.073	0.756	1.247	0.900	0.597	0.071
3600	0.25	4.803	0.070	0.132	0.341	0.197	0.208	0.049	4.555	0.045	0.092	0.477	0.163	0.189	0.076
	0.50	4.796	0.405	0.779	0.356	0.194	0.256	0.037	4.988	0.220	0.519	0.489	0.189	0.188	0.055
	0.75	4.958	0.384	1.000	0.466	0.248	0.381	0.025	5.456	0.223	0.990	0.688	0.218	0.235	0.040
10000	0.25	4.388	0.252	0.252	0.159	0.106	0.107	0.025	4.271	0.118	0.118	0.283	0.108	0.102	0.049
	0.50	4.040	0.954	0.962	0.196	0.140	0.128	0.021	4.175	0.742	0.764	0.289	0.109	0.101	0.033
	0.75	4.072	0.942	1.000	0.261	0.252	0.219	0.014	4.289	0.803	1.000	0.366	0.126	0.122	0.021
22500	0.25	4.400	0.338	0.338	0.114	0.082	0.072	0.018	4.369	0.175	0.175	0.199	0.081	0.079	0.034
	0.50	4.003	0.996	0.996	0.136	0.114	0.089	0.015	4.080	0.891	0.891	0.221	0.085	0.079	0.025
	0.75	4.001	0.999	1.000	0.180	0.240	0.170	0.010	4.034	0.980	1.000	0.283	0.120	0.097	0.016

Table S-3: Characteristic of MARS and RMSE for NL-SDM-MARS ($\sigma = 1$)

n	ρ	Regular lattice								Irregular lattice									
		Charact. FPM				RMSE				Charact. FPM				RMSE					
		mnc	cs	W_{yt}	β_0	β_1	β_2	β_3	β_4	ρ	mnc	cs	W_{yt}	β_0	β_1	β_2	β_3	β_4	ρ
400	0.25	6.183	0.000	0.286	-	-	-	-	-	5.819	0.001	0.228	-	-	-	-	-	-	
	0.50	6.423	0.013	0.805	0.935	0.644	0.429	5.656	4.332	0.078	6.061	0.001	0.526	0.935	0.644	0.429	5.656	4.332	0.078
	0.75	6.688	0.009	0.999	2.390	1.117	0.359	7.859	4.695	0.089	6.562	0.003	0.933	2.947	1.201	0.492	21.167	2.425	0.146
900	0.25	6.602	0.005	0.359	0.647	0.316	0.572	2.235	0.182	0.077	6.293	0.000	0.246	-	-	-	-	-	-
	0.50	7.047	0.101	0.959	0.903	0.655	0.450	4.274	1.100	0.087	6.760	0.008	0.690	1.174	0.456	0.208	8.387	1.395	0.078
	0.75	6.922	0.081	1.000	1.355	0.784	0.438	6.929	2.291	0.064	7.395	0.010	0.997	2.660	1.103	0.414	6.785	5.983	0.131
3600	0.25	5.854	0.338	0.565	0.508	0.193	0.276	1.083	0.479	0.075	5.351	0.011	0.226	0.613	0.162	0.211	23.573	1.340	0.143
	0.500	6.248	0.647	1.000	0.488	0.200	0.313	0.934	0.489	0.053	6.005	0.226	0.992	0.803	0.218	0.210	10.668	1.377	0.086
	0.750	6.029	0.507	1.000	0.519	0.308	0.354	2.929	0.622	0.028	6.205	0.185	1.000	0.792	0.223	0.276	12.868	2.050	0.043
10000	0.25	5.634	0.633	0.634	0.275	0.120	0.123	0.293	0.368	0.042	5.025	0.043	0.095	0.749	0.116	0.111	2.176	1.282	0.137
	0.50	5.998	0.996	1.000	0.296	0.166	0.155	0.437	0.407	0.031	5.826	0.792	1.000	0.431	0.110	0.104	2.266	0.623	0.050
	0.75	5.973	0.967	1.000	0.314	0.270	0.313	0.787	0.652	0.016	5.753	0.601	1.000	0.425	0.134	0.126	1.926	1.037	0.025
22500	0.25	5.701	0.701	0.701	0.187	0.073	0.083	0.201	0.261	0.028	5.020	0.020	0.028	0.753	0.084	0.060	0.572	0.704	0.136
	0.50	6.000	1.000	1.000	0.194	0.117	0.118	0.365	0.343	0.021	5.960	0.960	1.000	0.336	0.094	0.081	1.530	0.522	0.038
	0.75	5.999	0.999	1.000	0.218	0.263	0.256	0.757	0.703	0.011	5.851	0.840	1.000	0.327	0.146	0.095	1.117	0.753	0.019

Table S-4: Characteristic of MARS and RMSE for NL-SEM-MARS ($\sigma = 0.5$)

n	Regular lattice							Irregular lattice					
	Charact. FPM				NL-SEM-MARS			Charact. FPM			NL-SEM-MARS		
	ρ	<i>mnc</i>	<i>cs</i>	Wyt	β_0	β_1	β_2	<i>mnc</i>	<i>cs</i>	Wyt	β_0	β_1	β_2
400	0.25	4.750	0.291	0.033	0.217	0.335	0.313	4.788	0.292	0.035	0.253	0.416	0.345
	0.50	5.120	0.220	0.082	0.217	0.335	0.313	5.220	0.196	0.087	0.253	0.416	0.345
	0.75	5.699	0.133	0.187	0.297	0.537	0.370	6.057	0.092	0.215	0.337	0.556	0.408
900	0.25	4.165	0.361	0.018	0.129	0.185	0.190	4.337	0.309	0.020	0.126	0.181	0.181
	0.50	4.455	0.293	0.045	0.146	0.204	0.209	4.676	0.238	0.050	0.150	0.197	0.203
	0.75	5.092	0.209	0.117	0.179	0.262	0.252	5.717	0.104	0.200	0.209	0.303	0.242
3600	0.25	3.030	0.978	0.000	0.060	0.088	0.087	3.054	0.963	0.000	0.059	0.088	0.086
	0.50	3.090	0.933	0.002	0.066	0.094	0.093	3.172	0.878	0.000	0.065	0.094	0.092
	0.75	3.488	0.691	0.022	0.086	0.111	0.114	3.915	0.466	0.037	0.090	0.116	0.120
10000	0.25	3.000	1.000	0.000	0.034	0.053	0.050	3.000	1.000	0.000	0.034	0.053	0.050
	0.50	3.000	1.000	0.000	0.037	0.057	0.053	3.000	1.000	0.000	0.038	0.058	0.054
	0.75	3.007	0.994	0.001	0.049	0.069	0.065	3.049	0.969	0.004	0.052	0.075	0.069
22500	0.25	3.000	1.000	0.000	0.022	0.036	0.033	3.000	1.000	0.000	0.024	0.038	0.036
	0.50	3.000	1.000	0.000	0.025	0.039	0.036	3.000	1.000	0.000	0.026	0.041	0.039
	0.75	3.000	1.000	0.000	0.032	0.047	0.044	3.000	1.000	0.000	0.036	0.052	0.050

Results: NL-SEM

Table S-5: Characteristic of MARS and RMSE for NL-SEM ($\sigma = 1$)

n	Regular lattice							Irregular lattice					
	Characteristic FPM				NL-SEM			Charact. FPM			NL-SEM-MARS		
	ρ	<i>mnc</i>	<i>cs</i>	Wyt	β_0	β_1	β_2	<i>mnc</i>	<i>cs</i>	Wyt	β_0	β_1	β_2
400	0.250	4.757	0.213	0.038	0.480	6.213	0.505	4.767	0.192	0.037	0.519	9.037	0.473
	0.500	4.980	0.147	0.086	0.480	6.213	0.505	4.996	0.134	0.085	0.519	9.037	0.473
	0.750	5.334	0.092	0.202	0.582	10.615	0.626	5.489	0.057	0.230	0.661	6.437	0.488
900	0.250	5.038	0.237	0.035	0.305	1.078	0.368	5.117	0.215	0.039	0.301	0.992	0.373
	0.500	5.353	0.188	0.070	0.382	2.658	0.423	5.438	0.150	0.086	0.364	1.237	0.413
	0.750	5.597	0.131	0.136	0.408	3.190	0.404	6.158	0.074	0.226	0.512	1.866	0.486
3600	0.250	3.694	0.531	0.011	0.133	0.182	0.181	3.756	0.485	0.004	0.132	0.183	0.185
	0.500	3.953	0.411	0.021	0.153	0.195	0.205	4.116	0.341	0.024	0.145	0.194	0.210
	0.750	4.518	0.252	0.107	0.199	0.260	0.243	4.933	0.162	0.112	0.198	0.282	0.261
10000	0.250	3.013	0.991	0.000	0.070	0.107	0.101	3.008	0.995	0.000	0.071	0.109	0.102
	0.500	3.029	0.980	0.000	0.081	0.117	0.112	3.061	0.958	0.001	0.080	0.118	0.112
	0.750	3.132	0.902	0.004	0.107	0.142	0.139	3.426	0.739	0.025	0.110	0.151	0.142
22500	0.250	3.000	1.000	0.000	0.046	0.073	0.067	3.000	1.000	0.000	0.051	0.079	0.074
	0.500	3.000	1.000	0.000	0.050	0.078	0.072	3.000	1.000	0.000	0.056	0.085	0.080
	0.750	3.001	0.999	0.001	0.067	0.095	0.089	3.032	0.981	0.003	0.077	0.108	0.103